

Examen Transfert Learning

MASTER 2 CILS – 1h30

Document autorisé : article sur DASVM de L. Bruzzone & M. Marconcini.

1 Sur l'introduction

1. Quels sont les deux raisons qui expliquent le développement tardif des techniques d'adaptation de domaine ?
2. Quels sont les avantages, selon les auteurs, d'utiliser des SVM, en toute généralité ?

2 Sur DASVM

1. DASVM travaille à partir de deux ensembles d'apprentissage. Lesquels ? Que contiennent-ils ?
2. Quel type d'exemples sont utilisés pendant la phase d'initialisation ?
3. Le problème d'optimisation résolu dans la phase d'initialisation fait intervenir des variables ξ_i^s . Au delà des aspects mathématiques (régularisation), pourquoi introduit-on ces variables ?
4. On considère maintenant une itération de DASVM. Comment sont calculées les étiquettes des exemples non étiquetés ?
5. Quels sont les deux critères utilisés pour choisir les points non étiquetés à ajouter à l'ensemble d'apprentissage ?
6. Que désigne le paramètre ρ ?
7. Qu'est-ce qu'un exemple semi-étiqueté ?
8. Que deviennent les exemples semi-étiquetés à une itération qui sont mal étiquetés à l'itération suivante ?
9. Quel est le critère utilisé pour supprimer des exemples étiquetés de l'ensemble d'apprentissage ?
10. Combien d'exemples étiquetés sont supprimés de l'ensemble d'apprentissage à chaque itération ?
11. Au cours des itérations, le paramètre de régularisation associé à un exemple semi-étiqueté augmente. Pourquoi avoir fait ce choix ?
12. Quel est le choix fait concernant les paramètres de régularisation des exemples étiquetés ? Pourquoi ?
13. Les itérations de DASVM s'enchaînent. Quel est le critère qui devrait être utilisé pour arrêter ces itérations, dans l'idéal ?
14. Quel est le critère d'arrêt qui est utilisé en pratique ?
15. Dans le cadre d'un problème multi-classe, les auteurs suggèrent d'utiliser la stratégie One-Against-All. Quel en est le principe ?

3 Sur la stratégie de validation d'un algorithme d'adaptation

1. Quelle est l'idée générale développée par les auteurs pour déterminer si un algorithme est ou non capable d'adapter un classifieur appris sur un domaine source à un domaine cible ?
2. Pour quantifier la qualité d'un algorithme, les auteurs définissent 4 ensembles $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ de classifieurs. Pourquoi ces ensembles contiennent-ils plusieurs classifieurs, alors qu'il n'existe qu'une seule base d'apprentissage ?
3. Pour déterminer si un classifieur appartient à un des ensembles, on a besoin d'un paramètre Λ_{th} . Que représente ce paramètre ?
4. Comment détermine-t-on qu'un classifieur appartient aux ensembles \mathcal{A} ou \mathcal{C} habituellement ? Citer un exemple de technique habituellement utilisée.
5. Sous les hypothèses de travail des auteurs, peut-on déterminer si un classifieur appartient aux ensembles \mathcal{B} ou \mathcal{D} ? Si oui, comment ? Si non, pourquoi ?
6. La procédure de validation est modélisée sous la forme d'un diagramme d'états et de transitions, et l'algorithme d'apprentissage permet de passer des états \bar{A} ou \bar{C} , aux états \bar{B} ou \bar{D} . Intuitivement, pourquoi ne s'intéressent-on pas à un jeu de paramètres qui nous place dans l'état \bar{C} au départ ?
7. Décrire la procédure qui permet de passer des états \bar{B} ou \bar{D} , aux états \bar{A} ou \bar{C} .
8. Pourquoi ne faut-il pas que l'on puisse passer de l'état \bar{D} à l'état \bar{A} ? Est-il théoriquement possible de savoir si ça se produit ?
9. Supposons qu'un jeu de paramètres permette de passer de l'état \bar{B} à l'état \bar{C} . Pourquoi écarterez-vous ce jeu de paramètre (à tort) dans la procédure de validation ?

4 Sur les expérimentations

1. Pour faire les expérimentations, les auteurs disent utiliser un noyau gaussien. Où ce noyau intervient-il dans les différents problèmes de minimisation décrits par les problèmes (1), (8) et (12) ?
2. La Table 2 présente des valeurs. Comment sont-elles calculées ? Dans quel intervalle peuvent-elles théoriquement varier ? A quoi correspondent chacune des lignes ? Que peut-on déduire de ce tableau ?
3. Dans la Table 4, quelle est la différence entre la ligne DASVM^{best} et la ligne DASVM^{ave} ? Pourquoi les auteurs donnent-ils la seconde ?
4. La table 4 présente également $P(\bar{A} | \bar{B})$. Que représente cette probabilité ? Comment se fait-il que les auteurs arrivent à la calculer ? Comment interprétez-vous le fait qu'elle décroisse lorsque l'angle augmente ?
5. Expliquez, à l'aide de la Figure 5, comment les auteurs valident la pertinence de la méthode de validation circulaire qu'ils proposent.