

Comment définir un noyau semi-défini positif sur les graphes?

Cornelia Metzsig, Gilles Bisson, Cécile Amblard, Mirta Gordon

► **To cite this version:**

Cornelia Metzsig, Gilles Bisson, Cécile Amblard, Mirta Gordon. Comment définir un noyau semi-défini positif sur les graphes?. Laurent Bougrain. Conférence Francophone sur l'Apprentissage Automatique - CAp 2012, May 2012, Nancy, France. pp.223-237, 2012. <hal-00741459>

HAL Id: hal-00741459

<https://hal.archives-ouvertes.fr/hal-00741459>

Submitted on 12 Oct 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comment définir un noyau semi-défini positif sur les graphes?

Cornelia Metzgi¹, Gilles Bisson¹, Cécile Amblard¹, Mirta B. Gordon¹

Université Joseph Fourier / Grenoble 1 / CNRS
Laboratoire LIG - Bâtiment CE4
Allée de la Palestine
38610 Gières FRANCE

cornelia.metzig@imag.fr, gilles.bisson@imag.fr, cecile.amblard@imag.fr,
mirta.gordon@imag.fr

Résumé : Nous revenons sur le problème de l'application des Machines à Vecteurs Support (SVM) à la classification de graphes. Plusieurs mesures de similarité, appelées généralement noyaux, ont été proposées récemment. Cependant, certaines, comme le noyau d'affectation optimale (15), ne sont pas semidéfinies positives, ce qui limite leur applicabilité avec les SVM. Nous rappelons les conditions pour l'utilisation des SVM. Le théorème de Mercer donne les conditions nécessaires et suffisantes pour des données vectorielles. Dans le cas des graphes, l'espace de Hilbert doit être défini explicitement, car des conditions plus faibles ne sont pas suffisantes pour garantir la positivité de la mesure de similarité. Nous montrons que plusieurs noyaux proposés dans la littérature ont un espace de Hilbert sous-jacent, dont nous exhibons la base correspondante. Nos résultats sont illustrés avec des exemples de la littérature des noyaux de graphes.

1. Introduction

La comparaison de graphes est un problème qui existe dans de nombreux domaines de recherche, comme la bio- ou la chemo-informatique, où de grandes molécules avec plusieurs milliers d'atomes sont représentées sous forme de graphes. Des algorithmes de fouille de données ont été appliqués à des bases de données de tels graphes dans le but de découvrir de nouveaux médicaments. Un autre domaine d'application important est l'analyse de la structure des réseaux sociaux.

Récemment, plusieurs mesures de similarités de graphes ont été proposées et étudiées (14; 18; 20). Interprétées comme des noyaux, elles ont été utilisées

avec des SVM. Grâce au “kernel trick”, des noyaux semi-définis positifs dans l’espace des entrées peuvent être employés dans l’algorithme des SVM car ils sont un produit scalaire dans un certain *espace de caractéristiques* (feature space). Cependant, les graphes ne sont pas des données vectorielles ; il n’est pas évident que l’on puisse utiliser le “kernel trick” pour les comparer afin d’utiliser des SVM.

Dans cet article, nous clarifions quelles sont les conditions nécessaires et suffisantes pour que des mesures de similarité entre graphes représentent des produits scalaires dans un espace de caractéristiques, c’est-à-dire, soient des noyaux semi-définis positifs (ou non-négatifs). Nous montrons pourquoi certains satisfont la non-négativité et d’autres pas. Contrairement au travail de Shin et al. (11), notre raisonnement ne nécessite pas l’introduction de nouveaux formalismes. En effet, tenant compte du cas particulier de données non-vectorielles, une analyse approfondie du théorème de Mercer et de la définition des espaces de Hilbert à noyaux reproduisants (RKHS, pour “Reproducing Kernels Hilbert Spaces”) est suffisante pour déduire une règle simple permettant de vérifier la non-négativité. Toute la difficulté réside dans la généralisation du formalisme de données métriques à des objets structurés.

Le texte est organisé comme suit : un bref rappel des SVM est présenté dans la section 2. Les deux manières d’étendre les SVM à des graphes – soit par l’application du théorème de Mercer, soit par la définition d’un RKHS défini directement sur l’ensemble de graphes – sont décrites dans la section 3. Des exemples de RKHS définis sur des graphes, tirés de la littérature, sont discutés dans la section 4. Les possibles conséquences de nos résultats sont présentés dans la conclusion.

