Comparison of identity fusion algorithms using estimations of confusion matrices

G. Golino, A. Graziano, A. Farina, W. Mellano, F. Ciaramaglia
Selex ES
Rome, Italy
giovanni.golino@selex-es.com; antonio.graziano@selex-es.com

***Abstract*— Scope of this paper is to investigate the performances of different identity declaration fusion algorithms in terms of probability of correct classification, supposing that the information for combination of the inferences from the different classifier is affected by measurement errors. In particular, these information have been assumed to be provided in the form of confusion matrices. Six identity fusion algorithms from literature with different complexity have been included in the comparison: heuristic methods such as voting and Borda Count, Bayes' and Dempster-Shafer's methods and the Proportional Redistribution Rule n° 1 in the Dempster-Shafer's framework.**

***Keywords*—target classification, identity fusion, confusion matrix.**

I. Introduction

In a multi-sensor system the target classification performance can be improved by suitably combining the inferences generated by the autonomous classifiers of the single sensors (identity declaration fusion [1]). For this purpose it is desirable to use the available information about the classification performances of the single sensors. The confusion matrix, whose elements correspond to the likelihood of the different involved classes, is a compact and detailed way of representing the classification performance, from which the Probability of correct classification (Pcc) and the probability relative to the various misclassification errors can be derived. In particular the elements of the confusion matrix can be used to maximize the a-posterior Pcc according to Bayes' theory. In this case, if the numerical values of the confusion matrix were errorless, the performance of the identity fusion would be optimal. However, in practice, these values are estimated and affected by errors. In these conditions, the Bayes' rule does not always produce best results. In particular, in presence of strong-conflicting inferences and estimation errors, the application of Bayes' rule can be not effective. It can be better to apply simpler combination rules as some heuristic methods that are more robust to errors.

Dempster-Shafer's theory has been presented as a generalization of Bayes' theory in [2]. A recent work has disputed this claim, limiting its correctness to the case of uniform a-prior probabilities [3]. The problem of the Dempster-Shafer rule (and of Bayes' rule) in presence of conflicting inferences has been pointed out by the well-known Zadeh's paradox [4]. In this paper the performances of different algorithms that use estimated confusion matrices affected by errors to combine the inferences from the single classifiers are investigated. In particular the heuristic methods based on voting and on ranking (Borda count) are compared with the methods using Bayes' and Dempster-Shafer's rules. Moreover the effectiveness of the redistribution of the conflicting masses preserving the Dempster-Shafer framework, like Proportional Redistribution Rules (PCR) is evaluated.

The paper is organized as follows:

- in section II the identity fusion algorithms considered in this paper are briefly described;
- in section III, four simple but representative identity fusion problems are introduced as study cases and the corresponding results using different mean values of the estimation errors of the confusion matrices are reported and commented;
- section IV gives the conclusions.

II. Identity fusion algorithms

The algorithms for the identity fusion considered in this paper are:

- Majority Voting (MV),
- Weighted Voting (WV),
- Borda count,
- Bayes' rule,
- Dempster-Shafer's (D-S) rule with the following basic belief mass assignment: "q-least commitment",
- Proportional Redistribution Rule n°1 (PCR1) with the following basic belief mass assignment: "q-least commitment".

A brief description of the fusion algorithms follows.

Majority voting [5] is the simplest method for the combination of inferences: each inferred class corresponds to a single vote and the selected class after fusion is the most voted class: all the inferences matter the same. In the modified weighted version, the different votes are weighted by the estimated Pcc of the voter/classifier.

Voting methods use only the top choice of each classifier, but secondary choices often contain near misses that should not be overlooked. The Borda count [5] is a method in which classes are ranked in order of preference; it gives each class a certain number of points corresponding to the position in which it is ranked by each classifier. The class with the highest scoring is then selected after fusion.

Bayes' theorem [1,4] links the degree of belief in a proposition before and after accounting for evidence (a-priori and a-posteriori probabilities). The a-posteriori probability of a combination of two or more evidences is obtained by the multiplication of the likelihoods of the single evidences (the independence of the evidences is assumed).

