UNIVERSITÉ DE LA POLYNÉSIE FRANÇAISE

ÉCOLE DOCTORALE DU PACIFIQUE
ED 469

LABORATOIRE GePaSud

THÈSE
présentée et soutenue publiquement par

Gaël Mondonneix

le 11 décembre 2019

en vue de l'obtention du titre de
Docteur de l'Université de Polynésie française en
Département : Science, technologies et santé
Discipline : 27 Informatique

L'apprentissage automatique du rangement d'objets et son application à l'évaluation du lustre des perles de culture de Tahiti

---

Sous la direction de : Alban Gabillon et Jean Martial Mari

JURY

Pr. Alban GABILLON, GEPASUD, UPF	Co-directeur de thèse
Dr. Jean Martial MARI, GEPASUD, UPF	Co-directeur de thèse
Pr. Patrick CAPOLSINI, GEPASUD, UPF	Examinateur
Dr. Emmanuel BRUNO, Université de Toulon et du Var	Examinateur
Pr. Christophe ROSENBERGER, Laboratoire GREYC – ENSICAEN, FRANCE	Rapporteur
Pr. Massih-Reza AMINI, Laboratoire LIG – Université Grenoble Alpes	Rapporteur

Résumé

Les problèmes d'apprentissage automatique sont parfois résolus avec des méthodes inadaptées, menant à des résultats sous optimaux. La détermination de la méthode à appliquer à la résolution d'un problème d'apprentissage automatique n'est pas toujours aisée, d'un côté parce qu'elle nécessite d'identifier correctement le problème à résoudre, et d'un autre côté parce que même si le problème en question a été identifié correctement, il n'existe pas nécessairement de méthode spécifique pour le résoudre.

La présente thèse part du constat selon lequel il n'existe pas d'algorithme d'apprentissage automatique spécifiquement dédié au rangement d'objets (un type particulier de problèmes d'apprentissage ordinal consistant à apprendre un ordre total à partir d'un ensemble de couples d'objets).

Une méthode d'apprentissage automatique destinée à combler ce manque est proposée et implémentée sous forme d'une machine à noyau. Un méta-noyau est développé afin de l'adapter aux bibliothèques standard d'apprentissage automatique.

L'algorithme proposé est appliqué à un cas concret de problème de rangement d'objets, l'évaluation du lustre des perles de Tahiti, et comparé avec quatre autres méthodes. Cet algorithme donne de meilleurs résultats (93,6% ± 3,9% de prévisions correctes sans sélection de caractéristiques, 94,3% ± 3,4% avec sélection de caractéristiques) que les meilleures des autres méthodes testées (91,3% ± 3,4% et 92,6% ± 4,3% sans et avec sélection de caractéristiques), cette amélioration étant significative ($p < 0,01$ et $p = 0.02$ respectivement). En outre, cette méthode ne présente pas de différence significative entre les résultats obtenus avec et sans sélection de caractéristiques (valeur-p = 0.33), ce qui peut indiquer que son biais d'apprentissage correspond bien au problème à résoudre et peut ainsi réduire la charge de travail en phase de prétraitement des données.

Table des matières

Résumé 3
Table des matières 4
Introduction 7
Partie 1 : Les problèmes de rangement d'objets 12
Chapitre 1 : Revue des notions d'apprentissage automatique 13
I : Apprentissage automatique 13
II : Mode indirect de programmation 14
III : Recherche automatique des régularités présentes dans les données d'apprentissage 15
IV : Généralisation des régularités à de nouvelles données 16
V : Apprentissage des données bruitées 19
VI : Apprentissage supervisé / non supervisé 20
VII : Perte / Risque 21
VIII : Types d'apprentissage supervisé 23
IX : Problèmes de régression 25
X : Problèmes de classement 26
XI : Problèmes de classification 27
XII : Synthèse 30
Chapitre 2 : Etat de l'art de l'apprentissage automatique des données ordinales 31
I : Problèmes de classement ordinal 31
II : Problèmes de rangement 34
III : Problèmes de classification ordinale 38
IV : Synthèse 38
Chapitre 3 : Catégories hybrides de problèmes d'apprentissage automatique des données ordinales 39
I : Approche de l'apprentissage automatique des données ordinales tenant compte de la nature des informations 39
II : Problèmes implicites de rangement 46
III : Synthèse 50
Partie 2 : Proposition d'une méthode spécifique au rangement d'objets 52

Chapitre 4 : Description d'une méthode de rangement agnostique directe 53
I : Description de l'espace de recherche 53
II : Validité de l'espace de recherche 55
III : Expressivité de l'espace de recherche 55
IV : Stratégie de recherche 57
V : Synthèse 63
Chapitre 5 : Description d'un algorithme d'apprentissage automatique spécifique au rangement d'objets 64
I : Distinction entre rangement d'instances et rangement d'objets selon le critère du biais d'apprentissage 64
II : Solution optimale du rangement d'objets 66
III : Mise en œuvre par une machine à noyau 68
IV : Forme compatible avec les bibliothèques d'optimisation existantes 72
V : Expérience 76
VI : Synthèse 80
Partie 3 : Application à l'évaluation du lustre des perles de Tahiti 81
Chapitre 6 : De l'évaluation du lustre des perles de Tahiti 82
I : Propriétés physiques et optiques de la nacre 82
II : Définition du lustre 84
III : Aperçu de la pratique de l'évaluation du lustre des perles à Tahiti 85
IV : Synthèse 93
Chapitre 7 : Modélisation du lustre des perles 94
I : Le lustre d'une perle en tant que fonction de la surface de la perle 94
II : Le lustre d'une perle en tant que composition de caractéristiques d'images de la perle 99
III : Synthèse 102
Chapitre 8 : Proposition de descripteurs du lustre des perles de Tahiti 103
I : Qualification des attributs perceptuels du lustre 104
II : Quantification des attributs perceptuels du lustre 116
III : Synthèse 126
Chapitre 9 : Application de l'algorithme de rangement d'objets à l'évaluation automatique du lustre des perles de Tahiti 128
I : Jeu de données 128

II : Validation de l'algorithme d'apprentissage 130
III : Validation des caractéristiques du lustre 136
IV : Synthèse 141
Conclusion 143
Bibliographie 146

Introduction

Les problèmes d'apprentissage automatique sont parfois résolus avec des méthodes inadaptées, menant à des résultats sous optimaux. La détermination de la méthode à appliquer à la résolution d'un problème d'apprentissage automatique n'est pas toujours aisée, d'un côté parce qu'elle nécessite d'identifier correctement le problème à résoudre, et d'un autre côté parce que même si le problème en question a été identifié correctement, il n'existe pas nécessairement de méthode spécifique pour le résoudre.

La présente thèse part du constat selon lequel il n'existe pas d'algorithme d'apprentissage automatique spécifiquement dédié au rangement d'objets (un type particulier de problèmes d'apprentissage ordinal consistant à apprendre un ordre total à partir d'un ensemble de couples d'objets).

Ce constat est établi à partir d'une nomenclature des problèmes d'apprentissage ordinal ne tenant pas seulement compte de l'énoncé des problèmes mais aussi de la nature des informations sur lesquelles ils portent. Cette nomenclature étant plus fine que les nomenclatures fondées sur le seul énoncé des problèmes, elle facilite l'identification des problèmes implicites d'apprentissage (ceux dont la forme ne coïncide pas avec le fond, en raison d'une distorsion entre l'énoncé du problème et la nature des données sur lesquelles il porte) et permet de déterminer plus précisément les méthodes de résolution à appliquer, mettant ainsi en évidence l'absence éventuelle de méthodes pour résoudre certains problèmes.

Au niveau de la forme des problèmes d'apprentissage des données ordinales, l'absence de méthode de résolution spécifique au rangement d'objets n'est pas flagrante. En effet, tout problème de rangement d'objets peut être reformulé en termes de rangement d'instances (un autre type particulier de problèmes d'apprentissage ordinal, consistant à apprendre un ordre total à partir d'un ensemble d'instances de classes ordonnées), or il existe une méthode spécifique de résolution des problèmes de rangement d'instances. Au niveau du fond, en revanche, les deux problèmes sont bien distincts : le rangement d'instances suppose la présence de classes qui ont une existence propre tandis que le rangement d'objets suppose l'absence de

telles classes. Là où une méthode de rangement d'instances tend à apprendre l'ordre total qui accentue le plus les classes ou les couples sous la forme desquels sont donnés les éléments, une méthode de rangement d'objets doit tendre à apprendre l'ordre total qui atténue le plus ces classes ou ces couples.

Nous proposons une méthode d'apprentissage automatique destinée à combler ce manque. Alors qu'une méthode de rangement d'instances a pour objectif de trouver l'ordre qui maximise la distance minimale entre les instances de classes différentes, de manière à ce que les classes soient les plus marquées, la méthode de rangement d'objets que nous proposons a pour objectif de trouver l'ordre qui se tient le plus éloigné de tout ordre invalide, sans tenir compte de la distance entre les éléments de classes différentes.

La question de la nécessité d'élaborer de nouvelles méthodes pourrait se poser dans la mesure où il existe des méthodes permettant de résoudre des problèmes proches. La raison est que la performance d'une méthode dépend de sa pertinence au regard du problème à la résolution duquel elle est appliquée : moins les données d'apprentissage sont nombreuses, plus la qualité des prédictions dépend de la pertinence de la méthode.

Dans un contexte d'abondance des données ("Big Data" [1]), la perte d'une partie de l'information contenue dans des données ne pose pas de problème car davantage d'information peut être extraite à partir de données supplémentaires. Le défi n'est pas tant d'exploiter toute l'information contenue dans les données que de traiter toutes les données disponibles.

Au contraire, dans un contexte de rareté des données, si toute l'information contenue dans les données disponibles, et notamment la potentielle information ordinale, n'est pas exploitée, alors il n'est pas possible de la recouvrer à partir de données supplémentaires.

La méthode que nous proposons peut être implémentée sous forme d'une machine à noyau. Nous développons un méta-noyau afin de l'adapter aux bibliothèques standard d'apprentissage automatique. Nous testons ensuite l'algorithme sur des données artificielles et le comparons à un algorithme de rangement d'instances que nous prenons comme référence.

Un exemple concret de problème de rangement d'objets, et plus spécifiquement, de problème implicite de rangement d'objets, est posé en Polynésie française par le tri des perles de culture.

Les perles de culture de Tahiti sont triées en quatre classes, de A à D, par ordre décroissant de qualité.

D'après la forme du problème, il s'agit d'un simple problème de classement ordinal (la tâche consiste à distribuer les perles dans des classes ordonnées). D'après la nature des informations en jeu, il s'agit d'un problème de rangement d'objets ; en effet, il se trouve d'une part que les experts sont capables d'ordonner les perles totalement, ce qui en fait un problème de rangement, et d'autre part que les classes ne sont pas figées (leur configuration peut varier selon le lot de perles à trier ou selon le contexte de tri), ce qui en fait plus spécifiquement un problème de rangement d'objets. La forme du problème étant un classement ordinal et sa nature étant un rangement d'objets, il s'agit d'un problème implicite de rangement d'objets.

Le tri des perles est extrêmement important en Polynésie française. La perle de culture constitue la deuxième ressource de la Polynésie française après le tourisme et la première en termes d'exportations[1]. Afin d'aider la perle de Tahiti à être plus compétitive sur le marché international, l'accent a été mis sur le contrôle de la qualité par les autorités locales. Sur un marché aussi compétitif, une assistance automatisée pourrait apporter un avantage décisif à la Polynésie française.

Dans ce contexte, le projet RAPA[2] vise à apporter une assistance automatisée aux experts dans l'évaluation de la qualité des perles de Tahiti. Les recherches menées dans le cadre de ce projet ont porté sur la mesure automatique de l'épaisseur de la nacre [2], [3] et sur la caractérisation de la couleur des perles comme préalable à une classification des perles selon leur couleur [4], [5]. Les autres critères d'évaluation de la qualité des perles de culture de Tahiti sont la forme, l'état de la surface, la taille et le lustre.

---

[1] 8,8 milliards F.CFP en 2014, soit 69% des exportations de la Polynésie française [145].

[2] RAPA [146] signifie « Reconnaissance Automatique de la qualité des Perles de Tahiti », un projet parrainé par le ministère des Outre-mer.

Bien que des millions de perles soient triées chaque année, l'exercice du tri est opéré par différents experts sur des lots de petite taille (allant de quelques perles, voire quelques dizaines de perles pour le tri de détail à quelques centaines, voire quelques milliers de perles tout au plus pour le tri en gros). Le tri des perles de culture de Tahiti nécessite donc, pour être automatisé, l'utilisation d'une méthode pertinente.

Nous appliquons notre algorithme à l'évaluation du lustre des perles de Tahiti et le comparons à quatre autres méthodes. Cet algorithme donne de meilleurs résultats (93,6% ± 3,9% de prévisions correctes sans sélection de caractéristiques, 94,3% ± 3,4% avec sélection de caractéristiques) que les meilleures des autres méthodes testées (91,3% ± 3,4% et 92,6% ± 4,3% sans et avec sélection de caractéristiques pour l'approche par rangement d'instances), cette amélioration étant significative ($p < 0,01$ et $p = 0.02$ respectivement). En outre, cette méthode ne présente pas de différence significative entre les résultats obtenus avec et sans sélection de caractéristiques (valeur-p = 0.33), ce qui peut indiquer que son biais d'apprentissage correspond bien au problème à résoudre et peut ainsi réduire la charge de travail en phase de prétraitement des données.

La thèse est structurée en trois parties.

Dans la première partie, nous exposons la théorie de l'apprentissage automatique des données ordinales et proposons une nouvelle nomenclature des problèmes d'apprentissage des données ordinale, permettant d'identifier les problèmes implicites de rangement, de rangement d'instances et de rangement d'objets, et de relever ainsi l'absence de méthode propre au rangement d'objets. Dans la deuxième partie, nous proposons une méthode de résolution des problèmes de rangement d'objets afin de combler ce manque et la testons sur un jeu de données artificielles. Dans la troisième partie, nous étudions le tri des perles de Tahiti et l'identifions en tant que problème implicite de rangement d'objets, avant de le résoudre dans le cas du lustre avec la méthode proposée dans la deuxième partie.

Les parties sont elles-mêmes organisées en chapitres.

Dans le chapitre 1, nous passons en revue les notions d'apprentissage automatique qui seront utiles dans la suite de la thèse, en nous attardant notamment sur la notion de biais

d'apprentissage afin de souligner l'altération des performances entraînées par le fait de résoudre un problème avec une méthode inadaptée. Nous dressons ensuite un état de l'art de l'apprentissage des données ordinales dans le chapitre 2 et clôturons la première partie avec, dans le chapitre 3, l'exposition d'une nouvelle nomenclature des problèmes d'apprentissage ordinal et l'identification de nouveaux types de problèmes d'apprentissage ordinal : les problèmes implicites de rangement, de rangement d'instances et de rangement d'objets.

Les chapitres 4 et 5 forment le corps de la deuxième partie, relative à la résolution des problèmes de rangement. Nous décrivons dans le chapitre 4 une méthode générique de résolution des problèmes de rangement agnostique (c'est-à-dire de rangement qui ne soit ni spécifiquement un rangement d'instances ni spécifiquement un rangement d'objets) et, nous basant sur cette méthode générique, nous proposons dans le chapitre 5 une méthode spécifique au rangement d'objets. Nous développons cette dernière méthode sous la forme d'une machine à noyau et décrivons une manière de l'implémenter à l'aide des bibliothèques d'optimisation standard avant de la tester sur un jeu de données artificielles.

La troisième partie, relative à l'application de la méthode de rangement d'objets au tri du lustre des perles de Tahiti, débute avec le chapitre 6, dans lequel nous définissons la notion de lustre et dressons un état des lieux détaillé de la pratique de l'évaluation du lustre des perles de culture en Polynésie française. Nous discutons alors, dans le chapitre 7, des manières de modéliser le lustre afin de construire des vecteurs de caractéristiques aptes à appréhender le lustre à partir de photographies de perles. Dans le chapitre 8, nous examinons les attributs perceptuels du lustre et proposons des descripteurs mathématiques afin de les quantifier. Enfin, dans le chapitre 9, nous exposons et discutons les résultats de l'application de la méthode de rangement d'objets au tri du lustre des perles, à partir d'un ensemble de photographies de perles de culture de Tahiti dont le lustre est évalué par un expert en perles.

Partie 1 : Les problèmes de rangement d'objets

Chapitre 1 : Revue des notions d'apprentissage automatique

Il existe une grande variété d'algorithmes d'apprentissage automatique permettant de résoudre un grand nombre de problèmes, allant du diagnostic médical automatisé [6], [7] à la reconnaissance faciale [8], [9] en passant par la détection de courriers électroniques indésirables [10], [11] ou de fraudes à la carte bancaire [12]. De manière à mieux situer et définir notre travail dans le domaine de l'apprentissage des données ordinales, nous allons passer en revue les problématiques communes aux différents types d'algorithmes ainsi que leurs spécificités.

Nous exposerons dans un premier temps les problématiques communes, en introduisant de manière générale ce qu'est l'apprentissage automatique et en nous attardant sur les notions importantes que sont la recherche des régularités présentes dans les données et la généralisation de ces régularités à d'autres données. Dans un second temps, nous aborderons les différents types d'apprentissage automatique et préciserons la terminologie adoptée dans le reste de la thèse. Nous aborderons d'une part les types d'apprentissage traditionnels tels que la régression, le classement ("classification" en Anglais) et la classification ("clustering" en Anglais), et d'autre part les types d'apprentissage ordinal tels que le classement ordinal, le rangement ("ranking" en Anglais) et la classification ordinale.

I : Apprentissage automatique

L'apprentissage automatique ("machine learning" en anglais) est le domaine d'étude qui porte sur l'élaboration de programmes informatiques capables *d'apprendre*, c'est-à-dire de *s'améliorer par l'expérience* dans la réalisation d'une tâche déterminée [13]. Formellement, « un programme informatique est dit *apprendre* d'une expérience E par rapport à une classe de

tâches T et d'après une mesure de performance P si sa performance dans la réalisation de la tâche T, telle que mesurée par P, s'améliore avec l'expérience E. »[3] ([13] p. 2)

II : Mode indirect de programmation

Un programme informatique consiste en une implémentation d'une fonction. Or une fonction est un ensemble (possiblement infini) d'associations entre des entrées et des sorties et peut à ce titre être définie *en intension*, par les propriétés que vérifient les associations, ou *en extension*, par l'énumération des associations elles-mêmes [14]. Par exemple la fonction qui décide si un nombre est pair sera définie en intension par la propriété selon laquelle un nombre pair est un multiple de deux, et en extension par l'énumération (infinie) des nombres pairs $\{\dots, -6, -4, -2, 0, 2, 4, 6, \dots\}$.

La programmation informatique consiste généralement à implémenter une fonction à partir de sa définition en intension. Il peut arriver cependant que les propriétés de la fonction à implémenter ne soient pas connues mais que les associations $\langle \text{entrée}, \text{sortie} \rangle$ le soient ; il est alors possible d'implémenter cette fonction à partir de sa définition en extension, le plus souvent par l'intermédiaire d'un algorithme d'apprentissage.

Pour ce faire, l'algorithme procède à une recherche automatique des régularités présentes dans les données ([15] p. 1) : les données correspondent à la définition (partielle) en extension de la fonction ; les régularités correspondent à la définition en intension de la fonction. La fonction que recherche l'algorithme d'apprentissage est appelée « fonction cible » (et notée f) et celle qu'il trouve est appelée « hypothèse » (et notée h).

Définition : un algorithme d'apprentissage est un algorithme prenant en entrée une fonction partiellement définie en extension et retournant une hypothèse sur sa définition en intension. Cette définition reprend sous l'angle de la distinction intension/extension la définition donnée

[3] "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E." [13] p. 2.

par Hugo Larochelle : « Un algorithme d'apprentissage est un algorithme prenant en entrée un ensemble de données D et retournant une fonction f. » ([16] p. 1)

L'apprentissage automatique peut être vu comme un mode de programmation indirect, par les données [17][4]. Tout programme d'apprentissage automatique peut d'ailleurs être implémenté sous la forme d'un programme recevant des données en argument et retournant un programme qui leur est adapté.

Il y a trois fonctions en jeux dans un programme d'apprentissage automatique : la fonction cible, qui est définie partiellement (et parfois de manière bruitée) par les données et existe indépendamment du programme ; une fonction de recherche, directement programmée, qui recherche la fonction cible ; une fonction de prédiction, retournée par la fonction de recherche relativement aux données, qui représente une hypothèse sur la fonction cible.

III : Recherche automatique des régularités présentes dans les données d'apprentissage

La recherche automatique des régularités présentes dans les données s'effectue par l'exploration d'un espace de fonctions implicitement supposé contenir la fonction cible. Les fonctions de cet espace sont appelées « hypothèses » car elles représentent autant d'hypothèses possibles sur la fonction cible ; l'espace lui-même est appelé « espace des hypothèses ». En effet, l'algorithme n'est en général pas capable de trouver f et doit se contenter de retourner une hypothèse sur la fonction cible ([18] p. 10).

Un algorithme d'apprentissage dispose d'un système de représentation de l'information et recherche, parmi toutes les hypothèses que ce système lui permet de représenter, celle qui correspond le mieux aux données qui lui sont soumises. Les différents algorithmes d'apprentissage sont caractérisés par les différents systèmes de représentation qu'ils mettent

[4] "Programming computers [in the sense of making computers able] to learn from experience should eventually eliminate the need for much of this [playing checkers, but it could be any other task] detailed programming effort" [17]. La prétendue citation "Field of study that gives computers the ability to learn without being explicitly programmed", bien que répandue dans la littérature (par exemple [147]) et exprimant cette même idée de programmation implicite, par l'intermédiaire des données, ne figure pas dans l'article de Samuel de 1959.

en œuvre ; leurs noms s'y réfèrent d'ailleurs le plus souvent. Par exemple, un algorithme d'apprentissage par arbre de décision [19] présuppose que les données puissent être modélisées par un arbre de décision et explore ainsi l'espace des arbres de décision afin de sélectionner celui qui s'ajuste le mieux aux données.

Le choix de l'algorithme d'apprentissage est crucial car le choix du système de représentation détermine l'espace de toutes les hypothèses que le programme sera en mesure de représenter, et donc d'apprendre[5]. Chaque point de l'espace des hypothèses représente un réglage, un paramétrage différent du modèle. La recherche dans cet espace du point qui représente le meilleur modèle est un problème d'optimisation (par exemple, le réglage des paramètres d'un réseau de neurones artificiel est souvent réalisé par l'algorithme de *rétropropagation du gradient* [20]). Le fait de régler les paramètres d'un modèle afin d'accorder ce modèle avec des données n'est pas propre à l'apprentissage automatique, cependant l'apprentissage automatique effectue cet ajustement de manière automatique.

En résumé, un algorithme d'apprentissage reçoit un ensemble de données et en retourne un modèle. Ce modèle est dit « appris » par l'algorithme.

IV : Généralisation des régularités à de nouvelles données

La plupart du temps, la liste exhaustive de toutes les associations $\langle \text{entrée}, \text{sortie} \rangle$ de la fonction n'est pas disponible, et l'algorithme d'apprentissage doit inférer une définition en intension à partir d'une définition *partielle* en extension, c'est-à-dire généraliser à de nouvelles données les régularités découvertes sur l'ensemble d'apprentissage. Ce passage du particulier au général s'appelle l'induction et implique de faire des hypothèses tant sur les données d'apprentissage (hypothèse *I.I.D.*, pour *Indépendant et Identiquement Distribué*) que sur les régularités recherchées (biais d'apprentissage).

---

[5] "by selecting a hypothesis representation, the designer of the learning algorithm implicitly defines the space of all hypotheses that the program can ever represent and therefore can ever learn." [13] p. 23. (« en sélectionnant un système de représentation des hypothèses, le concepteur d'un algorithme d'apprentissage définit implicitement l'espace de toutes les hypothèses que le programme pourra jamais représenter, et en conséquence, apprendre. »)

A : Inférence par induction

L'inférence par induction suppose que « toute hypothèse approximant la fonction cible de manière satisfaisante sur un ensemble assez large d'exemples approxime cette fonction de la même manière sur d'autres exemples. »[6]

Cette inférence est appelée « généralisation » ([21] p. 6) car elle consiste à généraliser à l'ensemble des cas les régularités découvertes sur un de ses sous-ensembles. Les exemples du sous-ensemble sur lequel les régularités sont découvertes sont dits « d'apprentissage » ou « d'entraînement » ([21] p. 2). Les exemples disponibles qui ne sont pas d'apprentissage sont dits « de test » ([21] p. 10) dans la mesure où ils sont susceptibles de servir à tester la capacité de généralisation d'un modèle.

B : Hypothèse I.I.D.

L'inférence par induction repose sur l'hypothèse selon laquelle tous les exemples (qu'ils soient d'apprentissage ou de test) pourraient être considérés comme des variables aléatoires *indépendantes* et *identiquement distribuées*. Cette hypothèse est appelée « *hypothèse I.I.D.* » [22] : puisque les exemples d'apprentissage suivent tous une même distribution et sont indépendants, alors il est possible de s'appuyer sur ces exemples pour faire des prédictions relatives aux exemples de tests, c'est-à-dire d'inférer par induction.

C : Biais d'apprentissage

Quand l'espace des hypothèses est trop expressif, il arrive que plusieurs hypothèses s'accordent aussi bien les unes que les autres avec les exemples d'apprentissage. Il convient alors de réduire cet espace ou d'en restreindre son exploration afin de ne retenir qu'une seule hypothèse. En d'autres termes, il convient de sélectionner une généralisation parmi un ensemble de généralisations qui se trouvent être équivalentes au regard des données d'apprentissage : cette sélection ne pouvant s'appuyer sur les données d'apprentissage, elle

---

[6] "Any hypothesis found to approximate the target function well over a sufficiently large set of training examples will also approximate the target function well over other unobserved examples" [13] p. 23.

s’appuie nécessairement sur un élément extérieur. Étant extérieur aux données, cet élément biaise l’apprentissage ; aussi est-il appelé « biais d’apprentissage » [13].

Un biais d'apprentissage est un ensemble d'hypothèses qui sont faites, explicitement en définissant certains paramètres ou implicitement par le choix de la méthode, sur les données, et qui permettent de généraliser à de nouveaux exemples ce qui a été appris sur les exemples d'apprentissage ([13] p.42). Il peut être caractérisé selon deux critères, la force et la pertinence ([13] p.45) : un biais d’apprentissage plus fort confère un plus grand pouvoir de généralisation à de nouveaux exemples ; un biais d'apprentissage plus pertinent permet des prévisions plus justes. En ce sens, le choix du biais d'apprentissage est un aspect crucial du processus d'apprentissage.

Un exemple de biais d’apprentissage, appelé le rasoir d’Ockham [23], privilégie, parmi toutes les hypothèses capables de s’ajuster aux données d’apprentissage, l’hypothèse la plus simple. Ce biais peut s’illustrer dans le cas d’une modélisation de la fonction cible avec des hypothèses polynomiales. Si le degré du polynôme n’est pas borné, alors lorsqu’il est au moins égal au nombre d’exemples d’apprentissage, le polynôme peut s’ajuster parfaitement à ces exemples : le biais d’apprentissage initial (celui introduit par le choix d’une modélisation par polynômes sans autre contrainte) est trop faible. Si maintenant le degré des polynômes est borné supérieurement, alors plus la borne est basse, plus le biais d’apprentissage est fort ; à l’extrême, lorsque la borne est de 0, le modèle ne tient plus compte des données du tout : le biais d’apprentissage est trop fort. Entre les deux, le biais introduit dans l’apprentissage par le principe du rasoir d’Ockham consiste à sélectionner, parmi les polynômes compatibles avec les données, le polynôme de plus faible degré (lequel correspond à l’hypothèse la plus simple dans le cadre d’une modélisation par polynômes).

S’il est possible de mesurer la force d’un biais d’apprentissage [24] (plus il est fort, plus il restreint l’espace des hypothèses ; plus il est faible, moins il le restreint), il n’est pas possible de juger de la pertinence d’un biais d’apprentissage dans l’absolu [25], laquelle dépend du problème spécifique et de la connaissance qu’on en a [26] ; le rasoir d’Ockham n’y fait pas exception [27].

Le biais d'apprentissage peut être vu comme une forme de connaissance *a priori* de la fonction cible qui permet justement de décider du modèle dans une certaine mesure (la mesure correspondant à la part de la décision qui n'est pas couverte par les données) [13].

Plus largement, le choix même du système de représentation de l'algorithme d'apprentissage constitue un biais d'apprentissage dès lors qu'il simplifie les données ([13] p. 42) : la supposition implicite que la fonction cible soit contenue dans l'espace des hypothèses représentables par ce système repose sur une connaissance extérieure aux données.

Le théorème du "No Free Lunch" [28] affirme qu'il n'y a pas d'apprentissage possible sans biais d'apprentissage et qu'il n'y a pas de système de représentation de l'information qui soit meilleur que les autres pour tous les types d'information. Un système capturant efficacement un type d'information donné peut très bien être incapable d'en capturer un autre.

V : Apprentissage des données bruitées

Il peut arriver que certains exemples d'apprentissage ne correspondent pas exactement à la fonction cible, par exemple lorsque des mesures sont imprécise ou lorsque des erreurs se sont introduites dans la constitution du jeu de données ; dans ce cas les données sont dites « bruitées ».

Le choix du biais d'apprentissage est d'autant plus important quand les données sont bruitées. En effet, un modèle avec un biais trop faible modéliserait même le bruit ; cela reviendrait à apprendre les données « par cœur » en quelque sorte, à la manière d'une base de données. Le modèle serait dès lors incapable de généraliser à de nouveaux exemples. Dans ce cas, on parle de « sur-apprentissage » (pour une définition formelle du sur-apprentissage, cf. [13] p. 67).

Au contraire, lorsque le biais est trop fort, il ne permet plus au modèle de s'adapter aux données ; un modèle trop biaisé devient trop pauvre pour capturer la complexité des régularités présentes dans les données et n'exploite par conséquent pas toute l'information contenues dans les données. On parle dans ce cas de « sous-apprentissage » [29]. Dans le cas extrême d'un modèle totalement biaisé, il n'y aurait même aucun apprentissage et la fonction de prédiction retournée par le programme d'apprentissage pourrait aussi bien être programmée directement.

VI : Apprentissage supervisé / non supervisé

L'apprentissage de la fonction cible passe par la recherche de régularités dans les données d'apprentissage. Selon la forme de ces données, la fonction recherchée sera définie plus ou moins explicitement.

A : Apprentissage supervisé

Si les données définissent explicitement la fonction cible, ce qui est le cas lorsque les exemples d'apprentissage se présentent sous la forme d'associations $\langle \text{entrée}, \text{sortie} \rangle$, alors on parle d'apprentissage supervisé [30]. Les sorties sont appelées « étiquettes » et les données sont alors dites « étiquetées ». Une fois le modèle appris, il permet de prédire les étiquettes à associer à de nouvelles entrées.

Par exemple, l'apprentissage d'un réseau de neurones sur les données d'apprentissage, c'est-à-dire d'un modèle faisant correspondre les entrées des exemples d'apprentissage avec leurs étiquettes, permet de prédire, en appliquant le modèle obtenu à de nouvelles entrées, les sorties correspondantes.

B : Apprentissage non supervisé

Si les données ne spécifient pas explicitement ce que doit retourner la fonction cible, ce qui est le cas lorsque les exemples d'apprentissage ne consistent qu'en des entrées (les données ne sont pas étiquetées), alors on parle d'apprentissage non supervisé [31]. Une fois le modèle appris, une étape d'explicitation est nécessaire afin de lui donner la forme d'une fonction. En effet, le modèle appris n'est pas une fonction : la fonction cible exploite ce modèle mais ne se confond pas avec lui. La spécification de la fonction cible est extérieure aux données et dépend généralement du domaine d'application. Selon la manière dont il est exploité, un même modèle peut donner lieu à différentes fonctions cibles.

Par exemple, si un modèle de partition est appris sur l'ensemble d'apprentissage, alors ce modèle rend compte, au travers des sous-ensembles qu'il constitue, de la distribution gouvernant les données d'apprentissage, mais ne permet pas de calculer directement une sortie à partir d'une entrée.

Une explicitation typique de la fonction cible consiste à retourner, pour une entrée donnée, le sous-ensemble auquel appartient cet élément, sous la forme d'un identifiant quelconque se rapportant le plus souvent à la valeur moyenne des éléments de ce sous-ensemble (ainsi par exemple le partitionnement par k-moyennes [32]).

Rien dans les données n'interdit toutefois de spécifier la fonction cible différemment. Outre que celle-ci pourrait retourner différents attributs du sous-ensemble auquel appartient l'élément donné en entrée, telles la moyenne, la médiane, ou la fréquence relative des éléments de ce sous-ensemble, il pourrait aussi être convenu que celle-ci retourne des informations relatives au rapport entre l'élément et son sous-ensemble, telles la distance entre la valeur de l'élément et la valeur moyenne du sous-ensemble ou la différence de rang avec la valeur médiane.

VII : Perte / Risque

Afin de de pouvoir décider quel est le modèle qui ajuste le mieux la fonction cible, il faut disposer d'une fonction qui permette d'évaluer la différence entre un modèle et la fonction cible. Cette fonction s'appelle le *risque* (ou « erreur ») et est notée $R(h)$. Le risque est la *perte moyenne* sur l'ensemble de définition de la fonction cible. La *perte* (ou « coût », ou encore « fonction objectif ») est la fonction qui permet d'évaluer la différence entre un modèle h et la fonction cible f *à l'échelle d'un point* ; la perte en un point x se note $L\big(h(x), f(x)\big)$, où L se réfère à l'anglais "loss", pour « perte » [33].

L'équation (1) formalise le risque relatif à l'hypothèse $h \in H$ et la fonction cible $f : E \to F$.

$$R(h) = \mathbb{E}\big[L\big(h(x), f(x)\big)\big], x \in E \quad (1)$$

La perte peut avoir différentes formes selon le problème. Parmi toutes les formes possibles, les plus répandues sont les fonctions L_{0-1}, L_1 et L_2. La fonction L_{0-1} caractérise l'*existence* d'une différence entre le modèle et la cible [34] ; cette fonction n'étant cependant pas viable d'un point de vue calculatoire, des fonctions de substitution sont utilisées en pratique, telles les fonctions charnières ("hinge functions") ou sigmoïdes [35], [36]. Les fonctions L_1 et L_2 [37] caractérisent l'*importance*, le « poids » de cette différence, la première avec la distance

euclidienne entre le modèle et la cible, la seconde avec le carré de cette distance. Ces fonctions sont formalisées par les équations (2), (3), et (4), avec $f: E \to F$ la fonction cible, $h \in H$ une hypothèse et $x \in E$ un élément quelconque.

$$L_{0-1}\big(h(x), f(x)\big) = \begin{cases} 1 \; si \; h(x) \neq f(x) \\ 0 \; sinon \end{cases} \quad (2)$$

$$L_1\big(h(x), f(x)\big) = |h(x) - \; f(x)| \quad (3)$$

$$L_2\big(h(x), f(x)\big) = \big(h(x) - \; f(x)\big)^2 \quad (4)$$

Le risque basé sur la perte L_{0-1} (équation (2)) et, plus largement, sur ses substituts, calcule le taux d'erreur du modèle sur l'ensemble de points considéré.

Le risque basé sur la perte L_1 (équation (3)) calcule la distance moyenne entre le modèle et la cible.

Le risque basé sur la perte L_2 (équation (4)) calcule la distance quadratique moyenne entre le modèle et la cible.

D'un point de vue calculatoire, un risque basé sur L_2 est préférable à un risque basé sur L_1 car il est alors dérivable, ce qui facilite son optimisation.

Du point de vue de l'impact sur la sélection du modèle, le choix entre L_1 et L_2 dépend des contraintes que le modèle doit respecter : L_1 pénalise dans des proportions uniformes les différences entre le modèle et la cible alors que L_2 les pénalise dans des proportions variables, exacerbant les grandes différences et rendant négligeables les petites. Ainsi, entre un modèle différant légèrement de la cible sur tous les points considérés et un modèle s'ajustant parfaitement à la cible sauf en quelques points où les différences sont importantes, une perte basée sur L_1 aboutira à la sélection du second alors qu'une perte basée sur L_2 aboutira à la sélection du premier.

Une perte basée sur L_2 peut s'avérer problématique lorsque les données contiennent des valeurs aberrantes dans la mesure où les meilleurs modèles, fortement pénalisés par la

présence des valeurs aberrantes, seront le plus souvent rejetés. Une perte basée sur L_1 sera en revanche plus robuste aux valeurs aberrantes.

VIII : Types d'apprentissage supervisé

Les tâches d'apprentissage supervisé consistent à apprendre une fonction qui s'ajuste le mieux aux données dans le cas où les données sont constituées de couples $\langle entrée, étiquette \rangle$, c'est-à-dire une fonction qui modélise le lien entre les entrées et les étiquettes correspondantes.

Les tâches d'apprentissage supervisé peuvent être subdivisées selon la nature des valeurs prises par les étiquettes, c'est-à-dire selon le type de variable que constituent les étiquettes [38].

A : Catégories traditionnelles

Il existe plusieurs subdivisions : qualitatif/quantitatif, discret/continu [38], qui rendent compte des deux principales familles d'algorithmes d'apprentissage supervisé que sont la famille des algorithmes de classement et celle des algorithmes de régression.

Le problème est que ces subdivisions sont binaires et ne laissent par conséquent pas de place à d'autres types de tâches d'apprentissage. Or une troisième catégorie suscite de plus en plus d'attention avec par exemple le rôle grandissant des systèmes de recommandation sur le Web [39] : les tâches d'apprentissage des données ordinales. Ne trouvant cependant pas sa place dans ces dichotomies, l'apprentissage des données ordinales est souvent présenté comme un type composite, empruntant à la fois à la classification, car travaillant sur des valeurs discrètes, et à la régression, car impliquant des rapports quantitatifs. Cette présentation suggère que l'apprentissage des données ordinales serait un type secondaire pour lequel l'utilisation d'algorithmes issus des deux grandes familles d'algorithmes d'apprentissage serait satisfaisante [40], [41].

B : Catégories fondées sur l'information relative aux différences entre éléments

Il existe une subdivision ternaire qui se calque sur le type des données à prédire : données nominales (par exemple la nationalité d'un coureur donné), ordinales (par exemple le rang de ce

coureur à l'issue d'une course donnée) ou métriques (par exemple le temps qu'a mis ce coureur lors de cette course) [42], [43]. Cette subdivision prend appui sur les niveaux de mesure des variables définies par [44] : les variables nominales, ordinales, d'intervalle et de rapport.

- **Les variables nominales** prennent des valeurs qui ne peuvent être comparées que par le biais de l'égalité : deux variables sont soit égales soit inégales, aucune autre information n'est disponible. Les valeurs de ces variables peuvent être analysées comme des références à des classes : deux variables appartiennent à une même classe (leurs valeurs sont égales) ou à des classes différentes (leurs valeurs ne sont pas égales). Dans une tâche de classement, les valeurs prises par les étiquettes ne sont pas seulement discrètes, elles sont incomparables entre elles par tout autre opérateur que celui d'égalité.
- **Les variables ordinales** prennent des valeurs qui peuvent être comparées par le biais de l'égalité et de l'inégalité : une variable est supérieure, égale ou inférieure à une autre. Cette information relative au sens de la différence n'existe pas dans le cas des variables nominales. Cependant, la notion de distance entre les variables n'existe pas pour autant : les variables ordinales peuvent être ordonnées, mais la différence entre deux variables se cantonne ici à une information relative au sens, pas à la distance.
- **Les variables d'intervalle** permettent plus d'information encore sur leurs différences : non seulement la différence entre deux variables a un sens, mais elle a aussi un poids. Deux variables d'intervalle peuvent être plus ou moins différentes alors que les variables ordinales (et nominales) sont différentes ou ne le sont pas.
- **Les variables de rapport**, enfin, ne sont pas seulement comparables entre elles, mais sont aussi comparables à un repère, une origine. La présence d'un élément neutre permet de comparer des rapports, ce qui n'est pas possible avec les variables d'intervalle.

La distinction entre les différents types de variables se fonde sur la nature de la différence entre variables, à part entre les deux derniers types, où elle se fonde sur l'existence d'un élément neutre. Par exemple, le poids en kg est une variable de rapport alors que la température en degrés Celsius, ou encore le temps selon le calendrier grégorien, sont des variables d'intervalle.