2. Algorithme SVM et théorème de Mercer

L’algorithme d’apprentissage SVM détermine une fonction de décision linéaire $f : X \rightarrow Y$ à partir d’un ensemble de m données d’apprentissage étiquetées $\{\mathbf{x}_i, y_i\}_{i=1}^m$, où $\mathbf{x}_i \in \mathbb{R}^d$, $y_i \in \{\pm 1\}$. La classe attribuée à une donnée de test \mathbf{x} dépend de sa position par rapport à l’hyperplan séparateur,

$$f(\mathbf{x}) = \text{sgn}[\mathbf{w} \cdot \mathbf{x} + b], \quad (1)$$

où \cdot dénote le produit scalaire, \mathbf{w} est un vecteur perpendiculaire à l’hyperplan et b est la distance à l’origine (convenablement normalisée). La fonction f , qui maximise la marge entre les données et l’hyperplan, est la solution d’un problème quadratique d’optimisation sous contraintes. Le vecteur \mathbf{w} est une

combinaison linéaire de vecteurs de l'ensemble d'apprentissage proches de l'hyperplan, qu'on appelle vecteurs support :

$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i. \quad (2)$$

Dans l'équation 2, $\alpha_i > 0$ pour les vecteurs support, $\alpha_i = 0$ pour les autres exemples d'apprentissage. Donc, la fonction de décision s'écrit

$$f(\mathbf{x}) = \text{sgn} \left[\sum_i \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b \right]. \quad (3)$$

Comme seuls des produits scalaires sont nécessaires dans l'équation 3, l'algorithme SVM a été généralisé (procédé connu comme le "kernel trick") à des données vectorielles non séparables linéairement dans l'espace des entrées, grâce au théorème de Mercer.

2.1. Le théorème de Mercer

L'énoncé du théorème établi par J. Mercer en 1909 (4), est le suivant :

Soit C un sous-ensemble compact de \mathbb{R}^n . La fonction noyau $K(\mathbf{x}_1, \mathbf{x}_2)$, continue et symétrique en $L_2(C)$, représente un produit scalaire qui peut s'écrire comme suit :

$$K(\mathbf{x}_1, \mathbf{x}_2) = \sum_{k=1}^{\infty} a_k \Phi_k(\mathbf{x}_1) \Phi_k(\mathbf{x}_2), \quad (4)$$

avec des coefficients positifs $a_k > 0$, si et seulement si K est un noyau semi-défini positif (non-négatif), c'est-à-dire, il satisfait :

$$\int_C \int_C g(\mathbf{x}_1) g(\mathbf{x}_2) K(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \geq 0 \quad (5)$$

pour tout g de carré intégrable, autrement dit, $g \in L_2(C)$.

Les Φ_k de l'équation (4) sont intuitivement la base de l'espace des caractéristiques (1), correspondant à une application implicite de l'espace des entrées C .

L'intérêt de ce théorème réside dans le fait que l'application vers l'espace des caractéristiques n'a pas besoin d'être explicitée. Seul est nécessaire un

noyau non-négatif adéquat, défini dans l'espace des entrées ¹. Pour une introduction plus complète aux noyaux et au "kernel trick", voir (2) et (3).

En pratique, si les données ne sont pas linéairement séparables dans l'espace des entrées, nous pouvons soit définir explicitement une application vers un espace de caractéristiques, soit un noyau. Dans le premier cas il faut démontrer que l'espace des caractéristiques est un espace vectoriel (c'est-à-dire, doté d'un produit scalaire). Dans le deuxième cas, on doit prouver que le noyau est non-négatif.

2.2. Comment démontrer la non-négativité ?

La principale difficulté, quand on invoque le théorème de Mercer, est de démontrer que le noyau est semi-défini positif. Généralement l'intégrale dans l'équation (5) ne peut pas être évaluée explicitement. Dans ce cas la démonstration fait appel aux propriétés de fermeture des fonctions définies positives (voir (7; 8)). Les propriétés suivantes permettent la construction de nouveaux noyaux semi-définis positifs à partir d'autres noyaux ayant cette propriété. Soit \mathcal{X} un ensemble non vide.

- Fermeture par rapport à la somme : La somme de deux noyaux symétriques semi-définis positifs $K_A, K_B : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$:

$$K = K_A + K_B : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R} \quad (6)$$

est un noyau symétrique semi-défini positif.

- Fermeture par rapport au produit : Le produit de deux noyaux symétriques semi-définis positifs $K_A, K_B : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$,

$$K = K_A \cdot K_B : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R} \quad (7)$$

est un noyau symétrique semi-défini positif.

