



**UNIVERSITÉ
DE GENÈVE**

Archive ouverte UNIGE

<https://archive-ouverte.unige.ch>

Thèse

2005

Summary

Open Access

This file is a(n) Summary of:

Machine learning approach to semantic augmentation of multimedia
documents for efficient access and retrieval

Kosinov, Serhiy

This publication URL:

<https://archive-ouverte.unige.ch/unige:146923>

Publication DOI:

[10.13097/archive-ouverte/unige:146923](https://doi.org/10.13097/archive-ouverte/unige:146923)

© This document is protected by copyright. Please refer to copyright holders for terms of use.

Résumé

1.1 Préface

Cette thèse porte sur le développement théorique et l'application pratique de méthodes d'apprentissage par l'ordinateur à l'analyse et la récupération de multimédia par contenu. Afin d'être clair et pour éviter toute confusion, une attention particulière doit être portée à la manière dont cette thèse interprète et utilise le terme *multimédia*. Ce terme possède plusieurs significations et interprétations, ses définitions évoluant et se transformant en parallèle avec le progrès technologique.

D'après sa signification littérale, dérivant du latin *multus* = "plusieurs, multiple..." et *médium* = "un système de communication, information ou distraction", le multimédia représente l'utilisation d'ordinateurs pour présenter des textes, graphiques, vidéos, animations et sons de manière intégrée. Or l'utilisation du terme multimédia est tout à fait différente du point de vue d'un système d'apprentissage par ordinateur, ou de tout autre système d'information par multimédia en général. Un système d'information traite avec la même configuration binaire abstraite d'une information digitalisée et parfois codée, quelle que soit son origine sensorielle. Comme certains auteurs le disent, le multimédia dans ce contexte se réfère à n'importe quelle information visuelle, auditive ou textuelle, unique ou combinée. L'information multimédia digitale est immédiatement visible, audible, lisible et dans la plupart des cas compréhensible par l'utilisateur, mais pas pour le système. Cette divergence importante entre la représentation digitale et la représentation est connue sous le nom de *fossé sémantique*, la cible principale des techniques développées dans cette thèse.

1.2 Problématique

Dans la discussion contenue dans cette thèse, nous identifions les approches qui s'intéressent au fossé sémantique discuté plus haut, comme étant celles d'augmentation sémantique puisque la plupart d'entre elles portent spécifiquement sur le rapprochement de représentations visuelles multimédia de bas niveau et leurs significations pour améliorer l'efficacité d'accès et de récupération. Le choix d'un terme de référence aussi général que l'augmentation sémantique

nous permet d'inclure et analyser entièrement une grande quantité de différentes techniques qui ont le même but: diminuer le problème du fossé sémantique, résolu dans chaque méthode individuelle à travers un vaste éventail de paradigmes et formulations qui se rapportent à l'indexage, l'apprentissage, la classification, la catégorisation, la prédiction, etc. Ce choix de terminologie est lié à la manière dont le terme multimédia est perçu tout au long de cette thèse. Multimédia est compris comme étant n'importe quelle combinaison de représentation d'une information humainement compréhensible, dont les caractéristiques automatiquement calculables ne contiennent aucune d'expression directe d'une signification pouvant être recherchée par l'utilisateur.

