

# Calibration évidentielle conjointe de classifieurs SVM binaires fondée sur la régression logistique

## Evidential joint calibration of binary SVM classifiers based on logistic regression

P. Minary<sup>1,2</sup> F. Pichon<sup>1</sup> D. Mercier<sup>1</sup> É. Lefèvre<sup>1</sup> B. Droit<sup>2</sup>

<sup>1</sup> Univ. artois, EA 3026 LGI2A, Béthune, F-62400, France

<sup>2</sup> SNCF Réseau, Dpt des Télécommunications, La Plaine Saint Denis, France

{pauline.minary,benjamin.droit}@reseau.sncf.fr ; {frederic.pichon,david.mercier,eric.lefevre}@univ-artois.fr

### Résumé :

Dans le cadre de la combinaison de classifieurs, une étape de calibration, fondée sur la régression logistique, est généralement utilisée pour transformer indépendamment chaque sortie de classifieur en distribution de probabilité pour ensuite les combiner. Afin de mieux gérer les incertitudes cette calibration a récemment été redéfinie dans le cadre des fonctions de croyance. Dans cet article, nous proposons de calibrer conjointement les classifieurs en utilisant la régression logistique multiple, et d'étendre cette approche au cadre des fonctions de croyance. Notre approche évidentielle est testée sur des données simulées et réelles avec des classifieurs SVM binaires, et présente plusieurs avantages par rapport à sa version probabiliste et par rapport à l'approche évidentielle calibrant indépendamment chaque classifieur.

### Mots-clés :

Fonctions de croyance, fusion d'informations, calibration, régression logistique multiple.

### Abstract:

In a context of multiple classifiers, a calibration step based on logistic regression is usually used to independently transform each classifier output into a probability distribution, to be then able to combine them. This calibration has been recently refined, using the evidence theory, to better handle uncertainties. In this paper, we propose to calibrate jointly the classifiers using multiple logistic regression, and to extend this approach to the evidential framework. Our evidential approach is tested on generated and real datasets with binary SVM classifiers, and presents several advantages over its probabilistic version and the evidential approach based on independent calibration.

### Keywords:

Belief functions, information fusion, calibration, multiple logistic regression.

## 1 Introduction

Un moyen d'obtenir de meilleures performances pour un problème de classification donné consiste à fusionner les résultats de plu-

sieurs classifieurs [9]. Néanmoins, ces classifieurs peuvent avoir été entraînés sur des données d'apprentissage différentes ou reposer sur des méthodes de classification différentes. Ainsi, pour rendre comparables les sorties de ces classifieurs et pour pouvoir les combiner, une étape de calibration est nécessaire. Généralement, celle-ci permet de transformer la sortie du classifieur en une probabilité.

Une des méthodes de calibration les plus employées utilise la régression logistique [10]. Dans cette méthode comme pour les autres approches de calibration, un classifieur fournit en sortie une classe ainsi qu'un score de confiance associé à cette classe. À partir d'un ensemble d'apprentissage, cette calibration permet alors de fournir une probabilité sur la classe en se fondant sur l'a priori que plus le score de confiance est important, plus la probabilité doit être grande.

Récemment, Xu *et al.* [13] ont proposé un raffinement de cette calibration dans le cadre de la théorie des fonctions de croyance [11, 12]. Cette calibration porte sur un classifieur, et fournit une fonction de croyance au regard de la vraie classe à la place d'une probabilité. Cette méthode évidentielle permet d'obtenir de meilleures performances lorsque l'ensemble d'apprentissage contient en particulier peu de données. Plus précisément, elle permet de diminuer le problème de surapprentissage

notamment dans ces situations.

Au lieu de calibrer individuellement chaque classifieur puis de les combiner en cherchant la règle la plus adaptée, il est possible de considérer conjointement les scores retournées par un ensemble de classifieurs pour une même classe d'un objet donné, et d'apprendre la probabilité à attribuer à cette classe étant donné ces scores. Pour ce faire, une calibration utilisant la régression logistique à plusieurs variables (ou régression logistique multiple) [6] peut être employée.

Dans ce papier, nous proposons d'étendre au cas évidentiel une telle calibration conjointe et fondée sur la régression logistique multiple, avec les mêmes outils employés par Xu *et al.* pour étendre la calibration d'un classifieur fondée sur la régression logistique à une variable. Puis, nous montrons sur des expérimentations que cette approche peut s'avérer moins sensible au surapprentissage que sa version probabiliste. Nous montrons également comment, dans un problème de fusion de classifieurs, cette approche peut donner de meilleurs résultats qu'une fusion par la règle de Dempster de calibrations individuelles.