- Fermeture par rapport au produit tensoriel : Soient $K_A : \mathcal{A} \times \mathcal{A} \rightarrow \mathbb{R}$ et $K_B : \mathcal{B} \times \mathcal{B} \rightarrow \mathbb{R}$ des noyaux symétriques semi-définis positifs, $x_A \in \mathcal{A}$, $x_B \in \mathcal{B}$. Alors, leur produit tensoriel $K_A K_B : (\mathcal{A} \times \mathcal{B}) \times (\mathcal{A} \times \mathcal{B}) \rightarrow \mathbb{R}$, où

$$K_A K_B((x_{A1}, x_{B1}), (x_{A2}, x_{B2})) = K_A(x_{A1}, x_{A2}) \otimes K_B(x_{B1}, x_{B2}), \quad (8)$$

1. Plus tard, Dunford et Schwartz ont démontré que le théorème est valable dans un espace métrique compact (5). Dans cet article nous ne considérons pas d'autres généralisations, comme celle de (6), car elles ne sont pas nécessaires pour le problème qui nous occupe. Pour comprendre notre raisonnement, le lecteur peut imaginer que le noyau est défini dans un espace Euclidien au lieu d'un espace métrique compact.

est un noyau symétrique semi-défini positif.

- Fermeture par rapport à la concaténation de fonctions : La concaténation de deux noyaux symétriques semi-définis positifs $K_A : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, $K_B : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$,

$$K = K_B \circ K_A : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}, \quad (9)$$

est un noyau symétrique semi-défini positif.

Ces propriétés de fermeture sont utiles pour démontrer qu'un noyau K est semi-défini positif s'il est construit à partir de noyaux K_A, K_B définis sur \mathbb{R} ayant cette propriété. Ceci est souvent le cas des noyaux définis sur \mathbb{R}^n car les fonctions K_A, K_B peuvent s'exprimer comme des combinaisons linéaires du produit scalaire canonique sur \mathbb{R}^n , qui est défini positif. Par exemple, les propriétés précédentes permettent de démontrer que les noyaux qui admettent un développement en série, comme le noyau gaussien

$$K(\mathbf{x}_1, \mathbf{x}_2) = \exp(-(\|\mathbf{x}_1 - \mathbf{x}_2\|^2/2\sigma)), \quad (10)$$

ou le produit de noyaux, comme le noyau exponentiel,

$$K(\mathbf{x}_1, \mathbf{x}_2) = (\alpha \mathbf{x}_1 \cdot \mathbf{x}_2 + c)^d, \quad (11)$$

sont semi-définis positifs.

Considérons maintenant le problème posé par les algorithmes d'apprentissage. Ayant choisi un noyau, on peut calculer les valeurs qu'il prend pour un ensemble de m points spécifiques, ce qui définit la matrice de Gram. Rappelons que si le noyau est semi-défini positif, la matrice de Gram K_{ij} ainsi engendrée est symétrique et semi-définie positive. Elle satisfait

$$\sum_{1 \leq i \leq m, 1 \leq j \leq m}^m c_i c_j K_{ij} \geq 0 \quad (12)$$

pour tout $c_i \in \mathbb{R}$, ce qui implique que ses valeurs propres sont non négatives. Cependant, que la matrice de Gram soit semi-définie positive ne garantit pas que la fonction noyau utilisée pour l'engendrer possède cette propriété sur toutes les paires de points de l'espace. En d'autres termes, une matrice de Gram semi-définie positive sur la base d'apprentissage prouve que le noyau correspondant est semi-défini positif seulement si le noyau K est continu dans l'espace des entrées et le sous-ensemble de m points d'apprentissage est dense dans l'ensemble de test.

3. Généralisation des SVM à des graphes

Si les données sont définies dans un espace métrique, il y a des moyens de prouver qu'un noyau est semi-défini positif, et on peut donc appliquer les SVM. En effet, les propriétés de fermeture énoncées dans la section 2.2. permettent de démontrer que n'importe quel noyau *continu* est semi-défini positif *dans tout l'espace d'entrée*. Puisque tant l'ensemble d'apprentissage comme les données de test (inconnues) sont dans l'espace métrique (l'espace des entrées), le noyau – *quelques soient les données considérées* – engendre une matrice de Gram symétrique semi-définie positive.

Le défi est de prouver que des noyaux définis sur des graphes sont semi-définis positifs. Est-ce que le théorème de Mercer est utile pour cela ? A priori, s'il n'y a pas de métrique dans l'espace des graphes, la réponse est négative. Sans métrique on ne peut pas définir la continuité des fonctions, et on ne peut donc pas généraliser les résultats obtenus sur l'ensemble d'apprentissage à d'autres données (de test).

Le but de cette section est de définir des métriques dans l'espace des graphes, de façon à pouvoir utiliser les propriétés de fermeture afin de prouver sur l'ensemble d'apprentissage que les noyaux sont semi-définis positifs, et de généraliser cette propriété à n'importe quel autre élément de l'espace grâce à la continuité par rapport à la métrique choisie. Dans les sections 3.1. - 3.3. nous discutons trois différentes approches de la littérature, dont deux s'avèrent erronées.