Un bon contexte pour décrire la contribution de cette thèse est défini en considérant deux groupes de méthodes pour l'augmentation sémantique: interactives - les approches adaptatives qui sont guidées par un feedback de pertinence de l'utilisateur, et automatique - celles qui tentent de dériver des corrélations utiles entre des caractéristiques représentatives du multimédia et leurs aspects sémantiques en appliquant des techniques qui n'impliquent pas d'utilisateur. Les méthodes du premier groupe considèrent l'utilisateur comme étant la source ultime d'information sémantique. Elles corrigent leur solution de manière itérative en demandant et en incorporant un feedback de l'utilisateur et contiennent des contributions importantes telles que l'algorithme de Rocchio, le filtrage collaboratif, l'estimation de pertinence Bayésienne, l'apprentissage actif entre autres. D'un autre côté, les méthodes du second groupe agissent de manière complètement autonome et ne sollicitent jamais d'aide de l'utilisateur, comptant uniquement sur des techniques comme l'apprentissage par l'ordinateur, la classification, l'analyse discriminante, l'indexage sémantique latent, l'analyse de corrélation entre les langues, pour induire les significations recherchées à partir des données d'entraînement à leur disposition. Les contributions présentées dans cette thèse appartiennent au groupe des méthodes d'augmentation sémantique automatique et sont considérées du point de vue de l'apprentissage par ordinateur. De ce point de vue, les techniques proposées sont conçues pour être capables d'améliorer leur performance sur la base des expériences et résultats précédents de façon autonome, dans une tentative d'éliminer le besoin, ou alléger le poids placé sur, l'intuition humaine dans l'analyse d'un problème posé. Malgré le fait que le besoin de connaissance et d'intuition humaine ne pourra probablement jamais être complètement éliminé à cause de l'importance de décisions intelligentes dans la représentation et la caractérisation des données, l'approche de l'apprentissage par l'ordinateur présente un clair et incontestable avantage. En effet, en résolvant un problème plus général et probablement plus difficile en apprentissage par l'ordinateur, la même technologie peut s'appliquer à une grande variété de problèmes particuliers sans avoir à reconstruire la solution à partir de rien à chaque fois. Une fois résolue, la même méthode générale est applicable dans les scénarii d'augmentation sémantique pour les images digitales, les vagues de signaux auditifs, les films vidéo, et plein d'autres possibilités.

Cette thèse adopte la perspective de l'apprentissage par l'ordinateur dans l'approche des problèmes d'augmentation sémantique dans l'accès et la récupération efficaces de bases de

données multimédia et établit plusieurs contributions dans les domaines détaillés dans les chapitres suivants.

1.3 Analyse discriminante

Choix du domaine de l'analyse discriminante comme point de départ de cette étude reflète une tendance et une préférence délibérées pour une méthode d'apprentissage par ordinateur discriminante au-dessus d'une approche alternative générative. Cette décision est motivée par un certain nombre de raisons. D'abord, le modèle discriminante a beaucoup plus de flexibilité dans les parties de l'espace d'entrée où les probabilités postérieures diffèrent de manière significative de 0 ou de 1, alors que les approches génératives modèlent des détails de distribution de données de l'espace d'entrée ce qui peut être non pertinent pour déterminer des probabilités postérieures. En second lieu, les modèles discriminants sont en général très rapide à la fabrication de prévisions pour des données-tests, alors que les modèles génératifs exigent souvent une solution itérative. Troisièmement, toutes choses étant égales, on le prévoit ce les méthodes discriminants auraient une meilleure performance prédictive puisqu'ils sont formés pour prévoir l'étiquette de classe plutôt que distribution commune des données de l'espace d'entrée et des étiquettes. Ainsi, tenant compte les propriétés spécifiques de l'application destinée à l'augmentation sémantique, nous développons une technique d'analyse discriminante transformationnelle basé sur la distance, DDA. Un effort étendu est engagé à faire à la formulation de DDA pour affirmer un caractère *non paramétrique* avec prétentions minimales sur la distribution de données de l'espace d'entrée, *asymétrique* à apparier le scénario de déploiement le plus populaire "1-against-all", et basé sur la *transformation* afin de tenir compte de extensions, post-traitement et l'utilisation de la transformation dérivée pour fournir une métrique discriminante. Cette métrique explique les différences dans les échelles de différents dispositifs, retire corrélations globales et redondances parmi des dispositifs dans une certaine mesure, et s'adapte au fait que quelques dispositifs peuvent être beaucoup plus instructifs au sujet des étiquettes de classe que d'autres. Afin d'accomplir ces derniers conditions, la capacité de la DDA d'extraire les dispositifs distinctifs et réduire la dimensionnalité de données de l'espace d'entrée, tout en déterminant le numéro de les dimensions suffisantes automatiquement, est d'importance cruciale. Du point de vue de l'augmentation sémantique, DDA fournit une machine binaire d'apprentissage à utiliser pour distinguer entre une certaine concept sémantique et son complément, par exemple déterminer si l'image digitale montre une entité d'intérêt ou pas. Plus formellement, la formulation de la technique d'analyse discriminante basé sur la distance est récapitulé comme suit.

