Noyaux multiples : sélection de modèle appliquée à la détection de piéton

Frédéric Suard,Alain Rakotomamonjy

LITIS EA 4051, INSA/Université de Rouen,
avenue de l'université,
76800 Saint Etienne du Rouvray, FRANCE
frederic.suard@insa-rouen.fr

Résumé − Nous présentons une méthode de sélection de modèle appliquées à la détection de piétons par système de vision. Afin d'obtenir des performances optimales différents paramètres interviennent au niveau de la description d'image ou de la classification. Nous proposons d'utiliser l'approche des noyaux multiples afin de sélectionner automatiquement les meilleurs noyaux parmi un ensemble donné, cet ensemble correspondant ainsi aux différents paramètres possibles que nous souhaitons tester. Nous présentons briévement la théorie des noyaux multiples, puis nous appliquons cette méthode sur un problème de détection de piétons via la sélection des paramètres et une sélection de variable. Ces expériences nous permettent ainsi de démontrer l'intérêt de cette approche.

Abstract − This paper presents a pedestrian detection method based on the multiple kernel framework. One main problematic of pattern recognition resides in the pertinent characterization of the data. Depending on the descriptor, we sometimes have to tune the descriptor in order to be more efficient. Instead of accomplishing this tuning manually by testing and comparing all possible values we propose here to use the multiple kernel framework. The aim is to use a kernel as a linear combination of different kernels in order to combine and select automatically the best kernels within a set of kernels. This can be assimilated as model selection, where one kernel of the set corresponds to one model. We first introduce the MKL framework and finally apply this approach for a parameter tuning task and a feature selection problem.

1 Introduction

La détection de piétons a fait l'objet, ces dernières années, de différentes travaux de recherche ([3, 4, 1]). C'est un problème classique de la reconnaissance de forme avec deux étapes principales. Il convient tout d'abord de choisir une description pertinente des signaux, dans notre cas des images, afin d'extraire les données les plus représentatives de la présence d'un piéton. Dans ce domaine, les représentations à base d'histogrammes d'orientation de gradient ont démontré leur efficacité à différentes reprises ([4, 1]). Il s'agit en effet d'une description qui s'appuie sur une information de type contour grâce au calcul de différents histogrammes d'orientation de gradient, calculés sur des régions locales, c'est à dire une partie de l'image. Le descripteur proposé par Dalal et al. s'est révélé particulièrement efficace pour traiter le problème de la caractérisation d'images, mais possède un inconvénient majeur. Il nécessite en effet de régler différents paramètres interventant lors du calcul du descripteurs. Ce réglage peut alors être très couteux en temps de calcul afin d'obtenir l'ensemble de paramètres optimaux.

La seconde étape consiste à discriminer les descripteurs, c'est à dire à comparer les descripteurs et déterminer leur classe d'appartenance. Dans ce domaine, les machines à noyaux, tel que le classifieur *Support Vector Machines* [6] ont montré leur efficacité. Ici encore, des paramètres doivent être réglés afin d'obtenir des performances optimales.

Récemment, Lanckriet [2] a proposé de définir le noyau comme une combinaison linéaire de différents noyaux :

$$\mathbf{k}(\mathbf{x}, \mathbf{x}') = \sum_{k=1}^{K} \beta_k \mathbf{k}_k(\mathbf{x}, \mathbf{x}')$$

Chaque noyau peut être calculé sur différents sous-ensemble de la base d'apprentissage, par différentes méthodes de description des données ou selon différentes formulations. Le but de cette approche consiste à ne retenir parmi cet ensemble de noyau que les noyaux les plus pertinents. La sélection de modèle revient ainsi à choisir de manière optimale et automatiquement les valeurs des coefficients β.

Nous proposons ici d'appliquer cette formulation à la sélection de modèle pour la détection de piétons. Dans un premier temps, nous rappellerons le fonctionnement des noyaux multiples, puis nous illustrerons cette approche pour le choix de paramètres optimaux d'un descripteur et la sélection de variables.

2 Noyau multiple

Nous utilisons ici le classifieur binaire *Support Vector Machines* dont la fonction de décision est calculée à l'aide d'un ensemble d'apprentissage $\{\mathbf{x}_i, y_i\} \in \mathcal{X} \times \{-1, 1\}$, où $\mathbf{x}_i$ sont les données d'apprentissage et y_i les étiquettes associées aux données, pour $i = 1 : N$.