3.1. Généralisation de Haussler à des données non-vectorielles.

Le noyau de convolution de Haussler a été considéré avec beaucoup d'attention car il constitue une possible généralisation pour définir des noyaux positifs sur des données structurées comme des séquences, des arbres et des graphes. Dans son rapport technique (10) (page 30) Haussler propose une définition de métrique applicable à des ensembles arbitraires, avec le but de définir un noyau semi-défini positif. Si sa proposition était correcte, le présent article n'aurait pas lieu d'être. Son approche est la suivante :

1. définition d'une métrique sur l'ensemble, qui rend l'espace séparable,
2. sélection d'un noyau semi-défini positif dans cet espace,
3. démonstration que le noyau est continu dans l'ensemble par rapport à la métrique

4. conclusion : tous les autres éléments de l'ensemble sont des vecteurs dans l'espace de Hilbert défini par le noyau

Le raisonnement – comme celui qui fait appel aux propriétés de fermeture – utilise le fait que la continuité de la fonction noyau garantit qu'il est semi-défini positif dans tout l'ensemble (son domaine de définition). Hausler définit la métrique $d(x_1, x_2) = \delta(x_1, x_2)$. De son texte l'on déduit que cette notation signifie $d(x_1, x_2) = 0$ si $x_1 \neq x_2$ et $d(x_1, x_2) = 1$ si $x_1 = x_2$. Or, ceci n'est pas une métrique, car elle ne satisfait pas $d(x_1, x_1) = 0$. Cependant, elle permet apparemment de définir une fonction réelle dans cet espace, qui vérifie la définition de continuité suivante :

Une fonction $f : X \rightarrow Y$ ($X, Y \subset \mathbb{R}$) est continue si pour tout $\epsilon > 0$ existe $\rho > 0$ tel que $d(x_1, x_2) < \rho \Rightarrow d(f(x_1), f(x_2)) < \epsilon$.

Trivialement, si la distance entre une paire arbitraire d'éléments distincts x_1 et x_2 est nulle, $d(x_1, x_2) < \rho$ est toujours vrai pour n'importe quel ρ , indépendamment de ϵ . Donc, la définition de continuité est toujours vraie. Hausler utilise cet argument pour prouver la continuité de n'importe quel noyau défini sur un ensemble avec cette métrique. Cependant, dès qu'on utilise une définition de métrique discrète correcte, son raisonnement n'est plus valable. Donc, nous ne nous servons pas de sa démonstration.

3.2. Définition d'une métrique discrète sur l'ensemble des graphes.

Poursuivons l'idée originale de Hausler mais en considérant une métrique bien définie. Il est toujours possible de définir une métrique pour un ensemble dénombrable arbitraire, telle que la distance d entre deux éléments x_1, x_2 est $d(x_1, x_2) = 1$ si $x_1 \neq x_2$ et $d(x_1, x_2) = 0$ si $x_1 = x_2$. Dans cet espace tous les éléments sont équidistants. Pour appliquer le théorème de Mercer il est nécessaire que l'espace soit compact. Le concept de compacité est lié à la notion de voisinage. La compacité garantit que si l'on échantillonne un grand nombre de fois l'ensemble, on trouvera certainement un élément arbitrairement proche d'un élément quelconque. Ceci peut arriver de deux manières : soit parce qu'il y a un voisinage de l'élément choisi qui contient un nombre infini d'autres éléments, ou parce qu'on échantillonne l'élément lui-même – ce qui arrive nécessairement si l'espace est fini. Donc, des ensembles finis avec une métrique discrète sont compacts. Si l'ensemble n'est pas fini, la métrique qu'on vient de définir ne garantit pas la compacité : la probabilité d'échantillonner le même point dans un ensemble infini est nulle. Autrement dit, on ne trouvera pas de point arbitrairement proche du point choisi.

Etant donné que dans le domaine de l'apprentissage automatique tous les ensembles de données sont nécessairement finis, on pourrait soutenir que l'on n'a à traiter que des espaces métriques compacts. Cependant, cet argument est trop simpliste, car le but de l'apprentissage est de prédire la classe d'éléments qui n'appartiennent pas à l'ensemble d'apprentissage. Lorsque l'on définit le noyau, les données de test sont inconnues. C'est pourquoi l'on doit considérer que l'ensemble de données possibles est infini. Le noyau doit être semi-défini positif quelque soit la matrice de Gram définie par les données d'apprentissage et de test.

Cette section et la précédente nous montrent l'échec de tentatives de construire, de façon universellement valable, un espace métrique compact pour un ensemble discret pour l'apprentissage supervisé. Or, il nous reste à explorer la possibilité de définir un espace de Hilbert à noyau reproduisant (RKHS) sur un ensemble arbitraire, mais d'une façon moins générale. Dans la section suivante nous montrons comment on peut définir un tel espace pour les graphes.