Dempster–Shafer theory [1,4 and 5] allows to specify a degree of ignorance instead of being forced to supply probabilities that add to unity. In this formalism a degree of belief (also referred to as a Basic Belief Mass - BBM) is used rather than a Bayesian probability distribution. BBM values are assigned to sets of possibilities (union of one or more classes) rather than to a single class, probability is instead represented by intervals that are lower-bounded by the value "belief" (or "support") and upper-bounded by the value "plausibility". BBM values from different sources can be combined with Dempster-Shafer's rule of combination, assuming independent belief sources. There are more than one possible assignment for transforming probabilities into BBMs [7,8 and 9]. The "q-least commitment" basic belief mass assignment (that corresponds to the maximum compatible degree of ignorance) has been considered in this paper to transform the CM (Confusion Matrix) likelihoods and the a-priori probabilities into BBM values.

Proportional Redistribution Rules (PCR) is a family of fusion rules for the combination of uncertain information allowing to deal with highly conflicting sources. The PCR rules can be used as alternatives to the Dempster-Shafer's combination rule. Six PCR rules (PCR1-PCR6) have been defined [10,11 and 12]: from PCR1 up to PCR6 one increases in one hand the complexity of the rules, but in other hand one improves the accuracy of the redistribution of conflicting masses. The basic common principle of PCR rules is to redistribute the conflicting mass proportionally with some functions depending on the sum of the masses assigned by the single inferences. PCR1 is the least accurate combination rule of the PCR family, but it is the simplest to implement and it has been considered in this paper. PCR2-6 implementations are significantly more complex because the conflicting mass is redistributed only to the non-empty set that are involved in the conflict (extra computer memory is needed to keep track of the conflicting hypotheses and extra computation load is needed for combining them). A particular interesting action point for further investigation would be testing the most efficient PCR rule (PCR6) [12].

A. Combination rules

In this section, the rules of Bayes, Dempster-Shafer and PCR1 for the combination of two classifiers are briefly recalled. For further details and the generalization of the rules with more than two classifiers, see references [1], [4], [10] and [11]. Voting and Borda count combinations are not considered here because they consists simply in the sums of respectively the votes and the ranks.

Let consider a set Ω of possible exhaustive and mutually exclusive classes C_k, with *N* being the cardinality of this set:

$$\Omega = \{C_1, C_2, \cdots, C_N\} \qquad (1)$$

Let suppose that the independent classifiers 1 and 2 infer respectively the classes C_i and C_j; the a-posterior probability $P_{12}(C_k / A \cap B)$ of inferring the class C_k resulting from Bayes' rule of combination is:

$$P_{12}(C_k / C_i \cap C_j) = \frac{P_1(C_i/C_k) \cdot P_2(C_j/C_k) \cdot P_0(C_k)}{\sum_{h=1}^{N} P_1(C_i/C_h) \cdot P_2(C_j/C_h) \cdot P_0(C_h)} \qquad (2)$$

where:

- $P_0(\cdot)$ is the a-prior probabilities (without any information obtained by previous classifications) of the considered class;
- $P_1(C_y/C_x)$, $P_2(C_y/C_x)$ are the probabilities that classifiers 1 and 2 infer the class C_y assuming that the true class is C_x (likelihoods).

Let consider the power set 2^{Ω} of Ω as the set whose elements are all the possible subsets of Ω:

$$2^{\Omega} = \{F_k : F_k \subseteq \Omega\} = \{\emptyset, C_1, C_2, \cdots, C_N, C_1 \cap C_2, \cdots, C_{N-1} \cap C_N, \cdots, \Omega\} \qquad (3)$$

where $\emptyset$ is the empty set. The cardinality of 2^{Ω} is 2^N.

Let suppose that the independent classifiers 1 and 2 assign respectively BBMs $m_1(\cdot)$ and $m_2(\cdot)$ to the elements included in the power set 2^{Ω}; the combination BBM $m_{12}(F_k)$ of F_k resulting from Dempster-Shafer's rule of combination is:

$$m_{12}(F_k) = \sum_{i,j / F_k = F_i \cap F_j} \frac{m_1(F_i) \cdot m_2(F_j)}{1 - m_c} \qquad (4)$$

where m_c is the global conflicting mass, defined as follow

$$m_c = \sum_{p,q / F_p \cap F_q = \emptyset} m_1(F_p) \cdot m_2(F_q) \qquad (5)$$

In the case of the PCR1 rule the combination BBM $m_{12}(F_k)$ of F_k is instead:

$$m_{12}(F_k) = \sum_{i,j / F_i \cap F_j = F_k \neq \emptyset} m_1(F_i) \cdot m_2(F_j) + \frac{m_1(F_k) + m_2(F_k)}{\sum_{h=1}^{2^N} (m_1(F_h) + m_2(F_h))} \cdot m_c \qquad (6)$$

III. STUDY CASES

Four simple but representative study cases (three different classifiers for a three classes problem) have been investigated, as follows:

- complementary confusion matrices,
- supplementary confusion matrices,
- complementary conflicting confusion matrices,
- supplementary conflicting confusion matrices.