Dans le cadre d'analyse discriminante, on recherche à distinguer entre deux ou plus classes ou catégories sémantiques prédéfinis de documents multimédia. En prenant en considération pour l'instant le cas simple avec deux classes sémantiques seulement, nous recherchons une telle transformation des données visuelles de dispositif qui place les exemples de la certaine classe près l'un de l'autre, tout en replaçant les exemples de l'autre classe suffisamment

loin dans un certain espace de dispositif. En d'autres termes, la transformation linéaire recherchée $T \in \mathbb{R}^{m \times k}$ doit mettre en valeur la conformité des données à l'hypothèse de compacité avec le but d'améliorer l'exactitude de performance de la méthode du voisin le plus proche (NN). Formellement, une telle transformation est trouvée comme une solution d'un problème d'optimisation où le critère suivant est être réduit au minimum :

$$\log J(T) = \frac{2}{N_X(N_X - 1)} \sum_{i < j}^{N_X} \log \Psi(d_{ij}^W(T)) - \frac{1}{N_X N_Y} \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} \log d_{ij}^B(T) \quad (1.1)$$

où N_X, N_Y sont le nombre d'observations dans les sets de données X et Y de représentation des deux classes; $d_{ij}^W(T)$ dénote une distance entre les points i et j *en dedans* le set de données X transformé par T , et, de façon analogue, $d_{ij}^B(T)$ indique distances *entre* point i du set de données transformé X et point j du set de données transformé Y , $\Psi(d_{ij}^W(t))$ est une fonction robuste d'évaluation de Huber.

Le problème de réduire au minimum (1.1) est résolu par l'intermédiaire de majorization itérative, qui substitue la tâche d'optimiser une fonction d'objective compliqué par un ordre itératif de minimisation des problèmes plus simples en termes de membres de la famille de fonction spécial de l'auxiliaire. Etant donné les propriétés de ces fonctions auxiliaires, généralement appelé comme *fonctions de majorization*, la procédure itératif produit une sequence non-croissante de valeurs de fonction d'objective, convergeant à un point stationnaire qui est un minimizer local sous certaines contraintes. Pour le critère choisi d'optimisation, (1.1), nous dérivons une fonction de majorization approximative, jusqu'à un indépendant constant de, comme exprimé dedans (1.2):

$$\mu_{\log J}(T, \bar{T}) = \frac{\alpha}{2} \text{tr}(T^T X^T R X T) + \frac{\beta}{2} \text{tr}(T^T Z^T G Z T) - 2\beta \text{tr}(T^T Z^T G Z \bar{T}), \quad (1.2)$$

où \bar{T} est le point supportant de la transformation recherchée, c'est-à-dire, sa valeur à l'itération actuelle, R et G sont les matrices pour des calculs de distance, et Z est la matrice qui se retient les deux sets de données X et Y :

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix}. \quad (1.3)$$

À chaque itération, à la minimisation de (1.2) est résolu sujet à une contrainte de régularisation paramétrisé par Δ , produisant le résultat qui devient un point supportant de l'itération ultérieure, et ainsi de suite, jusqu'au la convergence est atteinte. Ce processus de dériver une transformation discriminante ainsi que la classification de NN constitue l'essence de la méthode d'analyse discriminante basé sur la distance (DDA), qui est récapitulée par l'algorithme ci-dessous :

Algorithme DDA.

1. assigner un premier point supportant $\bar{T} = \bar{T}_0 \in \mathbb{R}^{m \times k}$;

2. trouver point successeur :

$$T_s = \arg \min_T \mu_{\log J(T, \bar{T})} + \lambda (\mathbf{tr}(T^T T) - \Delta) ;$$

3. si $\log J(\bar{T}) - \log J(T_s) < \epsilon$, s'arrêter;

4. assigner $\bar{T} = T_s$, aller à 2.