3.3. Définition de l'espace RKHS par la transformation Φ

Quand on ne sait pas identifier une métrique adéquate, on peut définir l'espace de Hilbert à noyau reproduisant en exhibant directement la transformation Φ , car un noyau permettant de construire un RKHS peut être défini dans tout ensemble $(\mathcal{X}; \mathcal{F})$ – par exemple, dans l'ensemble des graphes. On peut le définir comme suit :

Soient \mathcal{X} un ensemble non vide, x ses éléments, et \mathcal{F} l'ensemble de fonctions $f : \mathcal{X} \rightarrow \mathbb{R}$. \mathcal{F} est un RKHS avec un produit scalaire $\langle \cdot, \cdot \rangle$ s'il existe une fonction $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ telle que K possède la propriété de reproduction suivante :

$$\langle f, K(x, \cdot) \rangle = f(x) \quad \forall f \in \mathcal{F}, \quad \forall K(x, \cdot) \in \mathcal{F} \quad (13)$$

Ceci implique, en particulier

$$\langle K(x_1, \cdot), K(x_2, \cdot) \rangle = K(x_1, x_2). \quad (14)$$

Autrement dit, pour chaque noyau semi-défini positif sur un ensemble, il existe un RKHS associé dans lequel la fonction noyau est un produit scalaire. Donc, le noyau peut être utilisé avec des SVM.

Comment peut-on démontrer qu'un noyau est semi-défini positif sur un ensemble arbitraire? On ne peut pas définir un noyau en termes d'une fonction continue s'il est défini seulement pour un ensemble discret de points. Les propriétés de fermeture dans ce cas ne sont pas utiles. La seule possibilité est de définir l'application $\Phi : \mathcal{X} \rightarrow \mathbb{R}^n$ de l'ensemble vers l'espace de caractéristiques, ce qui équivaut à exhiber la base du RHKS, et pas simplement le noyau qui la définit. Les descripteurs ainsi extraits peuvent alors être considérés comme les vecteurs de base d'un espace où chaque élément de l'ensemble à classer peut être vu comme un vecteur (voir des exemples dans (14)). La difficulté principale de cette approche est de trouver les descripteurs adéquats, car toute autre information sur les éléments de l'ensemble est perdue. Cette preuve de non-négativité s'applique à plusieurs noyaux définis dans la littérature. Dans certains cas le produit scalaire lui-même est utilisé comme mesure de similarité dans l'algorithme des SVM. Dans d'autres cas un noyau est défini dans l'espace des caractéristiques, la preuve qu'il est semi-défini positif est alors basée sur les propriétés de fermeture.

Le fait essentiel qu'un noyau défini sur une base de l'espace de caractéristiques est semi-défini positif a été formalisé autrement par Shin et Kuboyama (11) : leur condition nécessaire et suffisante pour que le noyau soit semi-défini positif est que ce qu'ils appellent le *système d'application* ("*mapping system*") soit symétrique et transitif, ce qui implique que les mêmes critères de comparaison doivent être utilisés pour tous les échantillons. En termes des SVM, ces critères sont les descripteurs qui engendrent les dimensions de l'espace de caractéristiques².

4. Exemples de noyaux de graphes

Nous démontrons ici que certaines classes de noyaux de graphes sont semi-définis positifs : nous montrons qu'ils sont des produits scalaires dans l'espace des caractéristiques en exhibant explicitement la base correspondante. Il s'avère que ceci est possible dans la plupart des cas. Cependant, quelques noyaux ont été introduits sans preuve qu'ils sont semi-définis positifs ou avec une preuve erronée (15). Nous considérons qu'un graphe est un ensemble de noeuds (ou sommets) reliés par des arêtes. Dans un graphe étiqueté, sommets et arêtes ont des descripteurs. Pour simplifier nous ne considérons que des exemples où seulement les sommets sont étiquetés.

2. Shin et Kuboyama ont étudié le cas particulier de noyaux d'arbres dans (12).

Très généralement, si le noyau utilise comme critère de comparaison le nombre de sous-structures communes aux paires de graphes, alors celles-ci constituent la base de l'espace des caractéristiques. Chaque graphe possède une représentation dans cet espace vectoriel³.

Les noyaux suivants ont été proposés pour des applications en chimio-informatique, où chaque graphe correspond à une molécule, les sommets étant des atomes ou des groupes fonctionnels, les étiquettes correspondantes étant les noms des atomes ou des groupes⁴.

4.1. Noyaux où les caractéristiques sont des chemins

Les noyaux considérés dans ce paragraphe comptent le nombre de chemins communs aux paires de graphes, autrement dit, le nombre de séquences identiques (de longueur donnée) d'étiquettes des sommets. Considérons la figure 1, et uniquement les chemins de longueur 3.