La fonction de décision est définie par l'équation :

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i \mathbf{k}(\mathbf{x}_i, \mathbf{x}) + b\right) \quad (1)$$

avec $\mathbf{k}(.,.)$ une fonction noyau, α et b des variables déterminées à partir de l'apprentissage.

Récemment, Lanckriet et al. [2], ont introduit l'utilisation de multiples noyaux au lieu d'un seul afin de combiner différentes sources d'informations. Le noyau est ainsi défini comme une combinaison linéaire convexe de K noyaux :

$$\mathbf{k}(\mathbf{x}, \mathbf{x}') = \sum_{k=1}^{K} \beta_k \mathbf{k}_k(\mathbf{x}, \mathbf{x}') \quad (2)$$

avec $\beta_k \geq 0$ et $\sum_k \beta_k = 1$. $\mathbf{k}_k(.,.)$ est un noyau déterminé sur tout ou partie de l'ensemble des données $\mathbf{x}$ selon différentes formulations de noyau.

Cette approche peut ainsi être assimilée à de la sélection de modèle. Les noyaux utilisés dans la combinaison linéaire appartiennent initialement à un ensemble. L'idée revient finalement à sélectionner et combiner les noyaux les plus pertinents parmi cet ensemble grâce à la pondération des coefficients β_k. Si un noyau s'avère peu pertinent, le coefficient associé aura alors une valeur nulle. Inversement, un noyau pertinent, verra la valeur de son coefficient β_k augmenter.

Ainsi, en constituant l'ensemble initial avec différents noyaux calculés sur les mêmes données, mais selon des formulations différentes ou des descripteurs différents, le choix des β_k remplacera le choix des représentations ou des formulations. Différentes applications sont possibles depuis le choix des paramètres des descripteurs, les paramètres des noyaux, les hyperparamètres du classifieur ou bien encore sélectionner les variables les plus pertinentes en calculant un noyau pour chaque caractéristique.

L'idée revient ainsi à reporter le choix des descripteurs et des paramètres sur le choix des coefficients β.

Les valeurs des coefficients α, b et β sont obtenues en résolvant le dual du problème d'optimisation suivant :

$$\begin{cases} \min\limits_{\mathbf{w},\beta,b,\xi} & \frac{1}{2}\left(\sum\limits_{k=1}^{K} \beta_k \|\mathbf{w}_k\|_2\right) + C\sum\limits_{i=1}^{N} \xi_i \\ \text{s.c.} & y_i f(\mathbf{x}_i) \geq 1 - \xi_i \quad \forall i = 1, \ldots, N \\ \text{et} & \sum\limits_{k=1}^{K} \beta_k = 1 \end{cases} \quad (3)$$

La solution optimale est obtenue en résolvant un problème de programmation linéaire semi-infinie, en suivant la formulation de Sonnenburg [5] :

$$\begin{cases} \max\limits_{\theta,\beta} & \theta \\ \text{s.c.} & \sum_{k=1}^{K} \beta_k = 1 \\ \text{et} & \sum_{k=1}^{K} \beta_k S_k(\boldsymbol{\alpha}) \geq \theta \end{cases} \quad (4)$$

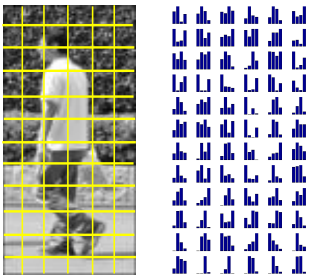
avec $S_k(\boldsymbol{\alpha}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j \mathbf{k}_k(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^{N} \alpha_i$.

Sonnenburg utilise l'algorithme appelé *Column Generation Technique* qui consiste à chercher les valeurs optimales de β et θ pour un sous-ensemble de contraintes puis de déterminer si $\boldsymbol{\alpha}$ satisfait la contrainte $\sum_{k=1}^{K} \beta_k S_k(\boldsymbol{\alpha}) \geq \theta$. Dans ce cas, la solution est optimale, sinon des contraintes sont ajoutées à l'ensemble et ce processus est itéré jusqu'à obtenir la convergence des valeurs des β.