La plan du papier est le suivant. Les notions de base nécessaires sur la théorie des fonctions de croyance sont exposées dans la section 2. La calibration conjointe fondée sur la régression logistique multiple ainsi que son extension dans le cadre évidentiel que nous proposons sont ensuite présentées à la section 3. Ces deux approches ainsi que l'approche calibrant indépendamment chaque classifieur, sont alors comparées expérimentalement dans la section 4. Enfin, une conclusion et des perspectives sont données dans la section 5.

## 2 Théorie des fonctions de croyance

Les notions de base de la théorie des fonctions de croyance sont rappelées dans la section 2.1, puis sont exposées dans la section 2.2 les applications en statistique de cette théorie à l'estimation et à la prédiction, qui sont des concepts uti-

lisés pour définir les calibrations dans le cadre des fonctions de croyance.

### 2.1 Concepts de base

La théorie des fonctions de croyance est un cadre permettant de raisonner sur des incertitudes. Avec  $\Omega$  un ensemble fini contenant l'ensemble des réponses possibles à une question  $Q$  d'intérêt, une information portant sur la réponse à  $Q$  peut être représentée par une *fonction de masse* (FM)  $m^\Omega$  définie de l'ensemble des parties de  $\Omega$ , noté  $2^\Omega$  dans l'intervalle  $[0, 1]$  telle que la somme des masses vaut 1, i.e.  $\sum_{A \subseteq \Omega} m^\Omega(A) = 1$ . La quantité  $m^\Omega(A)$  représente la part de croyance allouée au fait que la réponse à  $Q$  est dans l'ensemble  $A$  sans être plus spécifique. Tout ensemble  $A$  tel que  $m^\Omega(A) > 0$ , est appelé *élément focal* de  $m^\Omega$ . Quand les éléments focaux d'une FM  $m^\Omega$  sont emboîtés,  $m^\Omega$  est dite *consonante*. Une FM  $m^\Omega$  est en correspondance biunivoque avec une fonction de croyance  $Bel^\Omega$  et une fonction de plausibilité  $Pl^\Omega$  respectivement définies pour tout  $A \subseteq \Omega$  par

$$Bel^\Omega(A) = \sum_{B \subseteq A} m^\Omega(B), \quad (1)$$

$$Pl^\Omega(A) = \sum_{B \cap A \neq \emptyset} m^\Omega(B). \quad (2)$$

La fonction de plausibilité restreinte aux singletons est appelée *fonction de contour* et est notée  $pl^\Omega$ . On a  $pl^\Omega(\omega) = Pl^\Omega(\{\omega\})$  pour tout  $\omega \in \Omega$ . Quand une FM est consonante, la fonction de plausibilité peut être retrouvée à partir de la fonction de contour grâce à la relation suivante

$$Pl^\Omega(A) = \sup_{\omega \in A} pl^\Omega(\omega), \quad (3)$$

pour tout  $A \subseteq \Omega$ .

Différentes stratégies existent au moment de prendre une décision à partir d'une fonction de croyance [4]. Il est par exemple possible de choisir la réponse à la question  $Q$  minimisant les risques espérés inférieur ou supérieur.

En prenant des coûts de décision égaux à 1 en cas de mauvaise décision et égaux à 0 en cas de bonne décision, les risques espérés inférieur et supérieur sont respectivement définis par  $R_*(\omega) = 1 - m^\Omega(\{\omega\}) - m^\Omega(\Omega)$  et  $R^*(\omega) = 1 - m^\Omega(\{\omega\})$ . Choisir la réponse minimisant le risque inférieur  $R_*$  (resp. supérieur  $R^*$ ) est une stratégie optimiste (resp. pessimiste). Enfin, il est également possible d'introduire une décision ne faisant pas partie des réponses à la question  $Q$ , et qui est la *décision de rejet*, c'est-à-dire la décision de ne pas choisir une réponse à la question  $Q$ . Cette nouvelle décision, associée à un risque  $R_{rej} \in [0, 1]$ , permet de ne pas prendre la décision minimisant un risque  $R^*$  ou  $R_*$  lorsque celui-ci est supérieur à  $R_{rej}$ . Le risque  $R_{rej}$  correspond à un montant maximum de risque accepté, il permet d'éviter de prendre une décision trop risquée.

## 2.2 Estimation et prédiction statistiques

La théorie des fonctions de croyance peut être employée pour faire de l'estimation et de la prédiction statistiques [11, 5, 3, 8, 7].

