Bagging Improves Uncertainty Representation in Evidential Pattern Classification

Jérémie François[1,2], Yves Grandvalet[1], Thierry Denœux[1], and Jean-Michel Roger[2]

[1] Université de Technologie de Compiègne, Heudiasyc, UMR CNRS 6599, F-60205 Compiègne, France
[2] Cemagref, GIQUAL Research Unit, 361 rue Jean-François Breton, F-34033 Montpellier, France

Abstract. Uncertainty representation is a major issue in pattern recognition when the outputs of a classifier do not lead directly to a final decision, but are used in combination with other systems, or as input to an interactive decision process. In such contexts, it may be advantageous to resort to rich and flexible formalisms for representing and manipulating uncertain information, such as the Dempster-Shafer theory of Evidence. In this paper, it is shown that the quality and reliability of the outputs from an evidence-theoretic classifier may be improved using an adaptation from a resample-and-combine approach introduced by Breiman and known as "bagging". This approach is explained and studied experimentally using simulated data. In particular, results show that bagging improves classification accuracy and limits the influence of outliers and ambiguous training patterns.

Keywords: Supervised Pattern Recognition, K-Nearest Neighbor Rule, Decision Fusion, Dempster-Shafer Theory, Evidence Theory, Bootstrap, Bagging.

1 Introduction

In the last thirty years, the issue of uncertainty representation and management in supervised pattern recognition has received considerable attention. New theoretical frameworks have been proposed as alternatives to Bayesian Probability theory to describe, manipulate, and reason with partial knowledge and unreliable information. In particular, the so-called Dempster-Shafer (D-S) theory of Evidence, first proposed by Shafer [10] and further elaborated by many authors has been shown to constitute a rich and flexible framework, in which the concepts of a probability and possibility measures are recovered as special cases of the more general concept of *belief function*. This theory has been successfully applied in many areas such as diagnosis [12], sensor fusion [1] and pattern classification [9,3].

When applying D-S theory to classification tasks, the construction of belief functions from observation data is a crucial step. Typically, a training set of patterns with known classification is given, and one wishes to quantify one's beliefs concerning the category of a new pattern submitted to the

system. The evidential K-NN rule, previously introduced by one of the authors [3,15], is such a method for inferring a belief function by pooling the evidence from the nearest neighbors in the training set. In this paper, it is proposed to improve this method using a new variant of a technique, known as "bagging", proposed by Breiman [2] in a conventional statistical context to improve the stability of classification rules. This method is shown experimentally to provide a more "realistic" description of the uncertainty pertaining to the classification task, leading to improved classification performances.

The paper is organized as follows. Section 2 introduces the main concepts of Evidence Theory and their use in pattern recognition. Section 3 depicts the adaption of the bagging approach to evidential classifiers, followed by the presentation and discussion of experimental results obtained in an artificial learning task (Sections 4-6). Finally, Section 7 concludes the paper and presents directions for further research.

2 Background

2.1 Theory of Belief Functions

Only the main concepts of the Dempster-Shafer theory of belief functions that we use in this paper will be recalled here. The reader is referred to Shafer's book [10] for a detailed exposition of the mathematical background, and to more recent papers such as, e.g., Refs. [14,13] for up-to-date presentations of the latest developments in both the theoretical aspects and practical applications of belief functions. Although our approach is not tied to a particular interpretation of belief functions, we shall adopt the non-probabilistic view of Smets' Transferable Belief Model (TBM), which constitutes a particularly coherent and justified approach [14,13].

In short, the main assumptions underlying the TBM are that (1) degrees of belief are quantified by numbers between 0 and 1; (2) there exists a two-level structure composed of a *credal level* where beliefs are entertained, and a *pignistic level* where decisions are made; (3) beliefs at the credal level are quantified by belief functions, while decisions at the pignistic level are based on probability functions; (4) when a decision has to be made, beliefs are transformed into probabilities using the so-called pignistic transformation.