Ainsi, 2000 kg est deux fois plus lourd que 1000 kg car 0 kg signifie l'absence de poids ; 2000° C n'est pas deux fois plus chaud que 1000° C car 0° C ne signifie pas l'absence de chaleur ; l'an 2000 n'est pas deux fois plus avancé que l'an mille car l'an 0 ne signifie pas l'absence de temps. Trois niveaux d'information sont possibles relativement à cette différence : **existence** (variables nominales), **sens** (variables ordinales), et **distance** (variables d'intervalle et de rapport).

Lorsque la seule information relative à la différence entre les valeurs prises par les étiquettes est l'existence, nous parlerons de classement nominal ; lorsque l'information relative à la différence entre les valeurs est l'existence et le sens, nous parlerons d'apprentissage ordinal ; lorsque l'information consiste en l'existence, le sens et le poids, nous parlerons de régression[7]. Le Tableau 1 résume le lien entre les catégories de tâches d'apprentissage automatique et différents niveaux d'information relative à la différence entre les variables en jeu.

Tableau 1 : Relation entre les catégories de tâches d'apprentissage automatique, les types de variables en jeux, et la nature de la différence entre les variables.

		Type de variables en jeu	*Nature de la différence entre les variables*		
			Existence	**Sens**	**Distance**
Catégorie	**Classement**	Variables nominales	**X**	-	-
	Apprentissage ordinal	Variables ordinales	**X**	**X**	-
	Régression	Variables d'intervalle et de rapport	**X**	**X**	**X**

IX : Problèmes de régression

La régression consiste à apprendre une fonction prenant des valeurs métriques [29].

Par exemple, si le lien entre les entrées et les étiquettes correspondantes est supposé être de nature linéaire, alors l'algorithme d'apprentissage parcourt l'espace de toutes les droites possibles à la recherche de celle qui rend le mieux compte des données. De même, si la nature

[7] Ces termes correspondent respectivement aux termes anglais "classification", "ordinal regression" et "metric regression" [42], [43].

du lien est supposée quadratique, alors l'algorithme cherche à ajuster une fonction quadratique sur les données.

La régression peut être illustrée par le problème suivant dans lequel un producteur de films, par exemple, chercherait à prédire le montant des recettes que générerait un film avant de prendre la décision de le produire ou non.

Supposons qu'un film soit modélisé par un ensemble de n descripteurs, comme le nombre et l'âge des acteurs, le nombre et la durée des séquences, des éléments tirés de la bande sonore, du script, etc., et qu'un espace de description des films à n dimensions soit défini en associant chaque dimension de l'espace à un descripteur. Chaque film peut alors être représenté par un vecteur dans cet espace de description des films selon ses caractéristiques de ce film, c'est-à-dire selon les valeurs que prennent ses descripteurs. Etant donné un ensemble de films, chacun représenté par un vecteur de caractéristiques ainsi défini et associé aux recettes que ce film a générées, le problème de régression consiste à estimer les recettes que généreraient de nouveaux films en fonction de leurs vecteurs de caractéristiques.

Dans le cadre de la régression, la fonction de perte utilisée généralement est L_2 (carré de la distance euclidienne entre la valeur retournée par l'hypothèse et la valeur retournée par la fonction cible pour une entrée donnée).

X : Problèmes de classement

Le classement consiste à apprendre une fonction prenant des valeurs nominales [29].

Le classement peut être illustré par le problème suivant : étant donné un ensemble de films, chacun représenté par un vecteur de caractéristiques et associé à l'un des six genres parmi « action », « comédie », « drame », « policier », « science-fiction » ou « autre », déterminer les genres de nouveaux films en fonction de leurs vecteurs de caractéristiques.

Dans le cadre du classement, la fonction de perte utilisée est généralement L_{0-1} (existence d'une différence entre la classe retournée par l'hypothèse et celle retournée par la fonction cible pour une entrée donnée).

Le terme « classification » est souvent utilisé à la place de « classement » dans la littérature francophone relative à l'apprentissage automatique (par exemple [45], [46], [47]).

Le mot français « classification » exprime l'action de classifier, c'est-à-dire d'établir des classes[8]. Il correspond ainsi à l'anglais "clustering". Le mot anglais "classification", quant à lui, est un faux ami qui se traduit en français par « classement ». Le mot « classement » exprime l'action de classer, c'est-à-dire de « distribuer par classes » [48], ce qui correspond à la tâche visée ici.

Pour marquer la différence avec la classification au sens du terme anglais "clustering", il est alors courant de distinguer la « classification supervisée », qui correspond au terme anglais "classification", de la « classification non supervisée », qui correspond au terme anglais "clustering". Cette pratique ne paraît toutefois pas la plus opportune car au sein des problèmes dits de « classification non supervisée » ("clustering" en anglais), il en existe justement qui mettent en jeu des mécanismes de supervision ; le terme anglais correspondant à ces problèmes est d'ailleurs "supervised clustering". En poussant plus loin la logique, le terme français correspondant devrait être « classification non supervisée… supervisée ». Une pratique alternative consiste à traduire "clustering" par « regroupement de données » ou « partitionnement de données » ; ces termes se prêtent mieux à une déclinaison entre « supervisé » et « non supervisé ».

Afin de nous prémunir contre une éventuelle propagation à d'autres notions du décalage entre un terme et le sens du mot correspondant, nous prendrons le parti de désigner la tâche consistant à répartir des éléments dans des classes par le terme « classement » et nous réserverons le terme « classification » pour la tâche consistant à élaborer des classes (dans ce sens, voir [49], [50]).

XI : Problèmes de classification

La classification (parfois appelée « regroupement de données », « partitionnement de données », ou "clustering" en anglais) consiste à élaborer des sous-ensembles homogènes à partir d'un ensemble d'éléments, c'est-à-dire des sous-ensembles tels que, étant donné une mesure de similarité quelconque, un élément d'un sous-ensemble soit plus proche des autres éléments de ce sous-ensemble que des éléments d'un autre sous-ensemble [51], [52].

---

[8] « Emprunté d'une forme latine factice *classificare*, composé de *classis*, 'classe', et *facere*, 'faire'.» [148]

Dans la mesure où le terme « classification » est susceptible d'induire en erreur un lecteur habitué à la littérature anglophone, il serait tentant de lui préférer les termes « regroupement de données » ou « partitionnement de données ». Cependant, ces deux termes ne sont pas aussi neutres que le terme « classification », lequel se borne à évoquer la tâche consistant à élaborer des classes : ils évoquent aussi la manière dont ces classes sont élaborées et pourraient à ce titre être apparentés à des méthodes de classification. Ces méthodes relèvent de la classification hiérarchique [53] : le regroupement de données s'apparenterait à la méthode de classification ascendante hiérarchique ("agglomerative clustering"), consistant à partir des éléments pour arriver aux classes par agglomérations successives) ; le partitionnement de données s'apparenterait à la méthode de classification descendante hiérarchique ("divisive clustering"), consistant à partir de l'ensemble des éléments pour arriver aux classes par subdivisions successives).

A : Classification non supervisée

Lorsqu'elle ne procède d'aucune autre information que celles présentes dans les données à classifier, alors la classification est dite non supervisée. La plupart des tâches de classification étant non supervisées, la classification est parfois considérée comme étant non supervisée par définition [54]. Un exemple de classification non supervisée est la classification en k-moyennes ([55] p. 451).

Pour reprendre l'illustration avec les films, un problème de classification non supervisée pourrait être le suivant : étant donné un ensemble de films, chacun représenté par un vecteur de caractéristiques, répartir ces films en 6 groupes homogènes.

B : Classification semi-supervisée

La classification est le plus souvent non supervisée, cependant rien n'oblige à baser une approche exclusivement sur l'information intrinsèque aux données à classifier : lorsqu'une part des informations qu'elle exploite est extrinsèque aux données à classifier, alors elle est dite semi-supervisée [51]. C'est par exemple le cas lorsque sont ajoutées des contraintes sous forme de paires d'éléments indiquant si ces éléments doivent ou non figurer dans un même groupe. La

classification peut être dans ce cas considérée comme « guidée » par le biais d'interactions avec un oracle [56][9].

La classification semi-supervisée peut s'illustrer comme la classification non supervisée, mais avec des contraintes supplémentaires telles que « les films A et B sont dans le même groupe » ou « les films B et C sont dans des groupes différents ».

C : Classification supervisée

Le terme « supervisé » est réservé à la classification couverte par la définition utilisée par Finley et Joachims [57]. Selon cette définition, « la classification supervisée est le problème consistant à entraîner un algorithme de classification de manière à ce qu'il produise le regroupement désiré.[10] »

Cette définition s'inscrit en effet dans le cadre classique de l'apprentissage supervisé : les données d'apprentissage sont formulées par des couples $\langle entrée, étiquette \rangle$, où l'entrée est ici un ensemble d'éléments et l'étiquette associée est une classification des éléments de cet ensemble, avec pour but de prédire, pour tout nouvel ensemble, une classification des éléments de cet ensemble qui soit cohérente avec les données d'apprentissage.

La classification supervisée peut être vue comme une généralisation du classement, un genre de « classement total » : recevant en entrée des jeux de données consistant chacun en un ensemble d'éléments répartis dans un ensemble de classes, la tâche consiste à retourner pour tout nouvel ensemble d'éléments un ensemble de classes dans lequel sont répartis les éléments. La classification supervisée ne consiste pas à apprendre seulement, à l'instar du classement, la manière de distribuer les éléments dans les classes ; elle consiste à apprendre aussi la manière d'élaborer ces classes.

---

[9] Bien que l'article cité s'intitule "Supervised Clustering", la classification est ici rangée dans la classification semi supervisée car elle ne consiste pas à apprendre une manière donnée de regrouper les données mais se contente d'effectuer un regroupement de données en respectant des contraintes externes. De la même façon, et ce encore malgré la taxonomie utilisée dans certains articles, nous dirons d'une classification qui se contente de respecter des classes préétablies [149], [150] qu'elle est semi-supervisée.

[10] "Supervised clustering is the problem of training a clustering algorithm to produce desirable clusterings."

La classification supervisée peut être illustrée par le problème suivant : étant donné un ensemble de collections de films, chaque film étant représenté par un vecteur de caractéristiques (un même film pouvant appartenir à plusieurs collections) et chaque collection étant partitionnée en différents groupes de films, de manière possiblement variable selon la collection (mais selon un même principe au regard des films composant cette collection), partitionner de nouvelles collections de films selon ce même principe.

XII : Synthèse

L'apprentissage automatique est le domaine d'étude portant sur l'élaboration de programmes informatiques capables de s'améliorer par l'expérience dans la réalisation de tâches déterminées. A ce titre, l'apprentissage automatique peut être vu comme un mode indirect de programmation, c'est-à-dire permettant d'élaborer des programmes sans avoir à les coder explicitement.

L'apprentissage automatique consiste en une recherche automatique de motifs (au sens de régularités) présents dans les données et en une application de ces motifs à des données considérées similaires à celles dans lesquelles ces motifs ont été découverts (mécanisme de l'induction).

Le biais d'apprentissage est un ensemble d'hypothèses portant sur les motifs présents dans les données et permettant de généraliser à de nouvelles données ; sans biais d'apprentissage, les données elles-mêmes sont susceptibles d'être considérées comme des motifs et l'apprentissage se confond avec une simple mémorisation des données, ce qui le rend inexploitable, dès lors que les données sont différentes, pour permettre d'effectuer des prédictions.

Chapitre 2 : Etat de l'art de l'apprentissage automatique des données ordinales

L'apprentissage des données ordinales est la tâche d'apprentissage automatique dans laquelle les variables à prédire prennent des valeurs ordinales [43], [58]. Cette tâche est devenue incontournable dans un contexte de surabondance des données où il est nécessaire de filtrer automatiquement l'information selon sa pertinence [59]. A ce titre, il est éloquent que l'une des plus grandes entreprises du monde, Google, ait bâti son succès sur un algorithme d'apprentissage ordinal, le PageRank [60].

L'apprentissage des données ordinales peut se décliner en deux types de problèmes. Le premier consiste à distribuer les éléments dans des classes ordonnées. Le second consiste à établir un ordre entre les éléments. Nous appelons le premier type « classement ordinal » (comme dans [61] : "ordinal classification") et le second « rangement » (comme dans [62] : "ranking").

Le terme « classement ordinal » est parfois considéré comme équivalent au terme « régression ordinale » (voir par exemple [63], [64]), lui-même d'ailleurs parfois considéré comme équivalent à « rangement » (voir par exemple [58], [65]). Nous préférons limiter le terme « régression ordinale » aux méthodes de régression pour les valeurs ordinales [66], lesquelles peuvent être appliquées tant aux problèmes de classement ordinal qu'aux problèmes de rangement.

I : Problèmes de classement ordinal

Le classement ordinal peut être illustré par le problème suivant : étant donné un ensemble de films, chacun représenté par un vecteur de caractéristiques et associé à un nombre d'étoiles (de 0 et 5) selon le goût d'un spectateur donné, déterminer le nombre d'étoiles qu'attribuerait ce spectateur à de nouveaux films en fonction de leurs vecteurs de caractéristiques.

A : Définition

Les problèmes de classement ordinal sont des problèmes de classement où les classes dans lesquelles les éléments doivent être distribués sont ordonnées. En d'autres termes, le classement ordinal est la tâche d'apprentissage automatique qui consiste à apprendre des classes ordonnées [61]. Dans l'exemple ci-dessus, les classes correspondent aux nombres possibles de résidents.

Les algorithmes de classement traditionnels ne prennent pas en compte les possibles informations ordinales contenues dans les données [61]. Dans le classement multi-classes, cette absence de prise en compte est due à l'utilisation d'une fonction coût non différenciée : soit des classes ordonnées C_i, $C_i \prec_p C_{i+1}$ pour une préférence donnée p, alors une erreur de classement entre les classes C_i et C_{i+2} devrait être en quelque sorte plus pénalisée qu'une erreur de classement entre C_i et C_{i+1}.

B : Méthodes de classement ordinal

Plusieurs méthodes ont été proposées pour résoudre les problèmes de classement ordinal en adaptant ou en étendant des algorithmes traditionnels.

1 : Adaptation

La méthode de réplication de données [61], [67] permet d'utiliser des classeurs binaires pour résoudre les problèmes de classement ordinal. Le principe consiste à apprendre une fonction de distribution cumulative discrète : pour toutes les classes C_k à part la dernière, un classeur binaire est entraîné à décider si un élément appartient à l'union $\bigcup_{i=1}^{k} C_i$. Au stade de la prédiction, un élément est considéré appartenir à la classe C_k si les classeurs le placent dans l'union $\bigcup_{i=1}^{k} C_i$ mais pas dans l'union $\bigcup_{i=1}^{k-1} C_i$.

2 : Extension

Une variable ordinale peut être vue comme la manifestation discrète d'une variable continue latente : l'apprentissage de classes ordonnées consiste alors à ajuster à la fois une fonction continue et un ensemble de seuils représentant les frontières entre les classes (un seuil par paire de classes consécutives) [40], [68], [69].

Dans le classement binaire, la régression consiste à ajuster une fonction sigmoïde aux données ; par exemple, la régression logistique consiste à ajuster la fonction logistique. Ce principe a été étendu au classement ordinal [40] ; ainsi, le *logit ordonné* [69] est une extension de la *régression logistique*. Une autre méthode de classification ordinale est le *probit ordonné* [68], qui est similaire au *logit ordonné*, mais avec la fonction de de répartition de la loi normale au lieu de la fonction logistique.

Les méthodes de régression pour l'apprentissage ordinal impliquent une mise en correspondance de la fonction ordinale à apprendre avec une fonction continue latente, ce qui, analysé en termes de variables aléatoires, revient à sous-tendre des variables ordinales par des variables continues dont la distribution est inconnue et doit être conjecturée [40]. Or, si les données à apprendre sont véritablement ordinales, elles ne peuvent par définition être représentées par des données métriques et il n'existe par conséquent aucun principe permettant de conjecturer la distribution sous-jacente [61].

Une alternative à la régression consiste à appliquer le *principe de la vaste marge* à la classification ordinale. Cette approche permet de s'affranchir d'un recours à une distribution sous-jacente et d'éviter ainsi de devoir faire des hypothèses sur cette distribution [43].

Les SVM (Séparateurs à Vaste Marge, "Support-Vector Machines" en anglais) [70] peuvent être étendus au classement ordinal. Le principe consiste à apprendre des hyperplans séparateurs parallèles entre des classes ordonnées consécutives [43]. L'ensemble des hyperplans est choisi de manière à maximiser la marge de l'hyperplan dont la marge est la plus fine. Comme les hyperplans sont parallèles, les informations ordinales contenues dans les classes sont préservées. Il est aussi envisageable de maximiser la somme des marges [58]

C : Métriques pour le classement ordinal

Les résultats d'un classement ordinal sont souvent évalués à l'aide des métriques dédiées au classement nominal [71] : une erreur de classification est pénalisée de manière identique, que les classes entre lesquelles elle survient soient consécutives ou non, ignorant par conséquent les relations ordinales entre les classes.

Afin de prendre en compte ces relations ordinales, des métriques issues de tâches de régression sont parfois utilisées, telles que l'erreur quadratique moyenne ou l'erreur absolue moyenne [71]. Dans ce cas, cependant, une échelle éventuellement arbitraire est introduite dans le processus d'évaluation, ce qui contredit la nature ordinale des données.

Certaines métriques dédiées ont néanmoins été proposées pour la classification ordinale, comme une généralisation de la surface sous la courbe ROC (de l'anglais "Receiver Operating Characteristic", pour « caractéristiques de fonctionnement du récepteur ») [72] ou un coefficient d'erreur basé sur la matrice de confusion qui appréhende la distance entre les prédictions réelle et idéale [73].

II : Problèmes de rangement

Pour garder l'illustration avec les films, alors que les tâches de classement consistent à déterminer les *classes des films*, les tâches de rangement consistent à déterminer un *ordre entre les films*.

A : Définition

Le rangement est la tâche d'apprentissage automatique qui consiste à apprendre à ranger des éléments par ordre de préférence [74]–[76]. Il correspond au terme anglais "ranking"[11].

B : Catégories de rangements

Fürnkranz et Hüllermeier [62] distinguent trois catégories de problèmes de rangements : les problèmes de rangement d'objets, de rangement d'instances et de rangement de labels.

1 : Rangement de labels

Dans un problème de rangement de labels, étant donné un ensemble de labels et un ensemble d'éléments, chaque élément étant associé à un ordre partiel sur l'ensemble des labels, il s'agit d'induire une fonction prenant un élément quelconque et retournant une permutation sur les labels (voir [77], [78]).

[11] Pour un exemple d'utilisation du terme « rangement » comme traduction de "ranking", voir [151].

Par exemple, un système de recommandations de films prend en entrée un ensemble d'utilisateurs (chacun représenté par un vecteur de caractéristiques telles que l'âge, le genre, etc.), un ensemble de films avec, pour chaque utilisateur, un ordre de préférence entre ces films ; sa tâche est d'induire une fonction permettant de prédire la manière dont de nouveaux utilisateurs trieraient ces mêmes films afin de pouvoir leur recommander les films prédits en meilleure position.

2 : Rangement d'instances

Dans un problème de rangement d'instances, il s'agit d'induire, à partir d'un ensemble d'éléments appartenant à des classes ordonnées, une fonction prenant un ensemble d'éléments et retournant un ordre total sur cet ensemble (voir [79]).

Par exemple, si des films sont représentés par des vecteurs de caractéristiques et qu'un utilisateur les évalue en leur attribuant entre 0 et 5 étoiles, le rangement d'instances consiste à induire une fonction prenant un ensemble de films et les ordonnant comme le ferait cet utilisateur.

3 : Rangement d'objets

Dans un problème de rangement d'objets, il s'agit toujours d'induire une telle fonction, mais cette fois à partir d'un ensemble d'éléments partiellement ordonnés.

Un exemple de rangement d'objets peut être identique à celui de rangement d'instances, à la différence qu'un film ne serait pas noté avec des étoiles mais comparé à un ou plusieurs autres films pour former un ordre partiel (« le film A est préféré au film B », « le film C est préféré aux films D et E », etc.).

C : Méthodes de rangement

Il y a deux familles de méthodes de rangement : les méthodes qui utilisent une fonction d'utilité et les méthodes basées sur la comparaison des paires d'éléments. La première méthode considère les différents éléments individuellement ; la seconde considère les relations entre les éléments.

1 : Rangement par fonction d'utilité

Le rangement par fonction d'utilité consiste à apprendre une fonction d'utilité sur l'ensemble d'apprentissage et à trier les éléments de l'ensemble de test selon leurs images par la fonction d'utilité apprise afin d'expliciter un ordre total sur l'ensemble de test.

Comme pour le classement ordinal, le recours à une fonction sous-jacente nécessite de faire des hypothèses sur cette fonction [61], [40].

2 : Rangement par comparaisons sur les paires d'éléments

Le classement par comparaisons sur les paires d'éléments ne nécessite pas de faire de telles hypothèses.

Cette méthode consiste à apprendre directement une relation binaire à partir des couples d'éléments de l'ensemble d'apprentissage. Cette relation permet d'orienter les paires d'éléments de l'ensemble de test puis de construire un ordre total à partir des couples ainsi obtenus.

La relation peut être transitive ou non en fonction de la méthode utilisée pour l'apprendre.

a : Rangement en deux temps

Lorsque la relation binaire apprise n'est pas transitive (c'est-à-dire lorsqu'il existe trois éléments distincts x_1, x_2, x_3 tels que $x_1 \geq_p x_2$, $x_2 \geq_p x_3$ et $x_3 \geq_p x_1$, où $\geq_p$ représente la relation binaire), il convient de choisir l'ordre total le plus proche de la relation binaire apprise [79]. Puisque ce choix implique de résoudre le problème de Slater [80], lequel est NP-équivalent [81], un moyen d'approximer le bon ordre est nécessaire [79], [82].

b : Rangement en un temps

Il est aussi possible de réduire l'espace des hypothèses afin de n'apprendre que des relations d'ordres.

Bien qu'elle adopte un biais d'apprentissage adapté au classement ordinal, l'extension des SVM au rangement [43] peut représenter un moyen d'apprendre directement une relation d'ordre à partir d'un ensemble de couples d'éléments. En effet, puisque tous les hyperplans séparateurs sont parallèles, la relation d'ordre peut être représentée par tout vecteur normal à ces

hyperplans : la projection des éléments sur ce vecteur est réflexive, antisymétrique et transitive, et, dans la mesure où les éléments respectent les contraintes sur les couples explicitées par les hyperplans, leurs projetés sur un vecteur normal à ces hyperplans les respectent aussi.

D : Métriques pour le rangement

Les résultats d'un rangement d'instances peuvent être évalués en comptant le nombre de couples d'éléments contredisant les contraintes définies par les classes (généralement lorsque deux éléments obéissent à un ordre donné alors que leurs classes respectives obéissent à l'ordre inverse) [62], [72]. Cette métrique permet de présenter d'un autre point de vue la différence entre le classement ordinal (où il s'agit de minimiser l'erreur de classe) et le rangement d'instances (où il s'agit de minimiser l'erreur de rang[12]) : dans le cadre du classement ordinal, toutes les erreurs de classes sont pénalisées alors que dans le cadre du rangement d'instances, les erreurs de classes ne sont pénalisées que dans la mesure où elles dénotent une erreur sur l'ordre.

Le rangement d'objets et le rangement de labels peuvent être évalués en calculant leur corrélation avec le rangement attendu. Les coefficients de Kendall tau et de Spearman sont alors souvent utilisés [83]. Des métriques dédiées aux tâches de régression, telles que l'erreur quadratique moyenne ou l'erreur absolue moyenne, peuvent également être utilisées [13].

Dans un contexte de recherche d'informations, des métriques mettant l'accent sur les rangs les plus élevés ont été proposées, à l'instar du gain cumulatif actualisé normalisé [84], [85], [86]. Un exemple typique est celui des moteurs de recherche [87], qui affichent les résultats de la recherche par ordre de pertinence : la métrique utilisée doit prendre en compte le fait qu'il est moins grave d'avoir des erreurs sur l'ordre lorsqu'elles apparaissent à la dixième page que lorsqu'elles apparaissent à la première [88].

---

[12] "In ranking, one cares more about accurate ordering of objects, while in ordinal classification, one cares more about accurate ordered-categorization of objects" [63]

III : Problèmes de classification ordinale

La classification ordinale [89]–[91] consiste à élaborer des classes pour des données ordinales, c'est-à-dire des classes qui soient compatibles avec l'ordre organisant ces données.

La classification ordinale peut s'illustrer de la même manière que la classification non ordinale (qu'elle soit non supervisée, semi-supervisée ou supervisée), à la différence que les groupes qui sont formés entretiennent une relation d'ordre.

La classification ordinale ne se confond pas avec la classification hiérarchique ("hierarchical clustering") [53], qu'il s'agisse de la classification ascendante hiérarchique ("agglomerative clustering") ou de la classification descendante hiérarchique ("divisive clustering") : le terme « classification ordinale » désigne un type de problèmes tandis que le terme « classification hiérarchique » désigne une méthode de classification où les classes sont organisées en arborescence (par regroupements successifs des éléments pour la classification ascendante ou par partitionnements successifs de l'ensemble des éléments pour la classification descendante).

IV : Synthèse

L'apprentissage ordinal couvre les problèmes d'apprentissage dans lesquels les variables à apprendre sont de nature ordinale.

Selon le problème à résoudre, l'apprentissage ordinal peut être décomposé en classement ordinal, rangement et classification ordinale. Le classement ordinal ressemble au classement nominal traditionnel, à la différence que les classes dans lesquelles distribuer les éléments sont ordonnées ; le rangement consiste à apprendre un ordre total sur un ensemble d'éléments ; la classification ordinale ressemble à la classification nominale traditionnelle, à la différence que les classes à former sont ordonnées.

Chapitre 3 : Catégories hybrides de problèmes d'apprentissage automatique des données ordinales

Dans ce chapitre, nous proposons une nouvelle approche des problèmes d'apprentissage ordinal.

Si les problèmes d'apprentissage automatique des données ordinales sont abordés en considérant non seulement la forme des informations en jeu mais aussi leur nature, il est possible d'en avoir une vue plus fine. Cela permet de se rendre compte qu'il n'existe pas de méthode spécifique au rangement d'objets. En outre, cela permet de découvrir des catégories hybrides de problèmes, telle celle des problèmes implicites de rangement d'objets.

I : Approche de l'apprentissage automatique des données ordinales tenant compte de la nature des informations

Nous reconsidérons les problèmes d'apprentissage automatique des données ordinales sous un angle nouveau : non plus selon leur seule formulation mais en tenant compte aussi de la nature des informations en jeu. Ainsi, nous verrons qu'il ne faut pas se fier à la formulation d'un problème d'apprentissage pour distinguer un problème de classement ordinal d'un problème de rangement, ou un problème de rangement d'objets d'un problème de rangement d'instances, mais à la nature des informations que renferment les données sur lesquelles portent ces problèmes.

A : Insuffisance d'une identification des problèmes d'apprentissage fondée sur leurs formulations

Les problèmes d'apprentissage ordinal sont traditionnellement identifiés par leur formulation. Le Tableau 2 rappelle les définitions traditionnelles des problèmes de classement ordinal, de rangement d'instances et de rangement d'objets selon leur formulation. Cependant, si ces problèmes se distinguaient par leurs seules formulations, alors il serait possible de résoudre un problème en termes d'un autre.

Ainsi, il serait possible de résoudre un problème de classement ordinal en résolvant un problème de rangement d'instances sur les mêmes données d'apprentissage et en partitionnant, à l'aide des informations sur les classes issues des données d'apprentissage, l'ordre total prédit par le rangement d'instance.

De même, il serait possible de résoudre un problème de rangement d'instances en détaillant les couples d'éléments de l'ordre partiel à partir des classes ordonnées, de manière à expliciter l'ordre partiel exprimé par les classes, puis en résolvant le problème de rangement d'objets ainsi obtenu.

En combinant les deux procédés, il serait même possible de résoudre un problème de classement ordinal en résolvant un problème de rangement d'objets.

Enfin, il serait possible de résoudre un problème de rangement d'objets en élaborant des classes exprimant les mêmes contraintes que celles exprimées par les couples d'objets, puis en résolvant le problème de rangement d'instances ainsi obtenu. Une manière d'élaborer ces classes consiste à construire un graphe orienté à partir des couples d'éléments (chaque couple étant représenté par un arc), puis à placer chaque élément dans la classe dont le numéro correspond à la longueur du plus long chemin existant vers cet élément. Cette manière garantit un nombre minimal de classes puisque le nombre de classes est égal à la longueur du plus long chemin du graphe plus un (pour la classe 0).

La formulation des problèmes d'apprentissage ordinal ne paraît en conséquence pas un critère suffisant pour identifier les différents problèmes d'apprentissage ordinal.

Tableau 2 : Définition traditionnelle des problèmes de classement ordinal, de rangement d'instances et de rangement d'objets selon leur formulation.

	Forme des données d'apprentissage	*Forme de la prédiction*
Classement ordinal	Classes ordonnées	Classes ordonnées
Rangement d'instances	Classes ordonnées	Ordre total
Rangement d'objets	Ordre partiel	Ordre total

B : Découplage entre forme et nature de l'information ordinale

La forme des informations ordinales ne coïncide pas nécessairement avec leur nature (i.e. ce à quoi elles sont relatives).

Si les informations relatives aux classes ne peuvent être disponibles qu'au niveau des classes (i.e. sous forme de classes), les informations relatives aux rangs peuvent être disponibles tant au niveau des rangs (i.e. sous forme de couples d'éléments) qu'au niveau des classes. En effet, si des rangs différents n'impliquent pas des classes différentes, des classes différentes peuvent impliquer des rangs différents ; plus précisément, elles les impliquent dans le cas où les éléments de classes différentes sont comparables entre eux. En outre, dès lors qu'il est possible de dériver des rangs à partir de classes, rien n'interdit d'avoir des classes dont la seule raison d'être est de spécifier des rangs ; dans ce cas, l'information disponible au niveau des classes est exclusivement relative aux rangs.

Le Tableau 3 récapitule les différents cas de figure possibles. Rien n'interdirait d'avoir des informations à la fois au niveau des classes et des rangs, mais ce n'est pas considéré ici dans la mesure où cela ne correspond pas aux problèmes traditionnels d'apprentissage des données ordinales.

Tableau 3 : Configurations possibles de l'information selon son niveau de disponibilité et sa nature (ce sur quoi porte l'information).

		Nature de l'information		
		Classes seules	**Classes et rangs**	**Rangs seuls**
Niveau de disponibilité de l'information	**Classes**	**X**	**X**	**X**
	Rangs	-	-	**X**

Dans le cas où les classes ne contiennent que des informations ordinales relatives aux classes, alors les éléments de classes différentes ne sont pas comparables entre eux (sinon les classes contiendraient nécessairement des informations relatives aux rangs). Le contenu de chaque classe peut être ordonné, mais dans ce cas, les ordres internes aux différentes classes ne peuvent pas former un ordre cohérent. En d'autres termes, les classes peuvent être ordonnées et les éléments peuvent être ordonnés à l'intérieur de chaque classe, mais il n'existe pas d'ordre total sur l'ensemble des éléments qui soit cohérent avec les classes.

Par exemple, si les éléments représentent les jours d'une année et les classes en représentent les saisons sous un climat tempéré, elles peuvent être triées par température (en moyenne, froid pour l'hiver, intermédiaire pour le printemps et l'automne, chaud pour l'été). Leurs éléments peuvent également être triés par température. Mais aucun ordre total ne peut être déduit des classes car rien n'empêche un élément d'une saison donnée d'avoir une température supérieure à celle d'un élément de la saison consécutive (au sens de la séquence des saisons selon leurs températures moyennes). Néanmoins, les classes existent en elles-mêmes : leur contenu peut changer sans modifier leurs limites. Un exemple similaire serait celui des étudiants se trouvant dans différents niveaux.

Un exemple de cas où les classes contiennent des informations relatives à la fois aux classes et aux rangs pourrait être celui où les éléments représentent des échantillons d'eau et les classes représentent l'état de l'eau dans des conditions de pression constantes. Les éléments peuvent être triés par températures, de même que les classes : l'état solide se trouve à des températures inférieures à l'état liquide, lui-même à des températures inférieures à l'état gazeux. Un ordre total peut être dérivé puisque tout élément d'un état donné est d'une

température inférieure à tout élément de l'état consécutif. Néanmoins, les classes existent en elles-mêmes : leur contenu peut changer sans modifier leurs limites. Un exemple similaire serait celui des fréquences du spectre visible réparties dans des classes représentant les différentes couleurs.

Un exemple de cas où les classes ne contiennent que des informations ordinales relatives aux rangs serait celui où des individus seraient répartis selon leurs poids en trois classes de même taille (à plus ou moins un individu). Les classes n'existent pas en elles-mêmes : leurs limites sont susceptibles d'être modifiées dès que leur contenu change.

C : Redéfinition des problèmes d'apprentissage ordinal en fonction de la nature de l'information

Nous proposons de redéfinir les notions de classement ordinal, de rangement d'instances et de rangement d'objets en fonction non plus de leur formulation mais de la nature des informations en jeu : informations relatives au rang ou à la classe. En substance, le rangement exploite les informations relatives aux rangs des éléments tandis que le classement exploite les informations relatives aux classes des éléments.

Plus précisément, lorsque l'information ordinale n'est relative qu'aux classes, alors il s'agit d'un problème de classement ordinal. Lorsque l'information est relative aux classes et aux rangs, alors il s'agit d'un problème de rangement d'instances. Lorsque l'information n'est relative qu'aux rangs, alors il s'agit d'un problème de rangement d'objets, peu important que l'information se trouve au niveau des rangs ou au niveau des classes.

Cette redéfinition permet notamment de tenir compte du fait que l'information se trouvant au niveau des classes peut tout à fait être exclusivement relative aux rangs. Le Tableau 4 synthétise la définition des problèmes d'apprentissage ordinal selon la nature des informations en jeu.

Tableau 4 : Redéfinition des problèmes d'apprentissage ordinal selon la nature des informations en jeu.

	Information relative aux classes	Information relative aux rangs
Classement ordinal	**X**	-
Rangement d'instances	**X**	**X**
Rangement d'objets	-	**X**

Ces définitions privilégient la nature du problème sur sa présentation, tenant ainsi mieux compte du biais d'apprentissage. Si aucune information relative à la classe n'est disponible, rien ne distingue le rangement d'instances du rangement d'objets. En effet, la seule information utilisée pour rechercher un rangement est alors exactement la même, à savoir un ensemble de couples d'éléments, bien qu'elle soit exprimée de manière différente : une liste explicite de couples pour le rangement d'objets ; une liste compactée, à travers le formalisme des classes, mais développable en une liste explicite, pour le rangement d'instances). Contrairement au rangement d'objets, le rangement d'instances suggère l'idée de classes, d'où le terme « instances » ; son biais inductif a tendance à accentuer une structuration des éléments en classes alors que celui du rangement d'objets a tendance empêcher une telle structuration. Dans les problèmes de rangement d'instances, les classes initiales ont une signification intrinsèque, c'est-à-dire qu'elles existent pour elles-mêmes, indépendamment des éléments qu'elles contiennent. Dans les problèmes de rangement d'objets, les éventuelles classes initiales n'ont aucune signification intrinsèque, c'est-à-dire qu'elles sont formées par des contraintes extrinsèques et dépendent des éléments qu'elles contiennent.

D : Absence d'algorithme de rangement d'objets

Nous avons vu au chapitre précédent que les méthodes de rangement par comparaison sur les paires d'éléments sont préférables aux méthodes de rangement par fonction d'utilité du fait qu'elles ne nécessitent pas de faire d'hypothèse sur une quelconque fonction sous-jacente. Cependant, il n'existe pas de méthode de rangement par comparaison sur les paires d'éléments

qui soit spécifique au rangement d'objets tel qu'il a été redéfini selon la nature des informations en jeu.

Le Tableau 5 présente les méthodes de rangement par comparaison sur les paires d'éléments en fonction du type de rangement.

D'une part, la méthode de rangement en un temps par extension des SVM sélectionne l'ordre correspondant au vecteur résultant de l'ensemble des hyperplans séparateurs parallèles qui maximise la marge la plus mince : l'accent est mis sur les classes en ce sens que l'ordre sélectionné dépend de la configuration de ces classes. Cette méthode implémente donc un biais d'apprentissage de rangement d'instances. L'application de cette méthode à la résolution d'un problème de rangement d'objets aurait pour effet de mettre l'accent sur les relations données en entrée.

D'autre part, les méthodes de rangement en deux temps permettent de respecter les contraintes sur les paires d'éléments données en entrée (qu'elles soient données directement sous forme de couples ou qu'elles soient dérivées des classes) sans nécessairement les mettre davantage en évidence. A ce titre, elles pourraient être qualifiées « d'agnostiques ». Elles ne garantissent toutefois pas de ne pas retourner un ordre total mettant l'accent sur les classes : en ce sens, leur biais d'apprentissage est plus faible que ne le serait celui d'une méthode spécifique au rangement d'objets.

Une méthode spécifique au rangement d'objets serait une méthode qui retournerait l'ordre qui, tout en respectant les contraintes données en entrée, les lisserait le plus.

Tableau 5 : Méthodes de rangement par comparaison sur les paires d'éléments en fonction du type de rangement. Il n'existe pas de méthode spécifique au rangement d'objets.

Type de rangement	*Méthode de résolution par comparaison sur les paires d'éléments*
Rangement agnostique	Méthode en deux temps
Rangement d'instances	SVM à hyperplans séparateurs parallèles
Rangement d'objets	-

II : Problèmes implicites de rangement

Parfois, si deux éléments appartiennent à une même classe, cela ne signifie pas qu'ils soient ordinalement équivalents ; cela signifie simplement qu'aucune information explicite sur leur relation ordinale n'est disponible.

Supposons que l'on nous donne un ensemble d'éléments, chacun étiqueté par un entier compris entre 0 et 5, et que nous devions prédire les étiquettes de nouveaux éléments. Cela pourrait être vu comme un problème de classement et, si les entiers expriment un ordre entre les classes, comme un problème de classement ordinal.

Supposons maintenant qu'un élément soit noté par une personne selon ses préférences en lui attribuant de 0 à 5 étoiles. Si nous présentons à cette personne sept éléments, au moins deux d'entre eux se verront attribuer le même nombre d'étoiles. La raison a cependant moins à voir avec les préférences de cette personne qu'avec les contraintes externes que nous avons imposées sur sa façon de les exprimer ; ce problème n'est donc pas un problème de classement ordinal mais un problème implicite de rangement.

A : Définition

Un problème implicite de rangement est un problème de rangement formulé en termes de problème de classement.

Il n'existe pas de différence formelle entre un problème de classement ordinal et un problème implicite de rangement ; ce dernier peut d'ailleurs être résolu en tant que problème de classement ordinal. Le biais d'apprentissage est néanmoins différent : dans le classement, les éléments partageant une classe peuvent obéir à n'importe quelle relation, tandis que dans le rangement implicite, ces éléments obéissent à une relation d'ordre, que celle-ci soit connue ou non.

La principale différence réside dans le fait que la relation ordinale se situe au niveau de la classe de l'élément (classement) ou au niveau du rang de l'élément (rangement). Si la relation ordinale entre deux éléments est relative à leurs rangs tout en se trouvant au niveau de leurs classes,

c'est-à-dire exprimée sous forme d'appartenance des éléments à des classes différentes, alors il s'agit de rangement implicite.

Le Tableau 6 indique la possibilité d'existence d'un problème en fonction à la fois de la nature des données en jeu et de la forme du problème. Sur la diagonale descendante se trouvent les problèmes traditionnels dans leur forme non altérée : l'énoncé du problème correspond à la nature des informations à traiter. En dessous de la diagonale se trouvent les problèmes implicites : un problème de rangement d'instances peut très bien se cacher derrière un problème de classement ordinal ; de même, un problème de rangement d'objets peut très bien se cacher derrière un problème de classement ordinal, voire même derrière un problème de rangement d'instances.

Tableau 6 : Possibilité d'existence d'un problème ordinal en fonction de la nature et de la forme du problème.

		Forme du problème		
		Classement ordinal	**Rangement d'instances**	**Rangement d'objets**
Nature du problème	**Classement ordinal**	**X**	-	-
	Rangement d'instances	**X**	**X**	-
	Rangement d'objets	**X**	**X**	**X**

Le biais d'apprentissage d'un problème de rangement étant plus fort que celui d'un simple problème de classement ordinal, il est préférable de résoudre les problèmes implicites de rangement en tant que problèmes de rangement plutôt qu'en tant que problèmes de classement ordinal.