By complementary CMs it has been meant that the single classifiers show a complementary expertise in the recognition of the different classes. By supplementary CMs the different classifiers show similar behaviors. By conflicting CMs a possible overestimation of the performance of the single classifiers can make harder an effective combination of the contradictory inferences from different classifiers when they occur. A quantitative definition of complementary and supplementary CM can be found in [13].

In the following sub-sections, the estimated confusion matrices that have been selected for the four study cases are reported. The columns of the matrices represent the true classes, while the rows correspond to the inferred classes, so the element (k,h) of a matrix is an estimation of the probability of declaring k^{th} class when the true class is the h^{th} one:

$$\hat{M}_i(k,h) = \hat{P}_i(D \equiv k / T \equiv h) \qquad (7)$$

The a-prior probabilities of the different classes are assumed equal. A block diagram of the fusion system is shown in fig. 1.

The performances of the six algorithms in correspondence of the identity fusion of six inferences (two independent inferences for each classifier) have been considered. The performances of the different algorithms have been computed with 1000 Monte Carlo trials, each generating independent samples of the true confusion matrices and a-prior probabilities.

The results of the Monte Carlo trials are represented by the curves corresponding to Empirical Cumulative Distribution Function (ECDF) versus the Pcc. The x-axis values of Pcc have been computed exactly, that is the contribution of all the possible permutations of the single sensor inferences has been considered.

A. *Complementary confusion matrices*

The following confusion matrices corresponding to three different classifiers have been considered:

$$\hat{M}_1 = \begin{bmatrix} 0.90 & 0.30 & 0.30 \\ 0.10 & 0.40 & 0.30 \\ 0 & 0.30 & 0.40 \end{bmatrix}$$
$$\hat{M}_2 = \begin{bmatrix} 0.40 & 0.30 & 0.10 \\ 0.30 & 0.40 & 0 \\ 0.30 & 0.30 & 0.90 \end{bmatrix} \qquad (8)$$
$$\hat{M}_3 = \begin{bmatrix} 0.40 & 0 & 0.30 \\ 0.30 & 0.90 & 0.30 \\ 0.30 & 0.10 & 0.40 \end{bmatrix}$$

The performance is dependent on the true target class (class 1, 2 or 3):

- the first classifier identifies correctly targets belonging to class 1 (on average it makes only one mistake in ten of its inferences), while it almost randomly infers in correspondence of targets belonging to class 2 or class 3,
- the second classifier identifies correctly targets belonging to class 3 (on average it makes only one mistake in ten of its inferences), while it almost randomly infers in correspondence of targets belonging to class 1 or class 2,
- the third classifier identifies correctly targets belonging to class 2 (on average it makes only one mistake in ten of its inferences), while it almost randomly infers in correspondence of targets belonging to class 1 or class 2.

The performances of the six different algorithms are reported in the fig. 2 and 3 in correspondence of an estimation of the confusion matrix by using respectively 30 and 10 independent samples for each class. The performances of the single classifiers correspond to the dotted curves (indicated as C1, C2 and C3 in the legends of the figures). Bayes' rule, Dempster-Shafer's rule, Borda count and PCR1 give similar results, the performance of PCR1 is barely the best. The voting algorithms present significantly worse performance.