L'intérêt de cette approche réside dans l'utilisation d'un classifieur SVM standard pour chaque itération afin de résoudre $S_k(\boldsymbol{\alpha})$. Seul le noyau est mis à jour puisque des contraintes ou des valeurs de β peuvent changer. Nous avons utilisé une version optimisée de cet algorithme, dont l'implémentation Matlab est disponible sur demande.

3 Descripteur HoG

Les descripteurs d'images utilisant des histogrammes d'orientation de gradient ont prouvé leur efficacité dans les travaux de Shashua et Dalal [4, 1]. L'idée consiste à découper une image en un ensemble de régions et de calculer pour chacune d'elle un histogramme d'orientation de gradient.



a b

FIG. 1 – Découpage d'une image de piéton en cellules (a) et calcul d'histogrammes d'orientation de gradient pour chaque cellule (b).

La description d'une image est effectuée en quatre temps :

1. Calcul de la norme et de l'orientation du gradient de l'image,
2. découpage de l'image en cellules (figure 1-a),
3. calcul des histogrammes d'orientation de gradient pour chaque cellule (figure 1-b),
4. normalisation des histogrammes au sein d'un bloc de cellules.

Le descripteur final est obtenu en ajoutant dans un même vecteur tous les histogrammes normalisés. Comme nous le constatons, plusieurs paramètres interviennent afin de régler la taille des cellules, la taille des blocs, les caractéristiques des histogrammes ou encore le facteur de normalisation. Afin d'obtenir des performances optimale, nous sommes confrontés à expérimenter différentes configurations possibles pour des ensembles de paramètres. L'intérêt des noyaux multiples se révèle ainsi en automatisant le choix du meilleur ensemble de paramètres.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
AUC	0.95	0.95	0.96	**0.97**	0.93	0.94	**0.97**	**0.97**	0.89	0.92	0.92	0.95	0.87	0.91	0.91	0.94
%	0.89	0.89	0.90	**0.91**	0.86	0.87	**0.91**	**0.92**	0.81	0.84	0.86	0.88	0.80	0.84	0.84	0.87
β	0.00	0.00	0.00	**0.01**	0.00	0.00	**0.24**	**0.75**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
variance(β)	0.00	0.00	0.00	**0.03**	0.00	0.00	**0.07**	**0.08**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TAB. 1 – Valeurs des coefficients β pour chaque ensemble de paramètre et valeur de l'AUC pour chaque noyau.

4 Résultats

Nous allons maintenant présenter quelques applications des noyaux multiples pour la sélection de modèles. Tout d'abord nous allons appliquer cette approche pour la sélection de paramètres, puis pour la sélection de variables.

Nous avons utilisé une base de 310 images urbaines acquises selon différentes conditions climatiques et lumineuses : de jour ou de nuit, avec un temps ensoleillé ou pluvieux. Nous avons ainsi extrait manuellement 1240 piétons et environ 6000 non-piétons comme l'illustre la figure 2. Les piétons sont positionnés au centre des images extraites et peuvent être sujets à des occultations ou être présentés selon différentes apparences et postures. Toutes les images ont été redimensionnées à la même taille de 128×64 pixels.

4.1 Choix des paramètres

Comme nous l'avons souligné lors de la présentation du descripteur, nous avons besoin de régler quelques paramètres de manière optimale : la taille des cellules, la taille des blocs, le facteur de normalisation des histogrammes, le nombre de niveaux des histogrammes.

Nous avons ainsi testé différentes valeurs :
- taille des cellules : 16 ou 32 pixels,
- taille des blocs : 1 ou 2 blocs,
- nombre de niveaux dans les histogrammes : 4 ou 8,
- facteur de normalisation : L_1 ou L_2.

Nous calculons un noyau pour chaque ensemble de paramètres, soit 16 noyaux au total. La formulation utilisée pour calculer chaque noyau est un produit scalaire qui ne nécessite par de paramètrage et permet donc une comparaison objective. Tous les noyaux sont ainsi ajoutés dans un même ensemble qui est ainsi utilisé par l'algorithme de noyaux multiples qui va ainsi déterminer la valeur optimale des coefficients β (voir équation 2). Nous retenons ainsi les noyaux dont la pondération associée est la meilleure, c'est à dire les coefficients β_k non nuls.