The Credal Level Let $\Omega = \{\omega_1, \ldots, \omega_M\}$ be a finite possibility space containing *all* the possible answers to a certain question (the truth lies necessarily somewhere in Ω). In the type of applications envisaged here, Ω is the set of possible classes for an object with unknown class membership. It is assumed that any item of evidence can be represented by a belief structure, or basic belief assignment, defined as a function m from 2^Ω (the power set of Ω) to

the [0,1] interval, verifying:

$$\sum_{A \subseteq \Omega} m(A) = 1. \tag{1}$$

and $m(\emptyset) = 0$. The value of $m(A)$ can be interpreted as the "mass" of belief that is given to A and that cannot be given to any other subset without further information; if $m(A) > 0$, A is called a focal element. For example, $m(\Omega) = 1$ represents total ignorance (m is then called the vacuous belief structure), $m(\{\omega_1\}) = 0.5$ stands for a moderated belief in hypothesis 1, and $m(\{\omega_1, \omega_2\}) = 1$ means complete certainty that either hypothesis 1 or hypothesis 2 is true (with no evidence in favor of any of one them individually).

A new reliable piece of information can be incorporated by use of the Dempster's rule of combination [11], if and only if the two sources of belief (denoted m_1 and m_2) are independent and non-totally contradictory. The combination results in a new belief structure $m = m_1 \oplus m_2$ on Ω that represents the new state of knowledge.

The pignistic level Given a belief structure, different criteria can be used to choose one hypothesis. We will use here the pignistic risk minimization as defined and justified by Smets [14] on an axiomatic basis.

Let `Pbet` be the so-called pignistic probability distribution, defined by uniformly distributing the mass of belief given to each subset of Ω among its elements:

$$\texttt{Pbet}(\omega) = \sum_{\{A \subseteq \Omega | \omega \in A\}} \frac{m(A)}{|A|} \qquad \forall \omega \in \Omega, \tag{2}$$

where $|A|$ is the number of elements in A.

In the TBM, the pignistic probability function is used for decision making according to the Bayes decision theory. Let $\mathcal{A}$ denote a set of actions, and $\lambda(\alpha|\omega)$ the loss incurred if action $\alpha \in \mathcal{A}$ is selected, $\omega \in \Omega$ being the true state of nature. Then, the expected cost (or risk) of choosing action α, relative to the pignistic distribution, is:

$$\texttt{Rbet}(\alpha) = \sum_{\omega \in \Omega} \lambda(\alpha|\omega)\ \texttt{Pbet}(\omega) \tag{3}$$

$$= \sum_{A \subseteq \Omega} \frac{m(A)}{|A|} \sum_{\omega \in A} \lambda(\alpha|\omega). \tag{4}$$

The Bayes decision theory then recommends the action α with the lowest expected cost $\texttt{Rbet}(\alpha)$.

2.2 Application to pattern classification

Recently, Denœux proposed an evidence-theoretic distance-based classifier [3] which takes fully advantage of the evidence theory, by staying free of any intermediate probabilistic representation. The outline of this approach is summarized below.

Let x be the sample to be classified, and let $\mathcal{L} = \{(x_i, y_i)\}_{i=1}^{N}$ be the learning set of known patterns, where $y_i \in \Omega$ is the class of pattern x_i.

First, the K-nearest neighbors of x in $\{x_i\}_{i=1}^{N}$ are selected according to the Euclidean distance. Each neighbor x_k is then considered as an item of evidence about the class of x. If $y_k = \omega_q$, this evidence induces a belief structure m_k with focal elements $\{\omega_q\}$ and Ω [3,5]:

$$m_k(A) = \begin{cases} \alpha \exp(-\gamma_q \|x_k - x\|^2) & \text{if } A = \{\omega_q\} \\ 1 - \alpha \exp(-\gamma_q \|x_k - x\|^2) & \text{if } A = \Omega \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $\|x_k - x\|$ is the Euclidean distance between x_k and x. Parameter $\alpha \in [0, 1]$ sets the minimum belief mass given to Ω, thus limits the certainty expressed by training patterns. Parameters $\gamma_q \in \mathbb{R}^+$ adjust the influence of the patterns of class q according to their distance to x. These coefficients can be determined from the data using a learning scheme proposed in Ref. [15].

Then, as they are independent from each other, the K belief structures m_k are combined into a single structure m by means of Dempster's rule. This structure summarizes the available information about the class of x, provided by its neighborhood in the training set.