Soient  $\theta \in \Theta$  un paramètre inconnu,  $x \in \mathbb{X}$  une donnée observée réalisation d'une variable  $X$ , et  $f_\theta(x)$  la fonction de densité générative des données. L'estimation (ou inférence) statistique consiste à estimer le paramètre  $\theta$  étant donné l'observation  $x$ . Shafer [11, Chapitre 11] propose (une justification est donnée par Denœux dans [5]) de représenter cette connaissance sur  $\theta$  par une fonction de croyance consonante  $Bel_x^\Theta$  définie à partir de la fonction de vraisemblance  $L_x : \theta \rightarrow f_\theta(x)$ , et dont la fonction de contour associée est égale à la fonction de vraisemblance normalisée :

$$pl_x^\Theta(\theta) = \frac{L_x(\theta)}{\sup_{\theta' \in \Theta} L_x(\theta')}, \quad \forall \theta \in \Theta. \quad (4)$$

Supposons maintenant qu'ayant observé  $x$ , des informations sont disponibles sur  $\theta$  sous forme d'une fonction de contour  $pl_x^\Theta$ . Le but de la *prédiction statistique* est de donner des éléments

sur une donnée non encore observée  $y \in \mathbb{Y}$  réalisation d'une variable  $Y$ , dont la distribution conditionnelle étant donné  $X = x$  dépend de  $\theta$ . Une manière d'obtenir une telle fonction de croyance définie sur  $\mathbb{Y}$  est d'utiliser le modèle d'échantillonnage de Dempster [3, 8, 7]. Dans ce modèle, la variable  $Y$  est exprimée comme une fonction du paramètre  $\theta$  et d'une variable auxiliaire non observable dont la distribution est indépendante de  $\theta$ . Considérons le cas particulier important suivant comme exemple.

**Exemple 1** Supposons  $Y \in \mathbb{Y} = \{0, 1\}$  une variable aléatoire suivant une loi de Bernouilli de paramètre  $\theta$ . Xu et al. [13] montre que dans ce cas, on a :

$$Bel_x^\mathbb{Y}(\{1\}) = \hat{\theta} - \int_0^{\hat{\theta}} pl_x^\Theta(u) du, \quad (5)$$

$$Pl_x^\mathbb{Y}(\{1\}) = \hat{\theta} + \int_{\hat{\theta}}^1 pl_x^\Theta(u) du, \quad (6)$$

avec  $\hat{\theta}$  maximisant  $pl_x^\Theta$ .

## 3 Une approche de calibration évidentielle conjointe

Supposons qu'un classifieur binaire SVM fournit un score de confiance  $s \in \mathbb{R}$  accompagnant chacune de ses sorties s'exprimant au regard de réalisations d'une variable  $Y \in \mathbb{Y} = \{0, 1\}$ . L'étape de calibration, en particulier celle basée sur la régression logistique [10], permet d'apprendre comment interpréter ce score en estimant la probabilité  $p^\mathbb{Y}(\cdot|s)$  à partir d'un ensemble d'apprentissage composé de données étiquetées. Moins l'ensemble d'apprentissage contient de données et plus les probabilités estimées sont incertaines. Pour modéliser plus finement ces incertitudes, Xu *et al.* propose dans [13] d'utiliser une calibration évidentielle, i.e. utilisant des fonctions de croyance, plutôt qu'une calibration probabiliste, i.e. utilisant des probabilités.

Dans ce papier, nous proposons d'étendre la régression logistique à plusieurs variables [6] au

cas évidentiel et de l'employer dans une calibration jointe des sorties de plusieurs classifieurs binaires SVM.

Avec  $J$  le nombre de classifieurs SVM et  $n$  le nombre d'objets dans l'ensemble d'apprentissage de la calibration, cet ensemble d'apprentissage est défini par  $\mathcal{X} = \{(S_{11}, \dots, S_{J1}, Y_1), \dots, (S_{1n}, \dots, S_{Jn}, Y_n)\}$ , où  $S_{ji}$ ,  $j \in \{1, \dots, J\}$ ,  $i \in \{1, \dots, n\}$ , correspond au score donné par le  $j^{\text{ième}}$  classifieur au regard du  $i^{\text{ième}}$  objet à classer, et  $Y_i$  est la vérité terrain de l'objet  $i$ .