B. Supplementary confusion matrices

The following confusion matrices corresponding to three different classifiers have been considered:

$$\hat{M}_1 = \begin{bmatrix} 0.70 & 0.10 & 0.20 \\ 0.20 & 0.70 & 0.10 \\ 0.10 & 0.20 & 0.70 \end{bmatrix}$$
$$\hat{M}_2 = \begin{bmatrix} 0.80 & 0.10 & 0.10 \\ 0.10 & 0.80 & 0.10 \\ 0.10 & 0.10 & 0.80 \end{bmatrix} \quad (9)$$
$$\hat{M}_3 = \begin{bmatrix} 0.60 & 0.20 & 0.20 \\ 0.20 & 0.60 & 0.20 \\ 0.20 & 0.20 & 0.60 \end{bmatrix}$$

In the second example, three classifiers with supplementary confusion matrices have been selected. A single classifier can recognize all the three classes with the same accuracy, but the accuracy differs from classifier to classifier:

- the first classifier has an estimated probability of correct classification equal to 70%,
- the second classifier has an estimated probability of correct classification equal to 80%,
- the third classifier has an estimated probability of correct classification equal to 60%.

The performances of the six different algorithms are reported in the fig. 4 and 5 in correspondence of an estimation of the confusion matrix by using respectively 30 and 10 independent samples for each class. The performances of the single classifiers correspond to the dotted curves (indicated as C1, C2 and C3 in the legends of the figures). All the algorithms give comparable performance. PCR1 and Bayes' rule performance are exactly the same and they are slightly better than the others, weighted voting performs better than Borda count and majority voting.

C. Complementary conflicting confusion matrices

The following confusion matrices corresponding to three different classifiers have been considered:

$$\hat{M}_1 = \begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 0.00 & 0.60 & 0.40 \\ 0.00 & 0.40 & 0.60 \end{bmatrix}$$
$$\hat{M}_2 = \begin{bmatrix} 0.60 & 0.00 & 0.40 \\ 0.00 & 1.00 & 0.00 \\ 0.40 & 0.00 & 0.60 \end{bmatrix} \quad (10)$$
$$\hat{M}_3 = \begin{bmatrix} 0.60 & 0.40 & 0.00 \\ 0.40 & 0.60 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$$

In this third example, the three classifiers can be affected by conflicting inferences. As consequence, the application of Bayes' rule to fusion leads to severe performance degradation with respect to heuristic methods. The problem arises from an overestimation of the performance of the single classifiers.

The performances of the six different algorithms are reported in the fig. 6 and 7 in correspondence of an estimation of the confusion matrix by using respectively 30 and 10 independent samples for each class. The performances of the single classifiers correspond to the dotted curves (indicated as C1, C2 and C3 in the legends of the figures). PCR1 gives the best result that is slight better than Borda count. Majority and weighted voting have coincident performance that are significantly worse than the one of PCR1. The performance of Bayes' rule and Dempster-Shafer rule are perfectly coincident and worse than all the others because of the presence of conflicting inferences.

D. Supplementary conflicting confusion matrices

The following confusion matrices corresponding to three different classifiers have been considered:

$$\hat{M}_1 = \hat{M}_2 = \hat{M}_3 = \begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix} \qquad (11)$$

In the forth example, three classifiers with identity confusion matrices as estimations have been selected: according to these estimations the single classifier is never wrong. If the classifiers disagree on the inferred class, the Bayes' rule of fusion leads to severe performance degradation with respect to heuristic methods.

The performances of the six different algorithms are reported in the fig. 8 and 9 in correspondence of an estimation of the confusion matrix by using respectively 30 and 10 independent samples for each class. The performance of the single classifiers correspond to the dotted curves (indicated as C1, C2 and C3 in the legends of the figures). The performances of the voting algorithms, Borda count and PCR1 are perfectly coincident and near to 100%. The performances of Bayes' rule and Dempster-Shafer rule are perfectly coincident and much worse than all the others because of the presence of conflicting inferences, even much worse than the performance of the single classifiers.

E. Summary results

In table I the average (over the 1000 Monte Carlo trials) Pcc is reported for all the investigated study cases. It has been reported also an intermediate case where 60 samples (20 samples per class) for the estimation of each confusion matrix have been considered. It can be noted than PCR1 always brings the highest Pcc of all the six considered combination rules.

IV. CONCLUSIONS

Simulation results show that in the considered study cases, the algorithm using the PCR1 rule of combination brings the best performance of all the six considered alternatives and definitely overcomes Bayes' and D-S's rules in the cases where the probability of conflicts between the inferences is high. This performance difference increases with the decrease of the number of samples used for the estimation of the confusion matrices. This behavior is a consequence of the poor performance of the latter two combination methods in presence of conflicting inferences from the different classifier, as claimed by the Zadeh's paradox. In two investigated study cases with conflicting confusion matrices Bayes' and Dempster-Shafer rules perform even worse than heuristic approaches.