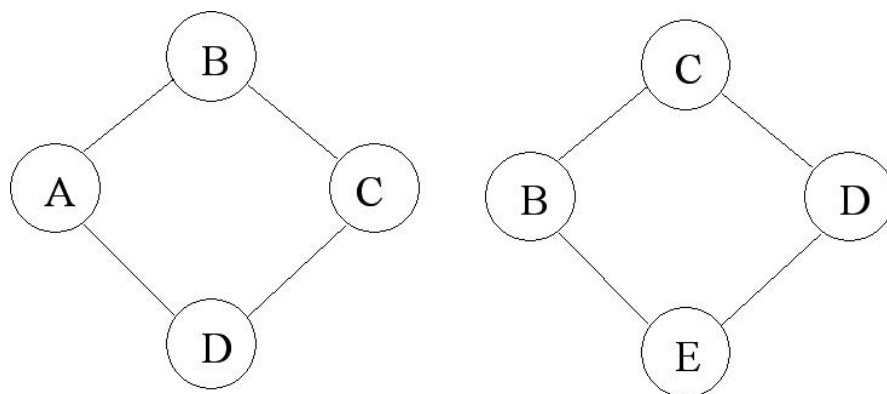


FIGURE 1: Deux graphes étiquetés : à gauche, exemple (1) ; à droite exemple (2)

Ces graphes contiennent les chemins de longueur 3 suivants : ABC, ADC,

3. Plus précisément, les graphes sont les éléments d'un semi-anneau avec une métrique Euclidienne, puisque leurs composantes ont des valeurs entières non-négatives. Cependant, l'hyperplan peut avoir des composantes non entières et de ce fait, il appartient à un semi-corps. L'espace des caractéristiques est donc un semi-espace, restreint à des vecteurs de composantes positives ou nulles. Un lecteur peu familier avec ces concepts peut penser qu'on est dans un espace vectoriel pour comprendre le raisonnement.

4. Nous ne nous focalisons pas sur les différences de mise en oeuvre, en particulier au fait qu'un formalisme de groupe de produits est utilisé – comme en (13) – ou non.

BAD, BCD, BED, CDE, CBE. Si on considère que chacun de ces 7 chemins engendre une dimension de l'espace des caractéristiques, les graphes peuvent être représentés par des vecteurs. L'exemple (1) de la figure 1 est alors le vecteur $(1, 1, 1, 1, 0, 0, 0)$, l'exemple (2) de la même figure est $(0, 0, 0, 1, 1, 1, 1)$.

Des graphes plus grands peuvent présenter les mêmes séquences à plusieurs endroits, leurs vecteurs représentatifs ayant alors des composantes > 1 . Le produit scalaire canonique dans cet espace est clairement semi-défini positif et peut donc être utilisé par les SVM.

Un noyau de graphes semi-défini positif fréquemment utilisé dans la littérature (par exemple (16)) est le noyau de Tanimoto, introduit dans (14). La base de l'espace de caractéristiques est exactement celle définie plus haut. Si $K(.,.)$ représente le produit scalaire dans cet espace, le noyau de Tanimoto K^t est défini par :

$$K^t(x_1, x_2) = \frac{K(x_1, x_2)}{K(x_1, x_1) + K(x_2, x_2) - K(x_1, x_2)}. \quad (15)$$

Le noyau de Tanimoto est le même que le coefficient de similarité de Jaccard ; cependant la distance calculée avec le noyau de Tanimoto dans l'espace de caractéristiques n'est pas la distance de Jaccard (17).

4.2. Noyaux avec caractéristiques correspondant à des chemins de longueur variable

Un autre noyau connu est le Noyau des Chemins Aléatoires (Random Walk Kernel (voir (13))), qui compte le nombre de chemins communs (chemins redondants, avec des visites répétées aux mêmes sommets) aux deux graphes. Les chemins possibles (de longueur variable) sont les vecteurs de base de cet espace de caractéristiques. Revenant à la figure 1, ces chemins sont A, B, C, D, E, AB, AD, BC, BE, CD, DE, ABA, ABC, ADA, ADC, BAB, BAD, BCB, BEB, BED, CBC, Dans cette base, l'exemple (1) est représenté par le vecteur $(1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, \dots)$; l'exemple (2) étant $(0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, \dots)$. Afin d'avoir une représentation unique, il faut faire certains choix. Par exemple, il faut décider si l'on admet des aller-retour (auquel cas les dimensions AB et ABA sont nécessairement identiques) ou si seulement les chemins non redondants sont admissibles (14). Dans cet espace le noyau est semblable à un produit scalaire standard, mais les composantes correspondant à des chemins de longueur k sont multipliées par un facteur d'atténuation $\lambda^k > 0$ ($0 < \lambda < 1$).

Il est important de bien choisir ce facteur pour assurer la convergence du produit scalaire⁵.