Étant donné un vecteur de scores  $s = (s_1, \dots, s_j, \dots, s_J)$ , où  $s_j$  est le score retourné par le  $j^{\text{ième}}$  classifieur, la calibration basée sur la régression logistique multiple (ou à plusieurs variables) est définie par

$$P^{\mathbb{Y}}(y = 1|s) \approx \frac{h_s(\sigma)}{1 + \exp(\sigma_0 + \sigma_1 s_1 + \sigma_2 s_2 + \dots + \sigma_J s_J)}, \quad (7)$$

où  $\sigma = \{\sigma_0, \dots, \sigma_J\} \in \mathbb{R}^{J+1}$  est obtenu par la maximisation de la fonction de vraisemblance

$$L(\sigma) = \prod_{i=1}^n p_i^{Y_i} (1 - p_i)^{1 - Y_i}, \quad (8)$$

avec

$$p_i = \frac{1}{1 + \exp(\sigma_0 + \sigma_1 s_1 + \dots + \sigma_J s_J)}. \quad (9)$$

Afin de mieux gérer les incertitudes, nous proposons d'étendre cette approche fondée sur la régression logistique multiple au cadre évidentiel en suivant le même raisonnement employé par Xu et al. dans [13] pour étendre au cas évidentiel la calibration basée sur la régression logistique.

La calibration d'un vecteur de score  $s$  avec la régression logistique peut être vue comme un problème de prédiction d'une variable de Bernoulli  $Y$  de paramètre  $\theta$ ,  $Y \sim \mathcal{B}(\theta)$ , avec  $\theta = h_s(\sigma)$ .

La fonction de croyance  $Bel^{\mathbb{Y}}(\cdot|s)$  peut être obtenue à partir de la fonction contour  $pl_{\mathcal{X}}^{\Theta}(\cdot|s)$  (cf. Section 2.2). Selon Xu *et al.* [13], cette fonction contour peut être calculée à l'aide de  $Pl_{\mathcal{X}}^{\Sigma}$  qui correspond à la fonction contour  $pl_{\mathcal{X}}^{\Sigma}$  définie par

$$pl_{\mathcal{X}}^{\Sigma}(\sigma) = \frac{L(\sigma)}{L(\hat{\sigma})}, \quad \forall \sigma \in \Sigma, \quad (10)$$

où  $\hat{\sigma} = (\hat{\sigma}_0, \dots, \hat{\sigma}_J)$  est l'estimateur du maximum de vraisemblance de  $\sigma$  et  $L$  la fonction de vraisemblance définie par l'équation (8).

Comme  $\theta = h_s(\sigma)$ , nous avons

$$pl_{\mathcal{X}}^{\Theta}(\theta|s) = \begin{cases} 0 & \text{si } \theta \in \{0, 1\}, \\ Pl_{\mathcal{X}}^{\Sigma}(h_s^{-1}(\theta)) & \text{sinon,} \end{cases} \quad (11)$$

avec

$$h_s^{-1}(\theta) = \left\{ \frac{\{(\sigma_0, \sigma_1, \dots, \sigma_J) \in \Sigma \mid 1 + \exp(\sigma_0 + \sigma_1 s_1 + \dots + \sigma_J s_J) = \theta\}}{1} \right\}, \quad (12)$$

$$= \{(\sigma_0, \sigma_1, \dots, \sigma_J) \in \Sigma \mid \sigma_0 = \ln(\theta^{-1} - 1) - \sigma_1 s_1 - \dots - \sigma_J s_J\}. \quad (13)$$

Des équations (11) et (13), on déduit alors la fonction de contour

$$pl_{\mathcal{X}}^{\Theta}(\theta|s) = \sup_{\sigma_1, \dots, \sigma_J \in \mathbb{R}} pl_{\mathcal{X}'}^{\Sigma}(\ln(\theta^{-1} - 1) - \sigma_1 s_1 - \sigma_2 s_2 - \dots - \sigma_J s_J, \sigma_1, \dots, \sigma_J), \quad (14)$$

pour tout  $\theta \in [0, 1]$ . Le vecteur de paramètres  $(\sigma_1, \sigma_2, \dots, \sigma_J)$  maximisant  $pl_{\mathcal{X}'}^{\Sigma}$  peut être trouvé en utilisant un algorithme de maximisation itératif, tel que l'algorithme de descente de gradient. Nous pouvons alors obtenir les fonctions de croyance et de plausibilité  $Bel^{\mathbb{Y}}(\cdot|s)$  et  $Pl^{\mathbb{Y}}(\cdot|s)$  en appliquant les équations (5) et (6).