Finally, m is used to compute pignistic probabilities $\texttt{Pbet}(\omega_k|x)$, from which class assignment can be performed, using the approach described in Section 2.1 [4]. We define $\mathcal{A} = \{\alpha_0, \alpha_1, \ldots, \alpha_M\}$ the set of actions, where α_i for $i = 1, \ldots, M$ is the decision to classify x in class ω_i, and α_0 denotes rejection. The loss is assumed to be 1 in case of a wrong classification and 0 for correct classification. The rejection loss is assumed to be constant, and equal to some value $\lambda_0 \in [0, 1]$. We thus have:

$$\lambda(\alpha_i|\omega_j) = 1 - \delta_{ij} \quad \forall i, j \in \{1, \ldots, M\} \tag{6}$$

$$\lambda(\alpha_0|\omega_j) = \lambda_0 \quad \forall j \in \{1, \ldots, M\}, \tag{7}$$

where δ_{ij} is the Kronecker symbol ($\delta_{ij} = 1$ if $i = j$, and 0 otherwise).

With these costs, the risks are defined, for each action, as follows :

$$\texttt{Rbet}(\alpha_i) = 1 - \texttt{Pbet}(\omega_i), \quad i = 1, \ldots, M \tag{8}$$

$$\texttt{Rbet}(\alpha_0) = \lambda_0 \tag{9}$$

Each pattern is thus assigned to the class with highest pignistic probability, provided that this probability is greater than $1 - \lambda_0$. Otherwise, it is rejected. Consequently, parameter λ_0 allows to control the rejection rate of the classifier.

3 Sampling, Learning and Uncertainty

Problem The basic belief assignment defined by Eq. 5 handles the uncertainty that stems from the possibly novel characteristics of the query sample. However, additional causes of uncertainty exist. First, the known instances x_k are usually not "prototypical" patterns, such as measurements obtained from some careful experimental design. They are records of past solved cases, which are supposed to be representative of future unsolved cases. In probabilistic terms, they may be considered as randomly sampled from the distribution of future cases. This random sampling is responsible for some uncertainty in the global belief assignment which cannot be taken into account by the basic belief assignment which is conditioned on a given realization of the training set. Additionally, when the parameters of the basic belief assignment are tuned by minimizing some performance criterion on the training set, the learned parameters are also random variables, whose variability is responsible for another part of uncertainty.

This is why we propose here the use of bagging, introduced in the probabilistic framework by Breiman [2] to limit the effects of sampling on a learned decision rule.

Bagging Decision Rules Bagging is a procedure for improving a classification using a resample-and-combine technique. Breiman argues that its main effect is to decrease the variance of the estimator, and advocates its use for unstable classification methods, i.e. methods which are sensitive to perturbations of the training set.

From the original decision rule, the bagged estimator is produced by aggregating using a majority vote on several replicates of the rule, trained on bootstrap resamples of the learning set. A bootstrap sample [7] is created by drawing with replacement N examples from the learning set $\mathcal{L} = \{(x_i, y_i)\}_{i=1}^{N}$. It has thus the same size as the original sample but may contain replicates of some given examples, while other ones are not represented. The drawing with replacement in $\mathcal{L}$ simulates the original sampling from the distribution that generated $\mathcal{L}$. Empirical evaluations showed that the method almost systematically compares favorably with the original predictor [2,6].

Bagging in the TBM In pattern classification, bagging is usually applied to the decisions. In this paper, however, we propose to use it upstream, at the credal level.

The main goal is to better take into account the uncertainty attached to the finite training set, in order to allow steadier decisions and, consequently, to improve the result of further combinations when new sources are available.