Un argument combinatoire peut appuyer cette remarque : le rangement suppose l'existence d'un ordre total sur l'ensemble des éléments alors que le classement ne le suppose pas ; les éléments d'une classe peuvent donc former n'importe quel tournoi dans un problème de classement alors qu'ils ne peuvent former qu'un ordre total dans un problème de rangement.

Supposons qu'une classe contienne n éléments. Posons U_n le nombre de tournois de cette classe et V_n le nombre d'ordres totaux de cette classe et calculons les termes à $n+1$. Dans le

cas du tournois, ajouter un élément revient à ajouter n arêtes, avec pour chacune 2 sens possibles, ce qui revient à multiplier le nombre de tournois à l'ordre n par 2^n : $U_{n+1} = 2^n U_n$. Dans le cas de l'ordre total, ajouter un élément revient à l'insérer dans un ordre total de n éléments, c'est-à-dire à $n+1$ places possibles, ce qui revient à multiplier le nombre d'ordres totaux à l'ordre n par $n+1$: $V_{n+1} = (n+1)V_n$. Le coefficient du premier croît exponentiellement alors que celui du second croît linéairement ; le rapport $\frac{U_n}{V_n}$ croît donc exponentiellement avec le nombre d'éléments.

B : Méthode de résolution

Pour pouvoir résoudre un problème de classement ordinal comme un problème de rangement, il faut le reformuler en termes de problème de rangement. En outre, le problème n'étant plus le même, la question des métriques à utiliser peut être reconsidérée.

1 : Reformulation

Les relations entre classes peuvent être extraites et utilisées pour apprendre une fonction de rangement. Au stade de la prédiction, les éléments du jeu d'apprentissage situés aux extrémités des classes servent de délimiteurs pour l'attribution des classes.

Un problème peut survenir si la fonction de rangement place un élément entre les limites de deux classes consécutives. Si les classes transmettent plus d'informations que le seul rangement des éléments, c'est-à-dire si elles transmettent à la fois des informations ordinales sur les éléments et des informations sur des classes, ces informations supplémentaires peuvent être utilisées pour déterminer la classe d'un élément (par exemple en combinant rangement et classement). Dans le cas contraire, le choix de l'une des deux classes peut être considéré comme étant déterminé au hasard.

2 : Métriques pour les problèmes implicites de rangement

Les problèmes implicites de rangement se situant entre classement et rangement, tant les métriques du classement que celles du rangement semblent pouvoir être utilisées.

Lorsque les informations véhiculées par les classes débordent les seules informations ordinales relatives aux rangs et comprennent également des informations relatives aux classes, les métriques du classement peuvent être utilisées.

Lorsque seules les informations ordinales relatives aux rangs sont rendues disponibles par la structure des classes, les métriques du classement paraissent sévères car les éléments rangés entre les éléments extrêmes de classes consécutives ont une probabilité de moitié d'être rattachés à la mauvaise classe.

Evaluer les résultats en utilisant le τ-b de Kendall [83] ne semble pas plus satisfaisant car le τ-b de Kendall considère les éléments partageant une classe comme ordinalement équivalents, or seules les contradictions entre rangs et classes doivent être pénalisées, non les éléments qui partageraient une classe sans être pour autant égaux.

Une façon acceptable consisterait à ne pénaliser que les éléments qui contrediraient les ordres compatibles avec les classes, c'est-à-dire les ordres dont les couples explicites correspondent aux frontières entre classes consécutives.

C : Problèmes Implicites de Rangement d'Objets (PIRO)

Nous appelons « Problème Implicite de Rangement d'objets » (PIRO) un problème de rangement d'objets exprimé sous la forme d'un problème de classement ordinal. Un PIRO a l'apparence d'un problème de classement ordinal : les données de formation consistent en un ensemble d'éléments répartis dans des classes ordonnées et la tâche consiste à répartir les nouveaux éléments de ces classes de manière cohérente avec les données de formation. Cependant, un PIRO n'est pas un problème de classement ordinal en ce sens que si deux éléments appartiennent à une même classe, cela ne signifie pas qu'ils soient ordinalement égaux ; cela signifie simplement qu'aucune information explicite sur leur relation ordinale n'est disponible. Dans un PIRO, les classes n'ont pas de signification intrinsèque ; ce sont des contraintes externes destinées à capturer la tendance d'un ordre total en l'exprimant de manière compacte, mais le nombre de classes, de même que leurs limites, peuvent changer de manière arbitraire sans avoir aucun impact sur l'information ordinale : les relations ordinales entre objets restent les mêmes.

De même, nous pourrions appeler « Problème Implicite de Rangement d'Instances » (PIRI) un problème de rangement d'instances exprimé sous la forme d'un problème de classement ordinal : dans un PIRO, les limites des classes sont susceptibles de changer avec les éléments de ces classes tandis que dans un PIRI, les frontières des classes ne changent jamais.

III : Synthèse

La prise en compte de la nature des informations ordinales permet une approche plus fine des problèmes d'apprentissage automatique des données ordinales que lorsque seule la forme des données est considérée.

Le rangement d'objets et le rangement d'instances ne se distinguent pas par la forme des données d'apprentissage (le rangement d'objets pour des données d'apprentissage sous forme d'un ensemble de couples d'éléments ; le rangement d'instances pour les données d'apprentissage sous forme de classes d'éléments) car il est toujours possible d'exprimer un ensemble de couples d'éléments par un ensemble de classes arbitraires d'éléments. Ils se distinguent en revanche par la nature des données d'apprentissage, selon que des classes existent en elles-mêmes ou non.

Les problèmes implicites de rangement sont des problèmes de rangement posés en termes de classement (les données d'apprentissage sont sous forme de classes d'éléments et les prédictions portent sur l'appartenance d'un élément à une classe) mais où il existe un ordre total sur l'ensemble des éléments. Le biais d'apprentissage du rangement étant plus fort que celui du classement ordinal, et *a fortiori* du classement nominal, les problèmes implicites de rangement sont résolus plus efficacement en tant que problèmes de rangement qu'en tant que problèmes de classement ordinal.

Un problème implicite de rangement d'objets (PIRO) est un problème de rangement implicite où les informations ordinales à exploiter sont exclusivement liées aux rangs. Contrairement à un problème de rangement d'instances, qu'il soit implicite (PIRI) ou non, l'ordre entre les éléments est indépendant des classes sous la forme desquelles sont donnés les couples d'éléments. Les classes étant arbitraires, leurs frontières doivent être considérées comme susceptibles de varier sans altérer l'ordre entre les éléments (invariance de l'ordre par rapport aux classes). Une

méthode de résolution des problèmes implicites de rangement d'objets doit être une méthode permettant de trouver un ordre variant aussi peu que possible lorsque les classes varient, comme si les frontières entre classes consécutives pouvaient glisser le long des éléments ordonnés de manière transparente pour l'ordre entre les éléments.

Partie 2 : Proposition d'une méthode spécifique au rangement d'objets

Chapitre 4 : Description d'une méthode de rangement agnostique directe

Nous appelons *méthodes de rangement agnostiques* les méthodes de rangement qui ne font aucune hypothèse sur ce que doit être le meilleur ordre parmi les ordres candidats et se contentent de retourner un ordre parmi les ordres compatibles avec les contraintes données en entrée, peu important lequel. Le biais d'apprentissage implémenté par ces méthodes est le plus faible des biais de rangement et se borne à supposer que les éléments puissent être totalement ordonnés.

Les méthodes de rangement par comparaisons sur les paires d'éléments procédant en deux temps (apprentissage d'un tournoi à partir des couples d'éléments donnés en entrée puis calcul de l'ordre total le plus proche du tournoi appris ; cf. partie 1, chapitre 3, section II-A) sont agnostiques par nature en ce sens que l'ordre total sélectionné résulte pour partie d'un processus NP-équivalent qui interdit en pratique tout choix entre ordres candidats. Cependant, il est tout à fait possible de définir une méthode de rangement par comparaisons sur les paires d'éléments procédant en un temps qui soit elle aussi agnostique.

La définition d'une telle méthode pourrait à première vue paraître redondante en ce qu'elle n'apporterait rien de nouveau par rapport aux méthodes procédant en deux temps. Toutefois, le fait qu'elle soit directe la rend apte à recevoir un biais d'apprentissage plus fort que celui du simple rangement agnostique. A ce titre, elle peut être vue comme une méthode de rangement générique à partir de laquelle il est possible de définir des méthodes de rangement plus spécifiques, telle notamment une méthode de rangement d'objets.

I : Description de l'espace de recherche

A : Principe

Suivant la distinction entre espace de recherche et stratégie de recherche dans un processus d'apprentissage [92], il existe deux manières principales de redéfinir la fonction de prédiction.

Soit l'espace de recherche est restreint de telle sorte que toute relation binaire trouvée dans cet espace soit un ordre total, soit des contraintes liées à l'ordre total sont ajoutées à la fonction de perte de sorte que la stratégie de recherche conduise à la sélection d'un ordre total parmi des relations binaires qui n'en sont pas nécessairement. Un mélange des deux pourrait aussi bien s'appliquer, par exemple en choisissant un espace de recherche prenant en compte la totalité et l'antisymétrie (ou l'asymétrie, selon qu'un ordre large ou strict est recherché, respectivement) et laissant la transitivité sous la responsabilité de la fonction de perte.

Adoptant la première manière, nous restreignons l'espace de recherche aux seuls ordres totaux et utilisons des prédicats binaires basés sur les propriétés de l'espace vectoriel afin de garantir l'apprentissage d'un ordre total.

B : Formalisation

Soit i un élément représenté par un point x_i dans un espace de description à n dimensions et (i, j) un couple d'éléments représenté par un vecteur $x_{i,j} = x_j - x_i$ reliant l'élément i à l'élément j. Il est alors possible de définir un graphe orienté dans l'espace de description à partir de l'ensemble des couples (i, j) donnés en entrée.

Nous apprenons un vecteur de poids w minimisant la variation du signe de $x_{i,j} \cdot w$ sur l'ensemble des vecteurs $x_{i,j}$ dérivés de l'ensemble des couples (i, j). Soit p^+ le nombre de couples (i, j) pour lesquels $x_{i,j} \cdot w$ est positif et p^- le nombre de couples (i, j) pour lesquels $x_{i,j} \cdot w$ est négatif : minimiser la variation du signe de $x_{i,j} \cdot w$ sur l'ensemble des $x_{i,j}$ équivaut à maximiser $|p^+ - p^-|$.

Gardons trace du signe le plus fréquent sur tous les $x_{i,j} \cdot w$ (ou, si l'apprenant épouse parfaitement les données d'apprentissage, le seul signe qui apparaisse) en définissant la valeur $s = \begin{cases} 1 \; si \; p^+ \geq \; p^- \\ -1 \; si \; p^+ < p^- \end{cases}$.

Nous pouvons maintenant définir un prédicat binaire P qui décide si un couple donné appartient à la relation binaire que nous avons apprise : $P(i, j) \Leftrightarrow (x_{i,j} \cdot w . s > 0)$.

II : Validité de l'espace de recherche

Comme le prédicat est basé sur une relation stricte, il est possible de montrer que toute relation binaire appartenant à cet espace est un ordre strict total, c'est-à-dire une relation binaire irréflexive, transitive et totale (la propriété d'asymétrie, impliquée par la conjonction de la transitivité et de l'irréflexivité, n'a pas besoin d'être vérifiée) :

Irréflexivité : $\neg \mathbf{P(i,i)}$.

$P(i,i) \Leftrightarrow (x_i - x_i) \cdot w.s > 0 \Rightarrow 0 > 0$, ce qui est faux, donc $\neg P(i,i)$.

Transitivité: $\mathbf{P(i,j) \wedge P(j,k) \Rightarrow P(i,k)}$.

$P(i,j) \wedge P(j,k) \Leftrightarrow \left(x_j - x_i\right) \cdot w.s > 0 \wedge (x_k - x_j) \cdot w.s > 0$

$\Rightarrow \left(x_j - x_i\right) \cdot w.s + (x_k - x_j) \cdot w.s > 0$

$\Leftrightarrow \left(x_j - x_i + x_k - x_j\right) \cdot w.s > 0 \Leftrightarrow (x_k - x_i) \cdot w.s > 0 \Leftrightarrow P(i,k)$.

Totalité: $\mathbf{P(i,j) \vee P(j,i) \vee (i = j)}$.

Pour tous éléments donnés i et j, nous avons l'une des trois possibilités suivantes :

$$\begin{cases} x_{i,j}.w > 0 \Leftrightarrow P(i,j). \\ x_{i,j}.w < 0 \Leftrightarrow x_{j,i}.w > 0 \Leftrightarrow P(j,i). \\ x_{i,j}.w = 0 \Rightarrow \mathrm{i} = \mathrm{j}. \end{cases}$$

Sous l'hypothèse raisonnable d'indépendance des éléments, P permet de prédire, pour tous éléments donnés i et j représentés par les vecteurs de caractéristiques x_i et x_j, lequel des couples (i,j) ou (j,i) appartient à l'ordre strict total qui correspond le mieux aux contraintes d'entrée.

III : Expressivité de l'espace de recherche

Les relations linéaires de l'espace de description peuvent ne pas être suffisamment expressives pour s'adapter à la plupart des ensembles d'entrée. Un espace de redescription peut alors être défini en étendant les vecteurs de caractéristiques d'origine avec des valeurs linéairement indépendantes. Le passage dans un tel espace de redescription permet une expressivité supérieure tout en restant dans un espace vectoriel.

A : Définitions préalables

1 : Equivalence ordinale

Soient x et y deux séquences de n éléments $(x_1, \ldots, x_n)$ et $(y_1, \ldots, y_n)$, $x_i \in X$, $y_i \in Y$, avec des relations d'ordre $<_X$ et $<_Y$ sur X et Y respectivement. Nous définissons l'équivalence ordinale entre x et y, et écrivons $x \Leftrightarrow_{ord} y$, la propriété selon laquelle $\forall i \leq n$, $\forall j \leq n$, $x_i <_X x_j \Leftrightarrow y_i <_Y y_j$. Par extension, nous disons que deux dérivations sont ordinalement équivalentes si leurs membres de gauche sont ordinalement équivalents et que leurs membres de droite sont ordinalement équivalents.

Par exemple, $\begin{pmatrix}0\\1\\3\end{pmatrix} \rightarrow \begin{pmatrix}3\\0\\1\end{pmatrix} \Leftrightarrow_{ord} \begin{pmatrix}1\\2\\3\end{pmatrix} \rightarrow \begin{pmatrix}3\\1\\2\end{pmatrix}$ puisque $\begin{pmatrix}0\\1\\3\end{pmatrix} \Leftrightarrow_{ord} \begin{pmatrix}1\\2\\3\end{pmatrix}$ et $\begin{pmatrix}3\\0\\1\end{pmatrix} \Leftrightarrow_{ord} \begin{pmatrix}3\\1\\2\end{pmatrix}$.

L'équivalence ordinale est juste un moyen pratique d'ignorer les échelles : une équivalence ordinale entre deux séquences est la même que celle d'un tau de Kendall [83] de 1 entre deux variables ordinales statistiques.

2 : Extension de dimensionnalité

Le rang d'une matrice peut être augmenté en ajoutant des lignes (ou des colonnes) linéairement indépendantes des lignes (ou des colonnes respectivement) de la matrice d'origine.

Nous définissons la fonction φ_r telle que $\varphi_r(X)$ étende X au rang r en ajoutant des colonnes qui sont des combinaisons non linéaires de ses colonnes d'origine.

A titre d'illustration, si $X = \begin{bmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \\ x_{3,1} & x_{3,2} \end{bmatrix}$ avec $rang(X) = 2$, nous pouvons avoir par exemple $\varphi_3(X) = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,1}x_{1,2} \\ x_{2,1} & x_{2,2} & x_{2,1}x_{2,2} \\ x_{3,1} & x_{3,2} & x_{3,1}x_{3,2} \end{bmatrix}$.

B : Preuve de l'expressivité

Soit X un ensemble de n éléments représentés par des vecteurs x_i, $1 < i < n$. Soit y une permutation des n premiers nombres entiers positifs indiquant un ordre entre les éléments ($y_i = k$ signifie que l'élément représenté par le vecteur x_i est à la $k^{ième}$ position dans la séquence d'éléments).

Sans perte de généralité, nous supposons que tous les éléments sont distincts deux à deux et que l'application des éléments vers leurs vecteurs représentatifs est injective ($i \neq j \Rightarrow x_i \neq x_j$). Nous cherchons un vecteur w tel que $\varphi_r(X)w \Leftrightarrow_{ord} y$.

Notons $\varphi_r(X)|y$ la matrice X étendue au rang r, elle-même augmentée du vecteur colonne y.

Si $rang(\varphi_r(X)|y) = rang(\varphi_r(X))$, alors il existe nécessairement un vecteur w tel que $\varphi_r(X)w = y$; or $y \Leftrightarrow_{ord} y$; il existe donc un vecteur w tel que $\varphi_r(X)w \Leftrightarrow_{ord} y$. En d'autres termes, l'espace de recherche est suffisamment expressif pour contenir toute fonction cible de rangement d'objets.

IV : Stratégie de recherche

A : Recherche par extension

1 : Définition d'un algorithme glouton

Il est possible de s'appuyer sur l'espace de recherche pour définir un algorithme glouton (Algorithme 1) qui calcule w en augmentant itérativement le rang de X jusqu'à pouvoir résoudre $\varphi_r(X)\ w = y$.

Algorithme 1 : Approche gloutonne.

$r \coloneqq rang(X)$
tant que $r < rang(\varphi_r(X)|y)$
    $r \coloneqq r + 1$
fin tant que
retourner w t.q. $\varphi_r(X)w = y$

Pour n éléments, $rang(\varphi_r(X)|y) \leq n$. Puisque n est fini, l'algorithme est assuré de terminer (il y a au plus $n-1$ itérations).

2 : Limites de l'algorithme glouton

Nous pouvons apprendre un ordre sur un ensemble de données dès que nous trouvons un vecteur w tel que $\varphi_r(X)w \Leftrightarrow_{ord} y$, et nous avons montré dans le paragraphe précédent qu'un tel w existe dans l'espace de recherche pour tout X à condition que r soit assez élevé. Néanmoins, même si l'algorithme glouton est assuré de trouver un vecteur w tel que $\varphi_r(X)w \Leftrightarrow_{ord} y$, il n'est pas garanti qu'il trouve celui pour lequel r est minimal.

Nous pouvons illustrer ce point par l'exemple suivant : lorsque nous exécutons l'algorithme, nous remarquons qu'il faut aller jusqu'à $r = 4$ pour apprendre $\begin{pmatrix}1\\2\\3\\4\end{pmatrix} \rightarrow \begin{pmatrix}1\\2\\4\\3\end{pmatrix}$ alors qu'il suffit d'aller jusqu'à $r = 2$ pour apprendre $\begin{pmatrix}1\\2\\3\\4\end{pmatrix} \rightarrow \begin{pmatrix}27\\44\\51\\48\end{pmatrix}$, or $\begin{pmatrix}1\\2\\4\\3\end{pmatrix} \Leftrightarrow_{ord} \begin{pmatrix}27\\44\\51\\48\end{pmatrix}$.

Afin de pouvoir généraliser à de nouveaux éléments, l'algorithme doit trouver la solution avec le r le plus bas. La recherche du r le plus bas semble justifiée du point de vue de la théorie de l'information. En effet, plus le rang r est bas, plus l'exploitation des régularités qui se trouvent dans l'ensemble d'apprentissage est élevée, or si l'ensemble d'apprentissage est représentatif des données sur lesquelles sont faites les prédictions (hypothèse I.I.D.), alors ces données contiennent les mêmes régularités que l'ensemble de test, ce dans la même proportion. Par conséquent, les prédictions basées sur cette information seront correctes avec une probabilité égale à cette proportion.

B : Recherche par extension / adaptation

1 : Raffinement de l'algorithme

L'algorithme glouton prend un y déterminé et itère sur la dimensionnalité jusqu'à ce qu'il puisse résoudre l'équation : il est capable d'étendre la dimensionnalité de l'espace de recherche mais il n'adapte pas la cible y.

Nous modifions alors l'algorithme en ajoutant une phase d'adaptation à chaque étape (Algorithme 2).

Algorithme 2 : Approche par extension / adaptation.

$r \coloneqq rang(X)$
tant que $r < rang(\varphi_r(X)|z), z = argmin_\gamma\ rang(\varphi_r(X)|\gamma)$ t.q. $\gamma \Leftrightarrow_{ord} y$.
    $r \coloneqq r + 1$
fin tant que
retourner w t.q. $\varphi_r(X)w = z$

L'algorithme d'extension / adaptation conserve les propriétés de l'algorithme glouton mais assure un r minimal (voir Fig. 1).

Tout dépend maintenant de la façon de chercher $argmin_\gamma\ rg(\varphi_r(X)|\gamma)$ tel que $\gamma \Leftrightarrow_{ord} y$. Si pour r minimal aucune solution n'est privilégiée, alors il s'agit d'un rangement agnostique. Il est cependant possible de renforcer le biais d'apprentissage en sélectionnant plus avant parmi les ordres valides.

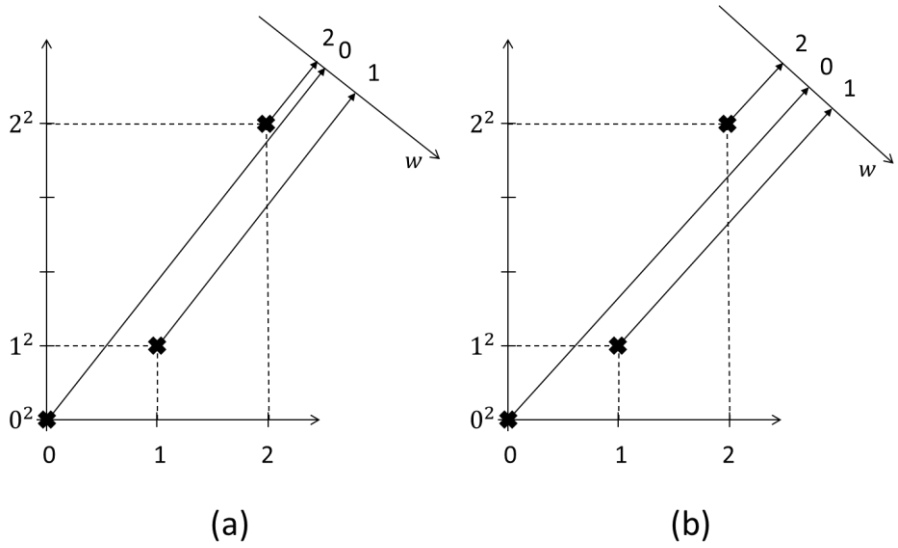


Fig. 1. Résolution de $\varphi_r\left(\begin{bmatrix}0\\1\\2\end{bmatrix}\right) w \Leftrightarrow_{ord} \begin{bmatrix}2\\0\\1\end{bmatrix}$ avec l'algorithme d'extension / adaptation. Cet algorithme permet de trouver w avec r minimal (ici $r = 2$) mais ne fait pas d'hypothèse sur ce qu'est la meilleure solution pour cette valeur de r. (a) et (b) représentent des hypothèses différentes mais toutes deux valides parmi une infinité d'hypothèses valides (du fait que les séquences sont à valeurs réelles).

C : Recherche par restriction de l'espace aux hypothèses valides

1 : Principe

Dans le cas d'un ensemble de données non bruitées, il existe un ensemble de vecteurs dont le produit scalaire avec n'importe quel vecteur x_{ij} est positif. Cet ensemble de solutions possibles peut être trouvé par élimination : chaque vecteur x_{ij} détermine la direction d'un hyperplan passant par l'origine et divisant l'espace en deux demi-espaces ; les solutions se situent nécessairement dans le demi-espace où se trouve le vecteur. Tant que nous supposons que le jeu de données est cohérent, nous pouvons laisser chaque vecteur x_{ij} réduire l'espace des solutions : cet espace ne sera jamais vide. La Figure 2 illustre ce principe sur la résolution de $\varphi_2\left(\begin{bmatrix}0\\1\\2\end{bmatrix}\right) w \Leftrightarrow_{ord} \begin{bmatrix}2\\0\\1\end{bmatrix}$: le vecteur allant du point 2 au point 0 permet de fermer le demi-espace figuré en dégradé de gris sur l'image (a) en ce que toute hypothèse compatible avec ce vecteur est représentée dans cet espace par un vecteur dont le produit scalaire avec le vecteur allant de 2 à 0 est positif ou nul. De même, en (b), le vecteur allant du point 0 au point 1 permet de

restreindre encore l'espace des hypothèses valides. L'image (c) illustre la conjonction des deux restrictions, et l'image (d), les hypothèses valides, w.

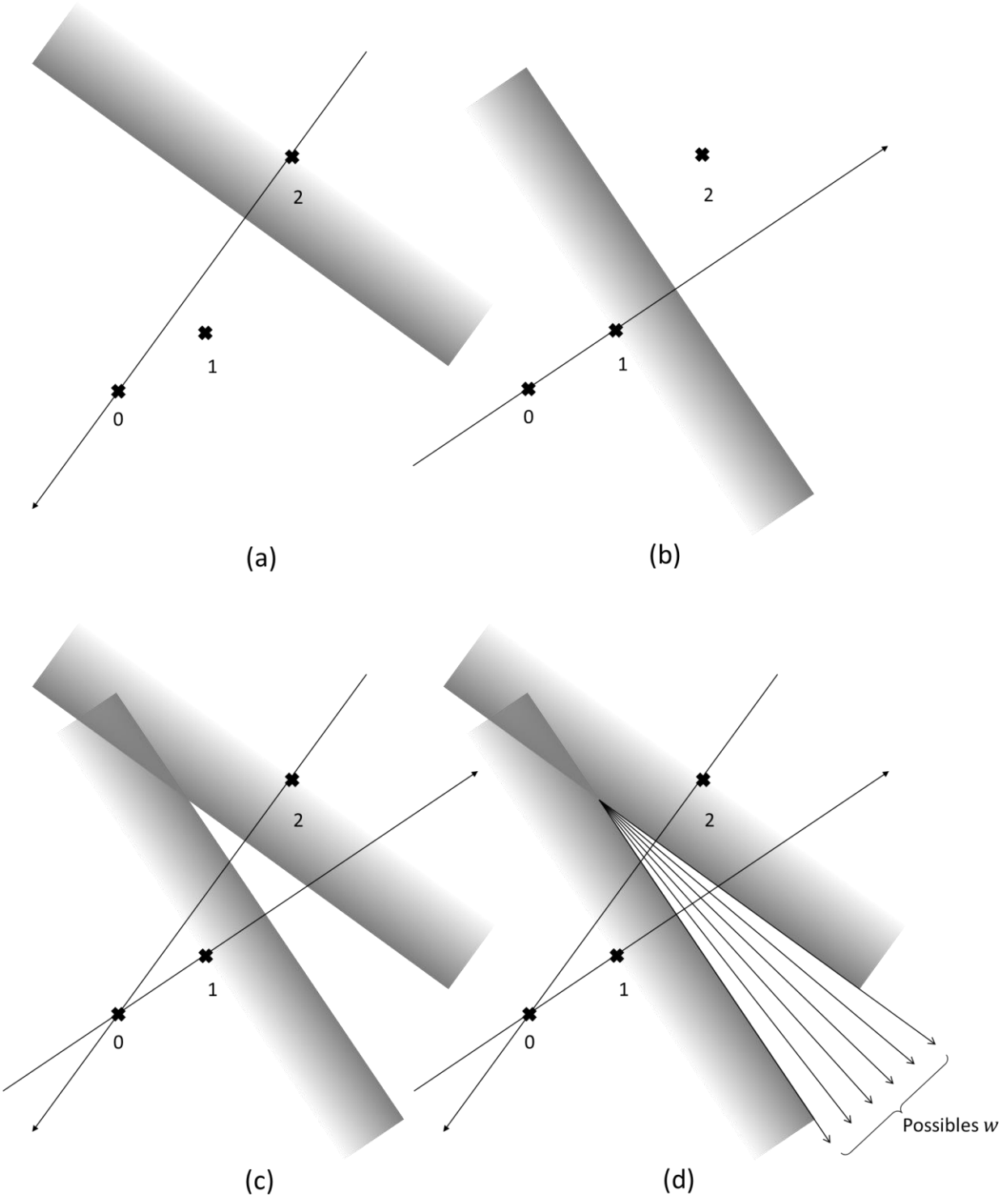


Figure 2 : Résolution de $\varphi_2\left(\begin{bmatrix}0\\1\\2\end{bmatrix}\right) w \Leftrightarrow_{ord} \begin{bmatrix}2\\0\\1\end{bmatrix}$ par restriction de l'espace de recherche. En (a), le vecteur allant du point 2 au point 0 permet de fermer un demi-espace (figuré en dégradé de gris). En (b), le vecteur allant du point 0 au point 1 permet de fermer un autre demi-espace. En (c), la conjonction des deux restrictions précédentes. En (d), les hypothèses valides, w, c'est-à-dire celles compatibles avec les vecteurs allant de 2 à 0 et de 0 à 1.

2 : Impraticabilité d'une résolution analytique

La restriction de l'espace aux hypothèses valides pourrait être effectuée analytiquement. Cependant, cela deviendrait vite fastidieux dès que le nombre de dimensions augmenterait. En outre, il faudrait l'adapter pour traiter les ensembles de données bruitées.

V : Synthèse

Dans ce chapitre, nous avons proposé une méthode de rangement agnostique. Cette méthode permet d'exprimer tous les ordres possibles avec un nombre minimal de dimensions. Dans le cadre d'un rangement agnostique, n'importe lequel des ordres valides peut être sélectionné. Cependant, il est possible de renforcer le biais d'apprentissage en ajoutant des contraintes sur l'ordre à sélectionner parmi les ordres valides. Dans le prochain chapitre, nous explorons les contraintes propres au rangement d'objets et proposons une méthode de résolution des problèmes de rangement d'objets.

Chapitre 5 : Description d'un algorithme d'apprentissage automatique spécifique au rangement d'objets

Une méthode de résolution en un temps des problèmes de rangement agnostique peut servir de base à des méthodes plus spécifiques, telle une méthode de résolution des problèmes de rangement d'objets. Tandis que le rangement agnostique fait seulement la supposition qu'il existe des régularités relatives à un ordre total sur les éléments, le rangement d'objets fait la supposition qu'il existe seulement des régularités relatives à un ordre total sur les éléments. Il est donc nécessaire de renforcer les contraintes sur l'hypothèse recherchée pour tenir compte de ce renforcement du biais d'apprentissage.

Nous proposons une machine à noyau recherchant une telle hypothèse, décrivons un méta-noyau permettant de l'adapter aux bibliothèques standard d'optimisation, puis la testons sur un jeu de données artificielles.

I : Distinction entre rangement d'instances et rangement d'objets selon le critère du biais d'apprentissage

Les méthodes de rangement agnostiques ne font aucune distinction au sein des ordres valides (c'est-à-dire les ordres qui sont compatibles avec les contraintes données en entrée). Leur biais d'apprentissage consiste en la supposition qu'il existe une relation d'ordre totale sur les éléments. Ce biais est certes plus fort que le biais d'apprentissage implémenté par les méthodes de classement ordinal, qui consiste en la supposition qu'il existe une relation d'ordre partiel sur les éléments, mais il est moins fort que les biais d'apprentissage des méthodes de rangement d'instances et d'objets, lesquels privilégient respectivement les ordres totaux qui accentuent les classes et ceux qui les atténuent.

Afin de bien cerner le biais d'apprentissage du rangement d'objets, il peut être utile de le contraster avec le biais d'apprentissage du rangement d'instances. Là où les contraintes données en entrée véhiculent des informations à la fois sur l'ordre et sur les classes dans le cadre d'un rangement d'instances, elles ne sont censées véhiculer que des informations sur l'ordre dans le cadre d'un rangement d'objets.

A : Biais d'apprentissage du rangement d'instances

Les méthodes de rangement d'instances opèrent une sélection plus fine de l'ordre à retourner que les méthodes de rangement agnostiques. L'ordre doit non seulement être compatible avec les contraintes données en entrée mais il doit aussi capturer les régularités dues aux classes. En effet, il s'agit de ranger les éléments en tant qu'instances, c'est-à-dire en tant que *réalisations particulières de classes* : le rangement d'instances fait la supposition qu'il existe non seulement un ordre total sur les éléments mais aussi des classes d'éléments.

Les régularités dues aux classes sont capturées en sélectionnant l'ordre qui les restitue le mieux, c'est-à-dire celui pour lequel les différences entre les instances de classes distinctes sont les plus marquées. L'extension du SVM au rangement présentée en [43] implémente le biais d'apprentissage du rangement d'instances en ce qu'elle sélectionne l'ordre qui maximise la marge minimale entre deux classes consécutives.

B : Biais d'apprentissage du rangement d'objets

Les méthodes de rangement d'objets opèrent elles aussi une sélection plus fine de l'ordre à retourner que les méthodes de rangement agnostiques. Cependant, leur cible est différente de celle des méthodes de rangement d'instances. L'ordre doit être compatible avec les contraintes données en entrée, mais il doit atténuer au maximum l'effet de ces contraintes pour le reste. C'est en tant qu'objets que les éléments doivent être rangés, non en tant qu'instances de classes. Les contraintes données en entrée n'expriment pas des différences de classes mais des différences d'objets : là où les méthodes de rangement d'instances cherchent à faire ressortir des classes à travers l'ordre, les méthodes de rangement d'objets privilégient l'ordre le plus neutre vis-à-vis des contraintes données en entrée.

II : Solution optimale du rangement d'objets

Le but est de trouver une solution qui évite toute forme de mise en valeur des contraintes données en entrée. Leur prise en compte doit être minimale, c'est-à-dire strictement limitée aux informations relatives aux objets : toute information supplémentaire doit être considérée, et traitée, comme étant parasitaire. En pratique, il faut tenter d'atténuer au maximum l'effet des contraintes données en entrée en ne s'en servant que pour établir un ordre entre les objets, mais l'ordre lui-même doit être le plus régulier possible.

A : Solution de principe

Dans le cas du classement par SVM, il a été statistiquement prouvé que la meilleure solution est celle qui correspond à l'hyperplan avec une marge maximale [93] ; dans les cas du classement ordinal et du rangement d'instances, il a été prouvé que la meilleure solution est celle qui correspond à l'ensemble d'hyperplans parallèles parmi lesquels l'hyperplan situé entre les classes les plus proches présente une marge maximale [43]. En s'appuyant sur l'existence de cette dernière preuve, nous pouvons affirmer que dans le cas du rangement d'objets, *la meilleure solution est le vecteur sur lequel la projection de tous les points aboutit à une séquence dans laquelle la distance entre les deux points les plus proches est maximale*.

En effet, nous pouvons considérer le rangement d'objets comme un cas spécifique de rangement d'instances où il n'y a qu'une seule instance par classe. La preuve nous indique alors que la meilleure solution correspond à l'ensemble des hyperplans parallèles parmi lesquels l'hyperplan situé entre les classes les plus proches, donc les points les plus proches, présente une marge maximale. Comme ces hyperplans sont parallèles, il est possible d'y passer un vecteur normal : les intersections entre ce vecteur et les hyperplans correspondent aux projections des points sur ce vecteur, donc les distances entre les hyperplans correspondent aux distances entre les points de la séquence en résultant.

B : Inapplicabilité de la solution de principe

À première vue, la méthode de résolution du rangement d'instances [43] pourrait être appliquée aux problèmes de classement ordinal d'objets en définissant chaque vecteur comme

une instance unique d'une classe. Cependant, cette méthode ne permet que de trouver des hyperplans avec une marge maximale entre des points *explicitement définis comme des couples* dans l'ensemble d'apprentissage.

Si le principe de la vaste marge se prête bien aux problèmes de rangement d'instances, ce n'est pas le cas pour les problèmes de rangement d'objets, où il faut trouver l'hyperplan ayant la marge maximale entre les deux points les plus proches du jeu de données même dans les cas où les informations relatives au rapport ordinal qu'ils entretiennent ne sont pas explicitement contenues dans ce jeu de données. En effet, ce problème est de complexité exponentielle par rapport au nombre de paires d'éléments dont le sens n'est pas spécifié. Cette méthode se prêterait cependant bien au cas particulier de rangement d'objets dans lequel les données d'apprentissage sont totalement ordonnées.

C : Solution alternative

A défaut de pouvoir sélectionner le meilleur ordre, i.e. celui qui dispose les objets de manière à ce que la plus petite distance entre deux objets quelconques soit maximale, une solution alternative peut consister à réduire au maximum l'impact des couples explicites sur l'ordre sélectionné. En d'autres termes, à défaut de pouvoir donner tout le poids aux couples implicites, une approche peut consister à en retirer le plus possible aux couples explicites, ne retenant que l'information strictement nécessaire à l'élaboration d'un ordre total.

Parmi tous les ordres totaux valides, le rangement d'instances consiste à rechercher l'ordre mettant l'accent sur les couples explicites en maximisant la distance entre les éléments du couple explicite dont les éléments sont le plus proche, ou, dans le cas où les couple sont exprimés par l'appartenance des objets à des classes différentes, en maximisant la marge la plus fine entre deux classes. A la place, nous recherchons l'ordre se tenant le plus éloigné de tout ordre invalide en maximisant la distance à l'ordre total non valide le plus proche : la seule information tirée des couples explicites est celle relative à ces ordres invalides.

Une fois réduit l'espace des hypothèses, la plupart des vecteurs x_{ij} n'auront aucun impact sur l'espace final : seuls les vecteurs correspondant aux frontières en auront. Le but est de les trouver de sorte que la meilleure solution, w^*, puisse être calculée : c'est le vecteur maximisant

l'angle avec les hyperplans normaux aux vecteurs frontières, ou de manière équivalente, minimisant l'angle avec ces vecteurs frontières.

Cette solution revient à minimiser l'espérance du risque sous l'hypothèse d'un risque uniformément distribué.

III : Mise en œuvre par une machine à noyau

Dans cette section, nous introduisons une formalisation de cette méthode en la décrivant sous une forme primale et en dérivant une forme duale permettant l'utilisation de noyaux.

A : Forme primale

Soit X un ensemble de vecteurs de m descripteurs à valeurs réelles tels que $x_i \in X$ représente l'élément i et $P \subseteq X^2$ un ensemble de couples d'éléments de X exprimant une relation d'ordre partielle sur X.

Soit $\varphi: \mathbb{R}^m \to \mathbb{R}^l$, $(m, l) \in \mathbb{N}^2$ une fonction non nécessairement linéaire ; définissons l'espace des hypothèses par $H = \mathbb{R}^l$ et représentons un ordre total par un vecteur $w \in \mathbb{R}^l$ sur lequel les points à ordonner sont projetés.

Nous définissons le sous-ensemble d'hypothèses cohérentes avec X par $H_X = \{w \in \mathbb{R}^l | \forall (x_i, x_j) \in P, w \cdot \varphi(x_i) < w \cdot \varphi(x_j)\}$. Si X est un ensemble d'apprentissage, alors tout $w \in H_X$ est une hypothèse empiriquement valide.

Choisissons φ et l tels que H_X soit non vide. La distance entre deux hypothèses (i.e. entre deux ordres totaux) peut être exprimée par l'angle entre leurs vecteurs représentatifs.