Le processus expérimental a été effectué de la manière suivante. Les bases d'apprentissage et de test contiennent chacune 1000 piétons et 1000 non-piétons. Elles sont choisies aléatoirement et sont utilisées pour un apprentissage et une validation. Afin de conforter les résultats obtenus, nous itérons plusieurs fois ce processus, en renouvelant systématiquement les bases d'apprentissage et de test. Les résultats donnés sont la moyenne obtenue sur 10 itérations. Pour comparer les performances, nous déterminons, d'une part, le taux de bonne classification, et calculons d'autre part la valeur de l'aire sous la courbe ROC. Cette courbe est obtenue en faisant varier le seuil pour la discrimination et en reportant ainsi le taux de vrais positifs en fonction du taux de faux positifs. L'hyperparamètre C permettant de régler la pondération des points mal classés durant l'apprentissge a été fixé à 1. Nous affichons les résultat obtenus sur le tableau 1.

Afin de vérifier si les valeurs obtenues par l'algorithme des noyaux multiples sont proches de la réalite, nous avons déterminé les performances de chaque noyau indépendamment. Les performances de chaque noyau correspondant à un ensemble de paramètres donné sont affichés sur la première ligne et la valeur correspondante au noyau des coefficients β sont affichés sur la deuxième ligne.

Nous constatons ainsi que les valeurs des β les plus élevées, c'est à dire les noyaux issus des représentations les plus discriminantes correspondent aux noyaux ayant obtenus les meilleurs performances individuelles. Si nous avions effectué le choix des paramètres en comparant tous les noyaux un à un, nous aurions abouti au même choix qu'en ajoutant tous les noyaux dans le même ensemble et en appliquant l'algorithme des noyaux multiples. Cependant, la sélection est effectuée automatiquement et se révèle plus rapide que la comparaison exhaustive de chaque noyau car l'algorithme nécessite moins de résolutions du problème SVM dual.

Les deux configurations principalement retenues sont les n°7 et 8, c'est à dire lorsque la taille des cellules est de 16 pixels, 4 cellules par bloc et 8 niveaux par histogrammes. La configure n°8 présente une normalisation L_2 et la n°7 une normalisation L_1. La configuration n°4 est identique à la n°8 mais avec 1 cellule par bloc. De plus, nous obtenons avec l'approche des noyaux multiples une AUC moyenne de 0.9735 un taux de bonne reconnaissance de 0.92%, ce qui nous permet d'améliorer légèrement les performances. Ce léger gain peut s'expliquer par le fait que tous les noyaux sont issus d'une représentation HoG selon différents paramètres. Nous avons donc une redondance entre les noyaux.

Cette application nous a donc permis d'appliquer les noyaux multiples pour la sélection de modèles. Nous avons ainsi effectué le même test afin de paramétrer les formulations appliquées pour calculer chaque noyau avec la configuration optimale du HoG. Nous avons retenu le noyau gaussien avec une largeur de bande égale à 5 comme le plus performant.

4.2 Sélection de variables

Un autre problème classique de la représentation de données concerne la sélection de variables. Cette application permet non seulement de résoudre le problème de la dimensionnalité des données, en réduisant le nombre de variables nécessaires pour décrire une image, mais permet également de ne conserver que les données pertinentes afin d'améliorer les performances.

Nous proposons donc d'appliquer les noyaux multiples pour la sélection de variable, c'est à dire de combiner et sélectionner les régions les plus discriminantes du découpage. Nous utilisons pour cela le descripteur HoG avec l'ensemble de paramètres suivant : des cellules de 16 pixels, des blocs de 4 cellules avec un recouvrement d'une cellule et 8 niveaux par histogramme. Cet ensemble est en effet le plus performance d'après l'expérience réalisée précédemment.

Nous opérons la sélection de variable de la manière suivante : comme le montre la figure (1), l'image est découpée en plusieurs régions. Cependant, certaines régions sont plus pertinentes lorsqu'elles sont positionnées à des emplacements significatifs du piétons tels que la tête, les jambes ou les bras. Les autres régions qui ne permettent pas de discriminer une image contenant un piéton seront ainsi écartées.