Une autre classe de noyaux représente les graphes comme des vecteurs, en extrayant les sous-graphes communs (10; 19). Les vecteurs de base de l'espace de caractéristiques correspondant sont tous les sous-graphes possibles.

4.3. Le noyau d'affectation optimale

Le "Optimum Assignment Kernel", que nous traduisons comme noyau d'affectation optimale, a été introduit dans (15) comme suit :

Soit \mathcal{X} l'ensemble de graphes x , \mathcal{Y} l'ensemble de sommets y . Soit $k : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ un noyau symétrique semi-défini positif, et π l'opérateur de permutation. Alors $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ avec

$$K(x_1, x_2) = \max_{\pi} \sum_{i=1}^{|x_2|} k(y_{1i}, y_{2\pi(i)}) \quad (16)$$

s'appelle "noyau d'affectation optimale". $|x|$ dénote la cardinalité (le nombre de sommets) du graphe x , et nous supposons que x_2 , constitué des sommets $\{y_{2i}\}$, est le graphe de cardinalité inférieure : $|x_2| \leq |x_1|$.

L'idée est de trouver le meilleur appariement du plus petit des deux graphes avec le plus grand, et ensuite calculer la somme des noyaux k entre les étiquettes des sommets appariés.

Ici il n'est pas possible d'identifier une base, comme dans les cas précédents. Les caractéristiques utilisées pour comparer deux graphes ne sont pas les mêmes pour toutes les paires de graphes, mais dépendent de quels graphes on considère. Il a été montré que le noyau d'affectation optimale n'est pas toujours semi-défini positif. Vert (21) a prouvé ceci en exhibant un contreexemple où la matrice de Gram a une valeur propre négative.

Maintenant, basés sur la théorie des RKHS, il est clair que le problème est dû au fait que l'espace de caractéristiques n'est pas bien défini. La démonstration que le noyau est semi-défini positif proposée dans (15) utilise la propriété de fermeture donnée par l'équation (6), qui affirme que la somme $K = K_1 + K_2$ de deux noyaux semi-définis positifs est encore une fonction semi-définie positive. Or, ceci n'est vrai que si K , K_1 and K_2 appartiennent au même espace. Ceci signifie que

5. Cependant, dans la pratique ceci est un problème mineur car les algorithmes ne peuvent considérer que des chemins de longueur bornée.

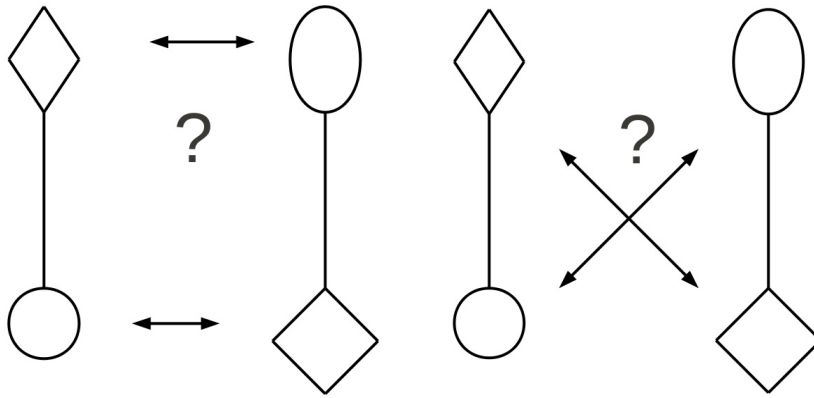


FIGURE 2: Deux possibles appariements : lequel doit-on choisir ?

- soit les trois noyaux sont des noyaux entre sommets, ou
- les trois noyaux sont des noyaux entre graphes.

Dans ce dernier cas, le noyau k entre sommets pourrait être interprété comme étant un noyau entre graphes, dont les caractéristiques sont le nombre de fois qu'une étiquette de sommet donnée existe dans les deux graphes. Mais alors, chaque étiquette de sommet engendrerait une dimension de l'espace de caractéristiques, l'information sur les arêtes serait perdue, et on ne pourrait pas faire d'appariement.

Dans le noyau d'affectation optimale il y a autant de descripteurs dans le noyau que de sommets dans le plus petit des deux graphes x_1 et x_2 , soit n . On pourrait penser que l'espace des caractéristiques a n dimensions, dans lequel chaque sommet – et non pas chaque étiquette de sommet – engendre une dimension différente. Mais alors il est clair qu'il existe un espace différent pour chaque paire de graphes ; un espace de caractéristiques unique, commun à tous les graphes, ne pourrait exister que dû au hasard.