Un ordre total qui maximise la distance avec l'ordre total invalide le plus proche est représenté par le vecteur $w^* \in \mathbb{R}^l$ qui maximise le plus petit angle avec un vecteur représentant un ordre total non valide ou, de la même manière, qui minimise l'angle maximal par rapport à un vecteur représentant un ordre total valide. Les équations (5), (6) et (7)

$$w^* = argmax_w \left\{ min_{i,j,(x_i,x_j)\in P} cos\left(\angle \left(w, \varphi(x_j) - \varphi(x_i) \right) \right) \right\} \quad (6)$$

$$w^* = argmax_w \, min_{i,j,(x_i,x_j)\in P} \left(\frac{w^t}{\|w\|} \frac{\varphi(x_j)-\varphi(x_i)}{\|\varphi(x_j)-\varphi(x_i)\|} \right) \quad (7)$$

formalisent cela en exprimant w^* de manière équivalente en termes d'angles, de cosinus et de produits scalaires respectivement. L'équation (8) introduit ϕ, la fonction qui calcule le vecteur unitaire de deux points dans l'espace de redescription déterminé par φ. Ceci permet de réécrire l'équation (7) de façon compacte (équation (9)).

$$w^* = argmin_w \left\{ max_{i,j,(x_i,x_j)\in P} \angle \left(w, \varphi(x_j) - \varphi(x_i) \right) \right\} \quad (5)$$

$$w^* = argmax_w \left\{ min_{i,j,(x_i,x_j)\in P} cos \left(\angle \left(w, \varphi(x_j) - \varphi(x_i) \right) \right) \right\} \quad (6)$$

$$w^* = argmax_w \left\{ min_{i,j,(x_i,x_j)\in P} \left(\frac{w^t}{\|w\|} \frac{\varphi(x_j) - \varphi(x_i)}{\|\varphi(x_j) - \varphi(x_i)\|} \right) \right\} \quad (7)$$

$$\phi(x_i, x_j) = \frac{\varphi(x_j) - \varphi(x_i)}{\|\varphi(x_j) - \varphi(x_i)\|} \quad (8)$$

$$w^* = argmax_w \left\{ min_{i,j,(x_i,x_j)\in P} \left(\frac{w^t}{\|w\|} \phi(x_i, x_j) \right) \right\} \quad (9)$$

Posons $w^t \phi(x_i, x_j) \geq 1 \; \forall (i,j) \; t.q. \; (x_i, x_j) \in P$ et maximisons $\frac{1}{\|w\|}$, ce qui revient à minimiser $\frac{1}{2}\|w\|^2$ (équation (10))

$$w^* = argmin_w \left\{ \frac{1}{2}\|w\|^2 \mid w^t \phi(x_i, x_j) \geq 1 \; \forall (i,j) \; t.q. \; (x_i, x_j) \in P \right\} \quad (10)$$

B : Forme duale

Extrayons la forme duale du problème en introduisant le lagrangien $L(w,\, a) = \frac{1}{2}\|w\|^2 - \sum_{\substack{(i,j)\ t.q.\\ (x_i,x_j)\in P}} \alpha_{ij}\left(w^t\phi(x_i,x_j) - 1\right)$, où $\alpha_{ij} \geq 0$. Le problème d'optimisation consiste maintenant à rechercher α^*, le vecteur optimal de coefficients des exemples d'apprentissage, tel qu'exprimé par l'équation (11).

$$\alpha^* = argmax_\alpha \left\{ \begin{array}{c} \sum_{\substack{(i,j)\ t.q.\\ (x_i,x_j)\in P}} \alpha_{ij} - \frac{1}{2} \sum_{\substack{(i,j)\ t.q.\\ (x_i,x_j)\in P}} \sum_{\substack{(k,l)\ t.q.\\ (x_k,x_l)\in P}} \alpha_{ij}\alpha_{kl}\phi(x_i,x_j)^t\phi(x_k,x_l) \\ t.q.\ \alpha_{mn} \geq 0 \end{array} \right\} \quad (11)$$

L'introduction de variables ressort [94] confère à cette méthode la capacité de traiter des jeux de données bruitées. L'expression de α^* devient alors telle que formalisée par l'équation (12), où C est le terme de régularisation (qui permet de se prémunir contre le sur-apprentissage en limitant la complexité du modèle).

$$\alpha^* = argmax_\alpha \left\{ \begin{array}{c} \sum_{\substack{(i,j)\ t.q.\\ (x_i,x_j)\in P}} \alpha_{ij} - \frac{1}{2} \sum_{\substack{(i,j)\ t.q.\\ (x_i,x_j)\in P}} \sum_{\substack{(k,l)\ t.q.\\ (x_k,x_l)\in P}} \alpha_{ij}\alpha_{kl}\phi(x_i,x_j)^t\phi(x_k,x_l) \\ t.q. 0 \leq \alpha_{mn} \leq C \end{array} \right\} \quad (12)$$

Introduisons la fonction $\dot{\varphi}$ (où le point figure le produit scalaire, non la dérivée d'après la notation de Newton), qui exprime dans l'espace de redescription déterminé par φ le produit scalaire de deux vecteurs dont les extrémités sont données en arguments (équation (13)).

$$\dot{\varphi}(x_i,x_j,x_k,x_l) = \varphi(x_j)^t\varphi(x_l) - \varphi(x_j)^t\varphi(x_k) - \varphi(x_i)^t\varphi(x_l) + \varphi(x_i)^t\varphi(x_k) \quad (13)$$

Il est alors possible de réécrire l'équation (12) en termes de $\dot{\varphi}$ (équation (14)).

$$\alpha^* = argmax_\alpha \left\{ \sum_{\substack{(i,j)\ t.q. \\ (x_i,x_j)\in P}} \alpha_{ij} - \frac{1}{2} \sum_{\substack{(i,j)\ t.q. \\ (x_i,x_j)\in P}} \sum_{\substack{(k,l)\ t.q. \\ (x_k,x_l)\in P}} \alpha_{ij}\alpha_{kl} \frac{\dot{\varphi}(x_i,x_j,x_k,x_l)}{\sqrt{\dot{\varphi}(x_i,x_j,x_i,x_j)}\sqrt{\dot{\varphi}(x_k,x_l,x_k,x_l)}} \right\} \quad t.q.\ 0 \le \alpha_{mn} \le C \qquad (14)$$

L'astuce du noyau [95] permet de remplacer le produit $\varphi(x_i)^t\varphi(x_j)$ par toute fonction noyau valide $K(x_i, x_j)$, c'est-à-dire par toute fonction noyau semi-définie positive [96]. Soit K une telle fonction ; il est alors possible de réécrire l'équation (13) en fonction de K (équation (15)).

$$\dot{\varphi}(x_i, x_j, x_k, x_l) = K(x_j, x_l) - K(x_j, x_k) - K(x_i, x_l) + K(x_i, x_k) \qquad (15)$$

Il en résulte que l'équation (14) peut être résolue en utilisant K, ce qui évite le calcule explicite des valeurs de φ.

Il s'agit d'un problème de programmation quadratique ; des algorithmes existent pour le résoudre (notamment la méthode d'optimisation minimale séquentielle (SMO) [97], qui peut le résoudre en $O(n^3)$, n étant le nombre de paires d'éléments interclasses).

C : Prédiction

Le vecteur optimal des coefficients des exemples d'apprentissage obtenu lors de la phase d'apprentissage, α^*, permet de calculer le projeté de n'importe quel vecteur le long de l'axe représentant l'ordre entre les éléments. Pour prédire l'ordre entre différents points selon cet axe, il suffit d'associer à chacun des points un vecteur allant de l'origine (ou de n'importe quel autre point, dès l'instant qu'il est le même pour tous les vecteurs) à ce point, puis de projeter les vecteurs ainsi obtenus sur l'axe en question : l'ordre entre les éléments correspond à l'ordre entre les projetés de leurs vecteurs respectifs le long de cet axe.

Le projeté d'un vecteur sur l'axe représentant l'ordre entre les éléments correspond au produit scalaire entre ce vecteur et le vecteur représentant l'axe, lequel se calcule comme la combinaison linéaire, selon α^*, des produits scalaires de ce vecteur avec les vecteurs unitaires formés à partir des éléments des classes consécutives de l'ensemble d'apprentissage. Ces produits scalaires sont eux-mêmes calculés par la fonction noyau utilisée lors de la phase d'apprentissage.

L'équation (16), développée dans l'équation (17), formalise le calcul du projeté $p(\boldsymbol{x})$ en prenant l'origine pour construire les vecteurs associés.

$$p(\boldsymbol{x}) = \sum_{\substack{(i,j)\ t.q. \\ (x_i,x_j)\in P}} \alpha_{ij}^* \frac{\dot{\varphi}(x_i, x_j, 0, \boldsymbol{x})}{\sqrt{\dot{\varphi}(x_i, x_j, x_i, x_j)}} \quad (16)$$

$$p(\boldsymbol{x}) = \sum_{\substack{(i,j)\ t.q. \\ (x_i,x_j)\in P}} \alpha_{ij}^* \frac{K(x_j, \boldsymbol{x}) - K(x_j, 0) - K(x_i, \boldsymbol{x}) + K(x_i, 0)}{\sqrt{K(x_j,\ x_j) - 2K(x_j,\ x_i) + K(x_i,\ x_i)}} \quad (17)$$

IV : Forme compatible avec les bibliothèques d'optimisation existantes

Dans cette section, nous proposons un moyen de mettre en œuvre l'algorithme proposé avec les bibliothèques d'optimisation existantes.

Il existe des bibliothèques implémentant SMO (notamment LIBSVM [98]), qui peuvent être utilisées pour résoudre le problème d'optimisation du rangement d'objets. Cependant, ces bibliothèques sont conçues pour les SVM, et de légères modifications sont nécessaires pour les utiliser dans le cadre du rangement d'objets.

Nous présentons le principe dans le cas linéaire puis le formalisons dans le cas général.

A : Illustration dans le cas linéaire

Supposons que nous soyons dans le cas linéaire. Il est alors possible d'adapter le problème pour le résoudre avec un SVM linéaire. La méthode est la suivante : nous normalisons la taille de tout vecteur x_{ij}, nous l'étiquetons comme positif et ajoutons son vecteur opposé, normalisé et étiqueté négativement, dans l'ensemble d'apprentissage. Comme les vecteurs ont tous la même magnitude et que les exemples positifs et négatifs sont symétriques, les vecteurs supports coïncideront avec les vecteurs déterminant les hyperplans réduisant l'espace des vecteurs valides. Il en résulte que le vecteur retourné par le SVM linéaire est la solution de notre problème d'origine (Fig. 3). En outre, il n'est pas nécessaire de prendre les vecteurs opposés étiquetés négativement dans la mesure où le SVM ne les considère qu'avec leurs étiquettes.

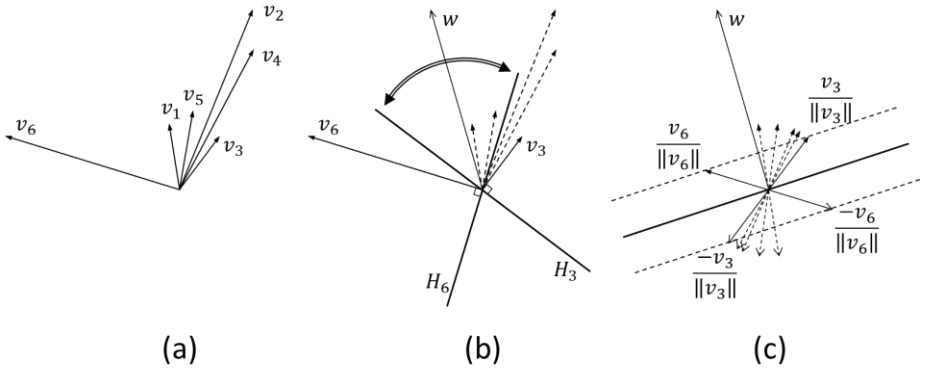


Fig. 3. (a) Six vecteurs v_h représentant les relations explicites $x_{ij} = \varphi_r(x_j) - \varphi_r(x_i)$ dans un espace à deux dimensions. (b) Résolution par élimination : v_3 et v_6 sont les vecteurs déterminant les hyperplans réduisant l'espace des vecteurs valides (les hyperplans passant par l'origine et normaux à ces vecteurs, H_3 et H_6, déterminent complètement l'ensemble des solutions possibles, indiqué par la double flèche) ; la solution, w, maximise l'angle minimal avec les hyperplans H_3 et H_6, ou de manière équivalente, minimise l'angle maximal avec les vecteurs arêtes v_3 et v_6. (c) Résolution avec la méthode pour SVM linéaire : une fois normalisés et symétrisés, les vecteurs déterminant les hyperplans réduisant l'espace des vecteurs valides peuvent être détectés en tant que vecteurs supports par un SVM linéaire.

Le SVM retourne l'hyperplan séparateur sous forme du vecteur normal à cet hyperplan et passant par l'origine ainsi que de la distance entre l'origine et le point d'intersection de l'hyperplan séparateur avec l'axe orienté selon ce vecteur. Cependant, seul le vecteur normal à l'hyperplan séparateur est utile au rangement. En effet, le rangement ne s'intéresse qu'aux positions relatives des éléments.

Pour trouver l'ordre total représenté par ce vecteur, il suffit de projeter les éléments sur l'axe qu'il représente.

B : Formalisation dans le cas général

Nous avons illustré comment résoudre un problème de rangement d'objets avec un SVM dans le cas linéaire. Pour que cette méthode soit intéressante en pratique, il faut être en mesure de la mettre en œuvre avec d'autres noyaux que le seul noyau linéaire. Nous décrivons un méta-noyau qui permet d'utiliser les solveurs de SVM pour résoudre les problèmes de rangement d'objets dans le cas général. Nous décrivons également un second méta-noyau, nécessaire pour la phase de prédiction.

1 : Méta-noyau pour la phase d'apprentissage

Dérivons de l'ensemble d'apprentissage un ensemble des paires interclasses d'éléments, $X' = \left\{x \mid \forall (i,j)\ t.q.\ (x_i, x_j) \in P, x = \left(x_i^{(1)} x_i^{(2)}, \dots, x_i^{(m)}, x_j^{(1)} x_j^{(2)}, \dots, x_j^{(m)}\right)\right\}$, où l'exposant figure la dimension et m le nombre de dimensions. Définissons $pre(x)$ et $post(x)$, qui retournent des vecteurs contenant respectivement les m premières et les m dernières valeurs de x. Nous pouvons alors décrire un méta-noyau K' valide (équation (18)) en fonction de $\left(x_i, x_j\right) \in X'^2$.

$$K'(x_i, x_j) = \frac{\dot{\varphi}(pre(x_i), post(x_i), pre(x_j), post(x_j))}{\sqrt{\dot{\varphi}(pre(x_i), post(x_i), pre(x_i), post(x_i))}\sqrt{\dot{\varphi}(pre(x_j), post(x_j), pre(x_j), post(x_j))}} \quad (18)$$

Ce méta-noyau est valide puisque les opérations sont effectuées dans l'espace de redescription sur un noyau valide K.

L'expression pour α^*est alors donnée par l'équation (19), où $N' = |X'|$.

$$\alpha^* = argmax_\alpha \left\{ \sum_{i=1}^{N'} \alpha_i - \frac{1}{2} \sum_{i=1}^{N'} \sum_{j=1}^{N'} \alpha_i \alpha_j K'(x_i, x_j)\ t.q.\ 0 \le \alpha_{mn} \le C \right\} \quad (19)$$

En utilisant K', le rangement d'objets peut être résolu avec les solveurs de SVM en mettant tous les labels à 1 (ou en les partageant entre 1 et -1 en inversant les couples correspondant aux -1 si le classement à une classe n'est pas prévu).

2 : Méta-noyau pour la phase de prédiction

Afin de pouvoir utiliser les fonctions de prédiction des bibliothèques traditionnelles pour prédire le rang d'un point $\boldsymbol{x}$ quelconque, il faudrait fournir ce point sous la forme $(0^m, \boldsymbol{x})$[13].

Cependant, ces fonctions de prédiction opéreraient le calcul exprimé par l'équation (20), ce qui, étant donné la forme du méta-noyau K', équivaut, une fois développée, à l'équation (21), ou, de manière équivalente, à l'équation (22), mais non à l'équation (16)

$$p\left((0^m, \boldsymbol{x})\right) = \sum_{i=1}^{N'} \alpha_i^* K'\left((0^m, \boldsymbol{x}), x_i\right), x_i \in X' \quad (20)$$

$$p(\boldsymbol{x}) = \sum_{\substack{(i,j)\ t.q. \\ (x_i, x_j) \in P}} \alpha_{ij}^* \frac{K(x_j, \boldsymbol{x}) - K(x_j, 0) - K(x_i, \boldsymbol{x}) + K(x_i, 0)}{\sqrt{K(x_j,\ x_j) - 2K(x_j,\ x_i) + K(x_i,\ x_i)}\sqrt{K(\boldsymbol{x},\ \boldsymbol{x}) - 2K(\boldsymbol{x},\ 0) + K(0,\ 0)}} \quad (21)$$

$$p(\boldsymbol{x}) = \sum_{\substack{(i,j)\ t.q. \\ (x_i, x_j) \in P}} \alpha_{ij}^* \frac{\dot{\varphi}(x_i, x_j, 0, \boldsymbol{x})}{\sqrt{\dot{\varphi}(x_i, x_j, x_i,\ x_j)}\sqrt{\dot{\varphi}(0, \boldsymbol{x}, 0, \boldsymbol{x})}} \quad (22)$$

[13] Dans le cas où les vecteurs à projeter sont construits à partir de l'origine. N'importe quel point peut être utiliser à la place ; la seule contrainte est que ce point soit le même pour tous les vecteurs à construire.

Il faudrait alors utiliser un autre méta-noyau que dans la phase d'apprentissage. L'équation (23 **)**, ou, de manière équivalente, l'équation (24), exprime un tel méta-noyau, K'', où $x_i \in X'$ et $\boldsymbol{x} \in X$.

$$K''(x_i, \boldsymbol{x}) = \frac{K(post(x_i), \boldsymbol{x}) - K(post(x_i), 0) - K(pre(x_i), \boldsymbol{x}) + K(pre(x_i), 0)}{\sqrt{K(post(x_i),\ post(x_i)) - 2K(post(x_i),\ pre(x_i)) + K(pre(x_i),\ pre(x_i))}} \quad (23)$$

$$K''(x_i, \boldsymbol{x}) = \frac{\dot{K}(pre(x_i), post(x_i), 0, \boldsymbol{x})}{\sqrt{\dot{K}(pre(x_i), post(x_i), pre(x_i), post(x_i))}} \quad (24)$$

Dans le cas où les bibliothèques ne permettent pas de modifier le noyau entre les phases d'apprentissage et de prédiction, alors il est toujours possible d'implémenter l'équation (16) dans la mesure où les bibliothèques permettent d'accéder aux vecteurs de supports et à leurs coefficients respectifs à l'issu de la phase d'apprentissage.

V : Expérience

A : Description

Nous comparons la méthode de rangement d'objets avec celle de rangement d'instances de [43], que nous prenons comme référence, en générant les observations de la manière décrite dans [43].

Une observation $x = (x_1, x_2)^T$ est générée aléatoirement dans le carré unitaire $[0, 1] \times [0, 1]$ selon une distribution uniforme et associée à la valeur $U(x) = 10\big((x_1 - 0.5) \cdot (x_2 - 0.5)\big)$.

Nous procédons à un tirage de 50 observations que nous rangeons par ordre croissant de leurs valeurs associées. Puisque nous sommes dans le cadre du rangement d'objets, où les frontières entre les classes n'ont pas de signification intrinsèques, nous testons pour cet ordre de référence tous les cas de figure possibles selon le nombre de classes. Soit n le nombre d'observations et c le nombre de classes, il y a $\binom{n-1}{c-1}$ combinaisons possibles. Nous faisons le

test avec 2 et 3 classes. Sur 50 observations, ceci revient à tester $\binom{49}{1} = 49$ et $\binom{49}{2} = 1176$ configurations respectivement.

Pour chacune des configurations, nous apprenons un ordre selon la méthode de rangement d'objets et selon la méthode de rangement d'instances, chacune avec les noyaux quadratique, cubique et gaussien, ce dernier étant paramétré avec $\sigma = 0.1$, $\sigma = 0.5$ et $\sigma = 1$. La fonction d'optimisation utilisée est la fonction fitcsvm de MATLAB, avec les paramètres par défaut, travaillant sur le méta-noyau défini précédemment (cf. section IV du présent chapitre).

Pour chaque ordre obtenu, nous calculons sa distance (en termes de nombre relatif de couples discordants) avec l'ordre de référence et en faisons la moyenne pour chaque méthode avec chaque noyau afin d'obtenir le coût pour chaque méthode et chaque noyau sur le tirage.

Nous répétons l'expérience sur 100 tirages indépendants et évaluons la significativité de la différence entre les résultats de méthodes différentes en calculant la valeur-p du test-t de Student [99]. Nous fixons le seuil de significativité à 0.05.

B : Résultats

Le Tableau 7Tableau 12 indique pour chaque noyau le coût moyen obtenu par les méthodes de rangement d'objets et de rangement d'instances sur deux classes, ainsi que la valeur-p de la différence entre les résultats des deux méthodes. Le Tableau 8Tableau 7Tableau 12 indique la même chose sur trois classes.

Tableau 7 : Coût moyen (en pourcentage de paires mal orientées) par noyau sur 2 classes selon le rangement d'objets (RO) et le rangement d'instances (RI), avec la valeur-p de la différence entre les résultats des deux méthodes. En gras, les meilleurs résultats pour chacune des deux méthodes.

Noyau	*RO*	*RI*	*Valeur-p des différences*
Quadratique	10.9% ± 1.7%	11.5% ± 1.6%	< 0.01
Cubique	**10.8%** ± 1.6%	**11.1%** ± 1.4%	< 0.01
RBF 0.1	22.7% ± 1.1%	24.3% ± 1.2%	< 0.01
RBF 0.5	12.1% ± 1.2%	12.7% ± 1.6%	< 0.01
RBF 1	12.1% ± 1.4%	14.2% ± 1.6%	< 0.01

Sur deux classes (Tableau 7)Tableau 12, le rangement d'objets donne de meilleurs résultats que le rangement d'instances quel que soit le noyau utilisé. La valeur-p des différences est inférieure à 0.01 pour tous les noyaux. Les meilleurs résultats sont obtenus avec le noyau cubique pour les deux méthodes (10.8% ± 1.6% pour le rangement d'objets et 11.1% ± 1.4% pour le rangement d'instances).

Tableau 8 : Coût moyen (en pourcentage de paires mal orientées) par noyau sur 3 classes selon le rangement d'objets (RO) et le rangement d'instances (RI), avec la valeur-p de la différence entre les résultats des deux méthodes. En gras, les meilleurs résultats pour chacune des deux méthodes.

Noyau	*RO*	*RI*	*Valeur-p des différences*
Quadratique	8.4% ± 1.2%	8.5% ± 1.3%	0.24
Cubique	7.5% ± 0.9%	**6.2%** ± 0.7%	< 0.01
RBF 0.1	15.6% ± 0.5%	17.9% ± 0.6%	< 0.01
RBF 0.5	**6%** ± 0.5%	6.3% ± 0.6%	< 0.01
RBF 1	6.9% ± 0.8%	9.8% ± 1.4%	< 0.01

Sur trois classes (Tableau 8), le rangement d'objets donne de meilleurs résultats que le rangement d'instances sur tous les noyaux sauf le noyau cubique. La valeur-p des différences est inférieure à 0.01 pour tous les noyaux sauf pour le noyau quadratique (p = 0.24). Les meilleurs résultats sont obtenus avec le noyau RBF (σ = 0.5) pour le rangement d'objets (6% ± 0.5%) et avec le noyau cubique pour le rangement d'instances (6.2% ± 0.7%). Un calcul de la valeur-p de la différence entre les deux donne un résultat inférieur à 0.01.

Enfin, un calcul de la valeur-p de la différence par noyau et par méthode entre les résultats sur deux classes et les résultats sur trois classes donne un résultat toujours inférieur à 0.01.

C : Discussion

Le coût des deux méthodes diminue avec tous les noyaux lorsque le nombre de classes passe de deux à trois, cette diminution étant statistiquement significative. Ceci peut s'expliquer par la quantité accrue d'information introduite par l'ajout d'une troisième classe : le nombre de

couples explicites est plus grand, sur un même nombre d'éléments, avec trois classes qu'avec deux classes. En effet, le nombre moyen de couples explicites par configuration est de $\frac{1}{\binom{n-1}{2}}\sum_{i=1}^{n-1}\sum_{j=1}^{n-i-1}(i(n-i)+j(n-i-j))$ dans le cas où il y a trois classes alors qu'il n'est que de $\frac{1}{\binom{n-1}{1}}\sum_{i=1}^{n-1}(i(n-i))$ dans le cas où il y en a deux. En prenant $n = 50$ (jeu de données de 50 éléments), cela fait 637.5 pour une configuration à trois classes contre 425 pour une configuration à deux classes, soit exactement 50% de plus.

Pour le cas où il y a 2 classes, la méthode de rangement d'objets a de meilleures performances que la méthode de rangement d'instances pour tous les noyaux, cette amélioration étant statistiquement significative ($p < 0.01$ pour toutes les différences).

Pour le cas où il y a 3 classes, il en va de même seulement pour les noyaux RBF. Il n'y a pas d'amélioration pour le noyau cubique, et l'amélioration pour le noyau quadratique n'est pas statistiquement significative ($p = 0.24$). Avec le noyau cubique, le rangement d'instances surpasse même le rangement d'objets (6.2% ± 0.7% pour le rangement d'instances contre 7.5% ± 0.9% pour le rangement d'objets), ce de manière statistiquement significative. Toutefois, tous noyaux confondus, le rangement d'objets donne de meilleurs résultats que le rangement d'instances (6% ± 0.5% pour le rangement d'objets avec le noyau RBF à $\sigma = 0.5$ contre 6.2% ± 0.7% pour le rangement d'instances avec le noyau cubique), ce également de manière statistiquement significative.

L'accroissement de la quantité d'information impliqué par le passage de deux à trois classes semble bénéficier plus à la méthode de rangement d'instances qu'à celle de rangement d'objets, du moins en ce qui concerne les noyaux quadratique et cubique.

Ceci pourrait s'expliquer par les considérations suivantes :

D'un côté, dans un espace de redescription donné, la solution de principe (cf. section II-A du présent chapitre) consiste à sélectionner l'ordre correspondant au vecteur sur lequel la projection de tous les points aboutit à une séquence dans laquelle la distance entre les deux

points les plus proches est maximale et si les couples auxquels appartiennent ces points sont explicites, alors la méthode de rangement d'instances est optimale (cf. section II-A du présent chapitre). Or plus le nombre de couples explicites augmente, plus l'ensemble de ces couples devient représentatif de l'ensemble de tous les couples, favorisant ainsi la méthode de rangement d'instances.

D'un autre côté, lorsque le nombre de couples explicites est insuffisant pour être représentatif de l'ensemble de tous les couples, la solution de principe n'est pas applicable (cf. section II-B du présent chapitre) et une solution alternative consiste à minimiser l'espérance du risque ; c'est ce que fait la méthode de rangement d'objets (cf. section II-C du présent chapitre).

Ainsi, la méthode de rangement d'instances donnerait de meilleurs résultats lorsque le nombre de couples explicites est assez élevé compte tenu de l'espace de redescription ; la méthode de rangement d'objets donnerait de meilleurs résultats dans le cas contraire.

Il se pourrait que le passage de deux à trois classes, c'est-à-dire de 425 à 637.5 couples explicites en moyenne par configuration dans le cas d'un ensemble de 50 éléments, suffise à ce que l'ensemble des couples explicites devienne assez représentatif de l'ensemble de tous les couples dans les espaces de redescription impliqués par les noyaux quadratique et cubique, mais pas dans l'espace de redescription impliqué par le noyau RBF, ce dernier espace étant bien plus vaste que les premiers.

VI : Synthèse

Dans ce chapitre, nous avons proposé une méthode de résolution propre aux problèmes de rangement d'objets en nous appuyant sur la méthode de résolution en un temps des problèmes de rangement agnostique et en en renforçant le biais d'apprentissage par l'ajout de la contrainte selon laquelle l'ordre sélectionné doit atténuer toute idée de classe. Nous avons mis en œuvre cette méthode par une machine à noyau et l'avons adaptée aux bibliothèques d'optimisation standard par le biais d'un méta-noyau, avant de la tester sur un jeu de données artificielles et de la comparer à la méthode de rangement d'instances.

Partie 3 : Application à l'évaluation du lustre des perles de Tahiti

Chapitre 6 : De l'évaluation du lustre des perles de Tahiti

Dans le chapitre précédent, nous avons proposé un algorithme de rangement puis l'avons testé sur un jeu de données générées artificiellement. Le présent chapitre introduit un exemple concret de problème implicite de rangement d'objets susceptible de servir de test à l'algorithme de rangement d'objets proposé : l'évaluation du lustre des perles de Tahiti. Nous abordons l'évaluation du lustre des perles de Tahiti tant sous un angle théorique (propriétés de la nacre et notion de lustre) que sous un angle pratique (pratique effective du tri des perles en Polynésie française), et montrons son caractère de problème implicite de rangement d'objets.

I : Propriétés physiques et optiques de la nacre

Une perle de culture consiste en un nucléus enrobé de nacre ; les aspects physiques du lustre empruntent donc aux propriétés optiques de la nacre. Certaines propriétés de la nacre sont communes à toutes les perles ; certaines autres sont spécifiques aux perles de Tahiti.

A : Propriétés communes aux perles

Une présentation complète de la structure de la nacre a été faite par Marin et al. [100]. Dans la perspective d'évaluer le lustre des perles, il convient de relever que la nacre est composée de plaquettes d'aragonite (cristal d'environ 0.5µm d'épaisseur) cimentées par de la conchyoline (une matière organique sécrétée par le mollusque). Ces plaquettes sont structurées en couches parallèles ; un calcul rapide montre que la nacre contient environ 2000 couches d'aragonite par millimètre d'épaisseur.

L'aragonite est transparente, tandis que la conchyoline contient des pigments. Leur structuration en couches parallèles détermine certaines des propriétés optiques de la nacre

[101]. Ainsi, l'orient (une des formes de l'iridescence[14] dans le cas des perles) résulte des multiples réflexions et réfractions de la lumière sur les interfaces entre les différentes couches d'aragonite et de conchyoline : plus il y a de couches, plus l'orient est prononcé. De même, les plaquettes d'aragonite sont intercalées avec de la conchyoline, or la conchyoline contient des pigments : plus il y a de couches, plus la perle est colorée.

L'aragonite est un matériau biréfringent, toutefois, un appareil photo équipé d'un filtre polarisant ne permet pas de détecter la polarisation due à la biréfringence sur une perle. La seule polarisation détectée réside sur les pourtours de la perle, là où la lumière se reflète à angle rasant ; cette observation s'explique cependant par les équations de Fresnel et n'est donc pas spécifique aux perles (le phénomène s'observe d'ailleurs aussi bien sur des perles artificielles).

Enfin, la nacre n'est pas nécessairement répartie de manière uniforme autour du nucléus ; le lustre n'est en conséquence pas nécessairement homogène sur la surface de la perle.

B : Particularité des perles de Tahiti

La perle de Tahiti est produite par un mollusque appelé pinctada margaritifera. D'une part, la conchyoline sécrétée par ce mollusque contient des pigments noirs [102] ; d'autre part, sa nacre a l'indice de texture le plus élevé de tous les mollusques [103], résultant en des couches de nacre plus compactes, c'est-à-dire en un plus grand nombre de couches pour une même épaisseur de nacre que la nacre des autres mollusques. En conséquence, la perle de Tahiti est plus sombre, plus colorée, et présente un orient plus soutenu que les autres perles pour une même épaisseur de nacre.

---

[14] L'interférence de la lumière, soit à la surface (« iridescence » dans le cas des perles) soit à l'intérieur d'un matériau (« orient » dans le cas des perles), produit une série de couleurs lorsque l'angle d'incidence varie. ("interference of light either at the surface or in the interior of a material that produces a series of colors as the angle of incidence changes" [152]).

II : Définition du lustre

Le lustre est communément défini soit comme le reflet de la lumière sur une surface[15] soit comme l'apparence de ce reflet[16]. Afin de lever toute ambiguïté pour la suite, nous qualifierons la première définition de *physique* et la seconde de *perceptuelle*.

Bien que synonyme de « brillant » [104], [105], « lustre » est préféré à « brillant » lorsque l'attention porte sur le matériau dont est composée la surface réfléchissante plutôt que sur le phénomène de réflexion lui-même (notamment en matière de gemmes [105], [106] et de tissus [107][108][109]) . Dans ce sens, concernant la perle, certains experts en perles définissent le lustre comme la capacité de réfléchir la lumière en profondeur, à travers les couches parallèles d'aragonite.

L'institut gemmologique américain (Gemological Institute of America, GIA) donne une définition incidente du lustre des perles en fournissant une liste d'adjectifs susceptibles de qualifier les réflexions d'une perle lors de son évaluation [110]. Un lustre de bonne qualité correspondrait à une réflexion lumineuse ("bright") et nette ("sharp", "distinct") tandis qu'un lustre de mauvaise qualité correspondrait à une réflexion terne ("weak", "dim") ou floue ("hazy", "blurred"). Plus spécifiquement, l'Assemblée de la Polynésie française (APF), dans un texte normatif [111], dispose que le mot « lustre » peut être indifféremment remplacé par « éclat » ou « brillance », et le définit comme « la réflexion plus ou moins parfaite de la lumière sur la surface de la perle. » Un lustre excellent est dit correspondre « à une réflexion totale de la lumière, donnant un effet miroir » tandis qu'« une perle sans lustre correspond à un aspect mat de sa surface. »

Tant selon le GIA que selon l'APF, la qualité du lustre dépendrait des paramètres physiques de la nacre : pour le premier, une épaisseur insuffisante de la couche de nacre entraîne un lustre de

---

[15] Par exemple : « brillant, poli naturel ou artificiel d'une surface » (dictionnaire de l'académie française [104]) ; « éclat que présente une surface brillante » ("brightness that a shiny surface has", Cambridge dictionnary [153]) ; « manière dont la surface reflète la lumière », "manner in which the surface … reflects light", Oxford dictionnary [105]) ; etc.

[16] Par exemple : « apparence d'une surface selon la manière dont elle réfléchit la lumière » ("appearance of a … surface in terms of its light-reflective qualities", Encyclopedia Britannica [106]).

mauvaise qualité ; pour la seconde, la qualité du lustre est déterminée à la fois par l'épaisseur et par la structure de la couche de nacre.

La notion de lustre paraît complexe, voire ambiguë (dualité physique/perceptuelle), et la présence d'une définition normative du lustre de la perle de Tahiti ajoute à la difficulté de décider quelle définition utiliser. Cependant, le but d'une étude sur l'évaluation du lustre des perles de Tahiti n'est pas de décider quelle signification doit revêtir le mot « lustre » mais d'appréhender ce qui est effectivement mis en œuvre lorsqu'une perle de Tahiti est évaluée selon un critère appelé « lustre. » À ce titre, la définition permettant d'acquérir le plus d'information sur l'opération d'évaluation du lustre est la plus large, c'est-à-dire celle selon laquelle *le lustre est l'apparence du reflet de la lumière sur la surface d'une perle*. De ce point de vue, le lustre est une réalisation particulière du brillant au sens de la Commission internationale de l'éclairage (CIE) [112].

III : Aperçu de la pratique de l'évaluation du lustre des perles à Tahiti

Les perles de Tahiti sont les perles cultivées en Polynésie française à partir du mollusque appelé pinctada margaritifera [111].

L'évaluation du lustre des perles de Tahiti se réfère au processus dans lequel des experts se voient présenter des perles de Tahiti, en tant qu'objets physiques, et évaluent leur lustre, c'est-à-dire délivrent des jugements de valeur liés à l'attribut des perles qu'ils appellent « lustre ».

Le terme « experts » désigne ici les personnes dont l'évaluation de la qualité des perles de Tahiti fait partie de leur activité professionnelle quotidienne, tels les producteurs de perles de Tahiti, les grossistes, les détaillants, les bijoutiers, et les agents de la Direction des Ressources Marines et Minières (DRMM). Dans cette section, nous étudions la pratique actuelle de l'évaluation du lustre des perles de Tahiti de différents points de vue. Cet aperçu est le résultat d'entretiens avec 9 experts exerçant à Tahiti, de l'observation de leurs façons de trier les perles de Tahiti, ainsi que d'une expérience à laquelle ils ont accepté de participer.

A : Environnement

1 : Conditions d'illumination

La plupart des experts que nous avons rencontrés évaluent la qualité des perles à la lumière du jour. Certains le font à l'extérieur, mais la plupart le font à l'intérieur. Parmi ces derniers, la plupart le font face à une fenêtre, mais pas nécessairement, et souvent avec un rideau blanc devant la fenêtre. Ces experts évaluent généralement la qualité des perles soit toujours le matin, soit toujours l'après-midi, en fonction de l'orientation du soleil par rapport à la fenêtre. Certains experts combinent la lumière du jour avec un éclairage au néon ; nous avons même rencontré un expert qui évalue la qualité des perles en utilisant exclusivement un éclairage au néon afin de garantir des conditions d'éclairage constantes.

Tous les experts que nous avons rencontrés posent les perles sur une surface blanche pour en évaluer la qualité. Il peut s'agir d'une assiette blanche (principalement chez les bijoutiers) ou d'une table recouverte d'une nappe blanche (principalement chez les grossistes). Certains experts portent même des vêtements blancs pour évaluer la qualité des perles.

2 : Mouvement et toucher

Il a été démontré que le mouvement améliore la perception du brillant [113]. En pratique, tous les experts font tourner les perles lors de l'évaluation de leur lustre. Comme la plupart d'entre eux les font tourner entre leurs doigts, le sens du toucher pourrait entrer en jeu ; en effet, le sens du toucher peut améliorer la perception visuelle [114]. Cependant, tous les experts ont rejeté cette hypothèse, et certains ont montré qu'ils pouvaient aussi manipuler les perles avec des outils sans perturber leur évaluation.

B : Stratégies d'évaluation

L'évaluation des perles consiste à les distribuer dans un ensemble de classes ordonnées. Cependant, le nombre de classes et leurs spécifications peuvent varier d'un expert à l'autre. De même, il existe différentes méthodes de tri des perles dans ces classes.

1 : Le tri traditionnel par classes

En Polynésie française, les prescriptions normatives [111] détaillent la nomenclature à utiliser pour l'évaluation de la qualité des perles. Chaque classe a quatre attributs : le diamètre, qui doit être exprimé en millimètres arrondis au millimètre près ; le poids, en grammes ; la forme, qui peut prendre des valeurs dans une liste limitative (« ronde », « semi-ronde », etc.) ; la qualité de la surface, qui peut prendre des valeurs de « A » à « D », « A » se rapportant à la plus haute qualité, « D » à la plus basse. Ce dernier attribut a trois composantes : le lustre, la couleur et la régularité de la surface. Cette nomenclature est suivie à peu près. Certains experts trient plus finement, rendant le lustre, la couleur et la régularité de la surface explicites. Il arrive aussi que certains utilisent d'autres libellés (par exemple « AA », « AA + », etc.).

2 : Un problème implicite de rangement d'objets

Le nombre de classes de qualité est traditionnellement de 4 (5 en considérant les rebuts comme formant une classe en soi), cependant tous les experts que nous avons consultés affirment pouvoir ordonner totalement les perles (certains en ont d'ailleurs fait la démonstration sur de petits jeux de perles). Il est rare cependant qu'ils aient à le faire, la classe représentant généralement un niveau satisfaisant de spécification de la qualité requise par le marché.

Il est fréquent qu'au cours du processus de tri d'un lot de perles, les experts changent de classes certaines perles préalablement classées au moment d'en classer de nouvelles ; la raison alors invoquée est l'homogénéité des classes et la relativité de leur signification au regard de la qualité des perles : les perles pouvant être ordonnées totalement, il est possible de faire glisser les frontières entre les classes le long de l'ordre formé par les perles sans incidence sur celui-ci.

Le fait que les perles puissent être totalement ordonnées par les experts indique qu'il ne s'agit pas d'un authentique problème de classement, même ordinal, mais d'un problème implicite de rangement. Le fait que les classes n'aient pas de signification intrinsèque en fait un problème implicite de rangement d'objets.

3 : Méthodes de tri

En pratique, le lustre n'est pas évalué directement mais avec la couleur et la régularité de la surface. Le classement des perles selon leur lustre seul n'est pas une opération courante pour

les experts ; ce critère intervient généralement seulement sur des ensembles de perles homogènes en ce qui concerne d'autres critères tels que la forme et la taille. Il existe plusieurs méthodes pour trier les perles ; ces méthodes peuvent être scindées en deux familles : tri individuel et tri par lots.

a : Tri individuel

Dans un mode de tri individuel, l'expert considère une perle et la compare à un standard (idéal ou réel) de chaque classe, puis range cette perle dans la classe dont la distance perceptuelle entre le standard et la perle évaluée est minimale. Avec cette méthode, chaque perle est observée une fois, et le résultat final dépend des standards utilisés.