Practically, as in decision rule bagging, B bootstrap learning sets $\mathcal{L}_b$ ($b = 1, \ldots, B$) are obtained by drawing with replacement N examples from the original learning set $\mathcal{L}$. Here, the bootstrap is balanced, which means that

each sample (x_i, y_i) is globally drawn B times over the B resamples. Then, for a given unknown sample x, each training set $\mathcal{L}_b$ produces a belief structure m_b through a given K-NN classifier. These are finally aggregated into the average structure $m_{\mathbf{B}}$, defined as:

$$\forall A \subseteq \Omega, \quad m_{\mathbf{B}}(A) = \frac{1}{B} \sum_{b=1,B} m_b(A). \tag{10}$$

In contrast to the usual bagging by majority vote on the B decision rules, aggregation thus takes place at the credal level, using the average operator. While it is clear that aggregation should operate on beliefs, other operators could be used, such as weighted averaging, or the selection of the median or some other particular element among the B structures[1]. Averaging was chosen as a good simple candidate as it is idempotent, commutative and linear: first, getting B times the same structure should lead to this same structure after aggregation (idempotency), second, the resulting structure should be independent from the aggregation order (commutativity), and third, the linear relationship between credal and probabilistic levels, introduced by Smets [14] in the decision process, also supports linear aggregation (linearity).

Remark: In our method, each bootstrap resample of the training set generates a belief structure for each x. These B structures are first aggregated by averaging, and the decision is then based on this average belief structure. The faithful transposition of the original proposition of Breiman would have been to perform a majority vote between the decisions provided by the B classifiers. Experimental results (not shown here) show that this strategy is a poor choice in the TBM framework. This suggests that the evidential K-NN procedure already provides stable decision rules, a finding in agreement with Breiman's results concerning the standard K-NN [2].

4 Experimental Settings

So as to investigate the benefits of bagging, we will focus on an artificial learning task. For easy problems, with well-separated classes and large training sets, many different algorithms yield similar results. A learning task of interest should therefore involve overlapping class distributions and a small learning set. Additionally, it should contain outliers as frequently encountered in real data sets. Finally, a bidimensional problem allows clear representation and interpretation of the results.

We thus consider three bidimensional Gaussian distributions with common covariance matrix $\Sigma = 2.25I$ and mean vectors $(0, 0)$, $(3, 0)$ and $(0, 5)$.

[1] The Dempster's rule of combination cannot be used because the belief sources are not independent.

Each training set $\mathcal{L}$ is constructed by drawing 15 points from each distribution. Additionally, to simulate the contamination of the training set by outliers, 6 points with randomly selected class labels are drawn from a uniform distribution on $[-5, 9] \times [-3, 8]$. To exhibit general trends, 15 training sets were generated from the same distribution. Fig. 1 shows an example of such a generated set.

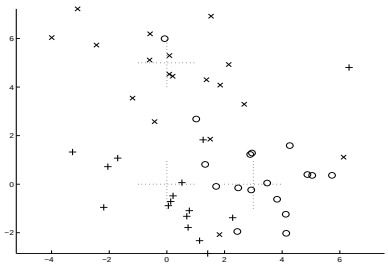

Fig. 1. Example of a generated learning set. The intersections of dotted lines indicate the class means.

Evaluation For each training set, the decision rule is evaluated on a single independent test set $\mathcal{T}$ generated from the same distribution as $\mathcal{L}$ with $N_{\mathcal{T}} = 2000 \times 3 + 800$ items. The mean classification cost $\mathbf{C}$ is estimated by the average of the classification costs on the $N_{\mathcal{T}}$ test points of $\mathcal{T}$:

$$\mathbf{C} = \frac{1}{N_{\mathcal{T}}} \sum_{(x,y)\in\mathcal{T}} \lambda(D(x)|y) \tag{11}$$

where $D(x) \in \mathcal{A} = \{\alpha_0, \ldots, \alpha_M\}$ denotes the decision made by the classifier for pattern x. The costs are defined according to Eqs. 6 and 7.

The classification error rate $\mathbf{E}$ is estimated by the proportion of bad predictions (rejection is not an error) and the rejection rate $\mathbf{R}$ is defined as the proportion of rejected items. We thus have the relation $\mathbf{C} = \mathbf{E} + \lambda_0\mathbf{R}$.

Finally, we will also make use of the mean quadratic difference between the pignistic probabilities $\texttt{Pbet}(\omega_i)$ and the class posterior probabilities $p(\omega_i|x)$:

$$\mathbf{Q} = \int \sum_i (\texttt{Pbet}(\omega_i|x) - p(\omega_i|x))^2 p(x)dx, \tag{12}$$

The mean classification cost and the error rate are also computed for the Bayes classifier, whose optimal solution provides a baseline to compare results with and without bagging. Its performances also characterize the intrinsic

difficulty of the task. For instance, the minimal rejection rates to achieve classification error rates of 10% and 5% are here respectively 23% (for $\lambda_0 = 0.34$) and 42% (for $\lambda_0 = 0.19$).