5. Conclusion

Dans ce travail nous avons étudié différentes façons de définir un noyau semi-défini positif sur l'ensemble de graphes, afin de l'utiliser dans des SVM. Après un rappel des SVM et du théorème de Mercer, nous avons essayé de généraliser leur application aux graphes en définissant une métrique dans le même esprit que (10). Notre conclusion est que ceci ne peut pas se faire

d'une manière générale, mais seulement au prix de perdre certaines informations des graphes. Nous montrons qu'il est néanmoins possible de définir, dans l'ensemble de graphes, un espace de Hilbert à noyaux reproduisant en définissant une base à partir des caractéristiques extraites. Avec cette méthode nous avons montré que plusieurs noyaux de graphes proposés dans la littérature récente sont effectivement semi-définis positifs. Dans le cas conflictuel du noyau d'affectation optimale, cette approche n'est pas valable parce qu'on ne peut pas identifier d'espace de caractéristiques et encore moins une base qui puisse lui être associée. On peut reformuler notre conclusion termes de logique : on ne peut démontrer qu'un noyau est semi-défini positif que si l'ensemble de données peut être exprimé en logique propositionnelle. Si l'on travaille directement en logique des prédicats, comme c'est le cas du noyau d'affectation optimale, on ne peut pas.

Références

- [1] Vapnik, V.N. : The Nature of Statistical Learning Theory, Springer (1995)
- [2] Vapnik, V.N. : Statistical Learning Theory, John Wiley and Sons (1998)
- [3] Schölkopf, B., Smola, A.J. : Learning with Kernels, The MIT Press (2002)
- [4] Mercer, J. : Functions of Positive and Negative Type, and their Connection with the theory of integral equations, Philosophical Transactions of the Royal Society of London, Series A 209, p. 415-446 (1909)
- [5] Dunford, N., Schwartz, J.T. : Linear Operators, Part II : Spectral Theory, John Wiley and Sons (1963)
- [6] Ferreira, J.C., Menegatto, V.A. : Eigenvalues of Integral Operators Defined by Smooth Positive Definite Kernels, Integral Equations and Operator Theory 64 , p. 61 - 81 (2009)
- [7] Berg, C., Christensen, J.P.R., Ressel, P. : Harmonic Analysis on Semigroups, Springer (1984)
- [8] Cortes, C., Haffner, P., Mohri, M. : Rational Kernels : Theory and Algorithms, Journal of Machine Learning Research 1, p. 1-50 (2004)
- [9] Aronszajn, N. : Theory of reproducing kernels, Transactions of the American Mathematical Society, Vol. 68, No. 3 p. 337-404 (1950)
- [10] Haussler, D. : Convolution Kernels on Discrete Structures, Technical Report UCSC-CRL-99-10 (1999)
- [11] Shin, K., Kuboyama, T. : A Generalization of Haussler's Convolution Kernel – Mapping Kernel, Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland (2008)

- [12] Shin, K., Kuboyama, T. : A Generalization of Haussler's Convolution Kernel – Mapping Kernel and its application to Tree Kernels, *Journal of Computer Science and Technology* Nr. 25, p. 1040 - 1054 (2010)
- [13] Kashima, H., Inokuchi, A. : Kernels for Graph Classification, *IEEE International Conference on Data Mining (ICDM 2002) Workshop on Active Mining*, Maebashi, Japan (2002)
- [14] Ralaivola, L., Swamidass, S.J., Saigo, H., Baldi, B. : Graph kernels for Chemical Informatics, *Neural networks, Special issue on neural networks and kernel methods for structured domains*, 18(8), p. 1093-1110 (2005)
- [15] Froehlich, H., Wegner, J.K., Sieker, F., Zell, A. : optimum assignment kernels for Attributed Molecular Graphs, *Proceedings of the 22nd International Conference on Machine Learning*, New York, NY, USA, p. 225-232. (2005)
- [16] Jacob, L., Vert, J.-P. : Protein–ligand interaction prediction : an improved chemogenomics approach, *Bioinformatics*, Vol. 24 no. 19, p. 2149 - 2156 (2008)
- [17] Lipkus, A. H. :A proof of the triangle inequality for the Tanimoto distance, *Journal of Mathematical Chemistry* 26, p.263-265 (1999)
- [18] Gärtner, T. : A Survey of Kernels for Structured Data, *SIGKDD Explorations* pp.49 - 58 (2003)
- [19] Mohr, J., Jain, B., Sutter, A., Laak, A.T., Steger-Hartmann, T., Heinrich, N., Obermayer, K. :A Maximum Common Subgraph Kernel Method for Predicting the Chromosome Aberration Test, *Journal of Chemical Information Modelling* no. 50, p. 1821 - 1838 (2010)
- [20] Vishwanathan, S.V.N., Borgwardt, K. M., Kondor, I.R., Schraudolph, N.N. :Graph Kernels, *Journal of Machine Learning Research* 9, p.1-41 (2010)
- [21] Vert, J.-P. : The Optimal Assignment Kernel is not positive definite, *Technical Report HAL-00218278*, (2008)