Cette méthode donne l'impression que chaque perle est directement évaluée, en un seul coup, car les experts sont entraînés et ont les standards en tête. Cependant, souvent, lorsque les classes commencent à contenir plusieurs perles, les experts affinent leur classification en déplaçant certaines perles d'une classe à l'autre, afin d'assurer l'homogénéité des classes.

b : Tri par lot

Dans un mode de tri par lots, l'expert considère un lot de perles et le filtre de manière itérative selon différents critères, jusqu'à obtenir les classes *in fine*. Avec cette méthode, chaque perle est observée plus d'une fois, et sa classe à l'issue de l'opération peut varier en fonction de l'ensemble initial à partir duquel la perle a été prise.

Pour certains experts, la forme et la taille sont les critères les plus importants car ils les disent influencer en grande partie le prix. Pour d'autres, le lustre et la couleur sont les plus importants car ils les disent influencer en grande partie la décision d'achat. Cependant, cet ordre « économique » ne dicte pas nécessairement l'ordre dans lequel les critères sont utilisés pour trier les perles : parmi les experts évaluant la qualité des perles par lot, tous sauf un commencent par la taille, puis la forme, avant de diverger quant à l'ordre des critères restants. L'expert différant des autres commence par le lustre, puis la couleur et l'état de la surface, lui permettant d'obtenir 4 classes de perles (A, B, C et D) ; il trie ensuite chaque classe par taille et termine avec la forme.

En général, les critères peuvent déboucher sur plus de deux valeurs (par exemple, le tri des perles après la forme peut donner « vert clair », « vert foncé », « paon », etc.). Certains experts font alors des subdivisions pour avoir à chaque étape un choix binaire à faire (par exemple « couleurs claires » contre « couleurs foncées », ou « très lustré » contre « peu lustré ») en affinant itérativement chaque catégorie.

C : Diversité des évaluations

Selon tous les experts que nous avons interrogés, l'évaluation de la qualité des perles est un processus hautement subjectif. D'après eux, les évaluations varient non seulement entre experts, mais aussi pour un même expert en fonction des conditions internes et externes. Par exemple, cela peut dépendre de l'humeur de l'expert, du but de l'évaluation, de la lumière, etc. Ces témoignages sont cohérents avec les idées théoriques sur la dualité physique / perceptuelle du lustre (supra).

1 : Expérience

Nous avons demandé à neuf experts d'évaluer le lustre de 30 perles. L'évaluation a consisté à placer chaque perle dans une des 4 classes de A à D (par ordre décroissant de qualité) ou dans une classe « rebuts » (lorsque la qualité est tellement basse que la perle est retirée du circuit). Puisque l'échelle perceptuelle n'est pas nécessairement linéaire, nous n'avons pris en compte que les rangs pour mesurer les similitudes et avons à cet égard utilisé le coefficient de corrélation τ-b de Kendall [83]. Nous reportons chaque résultat avec sa valeur-p et fixons pour cette dernière le seuil de significativité à 0.05.

À titre incident, nous reportons aussi l'environnement, la stratégie d'évaluation et la définition du lustre correspondants à chaque expert.

Chaque évaluation n'a pas toujours été faite par un seul expert. Par exemple, une évaluation a été faite par deux experts ensemble et une autre a été faite par un expert qui a parfois sollicité l'avis de ses collègues. Dans les deux cas, nous les notons comme un seul expert puisqu'ils donnent chaque fois une évaluation : seul compte le fait que chaque évaluation a été réalisée dans les mêmes conditions que celles dans lesquelles ces experts travaillent habituellement.

2 : Observations

Les experts qui ont évalué le lustre l'ont fait dans leur environnement de travail habituel. Alors même que l'évaluation du lustre ne porte que sur un seul paramètre, nous avons observé la même diversité dans les méthodes que pour l'évaluation de la qualité globale. Certains experts placent les perles directement dans les différentes classes de lustre, quitte à modifier les classes au cours de l'évaluation ; d'autres procèdent itérativement en divisant à chaque étape l'échantillon en sous-échantillons jusqu'à ce que les classes apparaissent (par exemple en supprimant dans un premier temps tous les rebuts, puis en écumant de manière réitérée les perles de moins bon lustre sur le résidu jusqu'à ce qu'il n'en reste plus). Ces méthodes correspondent respectivement à des méthodes de tri individuel et par lots.

Quelle que soit la méthode utilisée, et pour la plupart des experts, les frontières entre les classes ont évolué au cours du processus de tri (des perles étant changées de classe au fur et à mesure que de nouvelles perles étaient triées), à l'instar de ce qui se fait dans le cadre de l'évaluation globale de la qualité des perles.

3 : Résultats

Le Tableau 9 présente une mesure des similitudes entre les experts lors de l'évaluation de la qualité des perles selon leur lustre. La corrélation est calculée en utilisant le τ-b de Kendall ; chaque corrélation est présentée avec sa valeur-p (entre parenthèses).

Le Tableau 10 présente pour chaque expert les conditions d'éclairages dans lesquelles il travaille, la méthode qu'il utilise, et un mot ou une expression rendant compte en substance de sa définition du lustre.

Tableau 9 : Matrice de corrélation (avec les valeurs-p respectives) entre les évaluations par 9 experts d'un jeu de 30 perles selon leur lustre. L'évaluation consiste à placer chaque perle dans une classe de qualité de lustre (de A à D par ordre décroissant de qualité, plus une classe pour les rebuts). La corrélation entre deux évaluations est calculée en utilisant le τ-b de Kendall.

Expert	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
2	0.6306 p < 0.01							
3	0.3972 p = 0.01	0.3867 p = 0.02						
4	0.4165 p < 0.01	0.2735 p = 0.09	0.4332 p < 0.01					
5	0.4395 p < 0.01	0.4945 p < 0.01	0.5241 p < 0.01	0.3864 p = 0.02)				
6	0.5871 p < 0.01	0.5224 p < 0.01	0.6823 p < 0.01	0.4765 p < 0.01	0.5008 p < 0.01			
7	0.4220 p < 0.01	0.4678 p < 0.01	0.2943 p = 0.06	0.1819 p = 0.24	0.2008 p = 0.21	0.3705 p = 0.02		
8	0.6221 p < 0.01	0.5839 p < 0.01	0.3749 p = 0.02	0.2175 p = 0.17	0.3485 p = 0.03	0.5073 p < 0.01	0.5324 p < 0.01	
9	0.3857 p = 0.02	0.4184 p = 0.01	0.5438 p < 0.01	0.4365 p < 0.01	0.5696 p < 0.01	0.4941 p < 0.01	0.3059 p = 0.06	0.3418 p = 0.04

Tableau 10 : Indication, pour chaque expert, des conditions d'éclairage dans lesquelles l'évaluation du lustre a été effectuée, de la méthode qui a été utilisée ainsi que d'une définition succincte du lustre.

Expert	*Éclairage*	*Méthode*	*Définition succincte du lustre*
1	Naturel	Tri par lots	« Réflexion en surface »
2	Naturel + artificiel	Tri par lots	« Brillant »
3	Naturel + artificiel	Tri par lots	« Effet miroir »
4	Artificiel	Tri par lots	« Réflexion en surface »
5	Naturel + artificiel	Tri individuel	« Orient »
6	Naturel	Tri individuel	« Éclat »
7	Naturel + artificiel	Tri individuel	« Iridescence »
8	Naturel	Tri individuel	« Effet miroir »
9	Naturel	Tri par lots	« Réflexion en profondeur »

4 : Interprétation

a : Relation entre les évaluations

Tous les coefficients sont positifs, ce qui signifie que tous les experts ont classé les perles dans la même direction.

La force des relations entre les évaluations est généralement modérée (de 0.4 à 0.6) avec quelques relations faibles (de 0.2 à 0.4), ce qui signifie que les experts ne sont pas nécessairement d'accord dans le détail. Néanmoins, il se peut que certaines directions spécifiques émergent ; en effet, les experts 1 et 2, 1 et 8, ainsi que les experts 3 et 6, ont rendu des évaluations fortement corrélées (de 0.6 à 0.8), et pour chacune de ces corrélations, la valeur-p est inférieure à 0.01.

Pour résumer, les experts s'accordent sur une manière générale d'évaluer le lustre des perles, mais chacun a sa propre façon de l'évaluer en détail.

b : Relations entre éclairage, méthode, définition et évaluation

Il est remarquable que d'un côté certains experts qui évaluent le lustre dans différentes conditions d'éclairage et suivent des méthodes différentes donnent des résultats très proches (par exemple les experts 3 et 6) alors que, d'un autre côté, parmi les coefficients de corrélation les plus faibles, se trouvent des paires d'experts dont les environnements et les méthodes sont à peu près les mêmes (par exemple les experts 2 et 3). Par conséquent, il semble qu'une évaluation ne soit pas déterminée par l'éclairage ou la méthode choisis.

De même, il ne semble pas y avoir de dépendance entre les définitions du lustre données par les experts et la manière dont ils évaluent le lustre. Par exemple, les experts 1 et 8 ont des évaluations fortement corrélées malgré des définitions différentes du lustre ; les experts 3 et 8 ont des évaluations faiblement corrélées malgré une même définition du lustre.

5 : Conclusion

Les perles ont une dimension esthétique qui peut justifier la diversité relative des évaluations. La diversité des méthodes, pour sa part, peut s'expliquer par l'échelle à laquelle l'opération d'évaluation a lieu.

Tous les experts ayant de grandes quantités de perles à trier utilisent la méthode de tri par lots. Cela peut paraître surprenant à première vue, puisque chaque perle est observée autant de fois qu'il y a de critères. Néanmoins, tous s'accordent pour dire que cette voie est beaucoup moins fatigante et moins sujette aux erreurs. Les critères binaires permettent d'aller encore plus loin en renforçant la simplicité et la fiabilité de chaque opération. En outre, si le tri individuel est décomposé en une série de comparaisons entre une perle donnée et toutes les normes, alors il y a globalement plus de comparaisons que dans le tri par lots.

Le tri individuel nécessite un effort qui ne peut être maintenu sur une longue période, cette méthode est donc plutôt utilisée par les détaillants ou les bijoutiers, qui ont moins de perles à expertiser que les producteurs ou les grossistes.

IV : Synthèse

Dans ce chapitre, nous avons introduit les propriétés de la nacre des perles en soulignant les spécificités de la nacre des perles de Tahiti, se traduisant par une tendance à des perles plus sombres, plus colorées et présentant un orient plus soutenu que les autres perles.

Nous avons ensuite proposé une définition de la notion de lustre permettant de servir de base à l'acquisition des informations nécessaires à l'évaluation du lustre, à savoir le lustre comme *l'apparence du reflet de la lumière sur la surface d'une perle*.

Enfin, nous avons dressé un aperçu de l'évaluation du lustre des perles de Tahiti telle qu'elle est effectivement pratiquée en Polynésie française. L'étude des stratégies d'évaluation montre que l'évaluation du lustre des perles de Tahiti telle qu'elle est pratiquée en Polynésie française est un problème implicite de rangement d'objets ; à ce titre, il s'agit d'un problème concret auquel l'algorithme de rangement d'objets peut s'appliquer.

Les chapitres suivants s'attachent à définir une manière de représenter numériquement le lustre des perles de Tahiti afin de pouvoir appliquer à ce problème l'algorithme de rangement d'objets proposé au chapitre précédent.

Chapitre 7 : Modélisation du lustre des perles

Afin d'être en mesure d'appliquer un algorithme d'apprentissage automatique à l'évaluation du lustre des perles de Tahiti, il est nécessaire que le lustre de chaque perle de Tahiti soit représenté numériquement. Nous discutons dans ce chapitre des deux modèles possibles du lustre : le lustre comme fonction de la surface de la perle et le lustre comme composition de caractéristiques d'images de la perle.

I : Le lustre d'une perle en tant que fonction de la surface de la perle

La plupart du travail effectué sur la réflectance utilise des fonctions de distribution de réflectance bidirectionnelle (BRDF) [115] pour représenter la réflectance d'un matériau. Les BRDF se réfèrent à une classe de fonctions qui prennent deux directions comme entrée et renvoient un coefficient reliant le rayonnement entrant à un point dans une direction et le rayonnement sortant du même point dans l'autre sens (Figure 4).

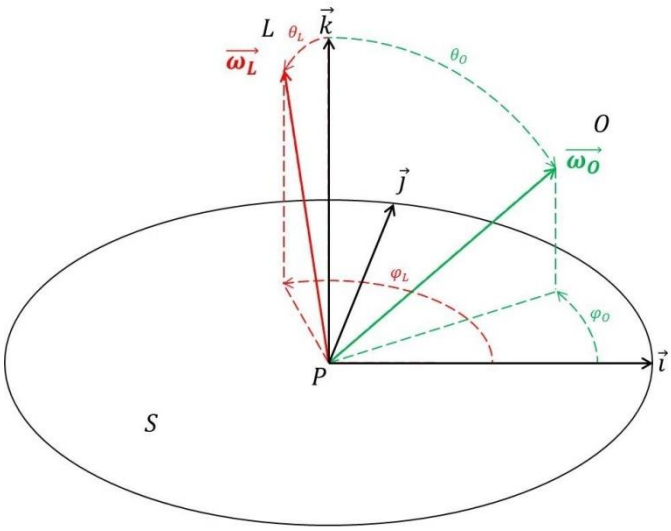


Figure 4 : Soit S une surface plane, P un point de S, L une source lumineuse et O an observateur. Soit ω_L et ω_O les directions de P vers L et O respectivement. BRDF(ω_L, ω_O), ou de façon équivalente BRDF(θ_L, φ_L, θ_O, φ_O), est la partie du rayonnement de L incident à S en P qui est réfléchie en O.

Les BRDF sont essentiellement des fonctions à quatre variables, chaque vecteur de direction étant spécifié par deux angles. Pour les matériaux isotropes (matériaux dont la réflectance est invariante par rotation autour de la normale), la différence entre les deux angles azimutaux est suffisante pour spécifier totalement la configuration, le nombre de variables peut donc être réduit à 3. Le plus souvent, les BRDF sont étendues afin de prendre en compte des paramètres supplémentaires, faisant croître le nombre de variables de manière adéquate ; par exemple, les BRDF peuvent prendre en compte la longueur d'onde de la lumière incidente, comprenant alors 5 variables (4 variables pour les matériaux isotropes). Comme toutes les fonctions, les BRDF peuvent être définies soit en extension (de manière descriptive), soit en intension (de manière prescriptive).

A : Les fonctions basées sur des mesures

Une BRDF peut consister en un ensemble de mesures brutes, sans modèle sous-jacent [116]–[118] ; appeler une telle fonction revient à interroger une base de données. Cette fonction est parfois appelée BRDF « basée sur le stockage », « basée sur la table » ou « basée sur la mémoire ».

Pour définir une BRDF basée sur des mesures, des mesures de réflectance pour tous les éléments d'entrée possibles doivent être fournies. Cependant, comme l'espace d'entrée n'est ni

fini ni discret, les BRDF basées sur des mesures ne peuvent être que partiellement définies, plus ou moins densément.

B : Les fonctions basées sur des modèles

Une BRDF peut également implémenter un modèle de réflectance. Plutôt que de stocker tout un ensemble de mesures décrivant la fonction en termes de couples <entrée, sortie>, la fonction s'appuie sur un modèle pour calculer les éléments de sortie directement à partir des éléments d'entrée. Les BRDF basées sur des modèles ont l'avantage sur les BRDF basées sur des mesures d'être totalement définies. Le modèle lui-même peut être prédéfini ou déduit des données.

1 : Modèles prédéfinis

Les BRDF peuvent implémenter des modèles préexistants. Ces modèles peuvent être empiriquement valides [118], [119] ou non [120], physiquement corrects [121]–[124], [118] ou non [120]. En effet, dans de nombreuses applications, la praticabilité est plus importante que la validité empirique ou la correction physique (selon Ward, cette raison explique le succès du modèle de Phong [120]), même si elles ne sont pas incompatibles [118]. Des heuristiques perceptuelles peuvent permettre de simplifier les calculs sans pour autant sacrifier la vraisemblance [125]–[129] cité par [130].

2 : Modèles inférés

Les modèles sur lesquels sont basées les BRDF peuvent eux-mêmes être déduits des données [119]. Cette « approche de modélisation de la réflectance pilotée par les données » [119] consiste à partir d'un ensemble de BRDF basées sur des mesures, lequel est alors considéré comme un échantillon tiré de l'espace des BRDF valides puis à y appliquer une réduction de dimensionnalité.

C : Paramétrage des fonctions bidirectionnelles de distribution de réflectance

Pour qu'une BRDF représente la réflectance d'un matériau donné, elle doit être correctement paramétrée.

Pour les BRDF basées sur des mesures, trouver les valeurs de paramètres adéquates consiste à mesurer la réflectance à n'importe quel point de l'espace d'entrée.

Lors de l'utilisation de BRDF basées sur un modèle, la seule information nécessaire serait en théorie l'information correspondant aux paramètres du modèle. Cependant, il peut être impossible de mesurer directement ces paramètres. Si tel est le cas, des informations relatives à la réflectance doivent être acquises pour un échantillon d'éléments d'entrée (comme pour le paramétrage des BRDF par mesures) et utilisées pour ajuster la BRDF basée sur un modèle afin de déduire les paramètres des modèles [131], [118]. Certains problèmes liés à cette approche sont soulignés par Matusik [119], notamment le choix d'une métrique d'erreur pertinente (« La distance euclidienne est une mauvaise métrique car elle tend à surestimer l'importance des pics spéculaires et à ignorer les valeurs hors spéculaire » [119]. Pour une métrique basée sur la signification perceptuelle, voir [132] cité par [130]), et le risque de bloquer sur des minima locaux compte tenu de la non-linéarité de la plupart des modèles. Alors qu'une BRDF basée sur des mesures fournit des informations exactes mais partielles sur la distribution de la réflectance décrite, une BRDF basée sur un modèle fournit des informations complètes mais approximatives.

D : Mesure de la réflectance

Les BRDF sont mesurées à l'aide d'instruments appelés gonioréflectomètres. Ces instruments permettent de tester tous les couples de directions en laissant la source lumineuse et le capteur bouger. Ces mesures peuvent être longues et coûteuses, notamment pour les matériaux anisotropes [118]. Pour surmonter ce problème, des techniques de mesure basées sur l'image ont été conçues [116], [118], [117], qui sont moins coûteuses, mais qui prennent encore beaucoup de temps [119].

E : Discussion : application au lustre des perles

Les BRDF capturent les propriétés de réflectance des matériaux, et non des objets, sauf à la rigueur ceux faits de matériaux uniques. Puisque la qualité de la nacre n'est pas nécessairement homogène sur la surface des perles, chaque point de la surface est susceptible d'avoir une BRDF différente. Afin de capturer la non-uniformité du matériau sur une surface, une fonction SVBRDF (fonction de distribution de la réflectivité bidirectionnelle variable spatialement) est nécessaire [133], [134]. Cette fonction a deux autres variables pour spécifier le paramètre d'emplacement de surface.

Même si les BRDF étaient bien adaptées pour décrire la réflectance des matériaux pour lesquels la lumière émergeant d'un point donné de la surface est déterminée par la lumière entrante à ce même point de la surface, une telle hypothèse ne peut être faite pour un matériau comme la nacre, où la lumière incidente à la surface à un endroit donné peut être diffusée par les couches successives d'aragonite avant de sortir à un endroit différent. Une telle fonction existe ; il s'agit de la BSSRDF (fonction de distribution de réflectance de diffusion de surface bidirectionnelle). Dans cette fonction, chaque direction est donnée comme une entrée avec le point sur la surface où l'interaction lumière-surface est considérée (sur les surfaces non opaques, la lumière incidente ne reflète pas totalement au point où elle frappe la surface mais est un observateur peut alors considérer des points distincts du point d'incidence. Une BRDF est juste un cas particulier de BSSRDF où les points incidents et observés coïncident). Cette fonction a deux autres variables pour spécifier l'emplacement du point observé de la surface par rapport à l'emplacement du point d'incidence.

Pour capturer la propriété de réflectance d'un objet dont le matériau varie sur la surface, une SVBRDF est requise ; pour capturer la propriété de réflectance d'un objet non opaque, une BSSRDF avec un paramètre de longueur d'onde explicite est requise. Il s'ensuit qu'une "SVBSSRDF", c'est-à-dire une fonction d'au moins 8 variables, serait nécessaire pour modéliser le lustre des perles. Si la modélisation et l'acquisition d'une telle fonction est théoriquement possible, elle est impraticable compte tenu du nombre de dimensions à prendre en compte. En effet, la mesure de chaque combinaison de valeurs de paramètres fait croître exponentiellement le temps de mesure avec le nombre de paramètres, or même une technique

de mesure à grande vitesse basée sur l'image nécessite plusieurs heures pour mesurer une simple BRDF [119].

II : Le lustre d'une perle en tant que composition de caractéristiques d'images de la perle

Des études scientifiques s'appliquant au lustre peuvent être trouvées dans la littérature scientifique portant sur le brillant ("gloss" en anglais[17]). Bien que cette littérature ne porte pas spécifiquement sur le lustre, et *a fortiori* sur le lustre des perles de Tahiti, elle s'applique à ce dernier en ce qu'il est une réalisation, quoique dans le cas particulier des perles de Tahiti, du brillant.

Le brillant est communément assimilé à la réflectance spéculaire et il n'est guère étonnant que l'instrument de mesure du brillant, le brillancemètre (ou "glossmeter" en anglais), se contente de mesurer la réflectance d'une surface à divers angles[18]. Toutefois, il résulte de la littérature scientifique sur le brillant que les aspects physiques et perceptuels du brillant évoluent dans des espaces distincts. Deux approches sont alors possibles : l'approche fechnérienne et l'approche gestaltiste.

A : Dualité du lustre

Malgré le sens commun, le brillant n'est pas équivalent à la réflectance spéculaire. En d'autres termes, la réflectance spéculaire n'est ni une condition suffisante ni même une condition nécessaire du brillant. D'une part, la présence de réflectance spéculaire ne suffit pas à déterminer la présence de brillant. En effet, il existe des matériaux pour lesquels la réflectance spéculaire n'est pas perçue comme brillante : la neige par exemple, lorsqu'elle est observée à

[17] Le terme anglais "gloss" est parfois équivalent à « brillant » (par exemple [104]) et parfois à « brillance » (par exemple [111]). Le choix fait ici est de le traduire par « brillant », conformément au vocabulaire de la Commission électrotechnique internationale [154].

[18] Pour des exemples de brillancemètre, voir [155], [156].

angle rasant, ne brille pas tout en ayant une forte réflectance spéculaire[19]. D'autre part, la présence de réflectance spéculaire n'est pas nécessaire pour déterminer la présence de brillant. En effet, le brillant peut être généré à partir de lumière diffuse [135].

Pour récapituler, il est possible que la réflectance spéculaire ne soit pas perçue comme brillante tout comme il est possible qu'une surface soit perçue comme brillante en l'absence de réflectance spéculaire. Ceci démontre que le brillant, et partant le lustre, ne peut pas être réduit *a priori* à la réflectance spéculaire. Il convient donc d'étudier le lustre des perles de Tahiti en tant que phénomène perceptuel.

Deux approches visant à appréhender la perception et à offrir un cadre d'étude des phénomènes perceptuels méritent d'être introduites : l'une est centrée sur l'objet de la perception ; l'autre est centrée sur le sujet de la perception.

B : Approche fechnérienne

L'approche fechnérienne appréhende la perception comme un problème inverse dont la solution est complètement déterminée par l'objet perçu [136]. En d'autres termes, cette approche suppose que le cerveau soit capable, étant donnée l'image rétinienne d'un objet, de calculer le modèle physique de cet objet. En conséquence, les êtres humains seraient capables d'évaluer ce qu'ils voient de manière objective. Dans le cadre d'une approche fechnérienne, une distinction entre les aspects physiques et perceptuels du lustre des perles de Tahiti n'a pas de sens car le monde perceptuel est calqué sur le monde physique.

Cette approche ne paraît toutefois pas satisfaisante : l'image rétinienne est le résultat d'une projection d'un espace à trois dimensions sur un espace à deux dimensions. Cette projection implique une perte d'information, et à toute image rétinienne correspond potentiellement une infinité d'objets. Inverser le processus, c'est-à-dire obtenir le modèle d'un objet en trois dimensions à partir d'une image en deux dimensions, est impossible sans contraintes

[19] [157] tel que cité par [137].

supplémentaires[20]. Concernant le lustre des perles, c'est-à-dire la perception de la réflectance spéculaire sur la surface des perles, le manque d'information relative à la géométrie en trois dimensions de la surface empêche de modeler la réflectance spéculaire et, partant, d'évaluer objectivement et sans autre information le lustre.

Il est concevable que le fait de disposer de l'objet lui-même puisse permettre de recouvrer l'information nécessaire l'évaluation du brillant en observant l'objet sous différents angles ; toutefois, l'être humain est capable de percevoir le brillant même sur une seule image (par exemple sur une photo de perle), ce qui suggère qu'un autre type d'information entre en jeu.

Une approche alternative a été proposée pour répondre à ce problème [136]. Cette alternative suppose l'existence d'une connaissance *a priori* permettant de résoudre le problème inverse, mais il n'est pas spécifié si cette connaissance est objective (fondée sur genre de patrimoine commun) ou subjective (fondées sur l'expérience personnelle de l'observateur).

C : Approche gestaltiste

Il est maintenant reconnu quasi-unanimement que le cerveau perçoit le brillant par l'intermédiaire d'une composition de caractéristiques relevées dans l'image, c'est-à-dire, comme une Gestalt [137].

Une Gestalt est une entité vue « comme un tout, de par sa configuration, c'est-à-dire de par la relation, l'interaction et l'interdépendance entre ses parties, plutôt que par la somme ou une

[20] « Il est quasi-consensuel que le cerveau ne procède pas – et ne pourrait d'ailleurs procéder – véritablement à un calcul de la solution d'un problème inverse d'optique ; il dispose tout simplement de trop peu d'information pour inverser le processus physique de formation des images. » ("There is nearly universal agreement that the brain does not – indeed, could not – literally perform an inverse optics computation; there is simply not enough information available to invert the actual physics of image formation." [158])

quelconque combinaison de celles-ci. »[21] C'est « différent de (et pas plus que) la somme de ses composants, les composants étant les stimuli reçus du monde extérieur. »[22]

De ce point de vue, le lustre est évalué au travers d'un faisceau d'indices. Le cerveau ne tente pas d'estimer les paramètres physiques des matériaux et des objets [138] ; à la place, chaque individu extrait différentes caractéristiques et les combine à sa manière [139]. La perception du brillant (et partant celle du lustre) varie donc d'un individu à l'autre [135]. La question de l'existence de règles qui mettraient en évidence des comportements communs n'est pas tranchée [140].

III : Synthèse

Dans la perspective de représenter le lustre des perles numériquement, nous avons discuté dans ce chapitre des deux modèles possibles du lustre : le lustre comme fonction de la surface de la perle et le lustre comme composition de caractéristiques d'images de la perle. Le premier modèle s'avère n'être pas viable en pratique compte tenu du temps de mesure nécessaire dans le cas du lustre des perles. Le second, faisant la distinction entre l'objet physique et la perception de cet objet, permet une approche moins coûteuse du lustre, pour peu que les caractéristiques à partir desquelles se compose la représentation du lustre soient définies. Cette définition fait l'objet du prochain chapitre.

---

[21] "As a whole, resulting from its configuration, i.e., the relationship, interaction, and interdependence between its parts, rather than the sum or random combination of its parts" [159].

[22] "Different from (not more than) the sum of its components, the components being the stimuli received from the outside world" [159].

Chapitre 8 : Proposition de descripteurs du lustre des perles de Tahiti

L'évaluation de la qualité Lustre est faite par des experts humains ; nous visons à imiter ces experts. Si nous supposons que le système perceptif humain n'essaie pas de construire une représentation de la scène réelle mais utilise à la place des indices, alors nous pouvons nous demander si passer par la reconstruction de la scène pour en extraire des indices pourrait être évité. Nous avons vu que, dans la pratique, les paramètres des BRDF doivent être inférés par une phase d'apprentissage ; en d'autres termes, un biais *a priori* est introduit (en ajustant le modèle) qui pourrait être évité en utilisant des indices extraits directement de l'image. En outre, si le temps de traitement est un point crucial, il est préférable d'éviter les calculs supplémentaires.

Les perles existent en tant qu'objets physiques ; à ce titre, leur image dépend de leurs propriétés optiques. Elles existent toutefois pour un expert par le vecteur de son système perceptuel. Les propriétés physiques et perceptuelles du lustre ne coïncidant pas nécessairement, il convient de fonder l'élaboration des vecteurs de caractéristiques du lustre sur les aspects perceptuels du lustre et non sur ses aspects physiques.

Nous répertorions dans un premier temps les attributs perceptuels du lustre et considérons dans un second temps la manière de les quantifier afin d'en former des vecteurs exploitables numériquement.

Dans une perspective d'apprentissage automatique, ces vecteurs de caractéristiques ont pour objet de constituer l'interface entre le monde sensible et la machine. Incidemment (dans une perspective de métrologie douce [141]), ils pourraient servir à étudier les mécanismes à l'œuvre dans la perception du lustre.

I : Qualification des attributs perceptuels du lustre

Nous inventorions ici les attributs susceptibles d'offrir une emprise sur le lustre en tant que phénomène perceptuel. Certains de ces attributs ont été identifiés en tant qu'attributs du brillant et s'appliquent au lustre dans la mesure où ce dernier est une réalisation particulière du brillant ; certains attributs sont plus spécifiques au lustre des perles.

A : Attributs du lustre en tant que réalisation du brillant

Une liste des aspects perceptuels du brillant a été dressée par Hunter [142]. Hunter lui-même reconnaissait que cette liste n'était pas forcément exhaustive. Ces aspects sont le « brillant spéculaire » ("specular gloss"), « l'éclat » ("sheen"), le « brillant de contraste » ("contrast gloss"), le « brillant de netteté de l'image réfléchie » ("distinctness-of-reflected-image gloss"), le « brillant d'absence de brouillard » ("absence-of-bloom gloss"), et le « brillant d'absence de texture » ("absence-of-texture gloss").

Deux des aspects sont définis négativement (l'absence de brouillard et l'absence de texture), sans doute par souci de faire correspondre leurs sens de variation avec celui du brillant. Par souci de simplicité, et sans altération de leur pouvoir descriptif, nous ne retiendrons de ces aspects que leurs critères : le brouillard et la texture.

Les attributs du lustre correspondants aux aspects du brillant relevés par Hunter seront exposés respectivement sous les appellations suivantes : *brillant spéculaire*, *éclat*, *contraste*, *effet miroir* (pour la netteté de l'image réfléchie), *brouillard* et *interférence de la surface* (pour la texture).

Bien que ces attributs aient été identifiés dans le cadre de travaux portant sur le brillant, ils restent pertinents dans le cadre d'un travail sur le lustre dans la mesure où le lustre n'est qu'une réalisation particulière du brillant.

1 : Brillant spéculaire

Le brillant spéculaire est la perception de la réflectance spéculaire à angle non rasant. La Figure 5 montre deux perles de brillant spéculaire différent. Il est possible d'utiliser un appareil photographique comme un outil rudimentaire de mesure de réflectance spéculaire à angle non

rasant en calculant les valeurs d'intensité maximale de la région d'intérêt (ici, la surface correspondant à la perle) dans un espace de couleurs HSV. Dans la Figure 5, elle est de 93,33% pour la perle A et de 100% pour la perle B.

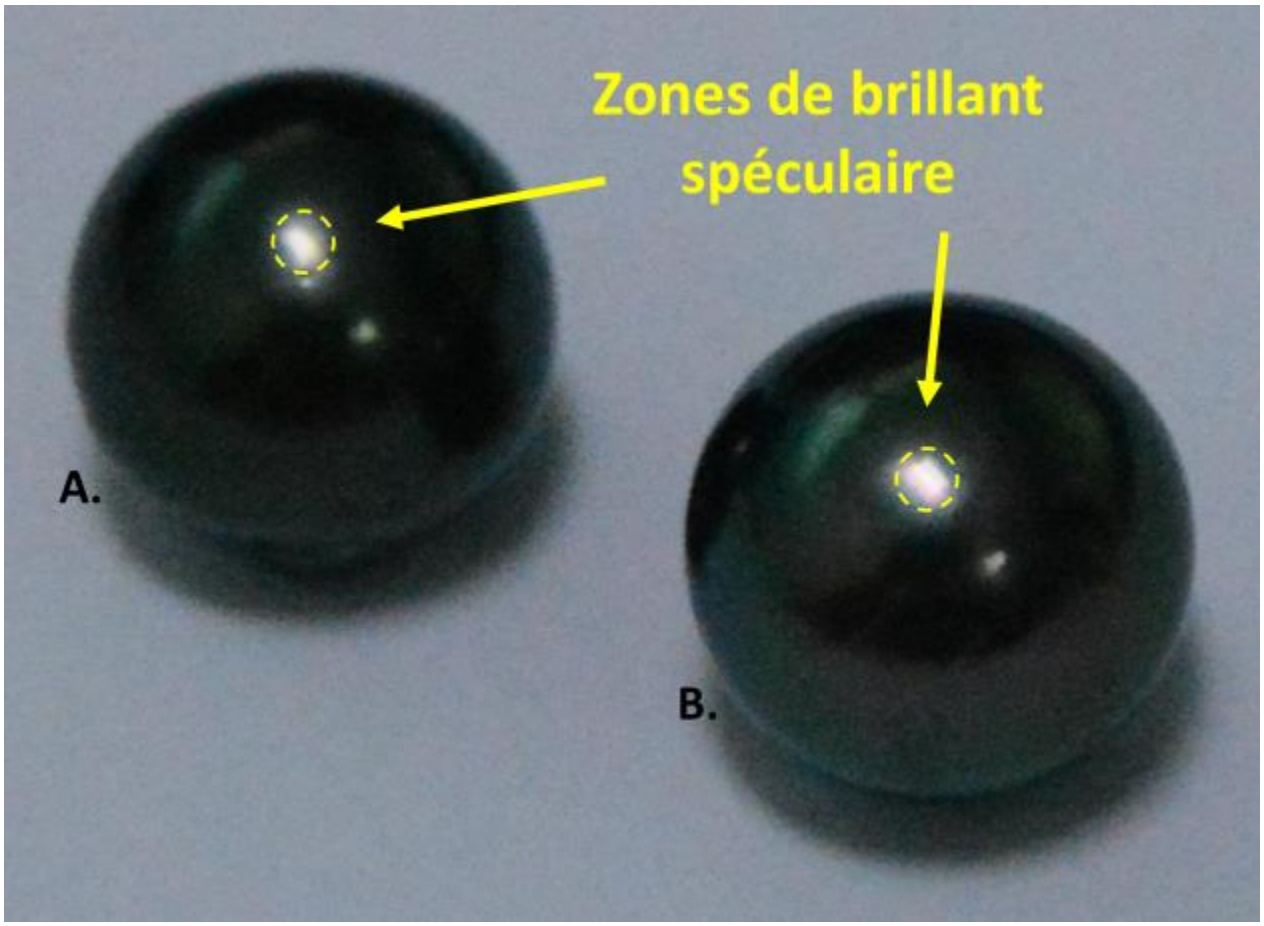


Figure 5 : Deux perles de brillant spéculaire différent (valeurs d'intensité maximale de l'image dans un espace couleur HSV : 93,33% pour la perle A, 100% pour la perle B).

2 : Eclat

L'éclat est la perception de la réflectance spéculaire à angle rasant. C'est un cas particulier de brillant spéculaire. La Figure 6 montre deux perles d'éclat différent (valeurs d'intensité maximale de l'image dans un espace de couleurs HSV : 96,47% pour la perle A, 99,22% pour la perle B).

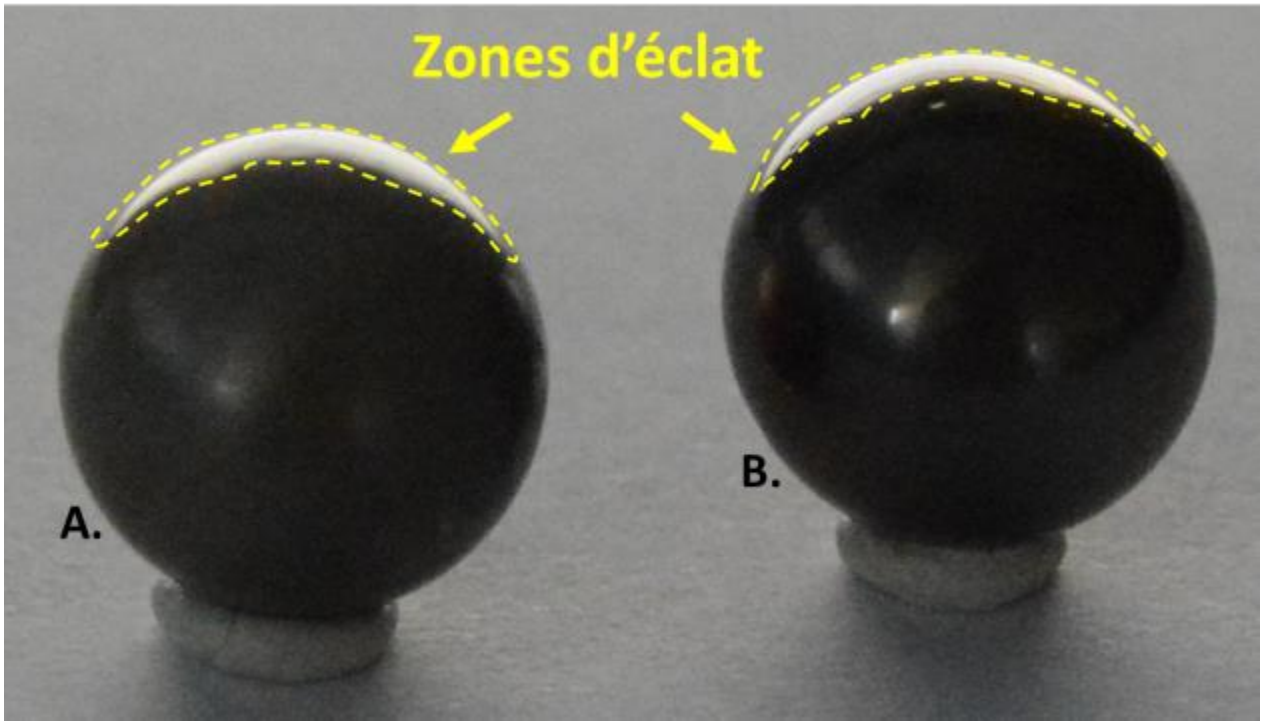


Figure 6 : Deux perles d'éclat différent (valeurs d'intensité maximale de l'image dans un espace couleur HSV : 96,47% pour la perle A, 99,22% pour la perle B).

Le brillant spéculaire et l'éclat, bien que tous deux exprimant la force de la réflectance spéculaire, ne sont pas nécessairement dépendants. Les cas où il y a à la fois le brillant spéculaire et l'éclat ou ni l'un ni l'autre sont triviaux. Le cas où il y a l'éclat sans le brillant spéculaire est justifié par les équations de Fresnel[23] ; par exemple, certaines peintures d'aspect mat à angle non rasant peuvent présenter une forte réflectance spéculaire à angle rasant. Enfin, le cas le plus contre-intuitif (car contredisant apparemment les équations de Fresnel), où il y a le brillant spéculaire sans l'éclat, peut se trouver dans « un certain nombre d'échantillons de fils et de papiers qui ont un aspect pelucheux » [142].

[23] L'éclat est lié à la lumière réfléchie à l'angle rasant (grand angle d'incidence), tandis que le brillant spéculaire est lié à la lumière réfléchie à un angle non rasant (petit angle d'incidence), cependant, selon les équations de Fresnel, plus l'angle d'incidence est grand, plus la fraction de lumière incidente réfléchie par la surface est élevée [160].

3 : Contraste

Le contraste correspond à la luminosité perçue d'une surface en raison du contraste entre la réflectance spéculaire et la réflectance diffuse. La Figure 7 souligne la différence entre le contraste et le brillant spéculaire : la perle B présente un contraste supérieur à celui de la perle A[24], tandis que la perle A possède un brillant spéculaire supérieur à celui de la perle B[25].

[24] Dans un espace colorimétrique HSV, la valeur d'intensité moyenne de a.2. est deux fois celui de a.1., tandis que celui de b.2. est 26 fois celui b.1.

[25] Valeurs maximales d'intensité de l'image dans un espace colorimétrique HSV: 93,73% pour la perle A; 93.33% pour la perle B.