In the cases where the conflict is less likely probable the performance of the PCR1 is comparable with the ones of Bayes' and D-S's rules (the same or slightly better).

The implementation of PCR1 slightly increases the computational complexity of D-S's rule. Future work may be addressed to the comparison of the performance resulting by the application of more complex PCR rules to the inferences of classifiers whose accuracies are represented by conflicting confusion matrices.

REFERENCES

[1] D. L. Hall, S. A. H. McMullen, "Mathematical Techniques in Multisensor Data Fusion", 2nd edition, Artech House, 2004.

[2] A. Dempster, "A generalization of bayesian inference", Journal of the Royal Statistical Society, series B, vol. 30, n° 2, pp. 205-247, 1968.

[3] J. Dezert , A. Tchamova, D. Han, J.M: Tacnet, "Why Dempster's fusion rule is not a generalization of Bayes fusion rule", Proc. of Fusion 2013 - 16th Int. Conference on Information Fusion, Istanbul, Turkey, July 9-12, 2013.

[4] A. Benavoli, L. Chisci, B. Ristic, A. Graziano, A. Farina, "Reasoning under uncertainty: from Bayes to Valuation Based Systems. Application to target classification and threat evaluation", Selex Sistemi Integrati, Università degli Studi di Firenze, September 2007.

[5] J. Llinas, "Fusion-based methods for target identification in the absence of quantitative classifier confidence", SPIE Signal Processing, Sensor Fusion and Target Recognition VI Conference, Orlando, FL, April 1997.

[6] A. Benavoli, B. Ristic, A. Farina, M. Oxenham, and L. Chisci, "An Application of Evidential Networks to Threat Assessment". IEEE Trans. of Aerospace and Electronic Systems, vol. 45, n° 2, pp. 620-639, April 2009.

[7] P. Smets, "Quantified epistemic possibility theory seen as a hyper cautious transferable belief model", Rencontres Francophones sur la Logique Floue et ses Applications, pages 343-353, Cepadues Editions, October 2000.

[8] P. Smets, "The transferable belief model for quantified belief representation", Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 1, pp. 267-301, Kluwer, Dordrecht, The Netherlands, 1998.

[9] B. Ristic, P. Smets, "Target classification approach based on belief function theory", IEEE Trans. of Aerospace and Electronic Systems, vol. 41, n° 2, pp. 574-583, April 2005.

[10] J. Dezert, F. Smarandache, "Information fusion based on new proportional conflict redistribution rules" , Proc. of Fusion 2005 - 8th Int. Conference on Information Fusion, Philadelphia, PA, USA, June 27 - July 1, 2005.

[11] J. Dezert, F. Smarandache (editors), "Advances and Applications of DSmT for Information Fusion (Collected works), second volume", American Research Press, 2006.

[12] J. Dezert, F. Smarandache, "On the consistency of PCR6 with the averaging rule and its application to probability estimation", Proc. of Fusion 2013 - 16th Int. Conference on Information Fusion, Istanbul, Turkey, July 9-12, 2013.

[13] S. R. M. Prasanna, B. Yegnanarayana, J. P. Pinto, H. Hermansky, "Analysis of Confusion Matrix to Combine Evidence for Phoneme Recognition", Idiap Resarch Institute, Switzerland, 2007.

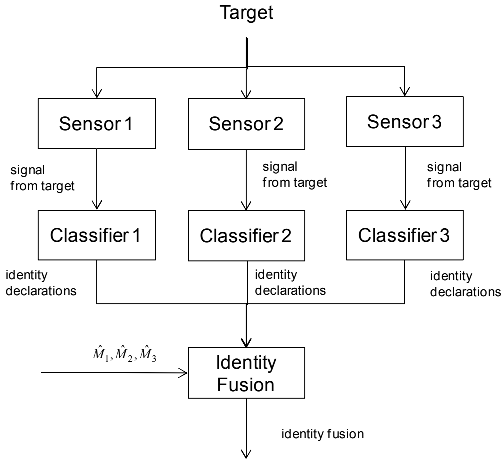

Fig. 1. Block diagram of the fusion system.