Results are reported according to two decision strategies. In the first one, the rejection cost λ_0 is fixed and may correspond, e.g., to the cost of the information needed to resolve the ambiguity. The performance is then measured by the mean classification cost $\mathbf{C}$. In the second one, a given classification error rate is required. Practically, the tuning parameter is still λ_0, but the interesting quantity is now the minimum rejection rate required to meet the criterion.

Implementation The method proposed by Denœux bears some resemblance with the Parzen method when the neighborhood is extended to the whole training set ($K = N$), because the influence of a neighboring vector decreases with its distance to the query point. Setting $K = 8$ achieves a near-asymptotic behavior while limiting the computational expense.

The influence of training patterns depends on parameters α and γ (see Eq. 5). As the influence of α on the classification is low [3], and in order to reduce the complexity of the analysis, it was set to the "standard" 0.95 value. Regarding γ, we will proceed here in two steps. First, all γ_q are fixed (Section 5); they are set to the same value (0.5) since the three classes have the same shape and the same number of items. Then, different learning strategies are tested in Section 6.

Finally, the average structure $m_{\mathbf{B}}$ estimates the expected structure over training sets. The expectation over training samples is ideally estimated by the expectation over bootstrap samples. Hence, the number B of bootstrap samples should tend towards infinity. In fact, the effect of bagging is quite visible for values as low as $B = 10$. We used $B = 50$, as the small improvement achieved by higher values is not worth the computation cost. Note that Breiman recommends values around 25.

5 Results without Learning

In this section, the results with and without bagging will be compared from two viewpoints: first, the quality of the decisions (measured by the mean classification cost or by the rejection rate needed to achieve a given error rate), and then the closeness of the pignistic probabilities to the class posterior probabilities.

5.1 Decision Level

The first plot of figure 2 shows mean classification costs vs. rejection costs for the 15 experiments. The horizontal segments in boxplots represent the lower quartile, median, and upper quartile results. Minimal and maximal values

are indicated by the whiskers, and the plotted curve itself is the average over experiments. Bagging clearly improves classification for low classification costs $\lambda_0 < 0.3$, *i.e.* higher rejection rates. Its cost is half-way between the original algorithm and the Bayes classifier. However, this benefit vanishes for high values of λ_0 (low rejection rates). The improvement due to bagging is thus linked to its higher capacity to reject truly ambiguous patterns. but the pignistic probabilities values may be significantly modified so that rejection is more common.

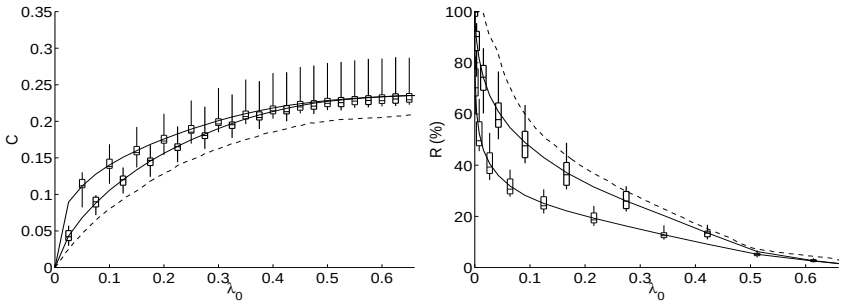

Fig. 2. Mean classification cost (left) and rejection rate (right) as a function of rejection cost for classic (thin line) and bagged (bold line) methods. The dotted line corresponds to the Bayes classifier

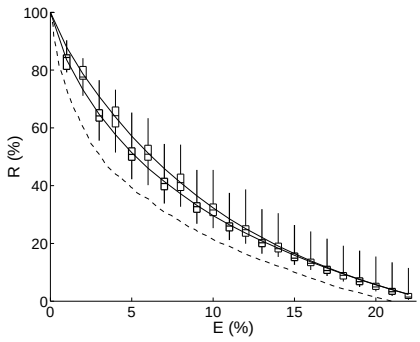

Fig. 3. Rejection rate **R** as a function of classification error rate **E** for original (thin line) and bagged (bold line) methods. The dotted line corresponds to the Bayes classifier