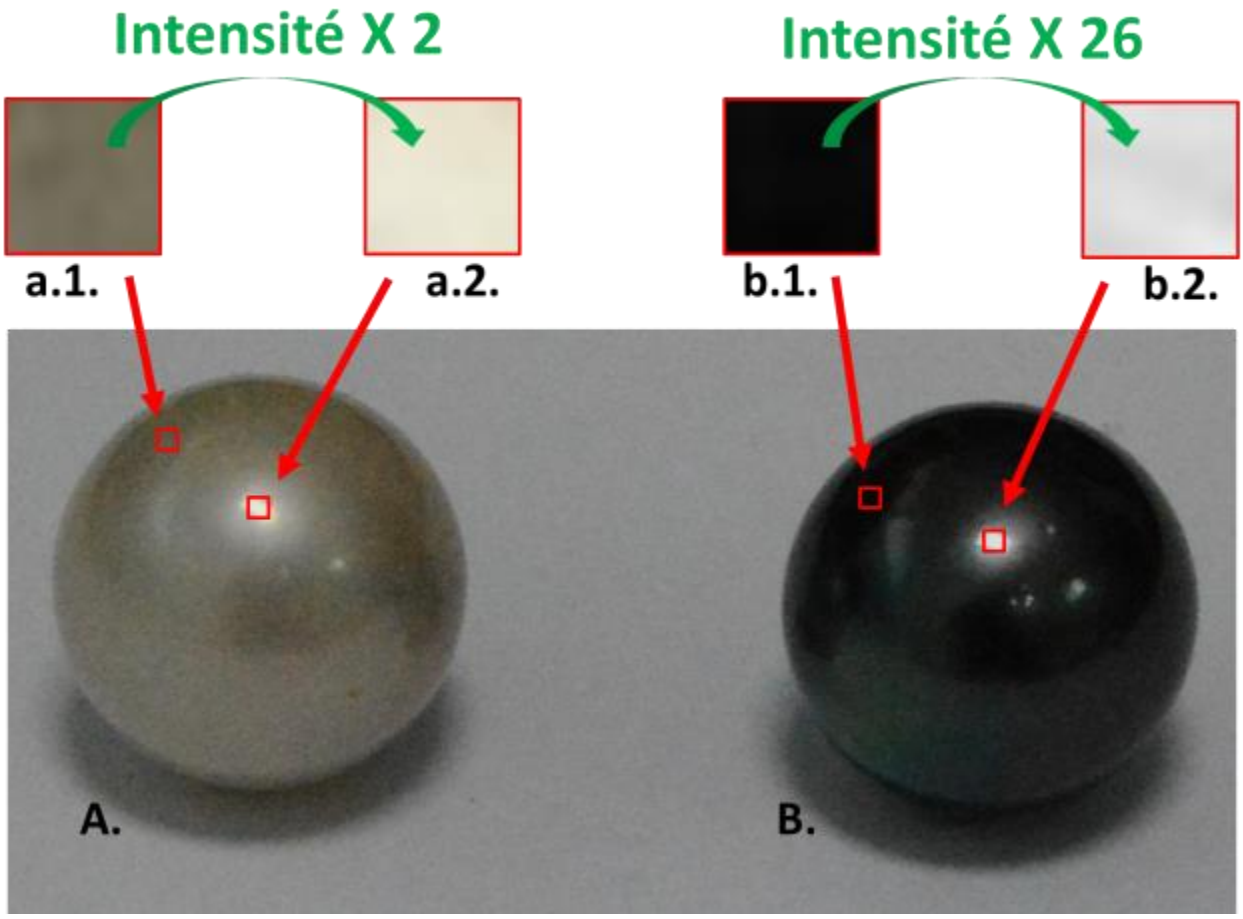


Figure 7 : Deux perles de contraste différent. La perle A présente un contraste inférieur à celui de la perle B (dans un espace colorimétrique HSV, la valeur d'intensité moyenne de a.2. est deux fois celui de a.1., tandis que celui de b.2. est 26 fois celui b.1.) alors que le brillant spéculaire est sensiblement le même (valeurs maximales d'intensité de l'image dans un espace colorimétrique HSV: 93,73% pour la perle A; 93.33% pour la perle B).

Le contraste et le brillant spéculaire se rapportent tous deux à l'intensité de la lumière spéculairement réfléchie (qu'elle le soit à angle rasant ou non). Cependant, ils ont des références différentes : le premier est défini par rapport à l'intensité de la lumière incidente tandis que le second est défini par rapport à l'intensité de la réflectance diffuse. Ces deux références expriment deux conceptions différentes du brillant comme perception : le brillant spéculaire implique la capacité d'évaluer la réflectance spéculaire étant donné une référence qui n'est pas nécessairement proche de l'angle spéculaire, ou même pas directement accessible dans la scène. Le contraste implique quant à lui la capacité d'évaluer la réflectance spéculaire à partir d'une référence située à proximité immédiate de l'angle spéculaire. En d'autres termes, à

moins de considérer le brillant spéculaire comme une simple réflectance spéculaire[26], auquel cas son évaluation serait basée sur des références indépendantes de la scène, l'évaluation du brillant spéculaire devrait être basée sur des références réelles ou même estimées d'une vision globale de la scène alors que l'évaluation du contraste pourrait être basée uniquement sur des références réelles provenant d'une vue locale de la scène.

4 : Effet miroir

L'effet miroir ("Distinctness Of reflected Image", ou "DOI") est le brillant perçu sur une surface en raison de la netteté des contours de la lumière réfléchie spéculairement. Plus la lumière réfléchie spéculairement présente des contours nets, plus la surface se comporte comme un miroir.

La Figure 8 montre 2 perles avec différentes qualités d'effet miroir. Le contour de la zone de réflectance spéculaire est plus distinct sur la perle B, présentant une meilleure netteté de l'image réfléchie[27].

[26] Ainsi, le brillant spéculaire serait simplement défini comme une lumière réfléchie dont l'intensité est supérieure à un seuil donné, quelle que soit l'intensité de la lumière incidente. Cette conception de base du brillant est parfois adoptée, même si une telle approche est critiquable (voir [137] sur [161]).

[27] Une démarcation est visible entre les parties gauche et droite de b., mais pas de a.

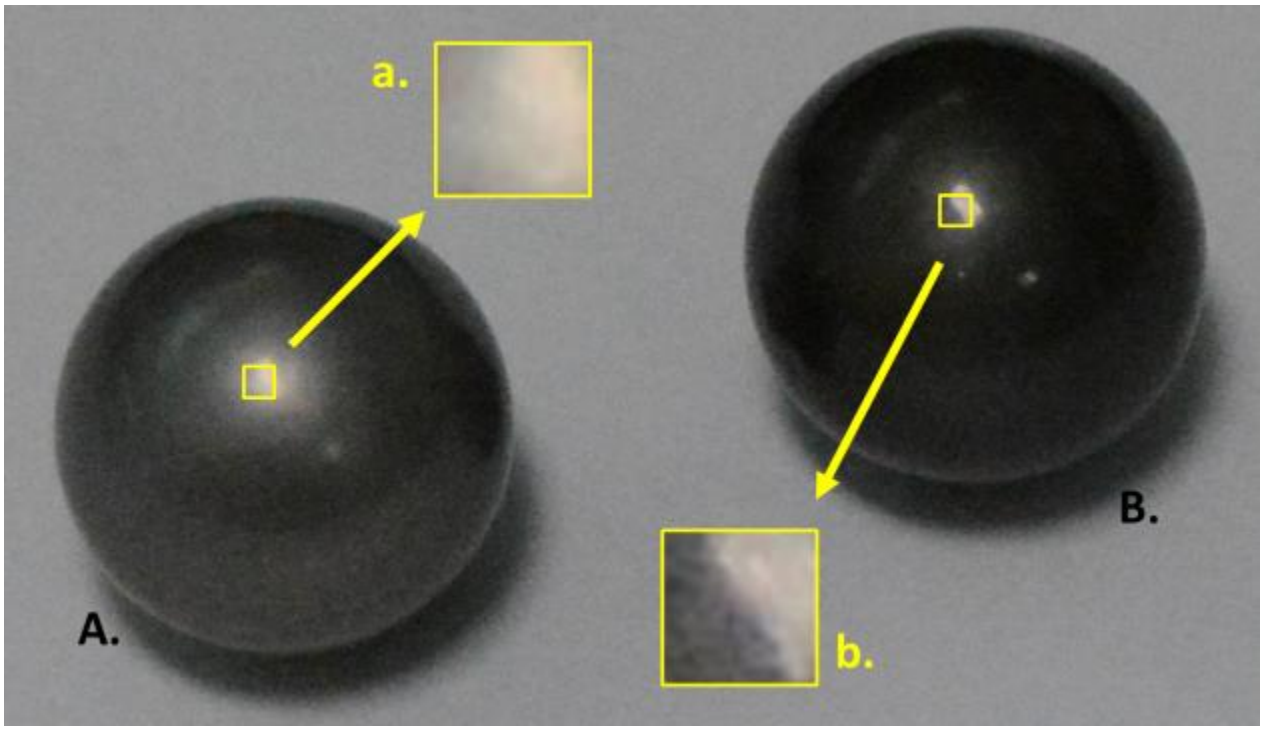


Figure 8 : Deux perles de d'effet miroir différent. La perle A possède un effet miroir inférieur à celui de la perle B (une démarcation est visible entre les parties gauche et droite de b., mais pas de a.).

De plus, la Figure 8 souligne la différence entre l'effet miroir et le brillant spéculaire : la perle B a un effet miroir plus élevé que la perle A tandis que la perle A possède un brillant spéculaire plus élevé que la perle B[28].

5 : Brouillard

Le brouillard est une tâche lumineuse pouvant se former autour des rayons lumineux réfléchis spéculairement. A titre d'illustration, dans la Figure 9, la perle B n'a presque pas de brouillard comparé aux deux autres perles.

[28] Valeurs maximales d'intensité de l'image dans un espace colorimétrique HSV: 87,84% pour la perle A ; 82.75% pour la perle B.

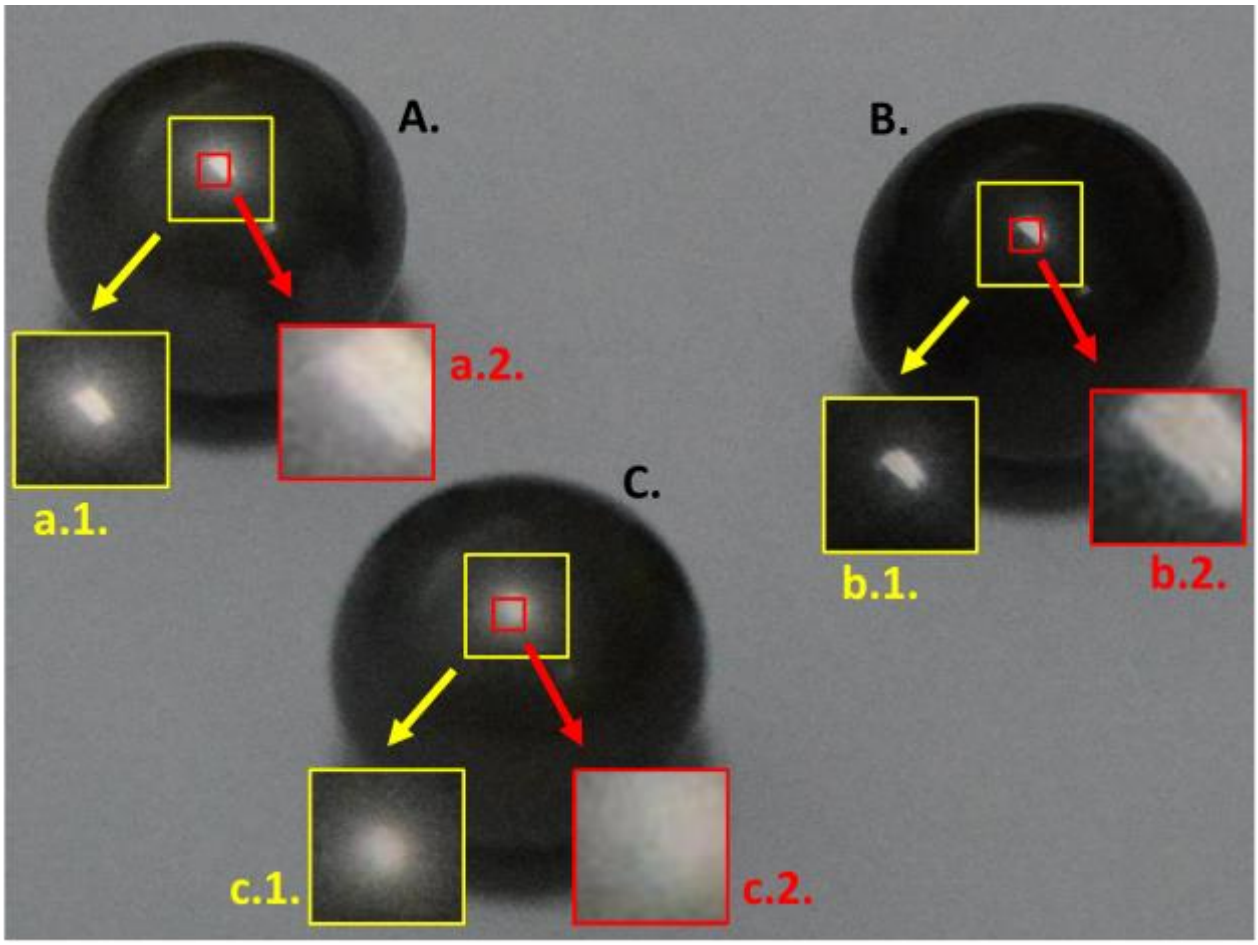


Figure 9 : Perles de brouillard différent. Les perles A et C présentent un brouillard supérieur à celui de la perle B. Les perles A et B présentent un effet miroir supérieur à celui de la perle C.

Le brouillard et l'effet miroir se rapportent tous deux à la qualité de l'image réfléchie. Cependant, ils en représentent deux aspects différents. Le premier rend compte (bien que négativement) de la façon dont l'image réfléchie préserve l'amplitude des intensités autour de l'angle spéculaire, tandis que le second rend compte du degré auquel l'image réfléchie préserve la netteté des frontières autour de l'angle spéculaire. Sur la Figure 9, les perles A et B présentent toutes deux un fort effet miroir par rapport à la perle C[29], mais la perle B présente un plus faible brouillard que le perle A[30].

[29] Une démarcation est visible entre les parties gauche et droite de a.2. et b.2., mais pas de c.2.

[30] La différence d'intensité entre les parties gauche et droite de a.2. est inférieur à b.2., mais le bord entre ces parties n'est pas plus net en b.2. qu'en a.2.

6 : Interférence de la surface

L'interférence de la surface est l'effet créé par les défauts de surface lorsqu'ils attirent l'œil, l'empêchant de se fixer sur l'image réfléchie.

Le brillant d'absence de texture de surface défini par Hunter est le brillant perçu sur une surface en raison de l'absence d'interférence entre l'image réfléchie et l'image de la surface elle-même.

La figure 7 illustre les deux images qui peuvent être formées à partir d'une même perle lorsque sa surface présente des irrégularités. La Figure 10A se concentre sur l'image réfléchie alors que la Figure 10B se concentre sur la texture de la surface. Dans la mesure où le regard peut passer d'une vue à l'autre, cela crée un effet qui diminue l'aspect brillant de la perle.

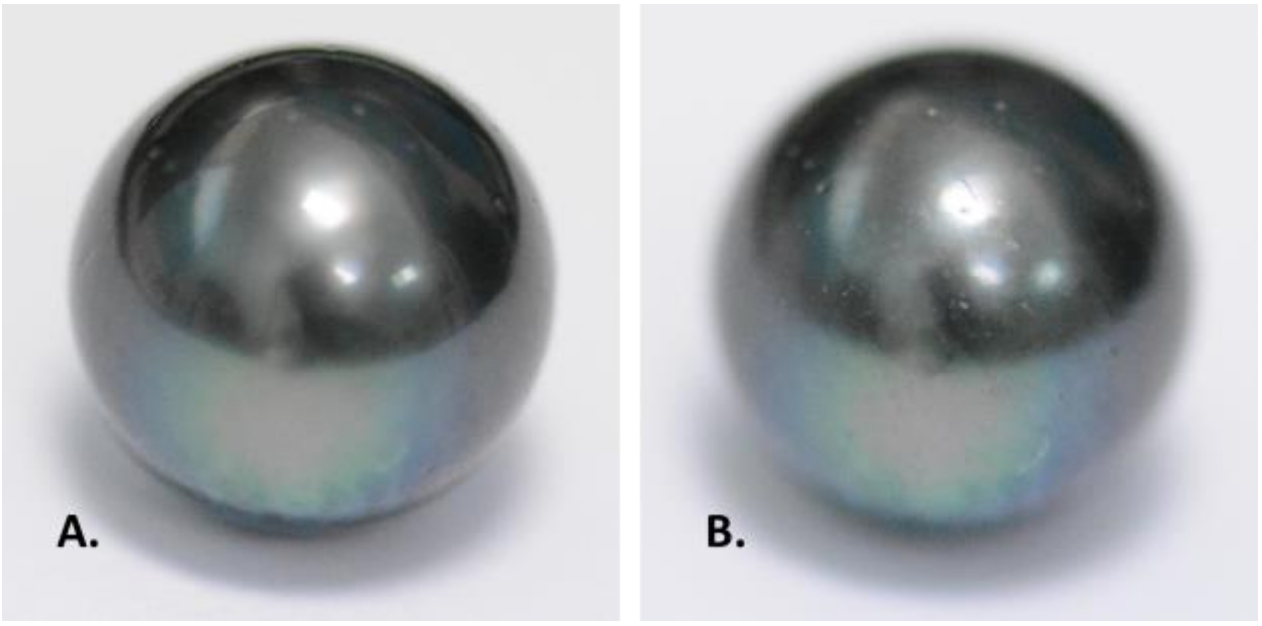


Figure 10 : Images d'une même perle prise avec deux distances focales différentes (A se concentre sur l'image réfléchie, B se concentre sur la surface de la perle).

B : Attributs du lustre spécifiques à la perle

Outre les attributs du lustre tirés de l'étude du brillant, deux attributs plus spécifiques au lustre des perles peuvent être mentionnés : l'iridescence et l'orient.

1 : Iridescence

Dans la mesure où les changements de longueur d'onde dus à la réflexion spéculaire sont négligeables pour les matériaux diélectriques, la réflectance spéculaire est habituellement étudiée comme étant achromatique, tandis que la couleur est liée à la réflectance diffuse seulement. Cependant, la nacre peut présenter un phénomène d'iridescence [101], or l'iridescence, bien que due à la réflexion spéculaire, se manifeste par des variations de couleur et fait ainsi obstacle à ce que la réflectance spéculaire soit étudiée dans le seul cadre achromatique. Du fait de cette propriété, il serait légitime de s'interroger sur la pertinence d'une prise en compte de l'iridescence dans le contexte d'une étude sur le lustre et si cette prise en compte ne serait pas plus appropriée dans le contexte d'une étude sur la couleur. Toutefois, tant que la question de savoir si le système visuel perçoit l'iridescence comme un attribut de la couleur ou comme un attribut du lustre (voire même s'il est justifié de cloisonner le lustre et la couleur) n'est pas tranchée, il semble prudent de le prendre en compte lors de l'étude du lustre, même s'il est possible qu'il s'avère *in fine* n'avoir pas d'impact sur la perception du lustre.

La Figure 11 illustre cet aspect, avec un agrandissement de la zone de réflectance spéculaire, où des zones colorées allongées parallèles sont visibles, comme superposées.

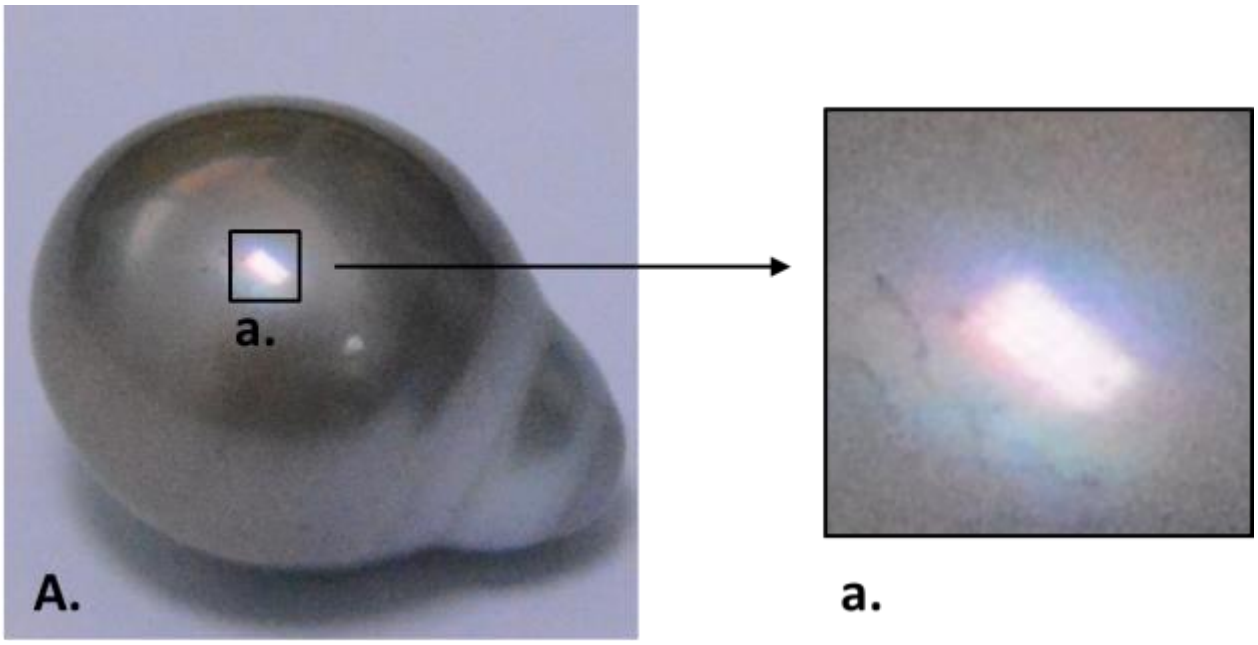


Figure 11 : Certaines perles présentent des changements de couleur autour de l'angle spéculaire en raison de l'iridescence.

Dans certains cas, la rotation de la perle fait tourner ces zones dans le même sens. Ce comportement anisotrope est observable sur les perles cerclées, comme celle de la Figure 11.

2 : Orient

La Figure 12 montre la différence de réflectance entre une fausse perle faite de plastique (A) et une perle de culture (B). Les deux images sont prises dans les mêmes conditions. La fausse perle présente seulement une zone de réflectance blanche tandis que la vraie présente deux zones de réflectance blanche et colorée.

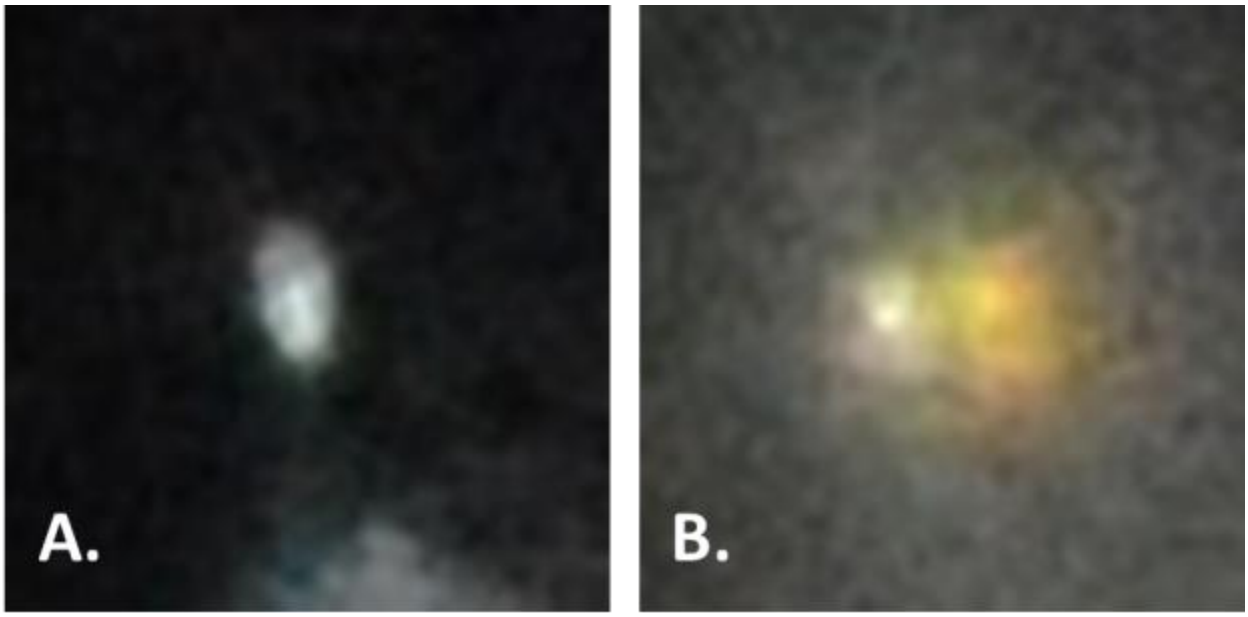


Figure 12 : Détail de la réflectance sur deux perles (en A, une perle artificielle en B, une vraie perle).

La distinction entre les deux zones à la surface de la perle de culture n'est pas clairement visible à l'œil nu, qui ne perçoit qu'un effet de réflectance « profonde » lors de la rotation de la perle. Prendre une photo de la perle en utilisant un temps d'exposition très court permet de le rendre visible. En effet, si le temps d'exposition est trop long, le capteur de la caméra absorbe trop de photons et les deux zones de réflectance fusionnent en une seule et même zone de réflectance plus grande.

Contrairement à l'iridescence, où les couleurs varient avec les angles d'illumination et d'observation, la couleur de la zone secondaire reste la même quel que soit l'angle d'éclairage ou d'observation et aura plutôt tendance à varier selon les perles. Une explication possible serait que, puisque la conchyoline contient des pigments, les différentes couches constituant la nacre agissent comme un filtre : seule une partie du spectre incident est réfléchie spéculairement à partir des couches plus profondes, le reste étant absorbé par les couches successives de conchyoline.

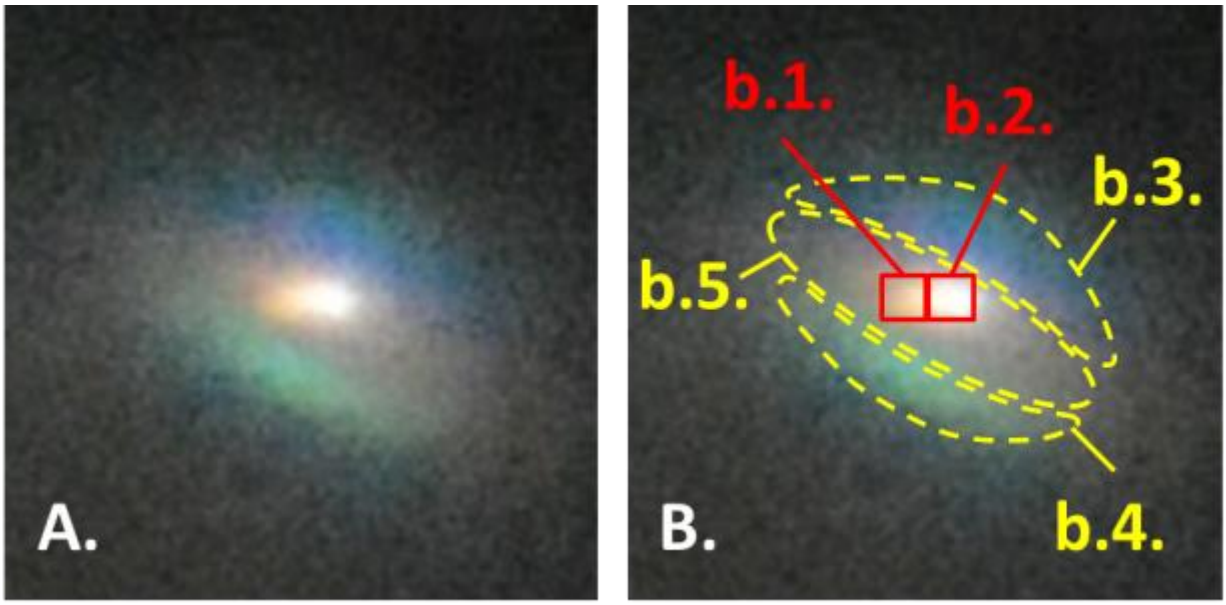


Figure 13 : Différence entre l'irisation (b.3 à b.5) et l'orient (b.1).

La Figure 13 montre la même perle que celle montrée sur la Figure 11. L'image est cependant prise avec un temps d'exposition très court. Il est possible d'observer à la fois l'iridescence (b.3 à b.5) et l'orient (b.1 pour la réflectance profonde et b.2 pour la réflectance de surface).

II : Quantification des attributs perceptuels du lustre

Dans la première section du chapitre, nous avons qualifié les attributs perceptuels du lustre. Nous proposons maintenant de les quantifier.

Dans ce qui suit, nous rappelons brièvement chaque attribut du lustre avant de discuter d'une proposition de quantification. Les attributs abordés ici sont le brillant spéculaire, le contraste, l'effet miroir, le brouillard, l'orient et l'iridescence. Toutes les images de perles présentées dans les figures suivantes ont été prises dans des conditions d'éclairage identiques.

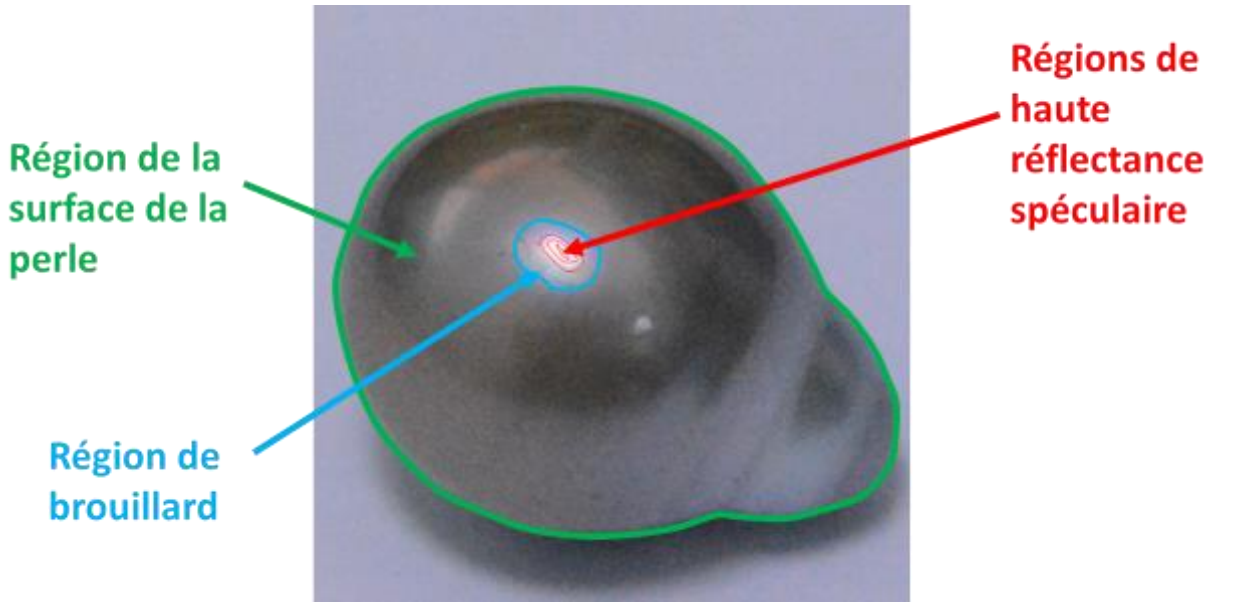


Figure 14 : Régions d'intérêt sur une image de perle (région de la surface de la perle, région de brouillard et régions de haute réflectance spéculaire).

Afin de simplifier la présentation des attributs perceptuels du lustre, nous définissons plusieurs régions d'intérêts sur les images de perles : la région de la surface de la perle, la région de brouillard et les régions de haute réflectance spéculaire (Figure 14). Nous détaillons les manières de déterminer automatiquement les régions de brouillard et de haute réflectance spéculaire dans les sections sur le brouillard et sur le brillant spéculaire respectivement.

1 : Brillant spéculaire

Le brillant spéculaire, que la lumière soit réfléchie à angle rasant ou non, se rapporte à la luminosité perçue d'une surface en raison de la réflectance spéculaire par rapport à la lumière incidente.

Une première proposition de quantification de cet attribut à partir de photographies pourrait être de prendre l'intensité maximale de la région de la surface de la perle sous l'hypothèse d'une lumière incidente constante.

Cette manière de faire est illustrée Figure 15. La perle A paraît avoir un brillant spéculaire inférieur à celui de la perle B. Dans l'espace de couleurs RGB, en prenant l'intensité d'un pixel comme la moyenne des trois composantes de couleurs R, G et B, l'intensité maximale de la perle A est 230 et celle de la perle B est 255. Il semble que les mesures reflètent les perceptions.

Cette quantification est cependant instable puisque l'estimation du brillant spéculaire repose sur un seul pixel.

Une solution rapide pourrait consister à lisser l'image afin de s'assurer que la caractéristique n'est pas déterminée par un pixel isolé. Néanmoins, cela ne fonctionnerait pas en cas de haute netteté de l'image : si des régions de fortes intensités sont étroitement imbriquées avec des régions de faibles intensités, une opération de lissage résulterait en une intensité maximale bien inférieure à l'intensité maximale réelle de l'image. En outre, il peut arriver que l'intensité maximale de l'image ne reflète même pas l'intensité maximale spéculairement réfléchie par la surface de la perle. C'est généralement le cas lorsque la lumière réfléchie sature le capteur de la caméra.

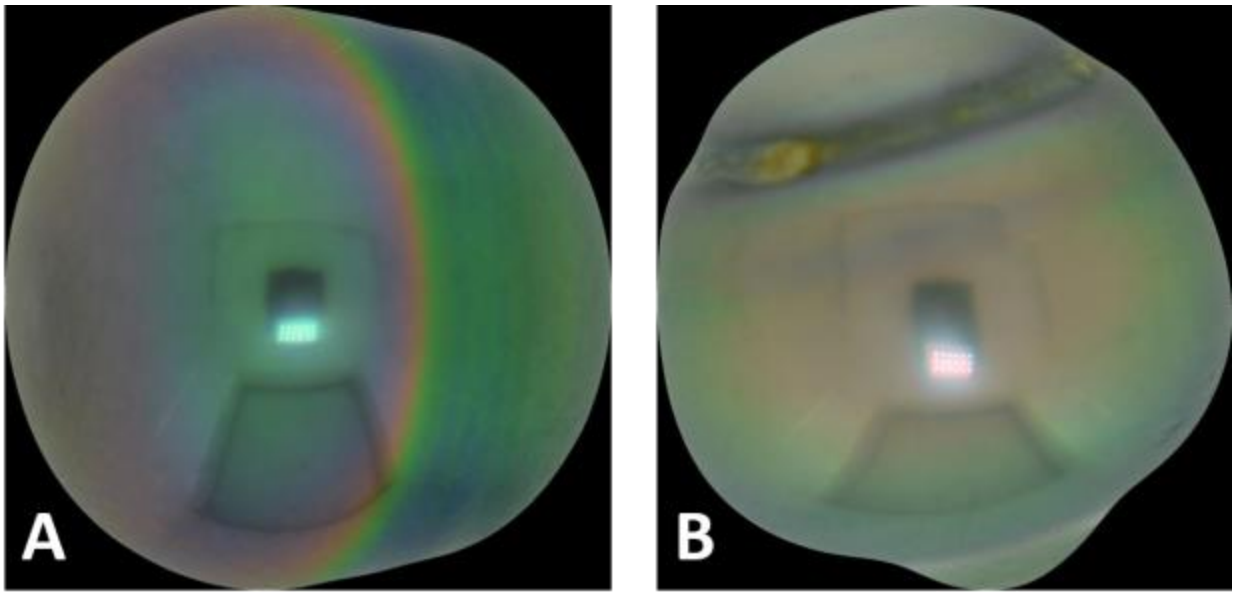


Figure 15 : Exemple d'images de perles prises dans des conditions de lumière similaires avec les valeurs de caractéristique d'intensité de réflectance spéculaire étant ici l'intensité maximale de l'image. La perle A a une intensité maximale (moyenne des canaux RGB : 230) inférieure à celle de la perle B (moyenne des canaux RGB : 255).

Soit le pixel d'intensité maximale est isolé et il est préférable qu'il ait peu d'impact sur la détermination du paramètre correspondant à l'intensité de réflectance spéculaire, soit ce pixel fait partie d'une région d'intensité comparable, et bien qu'aucune information sur l'intensité réelle ne soit disponible, il est possible de mesurer la taille relative de la région où la lumière réfléchie sature le capteur, ce qui représente une information, bien qu'indirecte et partielle, sur l'intensité de la réflectance spéculaire.

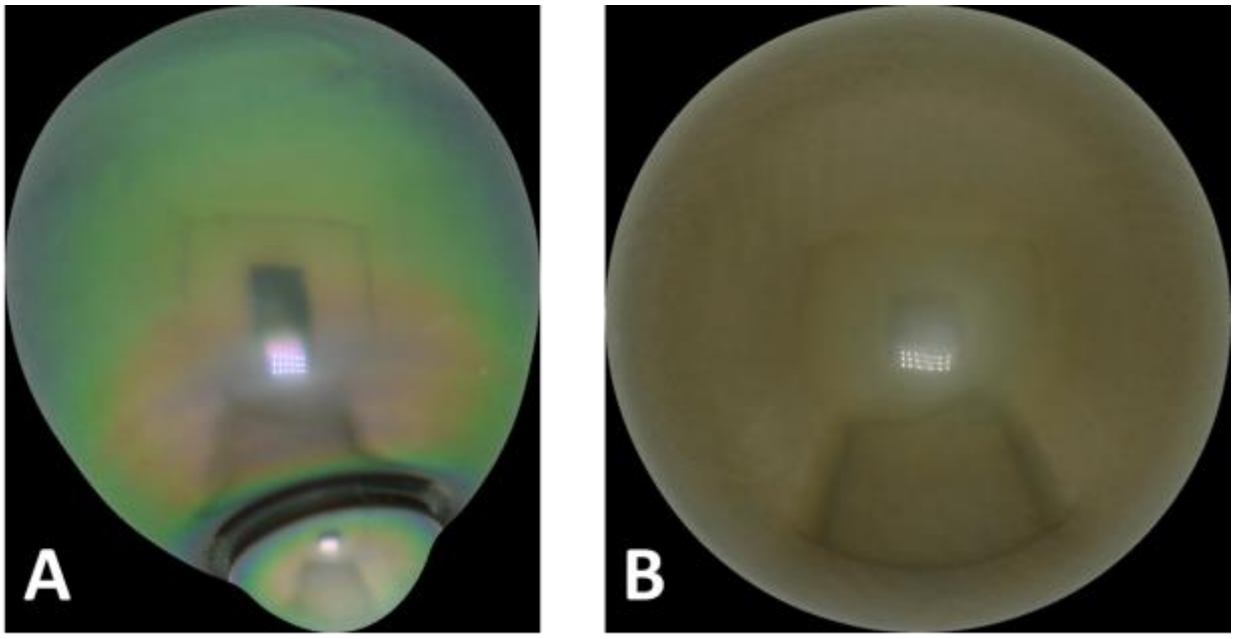


Figure 16 : Illustration de la différence entre la caractéristique de la figure 2 et une caractéristique estimant l'intensité de réflectance spéculaire avec les cases des intensités les plus élevées d'un histogramme normalisé (ici, une seule case est donnée). En A, la proportion des pixels d'intensité entre 230 et 255 (moyenne des canaux RGB) est de 0,0042 (l'intensité maximale est de 253.66). En B, la proportion des pixels d'intensité entre 230 et 255 (moyenne des canaux RGB) est de 0,0009 (l'intensité maximale est de 254).

Compte tenu de ces remarques, l'histogramme des intensités, normalisé par rapport au nombre de pixels dans la région de la surface des perles, pourrait être un meilleur estimateur du brillant spéculaire. Il est même possible de limiter la collecte d'informations aux plages des intensités les plus hautes.