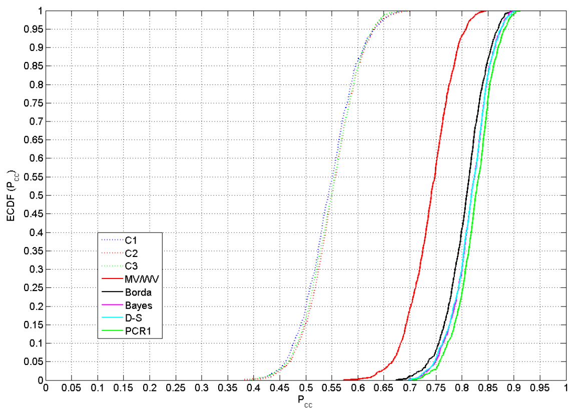

Fig. 2. ECDF versus Pcc with complementary confusion matrices (estimation from 30 samples per class).

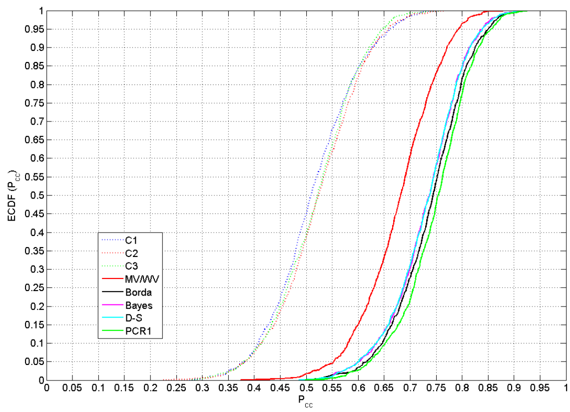

Fig. 3. ECDF versus Pcc with complementary confusion matrices (estimation from 10 samples per class).

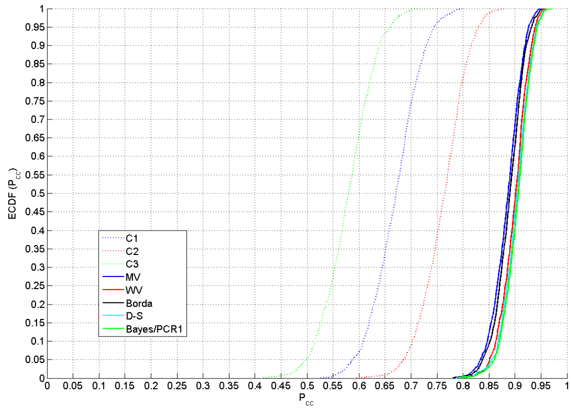

Fig. 4. ECDF versus Pcc with supplementary confusion matrices (estimation from 30 samples per class).

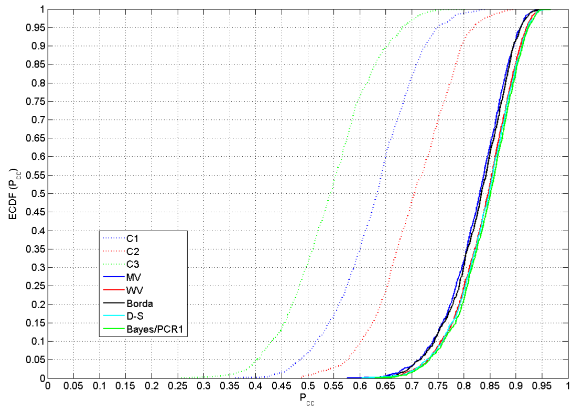

Fig. 5. ECDF versus Pcc with supplementary confusion matrices (estimation from 10 samples per class).

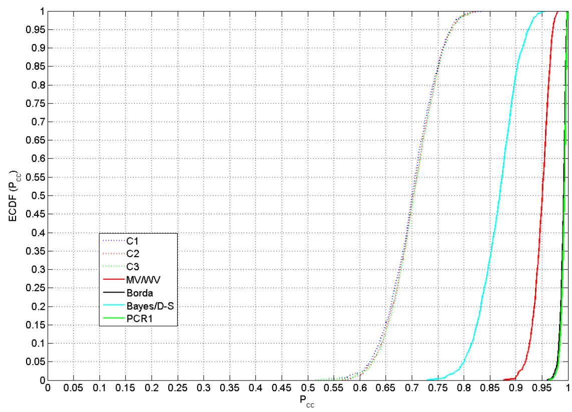

Fig. 6. ECDF versus Pcc with complementary conflicting confusion matrices (estimation from 30 samples per class).