## 4 Tests expérimentaux

Les expérimentations de cette section sont menées sur différents problèmes de classification

Base	SVM 1	SVM 2	SVM 3	Test
Données Simulées	20	40	10-50-200	100
<i>Australian</i>	30	70	10-60-190	400

Tableau 1 – Nombre d'exemples utilisés pour l'apprentissage de chaque classifieur et le test.

binaires issus des bases de données UCI [1] et sur des données simulées. Ces dernières suivent une loi normale bidimensionnelle de moyennes  $\mu_0 = (-1, 0)$  pour la classe 0 et  $\mu_1 = (1, 1)$  pour la classe 1, et de matrice de covariance  $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$  commune aux deux classes. Dans la section 4.1, la calibration (probabiliste) basée sur la régression logistique multiple ainsi que son extension évidentielle, que nous proposons, sont comparées à l'approche de Xu *et al.* utilisant des calibrations logistiques évidentielles individuelles [13]. Enfin, des expérimentations sont proposées dans la section 4.2 afin de comparer notre solution à sa version probabiliste.

#### 4.1 Comparaison entre régression logistique multiple et régression logistique individuelle

Dans cette première partie, nous reprenons la première expérimentation détaillée dans [13]. Pour chaque ensemble de données, trois classifieurs SVM sont entraînés en utilisant la librairie LIBSVM [2]. Le nombre d'exemples utilisés pour l'apprentissage et le test sont présentés dans le tableau 1 pour les données simulées et pour la base de données *Australian*<sup>1</sup>. Le nombre d'exemples d'apprentissage est fixé pour les deux premiers classifieurs et il varie pour le troisième classifieur. L'ensemble d'apprentissage est décomposé en deux sous-ensembles de taille égale. Le premier sous-ensemble est utilisé pour l'apprentissage du

classifieur.

Dans le cas de l'approche proposée par Xu *et al.*, le second sous-ensemble est utilisé pour l'apprentissage de la calibration. Ainsi dans la phase de test, chaque classifieur donne un score qui est calibré selon l'apprentissage réalisé et produit une fonction de masse de croyance. Les trois fonctions de masse sont alors fusionnées à l'aide de la combinaison de Dempster [11].

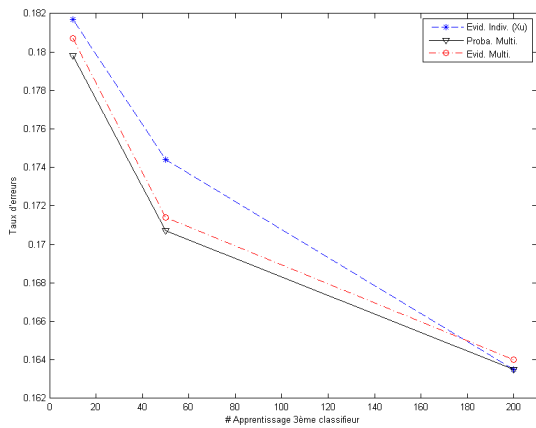
Pour le cas des approches de calibration jointe, l'apprentissage de la calibration est réalisé en concaténant les sous-ensembles des 3 classifieurs. Lors de la phase de test, les scores retournés par les trois classifieurs sont regroupés dans un vecteur qui est calibré à l'aide de la régression logistique multiple. Dans le cas évidentiel, une fonction de masse finale est retournée alors que dans l'approche probabiliste, une distribution de probabilité est obtenue.

Pour les approches évidentielles, la décision prise correspond au singleton ayant le maximum de croyance. Pour l'approche probabiliste, la décision correspond à l'hypothèse maximisant la probabilité.

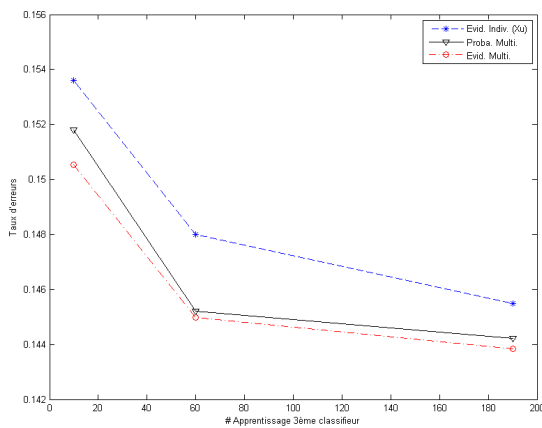
La figure 1 présente les taux d'erreur<sup>2</sup> de l'approche de régression logistique individuelle de Xu *et al.*, de la régression logistique multiple probabiliste et son extension évidentielle que nous proposons. L'approche proposée permet d'obtenir de meilleurs résultats que l'approche évidentielle individuelle aussi bien pour les données simulées (figure 1a) que pour les données *Australian* (figure 1b). Par ailleurs, on peut noter que les approches multiples (jointes) évidentielle et probabiliste donnent des résultats proches sur ces expérimentations. Dans la section suivante, d'autres expérimentations sont menées pour comparer ces deux approches.