The second plot of Fig. 2 shows that the rejection rate of the bagged rule is much higher than that of the original rule. This means that bagging increases

the uncertainty attached to the classification, that is, the uncertainty about conclusions. Rejection rate with bagging is also much closer to the Bayes classifier rate (for low λ_0 values), but however still lower. Bagging thus weakens the tendency of the original algorithm to over-estimate the confidence in classification.

These observations are confirmed when looking at the fixed error rate strategy in Fig. 3, where classification is visibly improved for the most stringent error rates requirements. Table 1 displays the mean rejection rates obtained with and without bagging for several error rates. The improvements are higher than what the boxplot in Fig. 3 suggests, because the boxplots show the global variability in performance for different training sets. The individual comparisons for each training set are summarized in Table 1. In this table, the second column (S-0-F) reports the number of trials for which the bagged version performed significantly better or significantly worse (first and third figure, respectively) than the non-bagged version at the 5 % significance level, according to the exact McNemar test for matched samples (see, for instance, [8]). The middle figure is the remaining number of cases, for which differences between the two methods were not significant. These results show that bagging never performs worse than the original classifier for error rates below 15%, and that it significantly improves the mean results for all error rates below 20%.

Table 1. Mean Rejection rates **R** (in %) for some given target error rates **E** (in %). Column S–0–F reports the number of significant success and failures (at the 5% level) of bagging for each training set. The p-values reporting the smallest level for which mean rejection rates differ significantly are all below 0.02%

E	S–0-F	Original	Bagged
2.5	12–3–0	74.2	68.9
5.0	15–0–0	57.0	51.5
10.0	15–0–0	32.3	29.6
15.0	7–8–0	16.4	15.9
20.0	6–6–3	5.8	5.6

For a classification task with a small number of classes, taking into account the uncertainty due to the finite size of the training sample hardly modifies the rank of the highest pignistic probability. Its value is however duly lowered, which is interpreted as a more uncertain outcome. Bagging is thus beneficial when the values attached to belief assignments are of interest. Besides rejection, all applications where a measure of uncertainty should be attached to the decision are concerned.

5.2 Pignistic Level

While it may be possible to display the effect of bagging at the credal level, there is no satisfactory criterion for measuring the relevance of a belief structure. We thus resort to the study of pignistic probabilities which give more information on beliefs than the decisions themselves. Results of the previous sections provided hints suggesting that, with bagging, the pignistic probabilities `Pbet` should be closer to the posterior probabilities p. Indeed, the mean quadratic errors $\mathbf{Q}$ on posterior class probabilities (Eq. 12) are about 40% lower on the whole space when bagging is applied to the K-NN rule.