Observons que le placement de la taille de colonne de T à une valeur arbitraire $k \ll m$ rend la méthode présentée de DDA une technique de réduction de dimensionnalité, laquelle peut être utilisée dans une variété d'applications telles que sélection de caractéristiques, visualisation bas-dimensionnelle de données, etc. D'ailleurs, la valeur de k , c'est-à-dire, le nombre exact de dimensions les données peut être réduit sans la perte de pouvoir discriminatoire selon (1.1), est avec précision déterminé par le nombre de valeurs singulières différent de zéro de T . En effet, les distances entre des observations transformé peuvent être considéré comme distances entre l'observations originals dans une métrique différente TT^T . Cela peut être exprimé en utilisant la décomposition de valeur singulière $TT^T = USV^T V S U^T = U_k S_k^2 U_k^T$. L'expression obtenue indique que l'effet de la transformation plein-dimensionnelle est capturé par la première k vecteurs gauche-singuliers multiplié par leur valeurs singulières différente de zéro. Ce nombre k donne une réponse à la question de combien de dimensions sont nécessaires dans l'espace transformé.

Tandis que la discussion ci-dessus est concentrée la plupart du temps sur la configuration de deux classes, il est facile pour généraliser la formulation présenté à une configuration d'analyse discriminante de plusieurs classes, pour le numéro de classes $K \geq 2$:

$$\log J_K(T) = \sum_{i=1}^{K-1} \left(\alpha^{(i)} S_W(T)^{(i)} - \beta^{(i)} S_B(T)^{(i)} \right). \quad (1.4)$$

Noter que (1.4) devient (1.1) pour $K = 2$.

En plus de la transformation explicitement recherchée, la technique de DDA peut également être vue en tant que fournir une métrique de distance discriminante qui tient compte des différences dans les échelles de dispositifs différents, enlève des corrélations et des redondances globales parmi des dispositifs dans une certaine mesure, et s'adapte au fait que quelques dispositifs peuvent être beaucoup plus instructif au sujet des étiquettes de classe que d'autres. Cette observation est facilement illustré par l'exemple de la machine de vecteur de support (SVM) avec noyau gaussian:

$$k_{\Sigma}(x_i, x_j) = e^{-(x_i - x_j)^T \Sigma^{-1} (x_i - x_j)}, \quad (1.5)$$

pour une certaine matrice Σ de covariance et des observations x_i, x_j représenté comme vecteurs de colonne. Un choix typique de σ ici est matrice d'identité multipliée par un certain facteur constant. Cependant, quand la technique de DDA est appliquée pour prétraiter les données de formation avant le traitement de SVM, le classificateur de SVM tire profit

entièrement du dispositifs distinctifs extraits par la méthode de DDA, puisque des produits noyaux peuvent maintenant être vus comme évalués dans une nouvelle métrique discriminante TT^T :

$$k_{\Sigma}(x_i, x_j) = e^{-(x_i - x_j)^T T T^T (x_i - x_j)}. \quad (1.6)$$

Par la suite comme conséquence SVM pouvant trouver une solution plus simple en impliquant peu de vecteurs de support et améliorer les propriétés de généralisation, ce qui mène naturellement à une amélioration de performance de classification, comme confirmé par les résultats expérimentaux. Davantage d'évaluation empirique a vérifié la performance de classification de la méthode proposée de DDA et ses extensions sur un certain nombre de sets de données de référence d'UCI et sur les tâches de recherche d'image sémantique. Les résultats encourageants ont démontré que l'approche proposée de DDA surpasse plusieurs méthodes populaires, et améliore l'exactitude de classification, quand combiné avec d'autres techniques, lui faisant un excellent candidat pour l'application destinée de l'augmentation sémantique automatique.

1.4 Méthodes basé sur noyaux

Beaucoup d'algorithmes linéaires d'apprentissage par ordinateur ont été mis en valeur à travers noyaux afin de pouvoir traiter des problèmes plus complex exigeant des fonctions non linéaires de décision. De même, la méthode de DDA est moulée comme technique *d'analyse discriminante non linéaire*, surmontant l'assumption de linéarités de la transformation discriminante recherché, en menant naturellement au développement de KDDA, une extension de noyaux de DDA. Ceci est accompli en reformulant le problème en termes de distances projetées de l'espace plus riche, probablement infini-dimensionnel, de dispositif induit par une fonction de noyaux choisie, comme détaillé ci-dessous.