1. Parmi les bases de données UCI, seule cette base est considérée par manque de place. Toutefois, les résultats obtenus avec d'autres bases de données UCI sont similaires.

2. Le taux de d'erreur correspond au nombre d'objets incorrectement classés divisé par le nombre d'objets total car il n'y a pas de rejet dans cette expérimentation.



(a) Données simulées.



(b) Base de données *Australian*.

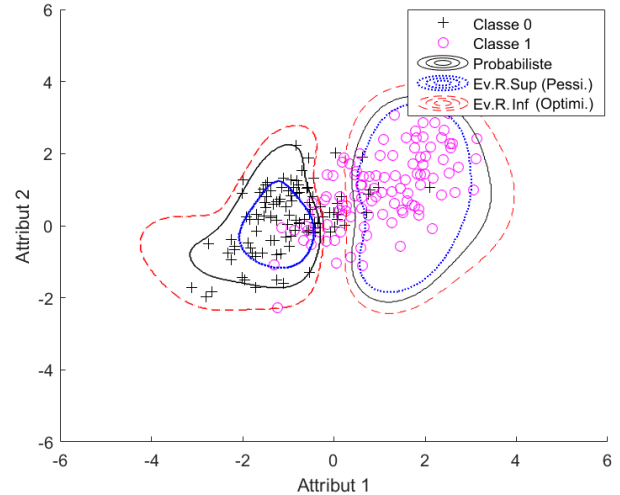
Figure 1 – Taux d’erreurs en fonction du nombre d’exemple d’apprentissage pour le 3ème classifieur.

## 4.2 Comparaison entre régression logistique multiple évidentielle et régression logistique multiple probabiliste

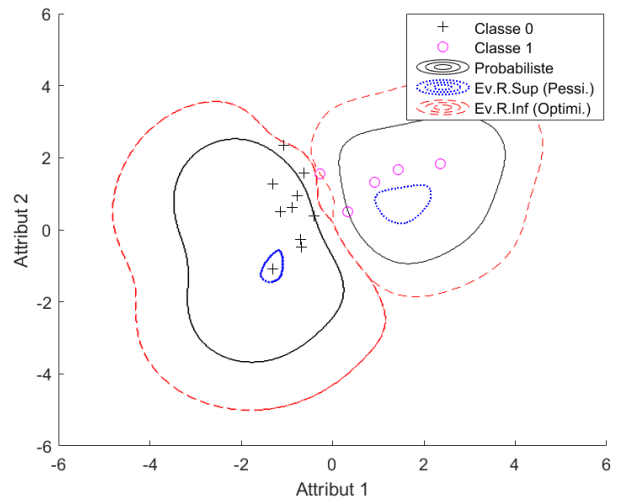
Pour cette expérimentation<sup>3</sup>, nous générons 290 réalisations suivant les lois décrites précédemment, dont 90 réalisations ont servi pour entraîner 3 classifieurs SVM sur 30 réalisations distinctes en utilisant la bibliothèque LIBSVM [2]. Dans un premier scénario, les 200 réalisations restantes servent de base d’apprentissage pour la régression logistique multiple.

3. Par manque de place, nous n’utilisons ici que les données simulées mais les résultats sont similaires avec les données UCI.

Dans un second scénario, seuls 15 réalisations sont choisies parmi les 200 restantes comme base d’apprentissage pour la régression logistique multivariable.



(a) Scénario 1 : apprentissage de chaque calibration avec 200 exemples



(b) Scénario 2 : apprentissage de chaque calibration avec 15 exemples

Figure 2 – Frontières de décision dans l’espace des attributs pour les calibrations utilisant une régression logistique multiple (cas probabiliste, évidentiel avec risque inférieur, évidentiel avec risque supérieur),  $R_{rej} = 0.2$ , apprentissage avec 200 réalisations (2a) et 15 réalisations (2b).

La figure 2 illustre les frontières de décision sur les scénarios 1 et 2 pour les calibrations utilisant une régression logistique multiple probabiliste, évidentielle avec risque inférieur, éviden-

tielle avec risque supérieur, pour  $R_{rej} = 0.2$ .