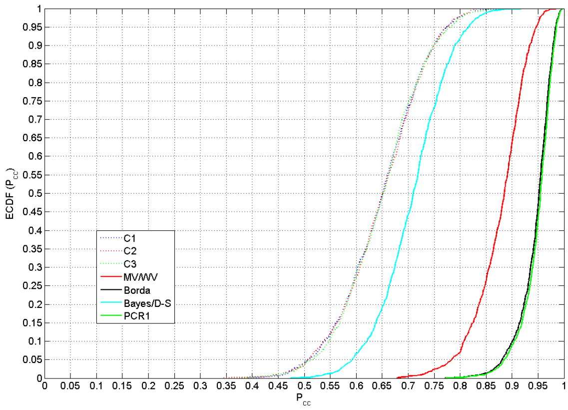

Fig. 7. ECDF versus Pcc with complementary conflicting confusion matrices (estimation from 10 samples per class).

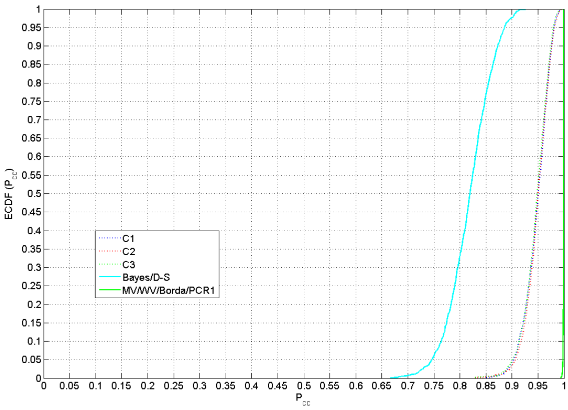

Fig. 8. ECDF of Pcc with supplementary conflicting confusion matrixes (estimation from 30 samples per class).

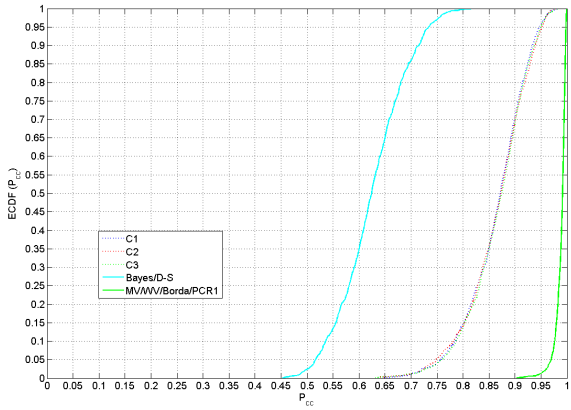

Fig. 9. ECDF of Pcc with supplementary conflicting confusion matrixes (estimation from 10 samples per class).

TABLE I. SUMMARY RESULTS (N IS THE TOTAL NUMBER OF SAMPLES).

Study cases	PROBABILITY OF CORRECT CLASSIFICATION (mean value, %)																	
	MAJORITY VOTING			WEIGHTED VOTING			BORDA COUNT			BAYES			DEMPSTER SHAFER			PCR1		
	N=30	N=60	N=90	N=30	N=60	N=90	N=30	N=60	N=90	N=30	N=60	N=90	N=30	N=60	N=90	N=30	N=60	N=90
Compl. CMs	67.8	72.4	73.9	67.8	72.4	73.9	73.9	78.9	80.7	73.1	79.3	81.6	73.0	79.4	81.6	74.8	80.5	82.4
Supp. CMs	82.3	86.9	88.4	83.9	88.4	89.9	82.7	87.4	88.8	84.3	88.9	90.4	84.0	88.7	90.2	84.3	88.9	90.4
Compl. confl. CMs	87.7	92.9	94.8	87.7	92.9	94.8	94.5	98.1	99.0	70.7	81.5	86.5	70.7	81.5	86.5	94.8	98.3	99.2
Supp. confl. CMs	98.8	99.8	99.9	98.8	99.8	99.9	98.8	99.8	99.9	62.5	75.5	81.8	62.5	75.5	81.8	98.8	99.8	99.9