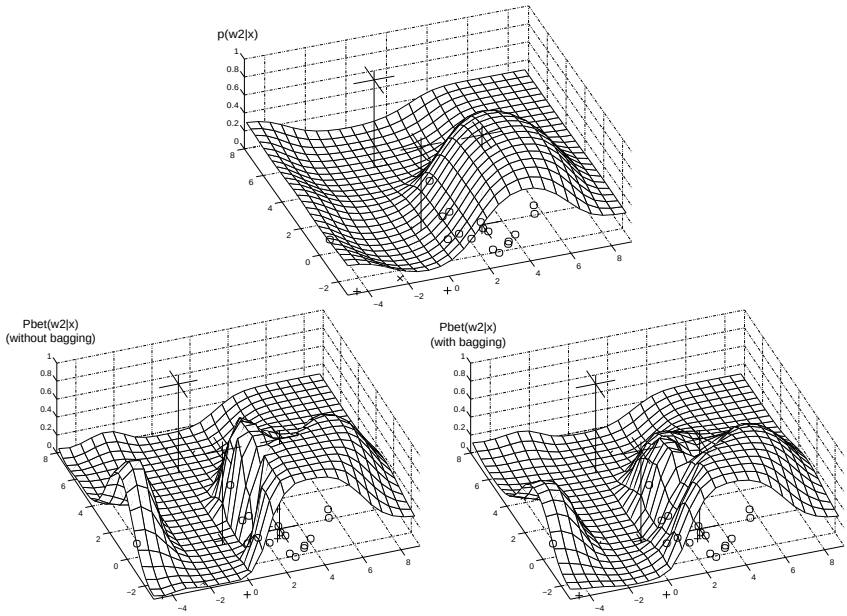

Fig. 4. Posterior probability $p(\omega_2|x)$ (top) and pignistic probability $\texttt{Pbet}(\omega_2|x)$ without bagging (bottom left) and with bagging (bottom right). Vertical lines locate the centers of the three class distributions

We can easily illustrate situations where bagging is the most beneficial by plotting the probability surfaces, An example is given in Fig. 4, which shows that the main improvements occur at class boundaries and for outliers (one is situated in the lower-left corner of the graph). Bagging thus yields a better representation of uncertainties, stemming either from ambiguity (where classes overlap) or from lack of information (in regions of low density of training patterns).

The correction of these two types of uncertainties does not have the same impact on the estimation of posterior class probabilities, and on the mean

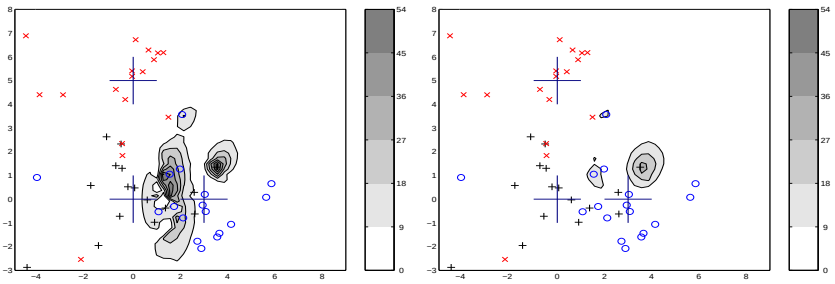

Fig. 5. Contours of the quadratic error on $p(\omega_2|x)$, weighted by the mixture density $p(x)$ ($\times 10^{-4}$). left: without bagging; right: with bagging.

classification rate (decision level). This is illustrated in Fig. 5, which gives the quadratic error contours weighted by the mixture density $p(x)$. For example, as the outlier is situated in a region of low density, the weakening of its influence by the bagging procedure results in a negligible contribution to the mean error rate improvement. However, this effect could be much more noticeable with other misclassification costs : in a medical diagnosis application, for example, an outlier in the "healthy" class can cause an absence of illness detection.

6 Remarks about Learning

In the previous sections, the parameters α and γ of the basic belief assignment were set at arbitrary values. The effect of bagging regarding uncertainty due to the finite sample size was thus isolated. This section depicts the effect of bagging regarding the uncertainty pertaining to the learning of parameters. Here, α is kept at 0.95 as it was shown to have only marginal influence on the classification results [3,15].

As explained in Section 2.2, the influence regions of training patterns are controlled by γ (Eq. 5). Fig. 6 shows the mean classification cost as a function of γ for the original classifier and its bagged version.

The bagged K-NN mean classification cost according to γ is always lower than that of the original algorithm for any given rejection cost. Thus, the results presented in the previous sections are representative of what would be obtained for any value of γ. The comparison of the two plots in Fig. 6 also shows that the differences between the two methods are larger for small rejection costs, regardless of γ.

Bagging is more effective in improving the original method for small values of γ, i.e., when all neighbors have the same influence, regardless of their distance to the query sample. In this case, the resulting belief is too confident, and bagging neatly corrects it.