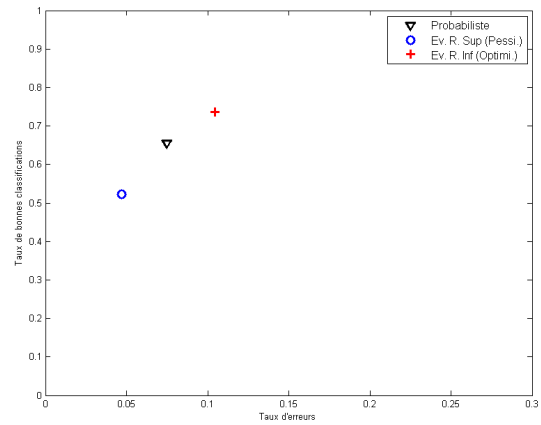
Nous pouvons observer que l'approche fondée sur la stratégie optimiste tend à décider plus et donc rejette moins que les deux autres approches. Ce constat est inversé pour la stratégie pessimiste.

En outre, nous pouvons également observer que l'approche évidentielle permet de mieux rendre compte de l'incertitude sur les zones de décision puisque lorsqu'il y a peu d'exemples, l'écart entre les frontières obtenues par une stratégie pessimiste et une stratégie optimiste est plus grand que lorsqu'il y a beaucoup d'exemples.

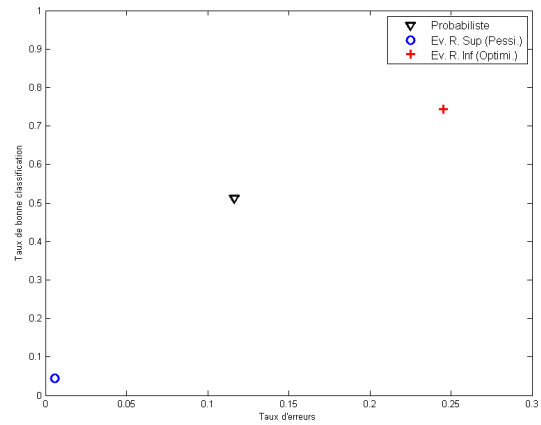
Nous nous intéressons maintenant aux taux de bonnes classifications et taux d'erreurs obtenus par ces mêmes trois calibrations pour 100 tests effectués avec les mêmes répartitions des données pour les ensembles d'apprentissage que dans les scénarios 1 et 2, toujours avec un risque de rejet  $R_{rej} = 0.2$ . Le taux de bonne classification correspond au nombre d'objets correctement classés divisé par le nombre d'objets classés (et pas divisé par le nombre d'objets à classer, car nous ne comptons pas les objets rejetés). Les résultats sont illustrés sur la figure 3. Nous pouvons observer que les points de fonctionnement des calibrations sont plus éloignés lorsqu'il y a peu d'exemples d'apprentissage pour les calibrations. Cet intervalle reflète l'incertitude qui est plus grande dans le scénario 2. La calibration probabiliste quant à elle peut donner des probabilités identiques quelque soit la taille de l'ensemble d'apprentissage ce qui ne permet pas d'intégrer d'incertitude.

Enfin, nous conduisons la même expérience que précédemment mais en considérant que le risque  $R_{rej}$  varie de 0 à 1. Notre calibration a été entraînée avec 45 puis 15 exemples. La figure 4 illustre les résultats obtenus.

Pour un taux d'erreur donné, les résultats obtenus avec notre proposition pour une stratégie pessimiste ont un taux de bonne classification supérieure (ou égale) à celui de l'approche



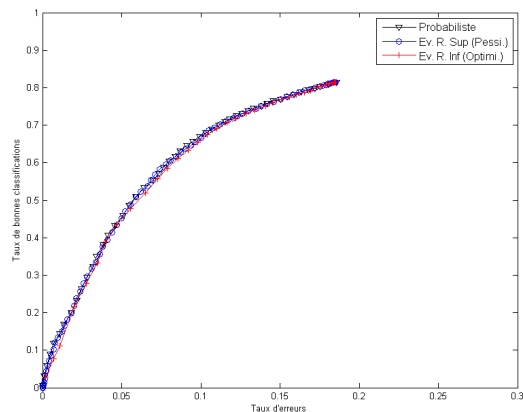
(a) Scénario 1 : apprentissage de chaque calibration avec 200 exemples