Supposons qu'il y a un espace \mathcal{F} d'où des échantillons des données de formation peuvent être tracées par l'intermédiaire de $\Phi : \mathbb{R}^m \rightarrow \mathcal{F}$, tels que là existe une fonction de noyaux $k(x, y) = (\Phi(x))^T \Phi(y)$, où $x, y \in \mathbb{R}^m$ et $k : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$. Nous supposerons également que la transformation distinctive est recherchée dans \mathcal{F} comme matrice ω de projection de la taille $[\mathcal{N}_{\mathcal{F}} \times d]$, où $\mathcal{N}_{\mathcal{F}}$ est la dimensionnalité de \mathcal{F} , et d est la dimension du sous-espace de projection discriminante dérivé, tel que les colonnes de ω se situent dans l'envergure de toutes les échantillons de formation tracé dans \mathcal{F} , en vertu du Théorème De Representer:

$$\omega = \left[\sum_i^N \alpha_i^{(1)} \Phi(z_i) \quad \sum_i^N \alpha_i^{(2)} \Phi(z_i) \quad \dots \quad \sum_i^N \alpha_i^{(d)} \Phi(z_i) \right], \quad (1.7)$$

où z_i est un des $N_X + N_Y$ échantillons provenant des données de formation de matrice composée Z , comme définie dedans (1.3). Les distances entre les images des échantillons x

et y projetées de \mathcal{F} par la solution ω sont ainsi exprimés comme :

$$\begin{aligned}
 \mathcal{D}_{xy}^2(\omega) &= (\Phi(x) - \Phi(y))^T \omega \omega^T (\Phi(x) - \Phi(y)) \\
 &= \mathbf{tr} (\omega^T (\Phi(x) - \Phi(y)) (\Phi(x) - \Phi(y))^T \omega) \\
 &= \sum_j^d \left(\sum_i^N \alpha_i^{(j)} (k(z_i, x) - k(z_i, y)) \right)^2.
 \end{aligned} \tag{1.8}$$

Dans la notation de matrice (1.8) peut être simplifié comme:

$$\mathcal{D}_{xy}^2(\omega) \equiv \mathcal{D}_{xy}^2(P) = \mathbf{tr} (P^T H_{xy} P) \tag{1.9}$$

où $P \in \mathbb{R}^{N \times d}$ est la transformation non linéaire recherché représentée comme matrice rassemblant tout les $\alpha_i^{(j)}$, $H_{xy} = (K_x - K_y)(K_x - K_y)^T$, et $K_s = [k(z_1, s), k(z_2, s), \dots, k(z_n, s)]^t$ dénote un vecteur des évaluations de noyaux pour l'échantillon s et tout le reste du set de données de formation.

En raison de (1.9), le logarithme du critère de l'optimisation de DDA (1.1) peut maintenant être exprimé en termes de distances projeté d'un espace plus riche et probablement infini-dimensionnel \mathcal{F} :

$$\begin{aligned}
 \log J(P) &= \frac{2}{N_X(N_X - 1)} \sum_{i=1}^{N_X} \sum_{j=i+1}^{N_X} \log \Psi (\mathcal{D}_{ij}^W(P)) \\
 &\quad - \frac{1}{N_X N_Y} \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} \log \mathcal{D}_{ij}^B(P)
 \end{aligned} \tag{1.10}$$

Le traitement du critère obtenu diffère seulement légèrement comparé au cas linéaire. De même à la voie il est fait dans la méthode de DDA, l'optimisation de (1.10) est exécutée par l'intermédiaire de majorization itérative. La fonction de majorization est dans ce cas-ci:

$$\begin{aligned}
 \mu_{\log J}(P, \bar{P}) &= \frac{1}{N_X(N_X - 1)} \mathbf{tr} (P^T \mathbb{K}_X B(\bar{P}) \mathbb{K}_X^T P) \\
 &\quad + \frac{1}{2N_X N_Y} \mathbf{tr} (P^T \mathbb{K}_{XY} C \mathbb{K}_{XY}^T P) \\
 &\quad + \frac{2}{N_X N_Y} \mathbf{tr} (P^T \mathbb{K}_{XY} G(\bar{P}) \mathbb{K}_{XY}^T \bar{P}) \\
 &\quad + \text{const},
 \end{aligned} \tag{1.11}$$

où \bar{P} est la solution actuelle, \mathbb{K}_X , \mathbb{K}_{XY} sont des matrices des produits intérieurs de noyaux évalués pour X et toutes les données, respectivement, et B , C et G sont matrices positives semi-défini indépendant de P .

