HAL
open science

Classification de données labellisées de manière imprécise et incertaine

Etienne Côme

To cite this version:

Etienne Côme. Classification de données labellisées de manière imprécise et incertaine. Journées MAS et Journée en l'honneur de Jacques Neveu, Aug 2010, Talence, France. inria-00496015

HAL Id: inria-00496015

https://inria.hal.science/inria-00496015

Submitted on 29 Jun 2010

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.

Session : Apprentissage statistique

Classification de données labellisées de manière imprécise et incertaine

par **Etienne Côme**

Cette intervention présente une solution aux problèmes de classification faisant intervenir des données d'apprentissage labellisées de manière imprécise et incertaine. Tous les exemples utilisés pour l'apprentissage sont donc décrits classiquement par un vecteur de descripteurs x_i , mais aussi par une étiquette imprécise et/ou incertaine m_i, spécifiant de manière douce l'appartenance de l'individu aux différentes classes d'intérêts. Nous proposons d'utiliser comme étiquette une fonction de masse de croyance. Le problème d'apprentissage est donc traité dans le cadre de la théorie des fonctions de croyance, ce qui permet une grande souplesse dans la définition des étiquettes. Cette approche généralise ainsi différents problèmes d'apprentissage couramment rencontrés dans la littérature : *supervisé, non-supervisé, semi-supervisé, partiellement supervisé*, que nous présenterons successivement. Pour résoudre le problème de classification, la solution décrite fait l'hypothèse que les données sont générées suivant un modèle de mélange. Nous montrerons que dans ce cadre il est possible de dériver grâce au théorème de Bayes généralisé un critère qui étend les critères de maximum de vraisemblance associés aux différents problèmes d'apprentissages sus-mentionnés. Nous présenterons également un algorithme de type *EM* dédié à l'optimisation de ce critère et permettant d'obtenir une estimation des différents paramètres du modèle. Finalement, nous donnerons des résultats de simulations mettant en évidence l'intérêt d'une telle approche, en particulier pour traiter élégamment les problèmes résultant d'erreurs d'étiquetages.

Adresse :
Etienne Côme
SAMM, Université Paris 1
90 Rue de Tolbiac
75013 Paris, France
E-mail : `Etienne.Come@univ-paris1.fr`
<`http://samm.univ-paris1.fr/COME-Etienne`>