D'après la configuration du descripteur utilisée, nous obtenons 84 histogrammes. Nous calculons donc un noyau pour chaque histogramme , c'est à dire un noyau pour chaque région, avec une formulation gaussienne de largeur de bande égale à 5. L'ensemble de noyaux est donc constitué de 84 noyaux. La sélection de variables consistera donc à ne retenir que les noyaux de l'ensemble pour lesquels le coefficient de pondération β est non nul. De plus, de par la pondération, nous auront ainsi la possibilité d'accorder davantage d'importance à des régions plus significatives.

La base d'apprentissage contient 500 données, la base de test en contient 1000. Par rapport au premier test, la taille de la base d'apprentissage est plus faible, car nous devons prendre en compte davantage de noyaux et sommes par conséquent limités par la mémoire nécessaire. Les bases sont renouvelées aléatoirement 10 fois et nous calculons la moyenne de ces 10 itérations pour comparer les résultats.

Descripteur complet		Sélection de variables		
AUC	%	AUC	%	$\#(\beta > 0)$
0.9332	0.8924	0.9793	0.9282	63.70 $\pm$6.74

Tab. 2 – Comparaison des performances sans et avec une sélection de variable.

Le tableau 2 affiche les résultats obtenus lorsque nous appliquons une sélection de variable à l'aide des noyaux multiples et les résultats obtenus sur les mêmes données sans effectuer de sélection de variable. Nous constatons que cette approche permet non seulement de réduire la taille du descripteur en conservant 63 noyaux au lieu de 84, tout en améliorant les performances. La sélection semble être peu efficace, puisque seulement 20% des histogrammes sont écartés du descripteur. Cela est du au fait que les piétons de la base d'apprentissage ne sont pas localisés exactement au même endroit dans l'image. En outre, les piétons apparaissent selon différentes postures et apparences. Pour approfondir la sélection de variable, il faudrait décomposer la base d'apprentissage en sous-ensembles.

Fig. 2 – Exemples d'images de piétons et de non-piétons.

5 Conclusion

Nous avons présenté une approche par noyaux multiples pour la sélection de modèles dans le cas de la détection de piétons. A partir d'un ensemble de noyaux obtenus en appliquant différents descripteurs ou différentes formulations sur les données, le but consiste à sélectionner et combiner automatiquement les noyaux les plus pertinents. Le noyau final est obtenu en combinant linéairement ces noyaux pertinents. Cette approche nous permet ainsi d'aborder le problème de sélection de modèle que nous avons illustré avec deux applications : la sélection optimale de paramètres d'un descripteur et une sélection de variable.

Actuellement, nous sommes encore limités par la mémoire nécessaire pour stocker l'ensemble des noyaux, il n'est donc pas encore possible de considérer une liste exhaustive de noyaux. Lorsque ce défaut sera pallié, il nous sera possible de tester de façon plus approfondie cette approche en considérant davantage de descripteurs, de paramètres, de variables. Il nous sera ainsi possible d'utiliser des informations complémentaires pour décrire les objets.

Références

[1] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, editors, *International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 886–893, June 2005.

[2] Gert R. G. Lanckriet, Nello Cristianini, Peter Bartlett, Laurent El Ghaoui, and Michael I. Jordan. Learning the kernel matrix with semidefinite programming. *J. Mach. Learn. Res.*, 5 :27–72, 2004.

[3] Constantine Papageorgiou and Tomaso Poggio. Trainable pedestrian detection. In *Proceedings of the 1999 International Conference on Image Processing*, pages 35–39, 1999.

[4] Amnon Shashua, Yoram Gdalyahu, and Gaby Hayon. Pedestrian detection for driving assistance systems : Single-frame classification and system level performance. In *Proceedings of IEEE Intelligent Vehicles Symposium*, 2004.

[5] Sören Sonnenburg, Gunnar Raetsch, and Christin Schaefer. A general and efficient multiple kernel learning algorithm. In *Advances in Neural Information Processing Systems 18*, pages 1273–1280, 2005.

[6] Vladimir Vapnik. *The Nature of Statistical Learning Theory.* Springer, N.Y, 1995.