Un aspect important distinguant KDDA d'autres méthodes basés sur noyaux est la convexité de la formulation qui se retient indépendamment si noyaux fondamental est positif défini. Ceci la propriété est évaluée empiriquement par l'incorporation des *noyaux indéfini* dans KDDA. Dans plusieurs domaines d'application ces noyaux sont connu pour correspondre aux mesures de distance qui améliorent la saisie de similitude perceptuelle. Supplémentairement,

une condition défavorable visée pendant que l’effet de faux projection positive est examiné et des stratégies d’élimination sont évaluées. En ce qui concerne l’augmentation sémantique, l’approche de KDDA étend non seulement la formulation précédente aux cas non linéaires, mais laisse également la machine d’apprentissage binaire de DDA être pertinemment utilisé avec des mesures non-métriques de dissimilitude par l’intermédiaire de noyaux indéfinis.

1.5 Ensembles des classificateurs

Rarement nous rencontrons un problème pratique de l’augmentation sémantique qui peut être traitée en sa totalité par une machine d’étude binaire simple. Par coïncidence, l’extension la plus populaire pour le cas avec plusieurs classes, les catégories ou les concepts sémantiques de cible est à dériver autant de machines d’étude binaires, une pour chaque étiquette de classe et, si nécessaire, exécuter un certain arbitrage parmi leurs prévisions. Dans la présente étude, on lui discute qu’une telle construction et ses variations sont désavantageux de la perspective de l’augmentation sémantique, en raison de la supposition limiteuse de l’ensemble de concepts sémantiques de cible étant *indépendant, non-recouvert, et approfondi*. Tout en fournissant a moyens d’étendre beaucoup de techniques existantes à un cas de multiple-catégorie, cette prétention peut mener aux résultats contradictoires, par exemple prévision de combinaisons assez peu probables des concepts tels que le “submersible” et de “sable de désert” pour un shot de vidéo, ou estimer une erreur de classification fausse de “fleuve” comme “lac” à être aussi important que une erreur de classification fausse de “fleuve” comme “fighter jet”. Afin d’adresser ces derniers limitations, on propose une approche pour modeler explicitement rapports sémantiques hiérarchiques parmi les classes de cible qui sont automatiquement dérivé et étendu par l’intermédiaire d’un lexique sémantique. Pratiquement, cette méthode est appliquée en tant que l’ensemble sémantique hiérarchique (HSE) de différents classificateurs, chaque réalisé comme machines d’étude binaire de DDA, qui agissent l’un sur l’autre en influençant des décisions de chacun par les liens exigés par la structure des relations parmi concepts sémantiques correspondants, comme récapitulés ci-dessous.

Donné un ensemble de formation de documents annotés de multimédia $\{I_t, K_t\}_{t=1}^n$, où I_t et K_t représente le vecteur de dispositif d’un document donné de multimédia et de son ensemble associé des mots-clés, respectivement, la structure de HSE est déterminés par la hiérarchie de concept $H = \{C_i\}_{i=1}^N$, qui est un graphique dirigé. Les arêtes de H sont définis par de rapports de type “hypernym-hyponym” qui connectent ses noeuds C_i , représentés par toutes les noms uniques de vocabulaire d’annotation $V = \bigcup_{t=1}^n K_t$ et leurs hyponyms dérivé de WordNet. Dans H , chaque concept C_i occupe un noeud séparé, et est associé à un classificateur binaire Θ_i conçu pour distinguer l’ensemble de concepts de feuilles englobés (directement ou indirectement) par C_i , dénoté comme $\mathbf{L}(C_i)$, de tous les autres. Un exemple d’une hiérarchie a dérivé pour un vocabulaire simple $V:\{beach, flower, grass, mountain, rock, sky, tree\}$ est montrée dans Figure 1.1. Afin d’exécuter l’augmentationn sémantique d’un document inconnu, par exemple une image digitale, représentée par un vecteur de dispositif