L'équation (25)**Erreur ! Source du renvoi introuvable.** formalise le calcul du brillant spéculaire pour la plage d'intensités se trouvant entre les intensités i_a et i_b, $a > b$, où P est l'ensemble des pixels de la région de la surface de la perle et $intensité(p)$ est l'intensité d'un pixel $p \in P$. Le calcul de cette intensité peut être obtenu en prenant la moyenne des trois canaux RGB, cette moyenne pouvant être pondérée (comme par exemple dans le cas du système HSV).

$$brillant\ spéculaire_{a,b} = \frac{Card(\{p \in P | a > intensité(p) > b\})}{Card(P)} \qquad (25)$$

La Figure 16 illustre le cas où l'intensité maximale des deux images est approximativement la même mais la distribution des intensités est différente. La perle A contient proportionnellement beaucoup plus de pixels dont l'intensité est dans la gamme 230-255 (moyenne des 3 canaux

RGB) que la perle B. Même si l'intensité maximale de l'image A est légèrement inférieure à celle de l'image B, la perle A semble avoir une intensité de réflectance spéculaire plus élevée.

2 : Contraste

Le contraste fait référence à la luminosité perçue en raison du contraste entre la réflectance spéculaire et la réflectance diffuse sur la surface de la perle.

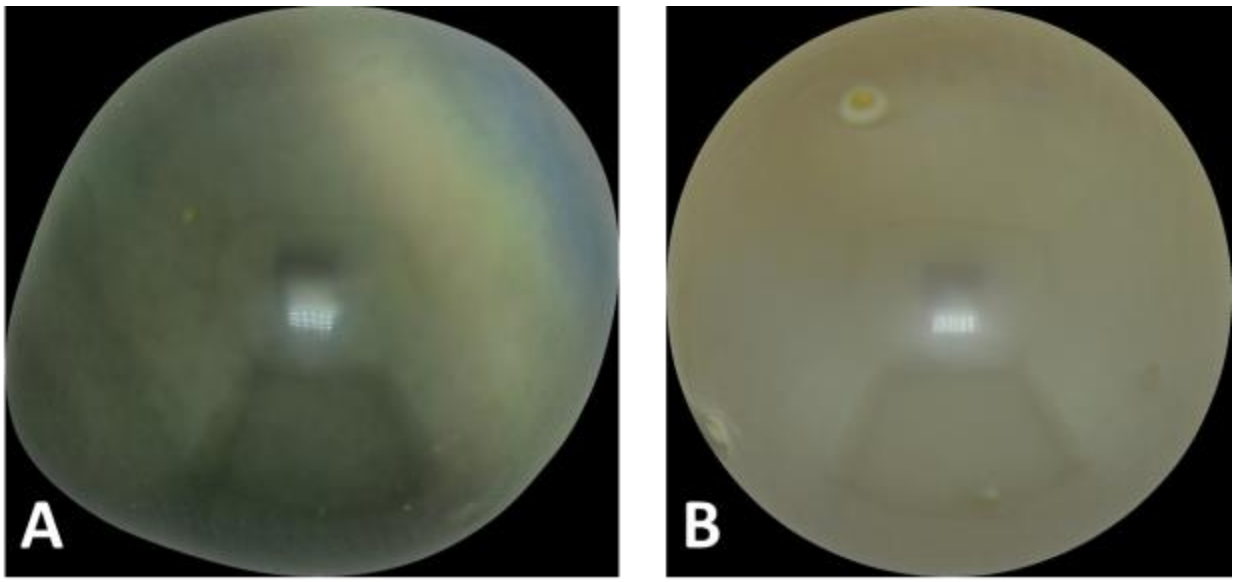


Figure 17 : Illustration d'une caractéristique de contraste de calcul comme variance de l'intensité dans la région de la surface de la perle. En A, le contraste est de 65.6198. En B, le contraste est de 6.2936.

La caractéristique peut être calculée en trouvant d'abord un seuil d'intensité approprié pour séparer les régions spéculaires et diffuses, puis en prenant la distance entre leurs d'intensités moyennes respectives. Le choix d'un seuil peut être évité en supposant une variance de faible intensité dans chacune de ces régions. Dans ce cas, la variance d'intensité dans la région de l'image correspondant à la surface de la perle peut représenter une estimation du contraste des caractéristiques. Puisque les perles ont une réflectivité élevée, il est probable que les variances respectives des intensités des régions spéculaires et des intensités des régions diffuses soient faibles relativement à la distance entre leurs moyennes, de sorte que nous prenons la variance d'intensité de toute la surface de la perle comme caractéristique du contraste. L'équation (26) formalise le calcul du contraste, avec P l'ensemble des pixels de la région de la surface de la

perle et $intensité(p)$ l'intensité d'un pixel $p \in P$. La Figure 17 illustre deux perles de contraste différent.

$$contraste = \left(\frac{1}{Card(P)}\sum_{p\in P} intensité(p)^2\right) - \left(\frac{1}{Card(P)}\sum_{p\in P} intensité(p)\right)^2 \quad (26)$$

3 : Effet miroir

L'effet miroir fait référence à la netteté des reflets spéculaires. Contrairement au contraste, qui se concentre sur l'amplitude des variations d'intensités, l'effet miroir se concentre sur la netteté des variations d'intensité.

La netteté des reflets spéculaires se rapporte à la force des contours de l'image réfléchie. Une caractéristique appréhendant cet attribut peut alors être calculée comme le gradient moyen d'intensité dans les régions de haute réflectance spéculaire. Ces régions peuvent elles-mêmes être dérivées de la caractéristique de brillant spéculaire (régions de haute intensité spéculaire) ou déterminées par un seuil fixe. La Figure 18 illustre le cas de deux perles d'effet miroir différent.

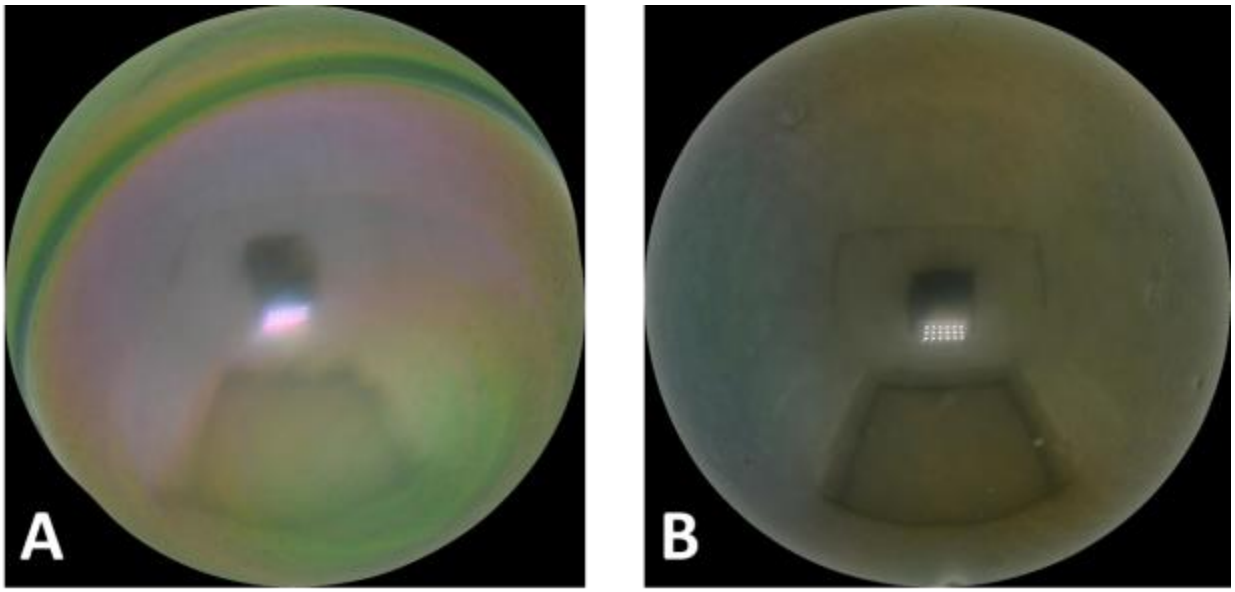


Figure 18 : Illustration d'une particularité de calcul de l'image réfléchie en tant que gradient moyen d'intensité dans la région de la réflectance spéculaire élevée. En A, l'effet miroir est de 20.7044. En B, l'effet miroir est de 162.0583.

L'équation (27) formalise une manière possible de calculer l'effet miroir, avec H l'ensemble des pixels se trouvant dans la région de haute réflectance spéculaire, cette région pouvant être délimitée par l'enveloppe convexe formée par les pixels de haute intensité.

$$effet\ miroir = \frac{1}{Card(H)} \sum_{p \in H} \nabla\ intensité(p) \qquad (27)$$

4 : Brouillard

La littérature sur la perception de la réflectance ne considère pas le brouillard directement, mais plutôt l'absence de brouillard [142]. Plus simplement, cet aspect du lustre peut être défini comme le brillant perçu en fonction du brouillard autour des hautes réflexions spéculaires. Par commodité, nous travaillons directement sur le brouillard et proposons une fonction appréhendant le brouillard ; l'information véhiculée est la même (on peut obtenir une valeur pour l'absence de brouillard en inversant la valeur du brouillard par exemple).

L'intensité décroît progressivement à mesure que la distance aux points d'intensités les plus élevées croît, jusqu'à ce que la réflexion spéculaire de la lumière directe n'ait plus d'impact perceptuel. Il n'est pas simple de déterminer automatiquement l'endroit où l'impact de la

réflexion spéculaire n'est plus perçu. Une segmentation s'appuyant sur la couleur utilisant le fait que dans le cas général les lumières incidente et réfléchie ont presque la même longueur d'onde n'est pas efficace dans le cas de perles à cause d'une iridescence possible à la surface de la perle (voir Figure 20, image B, où la région de brouillard est de différentes couleurs). De même, l'utilisation du signe de la dérivée première dans des directions éloignées des points d'intensités les plus élevées ne fonctionne pas toujours puisque l'intensité de la réflectance diffuse peut continuer à diminuer au-delà de ce point. La courbe des intensités lissées le long d'une ligne passant par le point d'intensité la plus élevée a une forme de cloche autour de ce point. Puisque la plupart des intensités les plus élevées se situent entre les deux points d'inflexion, nous supposons que les intensités au-delà de ces points ne sont pas perçues comme de la brume et prenons ces points comme limites de la région de brouillard le long de cette ligne. En termes de surface, *nous définissons la région de brouillard comme la région autour des points d'intensités les plus élevées où la distribution d'intensité peut être analysée comme une fonction concave des coordonnées (x, y)*.

Le brouillard n'est pas déterminé par la taille de sa région ; en effet, si la surface apparentée est presque plate, le brouillard est à peine perceptible, même sur une grande région.

De même, le brouillard n'est pas déterminé par l'intensité moyenne dans la région de brouillard ; en effet, la courbure et la valeur moyenne sont deux variables indépendantes.

En revanche, la variance de l'intensité dans la région de brouillard contient des informations sur la répartition du brouillard. Sachant que la surface représente une fonction concave, la distance entre les intensités extrêmes combinée à la taille de la région de brouillard pourrait aussi bien fournir des informations sur la répartition du brouillard ; néanmoins, cette information serait partielle (nous ne connaîtrions qu'une limite inférieure pour la variance) et pourrait être instable (elle pourrait dépendre de points aberrants).

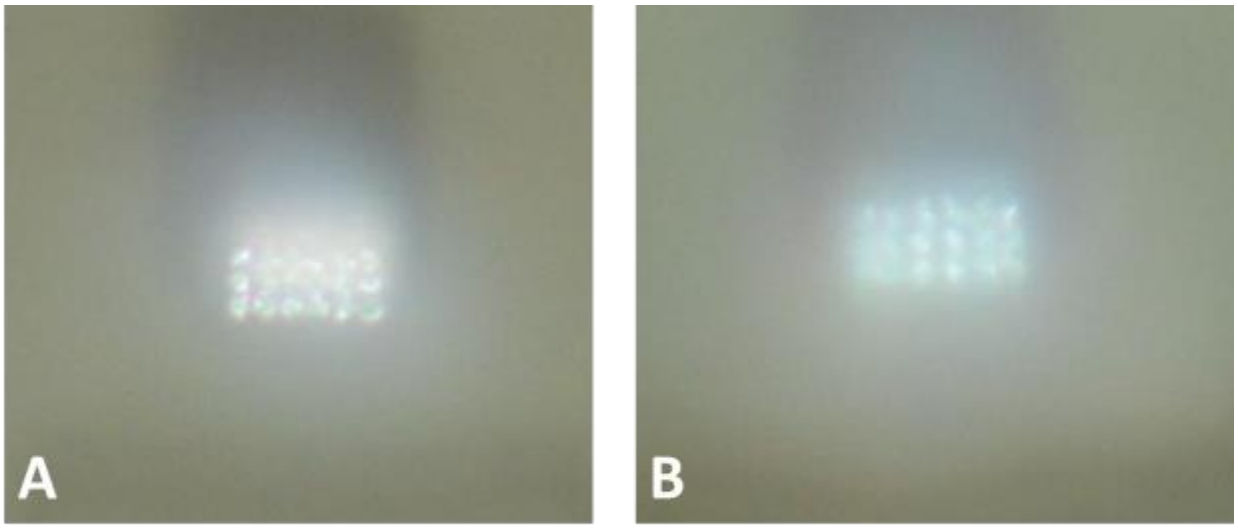


Figure 19 : Illustration d'une caractéristique calculant le brouillard comme étant la variance d'intensité dans la région autour des points d'intensités les plus élevées où la distribution d'intensité peut être analysée comme une fonction concave des coordonnées (x, y). En A, le brouillard est de 106.3080. En B, le brouillard est de 65.4035.

L'équation (28) formalise une manière possible de calculer le brouillard, avec B l'ensemble des pixels se trouvant dans la région de brouillard, cette région pouvant être définie comme la région autour des points d'intensités les plus élevées où la distribution d'intensité peut être analysée comme une fonction concave des coordonnées $(\boldsymbol{x}, \boldsymbol{y})$.

$$brouillard = \left(\frac{1}{Card(B)} \sum_{p \in B} intensité(p)^2 \right) - \left(\frac{1}{Card(P)} \sum_{p \in B} intensité(p) \right)^2 \qquad (28)$$

5 : Iridescence

L'iridescence correspond aux changements de couleur perçus sur la surface de la perle en raison de la réflectance spéculaire. La région de brouillard est bien adaptée à l'étude de l'iridescence car elle est censée correspondre à la plus grande région de la surface de la perle où la réflectance spéculaire est perçue. Si nous faisons l'hypothèse selon laquelle les changements de couleur perçus peuvent être appréhendés par une mesure de la largeur des plages de longueurs d'onde sur le spectre visible dans cette région, une façon de calculer la caractéristique correspondante serait de discrétiser le spectre visible en un nombre donné de plages de longueurs d'onde et de compter le nombre de plages dans lesquelles au moins une longueur d'onde est présente dans la région d'intérêt. Comme aucune relation d'ordre entre les

plages de longueurs d'onde n'est nécessaire, la dimension Hue de l'espace colorimétrique HSV peut être utilisée à la place du spectre visible. De plus, si nous faisons l'hypothèse additionnelle qu'une même distribution de gammes de Hue est perçue comme plus iridescente si ses couleurs sont plus saturées, nous pouvons pondérer chaque gamme de Hue avec sa saturation moyenne pour prendre en compte cette hypothèse.

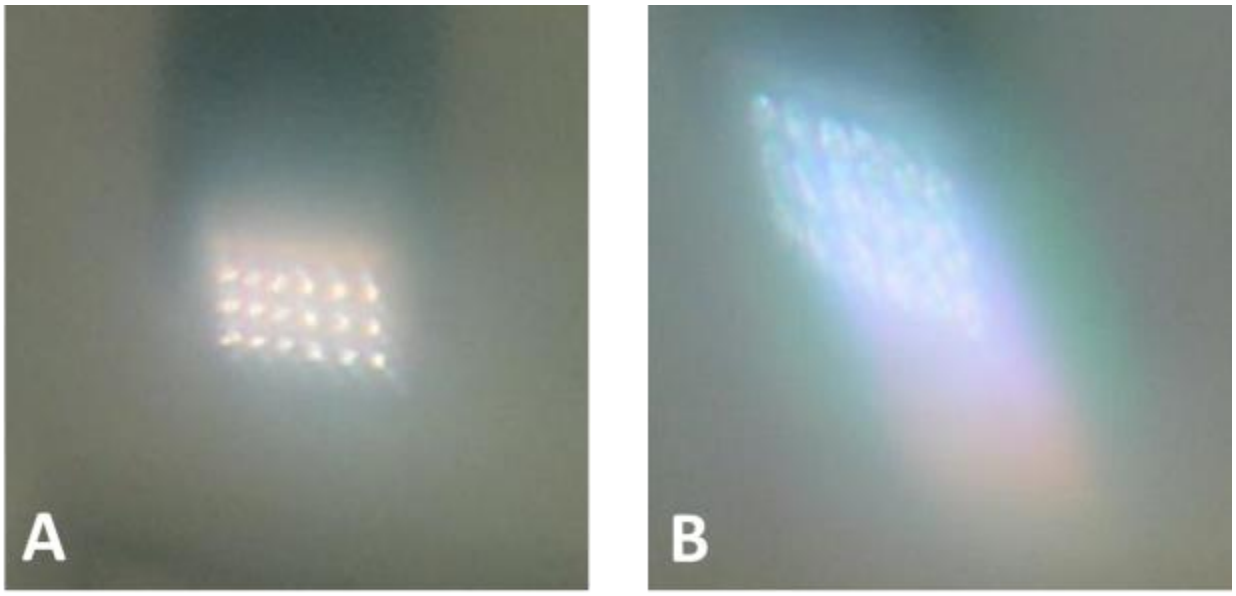


Figure 20 : Illustration d'une caractéristique calculant l'iridescence en tant que nombre de plages de Hue, pondérées avec leurs valeurs de saturation moyenne correspondantes, présentes dans la région d'intérêt. Nombre de plages = 1000. En A, l'iridescence est de 3.0236. En B, l'iridescence est de 16.9462.

La Figure 20 montre deux images de perles avec leurs valeurs d'iridescence correspondantes calculées comme proposé ci-dessus. Nous avons discrétisé la dimension Hue en 1000 plages de valeurs. La perle de l'image B a une iridescence plus élevée que la perle de l'image A.

L'équation (29) formalise le calcul de l'iridescence, avec $H_i = \left\{p \in B \mid \frac{i-1}{N} < Hue(p) \leq \frac{i}{N}\right\}$, B l'ensemble des pixels se trouvant dans la région de brouillard et N le nombre de plages en lesquelles est décomposée la gamme de Hue.

$$iridescence = \sum_{i=1}^{N} \left(\mathbb{1}(H_i \neq \emptyset) \frac{1}{Card(H_i)} \sum_{p \in H_i} saturation(p) \right) \quad (29)$$

6 : Orient

L'orient, ou réflectance profonde, est définie comme la perception d'une variation de couleur sous la surface de la perle (en raison de la réflexion sur les couches successives de nacre).

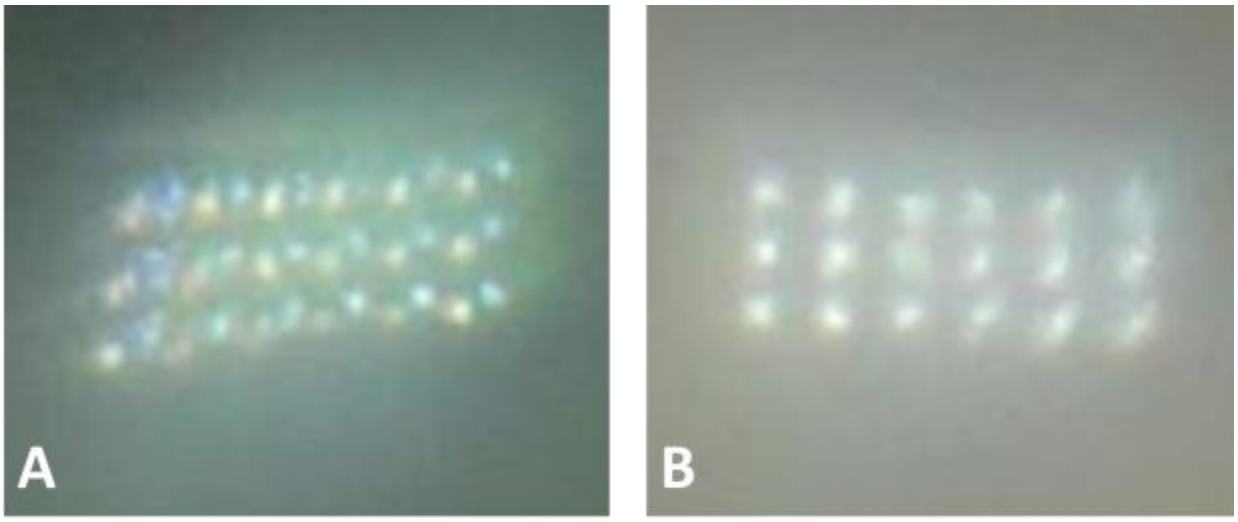


Figure 21 : Illustration d'une caractéristique calculant la réflectance profonde comme étant le gradient moyen de chromaticité dans la région des intensités de réflectance spéculaire élevées. En A, l'orient est de 0,0515. En B, l'orient est de 0,0204.

Contrairement à l'iridescence, qui peut se propager sur toute la zone de brouillard, la réflectance profonde est localisée dans la région de réflectance spéculaire la plus élevée. Une manière de capturer cet aspect peut être de prendre le gradient moyen de chromaticité (saturation dans le système HSV) dans les régions de haute intensité de réflectance spéculaire (Figure 21). Ceci est formalisé par l'équation (30), où H est l'ensemble des pixels de haute intensité.

$$orient = \frac{1}{Card(H)} \sum_{p \in H} \nabla\, saturation(p) \qquad (30)$$

III : Synthèse

Les propriétés physiques et perceptuelles du lustre ne coïncidant pas nécessairement, nous avons fondé l'élaboration des vecteurs de caractéristiques du lustre sur les aspects perceptuels du lustre et non sur ses aspects physiques.

Dans un premier temps, nous avons répertorié les attributs perceptuels du lustre ; dans un second temps, nous avons proposé une manière de les quantifier afin d'en former des vecteurs exploitables numériquement.

Dans une perspective d'apprentissage automatique, ces vecteurs de caractéristiques constituent une interface entre le monde sensible et la machine : il est maintenant possible de tester l'algorithme de rangement d'objets sur le cas concret de problème implicite de rangement d'objets que représente l'évaluation du lustre des perles de Tahiti. Ceci fait l'objet du prochain chapitre.

Chapitre 9 : Application de l'algorithme de rangement d'objets à l'évaluation automatique du lustre des perles de Tahiti

Le présent chapitre rapporte une expérience d'apprentissage automatique de l'évaluation du lustre des perles de culture de Tahiti. Cette expérience est destinée à tester l'algorithme de rangement d'objets à un cas concret de problème implicite de rangement d'objet. A titre incident, elle permet de tester les caractéristiques du lustre des perles de Tahiti.

I : Jeu de données

La qualité de lustre de chaque perle de l'ensemble de données a été évaluée par un expert confirmé. En raison de la difficulté d'obtenir des perles de lustre de classe A (lesquelles sont directement mises sur le marché) de la part des professionnels, notre ensemble de données ne contient que des perles des classes B à D.

Les perles de classe B étant les moins nombreuses (18 perles), nous sélectionnant au hasard 18 perles parmi les perles de classe C et 18 perles parmi les perles de classe D afin d'avoir des classes équilibrées. Nous obtenons ainsi un jeu de 54 perles réparties en trois classes équilibrées.

La surface des perles n'étant pas homogène et les zones à partir desquelles extraire les caractéristiques étant localisées, nous photographions chaque perle sous 16 angles distincts afin d'avoir un jeu de 864 photographies.

Toutes les photographies sont prises par le même appareil avec un paramétrage fixe et sous des conditions d'éclairage constantes. L'appareil photographique est un appareil numérique Nikon D7100 avec un objectif AF-S VR Micro-NIKKOR 105mm f/2.8G IF-ED. Le paramétrage de l'appareil est le suivant : temps d'exposition de 1/320 secondes ; ISO-100 ; F/8. Les conditions d'éclairage sont assurées par un caisson Digital Imaging Box 1612 de Universal Electronics

Industries Ltd (température de couleur : 5000K +/- 10%) avec les panneaux lumineux ajustés à l'intensité maximale pour l'avant et l'arrière-plan et à l'intensité minimale pour l'éclairage spéculaire.

Les caractéristiques sont calculées sur chaque photographie pour produire un vecteur de caractéristiques par image. Chaque exemple est représenté par un vecteur de caractéristiques destinées à appréhender les aspects visuels du lustre.

Ces vecteurs ont 10 dimensions : 3 pour le brillant spéculaire (rapport des pixels dont l'intensité est comprise entre 230-255, 205-230 et 180-205 respectivement), 5 pour les autres caractéristiques (contraste, effet miroir, brouillard, orient et iridescence) et 2 caractéristiques supplémentaires prenant en compte les couleurs de la région de réflectance diffuse (saturation moyenne et variance de la chromaticité dans la région de réflectance diffuse). Les deux dernières caractéristiques sont destinées à tester l'hypothèse selon laquelle le lustre serait totalement déterminé par la réflectance spéculaire.

Par commodité, nous avons associé les caractéristiques à des identifiants numérotés suivant la séquence de l'énumération ci-dessus. Le Tableau 11 synthétise les correspondances entre les identifiants et les noms des caractéristiques.

Tableau 11 : Correspondances entre les numéros de caractéristiques et leurs noms.

Identifiant	*Nom*
C1	Brillant spéculaire 1 (plage 230-255)
C2	Brillant spéculaire 2 (plage 205-230)
C3	Brillant spéculaire 3 (plage 180-205)
C4	Contraste
C5	Effet miroir
C6	Brouillard
C7	Orient
C8	Iridescence
C9	Saturation moyenne
C10	Variance de la chromaticité

II : Validation de l'algorithme d'apprentissage

Nous menons une expérience sur l'évaluation de la qualité des perles de culture de Tahiti.

Les experts classent les perles en quatre classes, de A à D, par ordre décroissant de qualité. À première vue, cela paraît être un problème de classement multi-classes, toutefois, l'observation du processus d'évaluation de la qualité des perles révèle qu'un certain nombre de perles peuvent changer de classe à mesure que de nouvelles perles sont évaluées. Le processus de tri se déroule comme si les classes n'étaient qu'un moyen compact d'évaluer la qualité des perles conformément aux exigences du marché, mais elles révèlent que les experts saisissent progressivement l'ordre total sous-jacent tout en réajustant la place de la délimitation entre classes.

En tant que telle, l'évaluation de la qualité des perles de culture de Tahiti est un problème typique de rangement d'objets défini en termes de problème de classement.

A : Evaluation des performances

La méthode d'estimation que nous utilisons est une validation croisée avec une partition en 9 sous-ensembles et répétée 16 fois sur des séries distinctes de 54 photographies, chaque photographie provenant d'une perle différente, de manière à utiliser la totalité des exemples d'apprentissage tout en assurant le meilleur compromis entre biais et variance de l'estimateur. En effet, il a été montré que le meilleur compromis entre biais et variance de l'estimateur était assuré par une partition en 10 sous-ensembles [143], or 9 est le diviseur entier de 54 qui se rapproche le plus du nombre préconisé. Le résultat moyen est calculé avec son écart type.

Les résultats sont comparés à ceux obtenus à l'aide des méthodes suivantes : classement nominal multi-classes (système de vote un contre un avec SVM), classement ordinal (méthode de réplication des données avec SVM), rangement agnostique (algorithme génétique avec population de 100 et un taux de mutation de 0.2) et rangement d'instances [43]. Pour toutes les méthodes, les résultats sont donnés avec et sans sélection de caractéristiques : nous opérons pour chaque méthode une sélection des caractéristiques qui permettent la meilleure généralisation ; puisque l'espace des caractéristiques a 10 dimensions, nous calculons les 210 combinaisons de caractéristiques possibles pour chacune des méthodes. Toutes les méthodes à

noyau sont testées avec un noyau quadratique et l'algorithme génétique est testé avec des fonctions de base assurant une expansion quadratique. Le noyau quadratique s'est en effet révélé être celui qui délivrait les meilleurs résultats sur ce jeu de données, le noyau linéaire ne permettant pas de capturer les régularités de l'ensemble d'apprentissage et les noyaux polynomiaux de degré supérieur à 2, de même que le noyau gaussien, donnant lieu à sur-apprentissage compte tenu de la petite taille du jeu d'apprentissage.

La fonction d'optimisation utilisée est la fonction fitcsvm de MATLAB, avec les paramètres par défaut, travaillant sur le méta-noyau défini au chapitre 5.

La significativité de la différence entre les résultats de méthodes différentes est évaluée en calculant la valeur-p du test-t de Student [99]. Nous fixons le seuil de significativité à 0.05.

Les méthodes de rangement agnostiques et de rangement d'objets ne disposent d'aucun principe permettant de déterminer la classe d'une instance. La classe d'un exemple de test est alors prédite à partir de la prédiction de son rang comme suit : si le rang prédit se situe entre les rangs de deux exemples d'apprentissage appartenant à une même classe, l'exemple de test prend cette classe comme étiquette; si le rang prédit est supérieur au meilleur exemple d'apprentissage d'une classe et inférieur au pire de la classe subséquente, l'exemple de test prend arbitrairement la classe inférieure comme label.

B : Résultats

Les principaux résultats sont présentés dans le Tableau 12, qui affiche pour chaque méthode le pourcentage d'instances correctement classées ainsi que l'écart type correspondant pour le classement nominal, le classement ordinal, le rangement agnostique, le rangement d'instances et le rangement d'objets.

Les valeurs-p des différences entre les résultats sont présentées dans le Tableau 13 pour toutes les caractéristiques et dans le Tableau 14 pour les caractéristiques sélectionnées.

Les valeurs-p des différences entre les résultats obtenus avec et sans sélection de caractéristiques sont présentées dans le Tableau 15.

Tableau 12 : Pourcentage des instances correctement classées et des écarts-types correspondants, ainsi que des valeurs-p respectives. Les extrema sont en gras.

	Toutes caractéristiques	*Meilleure sélection de caractéristiques*
Classement nominal	**74.7%** ± 6.5%	83.6% ± 6.4%
Classement ordinal	80.9% ± 5.8%	86.6% ± 6.1%
Rangement agnostique	89.7% ± 3.3%	91.6% ± 3.8%
Rangement d'instances	91.3% ± 3.4%	92.6% ± 4.3%
Rangement d'objets	93.6% ± 3.9%	**94.3%** ± 3.4%

Tableau 13 : Valeurs-p des différences entre les résultats obtenus sur l'ensemble des caractéristiques.

	Classement nominal	*Classement ordinal*	*Rangement agnostique*	*Rangement d'instances*	*Rangement d'objets*
Classement nominal	1				
Classement ordinal	< 0.01	1			
Rangement agnostique	< 0.01	< 0.01	1		
Rangement d'instances	< 0.01	< 0.01	0.02	1	
Rangement d'objets	< 0.01	< 0.01	< 0.01	< 0.01	1

Tableau 14 : Valeurs-p des différences entre les résultats obtenus sur les sélections de caractéristiques.

	Classement nominal	*Classement ordinal*	*Rangement agnostique*	*Rangement d'instances*	*Rangement d'objets*
Classement nominal	1				
Classement ordinal	0.02	1			
Rangement agnostique	< 0.01	< 0.01	1		
Rangement d'instances	< 0.01	< 0.01	0.19	1	
Rangement d'objets	< 0.01	< 0.01	< 0.01	0.02	1

Tableau 15 : Valeurs-p des différences, par méthode, entre les résultats obtenus sur l'ensemble des caractéristiques et sur les sélections de caractéristiques.

Classement nominal	< 0.01
Classement ordinal	< 0.01
Rangement agnostique	< 0.01
Rangement d'instances	0.09
Rangement d'objets	0.33

Dans le Tableau 12, les moins bons résultats (74,7% ± 6,5%) concernent le classement multi-classes avec toutes les caractéristiques, tandis que les meilleurs (94,3% ± 3,4%) concernent le rangement d'objets avec les caractéristiques sélectionnées.

Sans sélection de caractéristiques (Tableau 13), les valeurs-p des différences entre ces résultats sont toutes inférieures à 0,01, à l'exception de la différence entre le rangement d'instances et le rangement agnostique, qui a une valeur-p de 0,02. Avec la sélection des caractéristiques (Tableau 14), les valeurs-p des différences entre ces résultats sont toutes inférieures à 0,01 à part trois : entre le classement ordinal et le classement multi-classes (p = 0,02), entre le rangement d'instances et le rangement agnostique (p = 0,19), et entre le rangement d'objets et le rangement d'instances (p = 0,02).

Les valeurs-p des différences par méthode entre les résultats sur toutes les caractéristiques et les résultats sur les caractéristiques sélectionnées (Tableau 15) sont toutes inférieures à 0,01, à part dans le cas du rangement d'instances (0,09) et du rangement d'objets (0,33).

C : Discussion

Certains problèmes sont énoncés sous une forme qui peut conduire à l'application de méthodes mettant en œuvre un biais d'apprentissage qui ne correspond pas nécessairement à la nature réelle de ce problème. Dans cette étude, nous nous concentrons sur les problèmes implicites de rangement d'objets (PIRO) ; Les PIRO sont des problèmes de rangement d'objets qui peuvent être résolus par des méthodes inappropriées telles que le classement nominal, le classement ordinal, le rangement agnostique ou le rangement d'instances. Nous proposons une méthode pour les résoudre avec un biais d'apprentissage basé sur le rangement d'objets.

L'efficacité de la méthode proposée est comparée à celle du classement nominal, du classement ordinal, du rangement agnostique et du rangement d'instances, avec et sans sélection de caractéristiques (Tableau 12), et l'importance des résultats est évaluée à l'aide de la valeur-p (Tableau 13, Tableau 14 et Tableau 15), le seul de significativité étant fixée à 0,05.

Dans le Tableau 12, avec ou sans sélection de caractéristiques, les méthodes de rangement donnent de meilleurs résultats (respectivement 91,6 ± 3,8% et 89,7% ± 3,3% pour le pire des cas) que les méthodes de classement ordinal (86,6% ± 6,1% et 80,9% ± 5,8% respectivement pour le meilleur des cas). Parmi les méthodes de classement, le classement ordinal donne de meilleurs résultats (86,6% ± 6,1% et 80,9% ± 5,8% respectivement avec et sans sélection de caractéristiques) que le classement multi-classes (83,6% ± 6,4% et 74,7% ± 6,5% avec et sans sélection de caractéristiques respectivement). Parmi les méthodes de rangement, le rangement d'objets donne de meilleurs résultats (94,3% ± 3,4% et 93,6% ± 3,9% avec et sans sélection de caractéristiques, respectivement) que le rangement d'instances (92,6% ± 4,3% et 91,3% ± 3,4% avec et sans sélection de caractéristiques respectivement), lequel donne de meilleurs résultats que le rangement agnostique (91,6% ± 3,8% et 89,7% ± 3,3% avec et sans sélection de caractéristiques, respectivement). Dans les Tableau 13 et Tableau 14, les valeurs-p des différences entre les résultats obtenus indiquent que ces différences sont toutes significatives, à l'exception de celle entre le rangement d'instances et le rangement agnostique dans le cas de la sélection des caractéristiques (valeur-p = 0.19).

Ces résultats peuvent être interprétés à la lumière des biais d'apprentissage mis en œuvre par les différentes méthodes : comme indiqué au chapitre 1, des biais d'apprentissage plus forts ont un pouvoir de généralisation accru à de nouvelles instances, tandis que des biais d'apprentissage plus corrects produisent des prédictions plus correctes ([92] p.45).

Les performances augmentent entre classement nominal et le classement ordinal de même qu'entre le classement ordinal et le rangement. Cela semble refléter la force des biais d'apprentissage mis en œuvre par ces méthodes. Le classement nominal présente le biais d'apprentissage le plus faible, sans hypothèse sur les relations entre les éléments ni entre les classes. Le classement ordinal a un biais d'apprentissage plus fort en ce sens que même s'il ne fait pas de supposition sur les relations entre les éléments, il contraint les classes à être

totalement ordonnées. Le rangement présente le biais d'apprentissage le plus important, qui contraint les éléments à être totalement ordonnés (les classes étant donc nécessairement totalement ordonnées).

Parmi les trois méthodes de rangement, les performances augmentent entre le rangement agnostique et le rangement d'instances de même qu'entre le rangement d'instances et le rangement d'objets. La différence entre le rangement agnostique et les deux autres méthodes de rangement peut s'expliquer encore une fois par la force du biais d'apprentissage. En effet, le biais d'apprentissage du rangement agnostique contraint les éléments à être totalement ordonnés mais n'impose aucune contrainte sur l'ordre total lui-même, tandis que le biais d'apprentissage du rangement d'instances impose de choisir l'ordre total qui met le plus l'accent sur les classes et que le biais d'apprentissage du rangement d'objets impose de choisir l'ordre total qui met le moins l'accent sur les classes. Etant donné que le rangement d'instances et le rangement d'objets présentent des biais d'apprentissage aussi forts (les deux ramenant à un le nombre d'hypothèses parmi lesquelles l'optimum peut se situer), la force du biais d'apprentissage ne peut expliquer la différence entre leurs performances. Néanmoins, la justesse de leurs biais d'apprentissage le pourrait : les contraintes imposées par le biais d'apprentissage du rangement d'objets semblent plus pertinentes pour le problème que nous abordons que celles imposées par le biais du rangement d'instances ; en effet, les classes n'ont pas de signification intrinsèque dans l'évaluation de la qualité des perles de culture de Tahiti, il semble donc raisonnable qu'une méthode favorisant l'ordre total le moins dépendant de la classe donne de meilleurs résultats.

Les résultats semblent confirmer que la machine à noyau proposée implémente un biais d'apprentissage qui répond bien au problème implicite de rangement d'objets de l'évaluation de la qualité du lustre des perles de Tahiti.

Dans le Tableau 15, la sélection des caractéristiques améliore les résultats pour chaque méthode. Toutefois, l'amélioration n'est pas significative pour le rangement d'instances (valeur-p = 0.09) et le rangement d'objets (valeur-p = 0.33). Le fait que les résultats du rangement d'objets avec et sans sélection d'entités soient assez proches (respectivement 93,6% ± 3,9% et 94,3% ± 3,4%) peut être interprété comme un indice que le biais d'apprentissage intégré au

rangement d'objets s'inscrit bien dans le problème, ce qui rend un prétraitement tel que la sélection de caractéristiques inutile et permet un allégement de la charge de travail liée au prétraitement des données.

III : Validation des caractéristiques du lustre

Nous étudions ici la validité des caractéristiques utilisées pour appréhender le lustre des perles de Tahiti. Cette étude répond principalement à deux objectifs. Le premier objectif a trait aux vecteurs de caractéristiques seuls. Nous souhaitons avoir un aperçu de leur expressivité. Cela peut être fait en étudiant les relations entre les valeurs extraites. En effet, plus leur corrélation est faible, plus l'expressivité de leur combinaison est élevée. Le second objectif a trait aux vecteurs de caractéristiques avec leurs labels correspondants. Nous cherchons à nous faire une idée de la pertinence des différentes caractéristiques. Cela peut être fait en étudiant comment différents sous-ensembles de caractéristiques permettent de prédire la qualité du lustre. Enfin, une motivation incidente est de tester l'hypothèse selon laquelle le lustre serait totalement déterminé par la réflectance spéculaire.

A : Expressivité des caractéristiques du lustre

Le premier objectif a trait à l'expressivité du jeu de caractéristiques. Cette expressivité peut être testée en étudiant les relations entre les valeurs extraites, ce indépendamment des labels correspondants aux vecteurs : plus leur corrélation est faible, plus l'expressivité de leur combinaison est élevée.

1 : Méthode

Nous calculons les coefficients de corrélations entre les caractéristiques sur le jeu de données, ainsi que la valeur-p ; nous fixons le seuil de significativité à 0,05.

2 : Résultats

Les résultats sont présentés dans le Tableau 16.

a : Coefficients dont la valeur absolue est supérieure à 0.6

• Orient (C7) et effet miroir (C5) avec un coefficient de 0,69.

• Variance de la chromaticité (C10) et saturation moyenne (C9) avec un coefficient de 0,65.