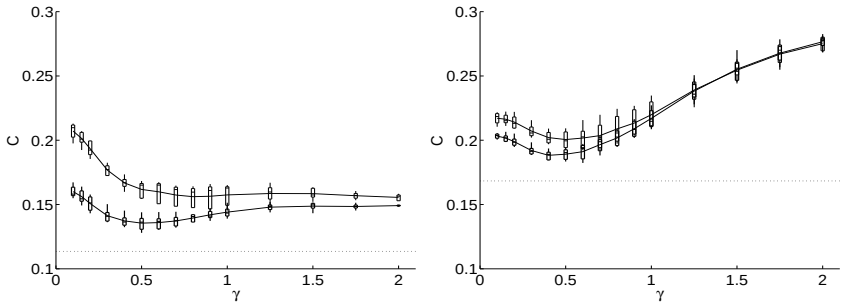

Fig. 6. Mean classification cost $\mathbf{C}$ as a function of γ for $\lambda_0 = 0.15$ (top) and $\lambda_0 = 0.3$ (bottom). The dotted line represents the Bayes classification cost, thin lines and bold lines represent respectively classic and bagged K-NN classifications

In comparing the two graphs, it may be noted that, for the bagged algorithm, the optimal γ value is identical for both rejection costs, while it depends on λ_0 for the standard algorithm. Indeed, these two values should ideally not interact, as beliefs should not be affected by the consequences of actions. These consequences should only be taken into account in the decision process.

Finally, the lower variability of $\mathbf{C}$ provides a steadier optimal γ value and a lower sensitivity to errors in γ, in terms of misclassification cost.

7 Conclusion

Standard classifiers are sensitive to ambiguous training items such as mislabeled patterns or outliers. Regarding this point, the evidential K-NN rule improves upon the original probabilistic rule, as the certainty expressed by training patterns can be limited to weaken the influence of ambiguous items. In this paper, we show that bagging the belief structure construction process further improves this robustness.

Classification error is shown to be significantly reduced for high to intermediate rejection rates, and is always observed to be lower than that of the non-bagged K-NN rule. Pignistic probabilities are much closer to posterior probabilities, which in turns supports the idea that bagging defines more relevant belief structures.

Beyond the evidential K-NN, this paper illustrates the necessity to build generic tools for inferring beliefs. It is probably the first attempt to take into account the uncertainty due to the presence/absence of an information source upon which beliefs are constructed. In the classical pattern recognition paradigm, where information sources are points assumed to be sampled from some fixed distribution, resample and combine techniques provide a fully automatic means to correct undue certainty in inferred beliefs.

Work in progress shows that the gain is more important for classifiers that make a more intensive use of data (with more learning parameters). More sophisticated inference methods such as decision trees or fuzzy K-means should thus also be improved. Investigations could be done on other operators to combine the belief structures in the bagging procedure in order to further improve the quality of belief representation at the credal level.

References

1. A. Appriou. Uncertain data aggregation in classification and tracking processes. In B. Bouchon-Meunier, editor, *Aggregation and Fusion of imperfect information*, pages 231–260. Physica-Verlag, Heidelberg, 1998.
2. L. Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996.
3. T. Denœux. A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Trans. on Systems, Man and Cybernetics*, 25(05):804–813, 1995.
4. T. Denœux. Analysis of evidence-theoretic decision rules for pattern classification. *Pattern Recognition*, 30(7):1095–1107, 1997.
5. T. Denœux. Application du modèle des croyances transférables en reconnaissance de formes. *Traitement du Signal*, 14(5):443–451, 1998.
6. T.G. Dietterich. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization. *Machine Learning*, 40(2):1–19, 2000.
7. B. Efron and R. Tibshirani. *An introduction to the bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1993.
8. B. D. Ripley. *Pattern Recognition and Neural networks*. Cambridge University Press, Cambridge, 1996.
9. G. Rogova. Combining the results of several neural network classifiers. *Neural Networks*, 7(5):777–781, 1994.
10. G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., 1976.
11. P. Smets. The combination of evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):447–458, 1990.
12. P. Smets. The application of the Transferable Belief Model to diagnosis problems. *International Journal of Intelligent Systems*, 13:127–158, 1998.
13. P. Smets. The Transferable Belief Model for quantified belief representation. In D. M. Gabbay and P. Smets, editors, *Handbook of Defeasible reasoning and uncertainty management systems*, volume 1, pages 267–301. Kluwer Academic Publishers, Dordrecht, 1998.
14. P. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66:191–243, 1994.
15. L. M. Zouhal and T. Denœux. An evidence-theoretic k-NN rule with parameter optimization. *IEEE Trans. on Systems, Man and Cybernetics C*, 28(2):263–271, 1998.