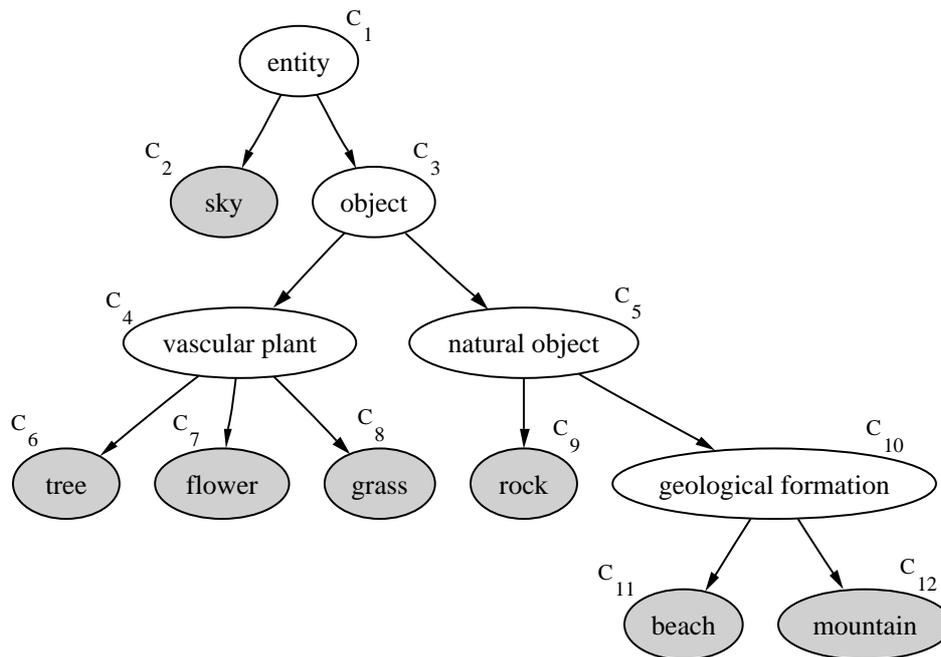


Figure 1.1: Exemple de hiérarchie. Les noeuds ombragés dénotent $C_i \in V$

I_U , chaque concept C_i est évalué en tant que candidat potentiel. Ainsi, le positionnement des annotations possibles n'est plus limité par V , que est le cas pour la majorité d'autres techniques semblables. La pertinence de C_i est vu comme différence entre, d'une part, à quel point la représentation des données I_U correspond à la description de la catégorie C_i de point de vue d'exactitude, et, d'autre part, comment détaillé ou non-ambigu l'ensemble de candidat de mots-clés $\mathbf{L}(C_i)$ est. Dans la méthode proposée de HSE, le premier de ces deux quantites est représentée par la probabilité postérieure d'un concept par rapport de données, $P(C_i|I_U)$, tandis que le second est estimé comme postérieur probabilité d'un concept donné la prétention qu'un mot-clé k de l'ensemble de tous les hyponyms de C_i est choisi correctement, dénoté comme $P(C_i|k)$. Formulé en termes bayésiens, la méthode de HSE apporte une indemnité supplémentaire de laisser pour déterminer exactement le numéro de catégories appropriées C_i à prévoir. Les résultats prometteurs de l'évaluation empirique confirment la viabilité de l'approche proposée, validé par rapport à plusieurs techniques d'ensemble apprenant, aussi bien qu'avec le type différent de classificateurs de base.