• Variance de la chromaticité (C10) et iridescence (C 8) avec un coefficient de 0,64.

• Iridescence (C8) et brouillard (C6) avec un coefficient de 0,61.

b : Coefficients dont la valeur absolue est entre 0.5 et 0.6

• Les troisième (C3) et deuxième (C2) caractéristiques pour le brillant spéculaire avec un coefficient de 0,58.

• Iridescence (C8) et orient (C7) avec un coefficient de 0,57.

• La deuxième (C2) et la première (C1) caractéristiques pour le brillant spéculaire avec un coefficient de 0,51.

c : Coefficients dont la valeur absolue est entre 0.4 et 0.5

• Variance de chromaticité (C10) et brouillard (C6) avec un coefficient de 0,47

• Distinction de l'image réfléchie (C5) et de la troisième caractéristique pour l'intensité de la réflectance spéculaire (C3) avec un coefficient de -0,44)

• Réflectance profonde (C7) et brume (C6) avec un coefficient de 0,43

• Variance de chromaticité (C10) et réflectance profonde (C7) avec un coefficient de 0,41.

Tableau 16 : Matrice de corrélation des caractéristiques. Les coefficients dont la valeur-p est inférieure à 0,01 sont en gras.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C1	1									
C2	**0.51**	1								
C3	**0.15**	**0.58**	1							
C4	**0.16**	**0.27**	**0.26**	1						
C5	**-0.07**	**-0.3**	**-0.44**	**0.06**	1					
C6	**0.34**	**0.35**	0.04	**0.29**	**0.19**	1				
C7	0.01	**-0.17**	**-0.34**	**0.15**	**0.69**	**0.43**	1			
C8	**0.18**	**0.21**	-0.01	**0.2**	**0.14**	**0.61**	**0.57**	1		
C9	-0.02	**-0.19**	**-0.27**	0.03	**0.23**	**0.2**	**0.25**	**0.34**	1	
C10	**0.14**	0.05	**-0.10**	**0.17**	**0.12**	**0.47**	**0.41**	**0.64**	**0.65**	1

3 : Discussion

La corrélation entre l'orient (C7) et l'effet miroir (C5) peut provenir de la façon dont ces caractéristiques sont calculées. Les deux reposent sur des amplitudes de gradient provenant d'une même région : gradient d'intensités pour le premier, gradient de chromaticité pour le second. Cependant, l'orient a un impact sur l'intensité et est ensuite pris en compte dans le calcul de la caractéristique capturant l'effet miroir. Cette caractéristique pourrait être améliorée afin d'ignorer les arêtes provenant de l'orient.

La corrélation entre la variance de la chromaticité (C10) et la saturation moyenne (C9) vient probablement de la façon dont la chromaticité est calculée : elle contient des informations non seulement sur la chromaticité mais aussi sur la saturation.

En ce qui concerne la corrélation entre la variance de la chromaticité (C10) et l'iridescence (C8), ces caractéristiques visent à capturer les changements de couleur, bien que le premier se

concentre sur la réflectance spéculaire et le second sur la réflectance diffuse. Ces caractéristiques ne sont pas censées être indépendantes (cette dernière étant une caractéristique supplémentaire pour tester l'hypothèse selon laquelle le lustre est totalement déterminé par la réflectance spéculaire), cependant, la restriction de cette dernière à la région diffuse réduirait probablement la corrélation.

La corrélation entre l'iridescence (C8) et le brouillard (C6) peut être due au fait que l'iridescence apparaît dans la région de brume.

La corrélation entre C2 et C1 ainsi que C3 et C2 (les trois classes d'intensités les plus élevées de l'histogramme des intensités) semble raisonnable puisque ces caractéristiques visent à capturer un même aspect.

La corrélation entre iridescence (C8) et orient (C7) peut provenir du fait que les deux caractéristiques visent à appréhender les changements de couleur dus à la réflectance spéculaire et reposent donc sur les informations chromatiques, mais pas de la même manière ni sur les mêmes régions. Une meilleure séparation de leurs régions respectives pourrait améliorer ces caractéristiques.

La corrélation entre la variance de chromaticité (C10) et le brouillard (C6) peut provenir du fait que le brouillard peut avoir la même couleur que la lumière incidente ou être iridescent, mais dans les deux cas, il a un impact sur la variance de chromaticité.

Les corrélations entre la réflectance profonde (C7), le brouillard (C6) et la variance de chromaticité (F10) pourraient être expliquées en considérant que la région de brouillard et la variance de la chromaticité englobent la région de l'orient ; l'orient ajoute alors à la variance d'intensité ainsi qu'à la variance chromatique.

Concernant l'anticorrélation entre l'effet miroir (C5) et la troisième caractéristique du brillant spéculaire (C3), une valeur élevée pour la troisième caractéristique du brillant spéculaire pourrait être interprétée comme une limite inférieure pour les intensités élevées, limitant indirectement l'amplitude des variations pour les intensités élevées. L'effet miroir devrait se concentrer sur la netteté de la variation d'intensité, et non sur son amplitude ; cette anticorrélation souligne la limite du procédé consistant à prendre le gradient moyen des intensités sur toute la région de la réflectance spéculaire élevée en tant que mesure de l'effet

miroir. La caractéristique pourrait être affinée en confinant la région d'intérêt à l'environnement des arêtes.

B : Pertinence des caractéristiques du lustre

Le second objectif a trait à la pertinence des différentes caractéristiques. Cette pertinence peut être testée en étudiant le pouvoir prédictif des sous-ensembles de caractéristiques.

1 : Méthode

Il ne s'agit pas ici de rechercher le meilleur algorithme d'apprentissage mais, pour un algorithme d'apprentissage donné, la meilleure combinaison de caractéristiques. Cette recherche s'apparente à une opération de sélection des caractéristiques.

2 : Résultats

Le Tableau 17 indique pour chaque méthode utilisée la combinaison de caractéristiques qui a permis de délivrer les meilleurs résultats (ceux de la colonne de droite du Tableau 12).

Les deux premières caractéristiques du brillant spéculaire (C1 et C2) sont présentes dans la meilleure sélection pour chacune des cinq méthodes, ainsi que le brouillard (C6), l'orient (C7) et la variance de la chromaticité (C10). Le contraste (C4) et l'effet miroir (C5) ne sont présents dans aucune d'entre elles.

Tableau 17 : Sélection des meilleurs ensembles de caractéristiques selon la méthode d'apprentissage.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Classement nominal	X	X	X	-	-	X	X	-	X	X
Classement ordinal	X	X	-	-	-	X	X	X	-	X
Rangement agnostique	X	X	-	-	-	X	X	X	-	X
Rangement d'instances	X	X	-	-	-	X	X	-	-	X
Rangement d'objets	X	X	-	-	-	X	X	-	X	X

3 : Discussion

Les résultats du Tableau 17 suggèrent que le contraste (C4) et l'effet miroir (C5) n'ont pas d'impact sur l'évaluation du lustre. En revanche, le brillant spéculaire (C1 et C2), le brouillard (C6), l'orient (C7) et la variance de la chromaticité (C10) sont sélectionnés pour toutes les méthodes, suggérant que ces caractéristiques sont pertinentes.

Le lustre est généralement considéré comme étant totalement déterminé par la réflectance spéculaire et souvent réduit à l'effet miroir. Ces deux propositions sont ici invalidées : la variance de la chromaticité, qui est indépendante de la réflectance spéculaire, a un impact sur l'évaluation de la qualité du lustre alors que l'effet miroir n'en a aucun. Néanmoins, cette expérience est basée sur l'évaluation d'un seul expert et devrait être étendue à davantage d'experts afin de vérifier ces invalidations.

IV : Synthèse

Dans ce chapitre, nous avons appliqué l'algorithme de rangement d'objets à l'évaluation du lustre des perles de Tahiti. L'évaluation du lustre des perles de Tahiti constitue un cas concret de problème implicite de rangement d'objets.

L'algorithme de rangement d'objets donne de meilleurs résultats (94.3% ± 3.4% ; $p < 0.01$) que l'algorithme de rangement d'instances (92.6% ± 4.3% ; $p < 0.01$). Ceci semble indiquer que l'algorithme de rangement d'objets implémente un biais d'apprentissage qui convient mieux au problème de l'évaluation du lustre des perles de Tahiti que le biais d'apprentissage implémenté par l'algorithme de rangement d'instances.

En outre, cette expérience a permis d'invalider deux idées communément admises selon lesquelles le lustre serait totalement déterminé par la réflectance spéculaire, voire assimilable à l'effet miroir.

Conclusion

Il est possible de redéfinir les problèmes d'apprentissage automatique des données ordinales d'après la nature des données qu'ils traitent plutôt que d'après la forme sous laquelle ils se posent.

Les problèmes de classement ordinal sont des problèmes de classement où il existe un ordre total sur les classes.

Les problèmes de rangement sont des problèmes où il existe un ordre total sur les éléments. Ces problèmes peuvent se décliner en trois catégories : rangement d'instances ; rangement d'objets ; rangement agnostique. Le rangement d'instances suppose l'existence de classes intrinsèques dans lesquelles les éléments ordonnés sont distribués. Le rangement d'objets suppose que de telles classes n'existent pas. Le rangement agnostique ne fait aucune supposition quant à l'existence de telles classes.

Enfin, les problèmes implicites de rangements sont des problèmes de rangement posés en termes de classement (les données d'apprentissage sont sous forme de classes d'éléments et les prédictions portent sur l'appartenance d'un élément à une classe) mais où il existe un ordre total sur l'ensemble des éléments. Un problème implicite de rangement d'objets (PIRO) est un problème de rangement implicite où les informations ordinales à exploiter sont exclusivement liées aux rangs ; un problème implicite de rangement d'instances (PIRI) est un problème de rangement implicite où les informations ordinales à exploiter sont dépendantes des classes.

Pour qu'une méthode d'apprentissage soit performante, il faut que son biais d'apprentissage coïncide avec la nature des données à laquelle elle s'applique. A ce titre, il n'existe pas de méthode spécifiquement dédiée à l'apprentissage des rangements d'objets.

La résolution des problèmes de rangement d'instances consiste à rechercher l'ordre total qui accentue les classes d'éléments. Cet ordre est celui qui maximise la distance minimale entre éléments de classes distinctes.

La résolution des problèmes de rangement d'objets consiste à rechercher l'ordre total qui au contraire atténue toute forme de classification des éléments. Nous avons proposé à cet effet l'ordre qui maximise la distance minimale avec tout ordre invalide.

Nous avons alors formalisé le problème d'optimisation correspondant puis dérivé une forme compatible avec les bibliothèques standard, avant d'appliquer l'appliquer à un problème concret de rangement d'objets : l'évaluation du lustre des perles de Tahiti.

Nous avons ensuite mené une étude du lustre des perles de Tahiti qui a permis de préciser les différents aspects du lustre des perles de Tahiti, tels l'éclat, l'orient ou encore l'iridescence puis proposé un procédé de quantification de ces aspects à partir de photographies de perles afin de permettre l'élaboration de vecteurs de caractéristiques offrant une représentation numérique du lustre des perles.

L'algorithme de rangement d'objets permet d'imiter un expert en perles à 94.3% (± 3.4% ; $p < 0.01$), ce qui est légèrement meilleur que l'algorithme de rangement d'instances (92.6% ± 4.3% ; $p < 0.01$). Ceci semble indiquer que l'algorithme de rangement d'objets implémente un biais d'apprentissage qui convient mieux au problème de l'évaluation du lustre des perles de Tahiti que le biais d'apprentissage implémenté par l'algorithme de rangement d'instances.

De manière incidente, cette expérience a permis d'invalider l'idée selon laquelle le lustre des perles serait totalement déterminé par la réflectance spéculaire et celle selon laquelle il serait assimilable à l'effet miroir.

Cette thèse jette les bases d'un développement d'une assistance automatisée au tri des perles de Tahiti. La méthode proposée n'a cependant été testée que sur le critère du lustre ; la prochaine étape serait de déterminer des vecteurs de caractéristiques qui offrent une représentation complète des perles de Tahiti.

Des problèmes liés à la complexité de la méthode proposée pourraient se poser en cas d'application de cette méthode à des données massives. En effet, l'apprentissage ne s'effectuant pas directement sur les éléments mais sur les paires d'éléments, la taille du jeu d'apprentissage effectivement traité est égale au carré de la taille du jeu d'apprentissage initial.

De ce point de vue, une prochaine étape serait de développer des heuristiques destinées à approcher à moindre coût les solutions de la méthode proposée. Ce problème de passage à l'échelle ne se pose pas seulement pour le rangement d'objets mais aussi pour le rangement d'instances ; en effet, lui aussi considère les paires d'éléments, non les éléments en eux-mêmes.

Les applications de la méthode proposée ne sont pas cantonnées à l'évaluation du lustre des perles de cultures de Tahiti ; elles peuvent s'étendre aux autres problèmes de rangement d'objets, qu'ils soient explicites ou implicites.

Un exemple d'application, qui pourrait intéresser les sciences sociales, est celui de l'exploitation des données recueillies dans les enquêtes de satisfaction ou autres sondages utilisant des outils ordinaux telle l'échelle de Likert [144]. L'échelle de Likert consiste à présenter une série de propositions à des individus et à recueillir le positionnement de chaque individu vis-à-vis de chaque proposition par l'intermédiaire d'un nombre limité de réponses possibles allant du rejet à l'adhésion.

La méthode de rangement d'objets permettrait de trier les différentes propositions par ordre d'adhésion pour chaque individu. En effet, le nombre de réponses possibles pourrait varier, cet ordre n'en serait pas modifié pour autant ; il s'agit donc bien d'un problème de rangement d'objets.

En outre, il pourrait être intéressant d'explorer dans quelle mesure cette méthode permettrait, appliquée à un ensemble d'individus, de dégager une tendance de groupe. En effet, même si les données sont bruitées par la présence éventuelle de contradictions entre individus, la méthode proposée pourrait être capable de retourner, en minimisant l'espérance du risque, une solution qui en moyenne générerait le moins de désaccord.

Bibliographie

[1] A. De Mauro, M. Greco, and M. Grimaldi, “What is big data? A consensual definition and a review of key research topics,” in *AIP Conference Proceedings*, 2015.

[2] M. Loesdau, S. Chabrier, and A. Gabillon, “Automatic nacre thickness measurement of tahitian pearls,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9164, pp. 446–455.

[3] M. Loesdau, S. Chabrier, and A. Gabillon, “Procédé de détermination de l’épaisseur de nacre sur une perle, notamment une perle de culture,” no. FR1670752[2]. French Polynesia, 2016.

[4] M. Loesdau, S. Chabrier, and A. Gabillon, “Hue and saturation in the RGB color space,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, vol. 8509 LNCS, pp. 203–212.

[5] M. Loesdau, S. Chabrier, and A. Gabillon, “Automatic Classification of Tahitian Pearls,” in *Advances in Intelligent Systems and Computing*, 2015, vol. 313 AISC, pp. 95–101.

[6] I. Kononenko, “Machine learning for medical diagnosis: History, state of the art and perspective,” *Artif. Intell. Med.*, 2001.

[7] S. K. Devalla *et al.*, “A deep learning approach to digitally stain optical coherence tomography images of the optic nerve head,” *Investig. Ophthalmol. Vis. Sci.*, 2018.

[8] N. B. Kachare and V. S. Inamdar, “Survey of Face Recognition Techniques,” *Int. J. Comput. Appl.*, 2010.

[9] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face Recognition: A Literature Survey,” *ACM Comput. Surv.*, 2003.

[10] K. Tretyakov, “Machine Learning Techniques in Spam Filtering,” *Computer (Long. Beach. Calif).*, 2004.

[11] T. S. Guzella and W. M. Caminhas, “A review of machine learning approaches to Spam filtering,” *Expert Systems with Applications*. 2009.

[12] S. Benson Edwin Raj and A. Annie Portia, “Analysis on credit card fraud detection methods,” in *2011 International Conference on Computer, Communication and Electrical Technology, ICCCET 2011*, 2011.

[13] T. M. Mitchell, “Mitchell1980,” vol. 1, 1990.

[14] L. Zhongwan, *Mathematical Logic for Computer Science*. 2013.

[15] C. M. Bishop, “Pattern Recognition and Machine Learning (Information Science and Statistics), 1st edn. 2006. corr. 2nd printing edn,” *Mach. Learn.*, 2006.

[16] H. Larochelle, “Étude de techniques d’apprentissage non-supervisé pour l’amélioration de l’entraînement supervisé de modèles connexionnistes,” Université de Montréal, 2008.

[17] A. L. Samuel, “Some Studies in Machine Learning,” *IBM J. Res. Dev.*, 1959.

[18] W. Chao, “Machine Learning Tutorial,” *DISP Lab, Grad. Inst. Commun. Eng. Natl. Taiwan Univ.*, vol. DISP Lab, 2011.

[19] V. Podgorelec and M. Zorman, “Decision Tree Learning,” in *Encyclopedia of Complexity and Systems Science*, 2015.

[20] Y. le Cun, “A Theoretical Framework for Back-Propagation,” *Proc. 1988 Connect. Model. Summer Sch.*, pp. 21–28, 1988.

[21] O. Simeone, “A Very Brief Introduction to Machine Learning with Applications to Communication Systems,” *IEEE Trans. Cogn. Commun. Netw.*, 2018.

[22] K. Jacobs, “Independent Identically Distributed (IID) Random Variables,” in *Discrete Stochastics*, Basel: Birkhäuser Basel, 1992, pp. 65–101.

[23] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth, “Occam’s Razor,” *Inf. Process. Lett.*, 1987.

[24] T. Dietterich and E. B. Kong, “Machine Learning Bias, Statistical Bias, and Statistical Variance of Decision Tree Algorithms,” 1995.

[25] C. Schaffer, “A Conservation Law for Generalization Performance,” in *Machine Learning Proceedings 1994*, 1994.

[26] D. R. Wilson and T. R. Martinez, “Bias and the probability of generalization,” in *Proceedings - Intelligent Information Systems, IIS 1997*, 1997.

[27] P. Domingos, “The role of Occam’s Razor in knowledge discovery,” *Data Min. Knowl. Discov.*, 1999.

[28] D. H. Wolpert, “The Lack of a Priori Distinctions between Learning Algorithms,” *Neural Comput.*, 1996.

[29] A. Ng, “CS229 Lecture notes Supervised learning,” no. 1, pp. 1–30, 2000.

[30] E. G. Learned-miller, “Introduction to Supervised Learning,” pp. 92–103, 2010.

[31] P. Dayan, “Unsupervised Learning,” *Encycl. Neurosci.*, pp. 4154–4154, 2008.

[32] K. R. Zalik, “An efficient k-means clustering algorithm,” *Pattern Recognit. Lett.*, vol. 29, no. 9, pp. 1385–1391, 2008.

[33] V. N. Vapnik, “An overview of statistical learning theory,” *IEEE Trans. Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.

[34] T. Nguyen and S. Sanner, “Algorithms for Direct 0-1 Loss Optimization in Binary Classification,” *Proc. 30th Int. Conf. Mach. Learn.*, vol. 28, pp. 1085–1093, 2013.

[35] W. Yang and M. Sim, “Yanga14,” pp. 1–9, 2014.

[36] T. Sypherd, M. Diaz, L. Sankar, and P. Kairouz, “A Tunable Loss Function for Binary Classification,” no. 3, 2019.

[37] L. Rosasco, E. De Vito, A. Caponnetto, M. Piana, and A. Verri, “Are Loss Functions All the Same?,” *Neural Comput.*, vol. 16, no. 5, pp. 1063–1076, 2004.

[38] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, “Supervised Machine Learning: A Review of Classification Techniques,” *Artif. Intell. Rev.*, vol. 26, pp. 159–190, 2006.

[39] F. Mansur, V. Patel, and M. Patel, “A review on recommender systems,” *Proc. 2017 Int. Conf. Innov. Information, Embed. Commun. Syst. ICIIECS 2017*, vol. 2018-Janua, no. November, pp. 1–6, 2018.

[40] C. Winship, “Regression models with ordinal variables,” *Am. Sociol. Rev.*, vol. 49, pp. 512–525, 1984.

[41] V. Torra, J. Domingo-Ferrer, J. M. Mateo-Sanz, and M. Ng, “Regression for ordinal variables without underlying continuous variables,” *Inf. Sci. (Ny).*, vol. 176, no. 4, pp. 465–474, 2006.

[42] W. Chu, “Gaussian processes for ordinal regression.,” *J. Mach. Learn. Res.*, vol. 6, pp. 1019–1041, 2005.

[43] R. Herbrich, T. Graepel, and K. Obermayer, “Support Vector Learning for Ordinal Regression,” *Science (80-.).*, 1999.

[44] S. S. Stevens, “On the Theory of Scales of Measurement,” *Science (80-.).*, vol. 103, no. 2684, pp. 677 LP – 680, Jun. 1946.

[45] R. Kessler, J. M. Torres-Moreno, and M. El-Bèze, “Classification automatique de courriers électroniques par des méthodes mixtes d’apprentissage,” *Ingénierie des systèmes d’information*, vol. 11, no. 2, pp. 93–112, 2007.

[46] A. Cornuéjols, L. Miclet, and V. Barra, *Antoine Cornuéjols Laurent Miclet-Vincent Barra 3 e édition Algorithmes*, 3rd ed. Paris: Eyrolles, 2018.

[47] R. Faiz and N. Naffakhi, “Apprentissage supervisé pour la classification des images basé sur la structure P-tree Introduction,” *Rev. des Nouv. Technol. l’Information*, vol. 2, no. December, pp. 469–480, 2005.

[48] “Classer,” *Dictionnaire de l’Académie française*. [Online]. Available: https://academie.atilf.fr/9/consulter/classer?page=1. [Accessed: 08-Nov-2018].

[49] R. Jalam, “Apprentissage automatique et catégorisation de textes multilingues,” UNIVERSITÉ LUMIÈRE LYON2, 2003.

[50] J. Ah-pine, “Apprentissage automatique,” Lyon, 2019.

[51] N. Grira, M. Crucianu, and N. Boujemaa, “Unsupervised and Semi-supervised Clustering: A Brief Survey,” *A Rev. Mach. Learn. Tech. Process. Multimed. Content, Rep. MUSCLE Eur. Netw. Excell.*, pp. 1–12, 2004.

[52] L. Rokach and O. Maimon, “Clustering Methods, Book Section,” *Data Min. Knowl. Discov. Handb.*, pp. 321–352, 2005.

[53] N. Rajalingam and K. Ranjini, “Hierarchical Clustering Algorithm - A Comparative Study,” *Int. J. Comput. Appl.*, 2011.

[54] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: a review,” *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999.

[55] J. P. Jiawei Han, Micheline Kamber, *Data Mining – Concepts & Techniques*. 2011.

[56] P. Awasthi and R. B. Zadeh, "Supervised Clustering," *Adv. Neural Inf. Process. Syst.*, pp. 91–99, 2010.

[57] T. Finley and T. Joachims, "Supervised clustering with support vector machines," *Proc. 22nd Int. Conf. Mach. Learn. - ICML '05*, pp. 217–224, 2005.

[58] A. Shashua and A. Levin, "Ranking with large margin principle: Two approaches," *Adv. Neural Inf. Process. Syst. 15*, pp. 937–944, 2002.

[59] F. Ricci, L. Rokach, and B. Shapira, "Recommender Systems: Introduction and Challenges," in *Recommender Systems Handbook*, Boston, MA: Springer US, 2015, pp. 1–34.

[60] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web," *World Wide Web Internet Web Inf. Syst.*, 1998.

[61] E. Frank and M. Hall, "A Simple Approach to Ordinal Classification," in *Machine Learning: ECML 2001*, 2001, pp. 145–156.

[62] J. Fürnkranz and E. Hüllermeier, "Preference Learning: An Introduction," in *Preference Learning*, 2010, pp. 1–17.

[63] H. Li, "A short introduction to learning to rank," *IEICE Trans. Inf. Syst.*, 2011.

[64] P. A. Gutiérrez, M. Pérez-Ortiz, J. Sánchez-Monedero, F. Fernández-Navarro, and C. Hervás-Martínez, "Ordinal Regression Methods: Survey and Experimental Study," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 1, pp. 127–146, 2016.

[65] W. Chu and S. S. Keerthi, "Support vector ordinal regression," *Neural Comput.*, 2007.

[66] L. Grilli and C. Rampichini, "Ordered Logit Model," *Encycl. Qual. Life Well-Being Res.*, pp. 4510–4513, 2014.

[67] J. S. Cardoso and J. F. Pinto, "Learning to Classify Ordinal Data : The Data Replication Method," *J. Mach. Learn. Res.*, 2007.

[68] R. D. McKelvey and W. Zavoina, "A statistical model for the analysis of ordinal level dependent variables," *J. Math. Sociol.*, vol. 4, no. 1, pp. 103–120, 1975.

[69] P. McCullagh, "Regression Models for Ordinal Data," *J. R. Stat. Soc. Ser. B*, vol. 42, no. 2, pp. 109–127, 1980.

[70] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[71] L. Gaudette and N. Japkowicz, “Evaluation methods for ordinal classification,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5549 LNAI, pp. 207–210, 2009.

[72] W. Waegeman, B. De Baets, and L. Boullart, “ROC analysis in ordinal regression learning,” *Pattern Recognit. Lett.*, vol. 29, no. 1, pp. 1–9, 2008.

[73] J. S. CARDOSO and R. SOUSA, “Measuring the Performance of Ordinal Classification,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 25, no. 08, pp. 1173–1195, 2012.

[74] X. Juret, “Conditions suffisantes de monotonie des procedures de rangement iteratives,” *Math. Sci. hum./ Math. Soc. Sci.*, vol. 161, pp. 59–77, 2003.

[75] D. Bouyssou and P. Vincke, “Relations binaires et modélisation des préférences,” in *Concepts et méthodes pour l’aide à la décision*, D. Bouysso, D. Dubois, M. Pirlo, and H. Prade, Eds. Lavoisier, Hermes Science Publications, 2006.

[76] N. Caspard, B. Leclerc, and B. Monjardet, *Ensembles ordonnés finis : concepts, résultats et usages*, vol. 60. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[77] J. Fürnkranz and E. Hullermeier, “Pairwise Preference Learning and Ranking,” *Mach. Learn. ECML 2003*, pp. 145–156, 2003.

[78] J. Fürnkranz and E. Hüllermeier, “Preference learning and ranking by pairwise comparison,” in *Preference Learning*, 2011, pp. 65–82.

[79] W. W. Cohen, R. E. Schapire, and Y. Singer, “Learning to order things,” *J. Artif. Intell. Res.*, 1999.

[80] P. Slater, “Inconsistencies in a Schedule of Paired Comparisons,” *Biometrika*, vol. 48, no. 3/4, pp. 303–312, 1961.

[81] O. Hudry, “On the complexity of Slater’s problems,” *Eur. J. Oper. Res.*, vol. 203, no. 1, pp. 216–221, 2010.

[82] D. Shmoys, “Cut problems and their application to divide and conquer,” in *Approximation algorithms for NP-hard problems*, 1996, pp. 192–235.

[83] A. Agresti, “Non-Model Based Analysis of Ordinal Association,” in *Categorical data analysis*, 2nd ed., New York: Wiley-Interscience, 2002, p. 68.

[84] Y. Wang, L. Wang, Y. Li, D. He, T.-Y. Liu, and W. Chen, “A Theoretical Analysis of NDCG Type Ranking Measures,” pp. 1–30, 2013.

[85] K. Järvelin and J. Kekäläinen, “Cumulated gain-based evaluation of IR techniques,” *ACM Trans. Inf. Syst.*, 2002.

[86] C. D. Manning, P. Raghavan, and H. Schutze, *Introduction to Information Retrieval*. 2008.

[87] T. Joachims, “Optimizing search engines using clickthrough data,” in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’02*, 2002.

[88] R. Fagin, R. Kumar, and D. Sivakumar, “ Comparing Top k Lists ,” *SIAM J. Discret. Math.*, 2003.

[89] M. Walesiak and A. Dudek, “Finding Groups in Ordinal Data: An Examination of Some Clustering Procedures,” in *Classification as a Tool for Research*, 2010, pp. 185–192.

[90] M. Giordan and G. Diana, “A Clustering Method for Categorical Ordinal Data,” *Commun. Stat. - Theory Methods*, vol. 40, no. 7, pp. 1315–1334, Mar. 2011.

[91] C. Biernacki, J. Jacques, and C. Biernacki Julien Jacques, “Model-Based Clustering of Multivariate Ordinal Data Relying on a Stochastic Binary Search Algorithm,” Springer Verlag, 2016.

[92] T. Mitchell, *Machine Learning*. 1997.

[93] V. N. Vapnik, *Statistical Learning Theory*. 1998.

[94] C. Cortes and V. Vapnik, “In Silico Log P Prediction for a Large Data Set with Support Vector Machines, Radial Basis Neural Networks and Multiple Linear Regression.,” *Chem. Biol. Drug Des.*, vol. 20, pp. 273–297, 1995.

[95] T. Segaran, “Kernel Methods and SVMs,” *Program. Collect. Intell.*, pp. 197–224, 2007.

[96] J. Mercer, “Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 209, no. 441–458, pp. 415–446, Jan. 1909.

[97] J. Platt, “Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines,” *URL citeseer. ist. psu. edu/platt98sequential. html*, pp. 1–21, 1998.

[98] C. Chang, C. Lin, and T. Tieleman, “LIBSVM : A Library for Support Vector Machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 307, pp. 1–39, 2008.

[99] D. G. Altman, “Practical statistics for medical research. Chapman and Hall,” *Stat. Med.*, 1991.

[100] F. Marin, G. Luquet, B. Marie, and D. Medakovic, “Molluscan Shell Proteins: Primary Structure, Origin, and Evolution,” *Current Topics in Developmental Biology*, vol. 80. pp. 209–276, 2007.

[101] M. R. Snow, A. Pring, P. Self, D. Losic, and J. Shapter, “The origin of the color of pearls in iridescence from nano-composite structures of the nacre,” *Am. Mineral.*, vol. 89, no. 10, pp. 1353–1358, 2004.

[102] S. Elen, “Identification of Yellow Cultured Pearls from the Black-Lipped Oyster Pinctada Margaritifera,” *Gems Gemol.*, vol. 38, 2002.

[103] D. Chateigner, C. Hedegaard, and H. R. Wenk, “Mollusc shell microstructures and crystallographic textures,” in *Journal of Structural Geology*, 2000, vol. 22, no. 11–12, pp. 1723–1735.

[104] “Lustre,” *Dictionnaire de l’Académie française*. [Online]. Available: https://academie.atilf.fr/consulter/lustre?page=1. [Accessed: 06-Jun-2018].

[105] “Lustre,” *Oxford Dictionnary*. [Online]. Available: https://en.oxforddictionaries.com/definition/lustre. [Accessed: 06-Jun-2018].

[106] “Lustre,” *Encyclopædia Britannica*. [Online]. Available: https://www.britannica.com/science/lustre. [Accessed: 06-Jun-2018].

[107] J. J. Kim, K. in Shin, H. S. Ryu, E. ae Kim, M. Lee, and K. Wha oh, “Luster Properties of Polyester Filament Yarn Woven Fabrics,” *Text. Res. J.*, vol. 74, no. 1, pp. 72–77, 2004.

[108] K. I. Shin, S. H. Kim, and J. J. Kim, “Image analysis of the luster of fabrics with modified cross-section fibers,” *Fibers Polym.*, vol. 6, no. 1, pp. 82–88, 2005.

[109] M. Hadjianfar, D. Semnani, and M. Sheikhzadeh, “A New Method for Measuring Luster

Index Based on Image Processing," *Text. Res. J.*, vol. 80, no. 8, pp. 726–733, 2009.

[110] Gemological Institute of America, "Pearl Quality Factors." [Online]. Available: https://www.gia.edu/pearl-quality-factor. [Accessed: 06-Jun-2018].

[111] Assemblée de la Polynésie française, "Délibération n° 2005-42 APF du 4 février 2005 portant définition des produits tirés de l'activité de la perliculture en Polynésie française et fixation des règles relatives à la classification, au transport, à la commercialisation et aux formalités d'expo." .

[112] Commision électrotechnique internationale, "Eclairage : Vocabulaire international de l'éclairage," in *Vocabulaire electrotechnique international*, Bureau central de la Commission électrotechnique internationale, Ed. Genève, 1987.

[113] G. Wendt, F. Faul, V. Ekroll, and R. Mausfeld, "Disparity, motion, and color information improve gloss constancy performance," *J. Vis.*, vol. 10, no. 9, pp. 7–7, Sep. 2010.

[114] W. M. Bergmann Tiest and A. M. L. Kappers, "Haptic and visual perception of roughness," *Acta Psychol. (Amst).*, vol. 124, no. 2, pp. 177–189, Feb. 2007.

[115] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, W. I. Ginsberg, and T. Limperis, "Geometrical considerations and nomenclature for reflectance," *Appl. Opt.*, vol. 9, pp. 1474–1475, 1977.

[116] S. R. Marschner, S. H. Westin, E. P. F. Lafortune, and K. E. Torrance, "Image-based bidirectional reflectance distribution function measurement," *Appl. Opt.*, vol. 39, no. 16, p. 2592, Jun. 2000.

[117] W. Matusik, H. Pfister, M. Brand, and L. McMillan, "Efficient Isotropic BRDF Measurement," 2003.

[118] G. J. Ward, "Measuring and modeling anisotropic reflection," *ACM SIGGRAPH Comput. Graph.*, vol. 26, no. 2, pp. 265–272, Jul. 1992.

[119] W. Matusik, H. Pfister, M. Brand, and L. McMillan, "A data-driven reflectance model," 2003, p. 759.

[120] B. T. Phong, "Illumination for computer generated pictures," *Commun. ACM*, vol. 18, no. 6, pp. 311–317, Jun. 1975.

[121] M. Ashikmin, S. Premože, and P. Shirley, “A microfacet-based BRDF generator,” 2000, pp. 65–74.

[122] X. D. He, K. E. Torrance, F. X. Sillion, and D. P. Greenberg, “A comprehensive physical model for light reflection,” *ACM SIGGRAPH Comput. Graph.*, vol. 25, no. 4, pp. 175–186, Jul. 1991.

[123] K. E. Torrance and E. M. Sparrow, “Theory for Off-Specular Reflection From Roughened Surfaces*,” *J. Opt. Soc. Am.*, vol. 57, no. 9, p. 1105, Sep. 1967.

[124] R. L. Cook and K. E. Torrance, “A Reflectance Model for Computer Graphics,” *ACM Trans. Graph.*, vol. 1, no. 1, pp. 7–24, Jan. 1982.

[125] R. Dumont, F. Pellacini, and J. A. Ferwerda, “Perceptually-driven decision theory for interactive realistic rendering,” *ACM Trans. Graph.*, vol. 22, no. 2, pp. 152–181, Apr. 2003.

[126] S. Gibson and R. J. Hubbold, “Perceptually-Driven Radiosity,” *Comput. Graph. Forum*, vol. 16, no. 2, pp. 129–141, Jun. 1997.

[127] D. Luebke and B. Hallen, “Perceptually Driven Simplification for Interactive Rendering,” in *Rendering Techniques 2001*, S. J. Gortler and K. Myszkowski, Eds. Vienna: Springer Vienna, 2001, pp. 223–234.

[128] K. Myszkowski, “Perception-based global illumination, rendering, and animation techniques,” 2002, p. 13.

[129] G. Ramanarayanan, J. Ferwerda, B. Walter, and K. Bala, “Visual equivalence: towards a new standard for image fidelity,” *ACM Trans. Graph.*, vol. 26, no. 3, p. 76, Jul. 2007.

[130] J. Wills, S. Agarwal, D. Kriegman, and S. Belongie, “Toward a perceptual space for gloss,” *ACM Trans. Graph.*, vol. 28, no. 4, pp. 1–15, Aug. 2009.

[131] E. P. F. Lafortune, S.-C. Foo, K. E. Torrance, and D. P. Greenberg, “Non-linear approximation of reflectance functions,” 1997, pp. 117–126.

[132] A. Ngan, F. Durand, and W. Matusik, “Experimental Analysis of BRDF Models,” 2005.

[133] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar, “Acquiring the reflectance field of a human face,” 2000, pp. 145–156.

[134] H. P. A. Lensch, J. Kautz, M. Goesele, W. Heidrich, and H.-P. Seidel, “Image-Based

Reconstruction of Spatially Varying Materials," 2001.

[135] R. Sève, "Problems connected with the concept of gloss," *Color Res. Appl.*, vol. 18, no. 4, pp. 241–252, Aug. 1993.

[136] Z. Pizlo, "Perception viewed as an inverse problem," *Vision Research*, vol. 41, no. 24. pp. 3145–3161, 2001.

[137] A. C. Chadwick and R. W. Kentridge, "The perception of gloss: A review," *Vision Res.*, vol. 109, pp. 221–235, Apr. 2015.

[138] R. W. Fleming, "Visual perception of materials and their properties," *Vision Res.*, vol. 94, pp. 62–75, 2014.

[139] F. B. Leloup, M. R. Pointer, P. Dutré, and P. Hanselaer, "Overall gloss evaluation in the presence of multiple cues to surface glossiness," *J. Opt. Soc. Am. A*, vol. 29, no. 6, p. 1105, 2012.

[140] Commission Internationale de L'Eclairage, "A framework for the measurement of visual appearance," Vienna, 2006.

[141] F. B. Leloup, G. Obein, M. R. Pointer, and P. Hanselaer, "Toward the soft metrology of surface gloss: A review," *Color Res. Appl.*, 2014.

[142] R. Hunter, "Methods of determining gloss," *Res. Pap. RP958*, vol. 18, Jan. 1937.

[143] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection 2 Methods for Accuracy Estimation," *Proc. of IJCAI'95*, pp. 1137–1145, 1995.

[144] R. Likert, "A technique for the measurement of attitudes," *Arch. Psychol.*, 1932.

[145] Institut de la statistique de Polynésie française, "Bilan de la perle 2014."

[146] "RAPA project." .

[147] M. Awad and R. Khanna, "Machine Learning," in *Efficient Learning Machines*, Berkeley, CA: Apress, 2015, pp. 1–18.

[148] "Classifier," *Dictionnaire de l'Académie française*. [Online]. Available: https://academie.atilf.fr/9/consulter/classifier?page=1. [Accessed: 08-Nov-2018].

[149] C. F. Eick, N. Zeidat, and Z. Zhao, "Supervised clustering - Algorithms and benefits," *Proc.* -

Int. Conf. Tools with Artif. Intell. ICTAI, no. December 2004, pp. 774–776, 2004.

[150] N. Zeidat, C. F. Eick, and Z. Zhao, “Supervised Clustering : Algorithms and Application,” *Analysis*, pp. 94–99, 2006.

[151] X. Juret, “Conditions suffisantes de monotonie des procédures de rangement itératives,” *Mathématiques Sci. Hum.*, 2003.

[152] “Iridescence,” *Encyclopædia Britannica*. [Online]. Available: https://www.britannica.com/science/iridescence-mineralogy. [Accessed: 06-Jun-2018].

[153] “Lustre,” *Cambridge Dictionnary*. [Online]. Available: https://dictionary.cambridge.org/dictionary/english/lustre. [Accessed: 06-Jun-2018].

[154] “Gloss/Brillant,” *Commission électrotechnique internationale*. [Online]. Available: http://www.electropedia.org/iev/iev.nsf/display?openform&ievref=845-04-73. [Accessed: 06-Jun-2018].

[155] R. Bazin, L. Chommeloux, G. Obadia, and H. Chardron, “Appareil destiné à permettre d’évaluer la brillance d’une surface, en particulier de la peau,” EP0475803A1, 1990.

[156] E. Rapaport, A. Nussinovitsch, and E. Mey-Tal, “Glossmeter,” US6018396A, 1995.

[157] W. E. K. Middleton and A. G. Mungall, “The Luminous Directional Reflectance of Snow*,” *J. Opt. Soc. Am.*, vol. 42, no. 8, pp. 572–579, 1952.

[158] B. L. Anderson, “Visual perception of materials and surfaces,” *Curr. Biol.*, vol. 21, no. 24, 2011.

[159] S. Sabar, “What’s a Gestalt?,” *Gestalt Rev.*, vol. 17, no. 1, pp. 6–34, 2013.

[160] J. Peatross and M. Ware, “Physics of Light and Optics: A Free Online Textbook,” 2010, pp. 65–68.

[161] S. Serikawa and T. Shimomura, *Method for measuring glossiness of plane surfaces based on psychological sensory scale*. 1993.