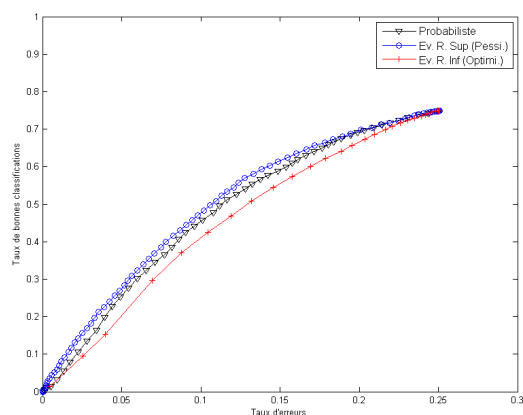
(b) Scénario 2 : apprentissage de chaque calibration avec 15 exemples

Figure 3 – Taux d'erreurs et de bonnes classifications obtenus par les 3 calibrations sur 100 tests avec  $R_{rej} = 0.2$ , 200 données d'apprentissage pour la calibration (3a) et 15 données d'apprentissage pour la calibration (3b).

probabiliste lorsqu'il y a peu d'exemples d'apprentissage disponibles (figure 4b). Nous pouvons également noter que pour un taux d'erreur donné, les taux de bonne classification pour les approches probabiliste et évidentielle sont obtenues pour des valeurs différentes de  $R_{rej}$  (comme nous avons pu le constater dans la première expérimentation de cette section dont les résultats sont donnés sur la figure 3). Par ailleurs, lorsque le nombre de données d'apprentissage est plus important (figure 4a) les résultats obtenus pour les approches probabiliste



(a) Données Simulées – 45 exemples d'apprentissage



(b) Données Simulées – 15 exemples d'apprentissage

Figure 4 – Taux d'erreur et de bonne classification avec 45 (en haut) et 15 exemples d'apprentissage (en bas) pour les données simulées.

et évidentielle deviennent similaires.

## 5 Conclusion

Dans cet article, une approche de calibration évidentielle conjointe et fondée sur la régression logistique multiple a été proposée. Cette approche permet de prendre en compte l'ensemble des scores retournés par plusieurs classifieurs pour un objet et l'utilisation de la théorie de l'évidence permet de mieux gérer les incertitudes que la version probabiliste lorsqu'il y a peu d'exemples pour l'apprentissage. Notre approche a également été comparée à une méthode évidentielle, proposée par Xu *et al.*, où les scores donnés par les classifieurs sont calibrés

individuellement puis combinés.

Seule l'extension de la régression logistique a été étudiée dans cet article mais d'autres approches de calibration individuelle peuvent être étendues au cas joint. Enfin, notre approche pourrait également être étendue au cas multi-classes, en suivant un raisonnement similaire à celui de Xu *et al.* pour la calibration individuelle [14].

## Références

- [1] K. Bache et M. Lichman. UCI machine learning repository. 2013. <http://archive.ics.uci.edu/ml>.
- [2] C.C. Chang et C.J. Lin. LIBSVM : A library for support vector machines. *Transactions on Intelligent Systems and Technology*, 2(3), article 27, 2011. Logiciel disponible ici <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [3] A.P. Dempster. New methods for reasoning towards posterior distributions based on sample data. *The Annals of Mathematical Statistics*, 37(2) : 355-374, 1966.
- [4] T. Denœux. Analysis of evidence-theoretic decision rules for pattern classification. *Pattern Recognition*, 30(7) : 1095-1107, 1997.
- [5] T. Denœux. Likelihood-based belief function : justification and some extensions to low-quality data. *International Journal of Approximate Reasoning*, 55(7) : 1535-1547, 2014.
- [6] D.W. Hosmer et S. Lemeshow. *Applied logistic regression*. Wiley & Sons, 2004.
- [7] O. Kanjanatarakul, T. Denœux et S. Sriboonchitta. Prediction of future observations using belief functions : A likelihood-based approach. *International Journal of Approximate Reasoning*, 72 : 71-94, 2015.
- [8] O. Kanjanatarakul, S. Sriboonchitta et T. Denœux. Forecasting using belief functions : an application to marketing econometrics. *International Journal of Approximate Reasoning*, 55(5) : 1113-1128, 2014.
- [9] L.I. Kuncheva. *Combining pattern classifiers : methods and algorithms*. John Wiley & Sons, 2004.
- [10] J.C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3) : 61-74, 1999.
- [11] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, 1976.
- [12] P. Smets et R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66 : 191-243, 1994.
- [13] P. Xu, F. Davoine, H. Zha et T. Denœux. Evidential calibration of binary SVM classifiers. *International Journal of Approximate Reasoning*, 72 : 55-70, 2016.
- [14] P. Xu, F. Davoine, H. Zha et T. Denœux. Evidential multinomial logistic regression for multiclass classifier calibration. *18th International Conference on Information Fusion*, pages 1106-1112, 2015.