1.6 Aspects théoriques

En plus de l'évaluation empirique de toutes les méthodes proposé, un effort exprès est faites pour établir connexions théorétiques et analogies avec le dernier cri en méthodes d'apprentissage par l'ordinateur. Ainsi, des comparaisons sont effectuées pour trouver des parallèles parmi les techniques de DDA, de SVM et d'ACM (machine centrale analytique), commençant par la traduction géométrique des contraintes de séparabilité qui provoque la

notion de l'espace de version, dont là où la solution obtenue par un SVM correspond à un centre de Tchebycheff et celui trouvé par un ACM est un centre analytique. En ce contexte, la formulation de méthode de DDA est montrée pour fournir l'approximation du critère recherché à être réduite au minimum par ACM. Résultats analogues obtenus par DDA et ACM dans le cas idéal quand les sets de données sont parfaitement séparables et presque équidistamment projetés peuvent être attribués au fait que les fonctions objectives des deux méthodes deviennent semblables dans ces conditions. En outre, la méthode de DDA peut être interprétée en tant que parent étroit d'ACM, en raison de l'adoption des fonctions potentielles de logarithme négatif, mais avec une distinction importante: les sets de données séparables et inséparables sont traités uniformément, depuis les distances sont non négatives et ainsi toujours dans le domaine admissible du logarithme, excepté la classe de noyaux indéfinis cités précédemment. La dernière distinction fait le DDA ressembler au SVM, puisque ni l'un ni l'autre des deux n'exige la séparabilité de données, tandis que ACM face un problème avec une région faisable vide quand les échantillons de données ne sont pas séparables. En conclusion, la taille de colonne d'unité de la transformation recherchée nécessaire pour effectuer l'analyse ci-dessus, est atypique pour le DDA. En effet, la meilleure performance empirique de la technique développée est observée quand la dimensionnalité de l'espace de cible est plus grande que un. Dans le contexte ci-dessus de SVM et d'ACM, ceci correspond à avoir plusieurs hyperplans de séparation utilisés dans un classificateur, au lieu d'un. Afin d'étudier plus rigoureusement des implications de la dernière configuration inhérente à DDA, nous considérons un classificateur de séparation optimal d'hyperplan avec une extension explicite au cas de multiples hyperplans (MH), démontrant les possibilités d'extraire d'autres prestations d'une telle formulation, et fournissant la dérivation de garantie de généralisation de performance en termes de la limite associée de dimension fat-shattering.

1.7 Perspectives futures

Plusieurs résultats contenus dans ce travail ont révélé des pistes intéressantes pour de prochaines recherches. Dans l'analyse binaire discriminante, nous avons vu le bénéfice de l'asymétrie et la nature transformative de la formulation qui chacune devraient être explorées plus avant. Les avantages apparents de la technique de majorisation itérative soulignés à travers des expériences pratiques grâce à l'optimisation des critères de la méthode développée pourraient être d'une grande aide dans le développement futur d'approches basées sur l'optimisation. Au moment de l'écriture de ce texte, cette proposition ci-dessus a déjà été confirmée par un certain nombre de publications récentes documentant la popularité croissante de la technique numérique ci-dessus. Dans le domaine des méthodes basées sur les noyaux, nous avons démontré une manière uniforme de traiter les noyaux indéfinis ainsi que les noyaux semi définis positifs traditionnels. Ce premier type de noyaux est lié aux mesures correspondantes de distance non métriques qui peuvent mieux capturer la similarité perceptuelle, d'où l'importance de cette contribution à l'application d'algorithmes d'apprentissage par l'ordinateur dans le

traitement de l'information perceptuelle. Du point de vue des classifications de catégories multiples, nous avons démontré la validité du mécanisme proposé pour expliquer les relations entre différentes catégories-cibles, qui peut être appliqué et facilement étendu à d'autres domaines. De plus, nous avons déjà commencé à explorer plus avant le contexte HSE afin d'améliorer l'approche proposée en prenant avantage de la structure donnée par la classification hiérarchique résultante, afin d'incorporer un feedback de l'utilisateur, étendant ainsi l'approche au domaine des méthodes de l'augmentation sémantiques interactives. Les connexions théoriques établies entre les techniques proposées d'analyse discriminante et le dernier cri en méthodes d'apprentissage par l'ordinateur ont souligné les propriétés uniques de la formulation proposée et déclenché le début d'une étude des configurations hyperplanes multiples dans le contexte des marges larges. L'expérience positive et les résultats encourageants décrits le long de cette thèse nous donnent à penser que des efforts supplémentaires pour faire avancer les technologies d'apprentissage par l'ordinateur pour l'augmentation sémantique sont à la fois viables et justifiables.