

Classifier Fusion using Triangular Norms

Piero Bonissone, Kai Goebel, and Weizhong Yan

GE Global Research, One Research Circle, Niskayuna, NY 12309, USA
{bonissone, goebelk, yan}@research.ge.com

Abstract. This paper describes a method for fusing a collection of classifiers where the fusion can compensate for some positive correlation among the classifiers. Specifically, it does not require the assumption of evidential independence of the classifiers to be fused (such as Dempster Shafer's fusion rule). The proposed method is associative, which allows fusing three or more classifiers irrespective of the order. The fusion is accomplished using a generalized intersection operator (T-norm) that better represents the possible correlation between the classifiers. In addition, a confidence measure is produced that takes advantage of the consensus and conflict between classifiers.

1 Introduction

Design of a successful classifier fusion system consists of two important parts: design of the individual classifiers, selection of a set of classifiers [9, 17], and design of the classifier fusion mechanism [14]. Key to effective classifier fusion is the diversity of the individual classifiers. Strategies for boosting diversity include: 1) using different types of classifiers; 2) training individual classifiers with different data set (bagging and boosting); and 3) using different subsets of features.

In the literature we can find many fusion methods derived from different underlying frameworks, ranging from Bayesian probability theory [18], to fuzzy sets [5], Dempster-Shafer evidence theory [1], group decision-making theory (majority voting [10], weighted majority voting, and Borda count [7]). Existing fusion methods typically do not address correlation among classifiers' output errors. For example, methods based on Dempster-Shafer theory assume evidential independence [16, p. 147].

In real-world applications, however, most individual classifiers exhibit correlated errors due to common data sources, domain knowledge based filtering, etc. Typically, they all tend to make the same classification errors on the most difficult patterns. Prior research has focused on relating the correlation in the classifiers' errors to the ensemble's error [17], quantifying the degree of the correlation [8, 13], and minimizing the correlation degree between the classifiers [11]. However, not much attention has been devoted to investigating fusion mechanisms that can compensate for the impact of the classifiers' error correlations.

This paper investigates the impact of parameter changes within a T-norm based framework for a set of n classifiers. Rather than using pairwise correlations of classi-

fiers' errors as a method to select a subset of classifiers, we focus on the overall behavior of the fusion. We use a single parameter in an attempt to provide a *uniform* compensation for the correlation of the classifiers' errors. We explore the impact of such compensation on the final classification by using four performance metrics. We empirically obtain the value of the parameter that gives us the best solution in this performance space.

2 Fusion via triangular Norms

We propose a general method for the fusion process, which can be used with classifiers that may exhibit any kind of (positive, neutral, or negative) correlation with each other. Our method is based on the concept of Triangular Norms, a multi-valued logic generalization of the Boolean intersection operator. To justify the use of the intersection operator it can be argued that the fusion of multiple decisions, produced by multiple sources, regarding objects (classes) defined in a common framework (universe of discourse) consists in determining the underlying degree of consensus for each object (class) under consideration, i.e., the intersections of their decisions. With the intersections of multiple decisions one needs to account for possible correlation among the sources, to avoid under- or over-estimates. Here we explicitly account for this by the proper selection of a T-norm operator.

We combine the outputs of the classifiers by selecting the generalized intersection operator (T-norm) that better represents the possible correlation between the classifiers. With this operator we will intersect the assignments of the classifiers and compute a derived measure of consensus. Under certain conditions, defined in section 2.2, we will be able to perform this fusion in an associative manner, e.g., we will combine the output of the fusion of the first two classifiers with the output of the third classifier, and so on, until we have used all available classifiers. At this stage, we can normalize the final output (showing the degree of selection as a percentage), identify the strongest selection of the fusion, and qualify it with its associated degree of confidence.

2.1 Triangular Norms

Triangular norms (T-norms) and Triangular conorms (T-conorms) are the most general families of binary functions that satisfy the requirements of the conjunction and disjunction operators, respectively. T-norms $T(x,y)$ and T-conorms $S(x,y)$ are two-place functions that map the unit square into the unit interval, i.e., $T(x,y):[0,1]\times[0,1]\rightarrow[0,1]$ and $S(x,y):[0,1]\times[0,1]\rightarrow[0,1]$. They are monotonic, commutative and associative functions. Their corresponding boundary conditions, i.e., the evaluation of the T-norms and T-conorms at the extremes of the $[0,1]$ interval, satisfy the truth tables of the logical AND and OR operators. They are related by the DeMorgan duality, which states that if $N(x)$ is a negation operator, then the T-conorm $S(x,y)$ can be defined as $S(x,y) = N(T(N(x), N(y)))$.

In Bonissone and Decker [4], six parameterized families of T-norms and their dual T-conorms were discussed and analyzed by the authors. Of the six parameterized families, one family was selected due to its complete coverage of the T-norm space and its numerical stability. This family, originally defined in [15], has a parameter p that spans the space of T-norms. By selecting different values of p we can instantiate T-norms with different properties. See [3, 15] for more detailed information.

2.2 Determination of a Combined Decision

Let us define m classifiers S_1, \dots, S_m , such that the output of classifier S_j is the vector I^j showing the normalized decision of such classifier to the N classes. In this representation, the last $(N+1)^{\text{th}}$ element represents the universe of all classes, U , and is used to indicate the classifier's lack of commitment, i.e. no-decision:

$$I^j = [I^j(1), I^j(2), \dots, I^j(N+1)], \text{ where } I^j(i) \in [0,1] \text{ subject to: } \sum_{i=1}^{N+1} I^j(i) = 1$$

We define the un-normalized fusion of the outputs of two classifiers S_1 and S_2 as:

$$F(I^1, I^2) = \text{Extraction}[\text{Outerproduct}(I^1, I^2, T)] = \text{Extraction}[A] \quad (1)$$

where the outer-product is a well-defined mathematical operation, which takes as arguments two N -dimensional vectors I^1 and I^2 and generates as output the $N \times N$ dimensional array A . Each element $A(i,j)$ is the result of applying the operator T to the corresponding vector elements, namely $I^1(i)$ and $I^2(j)$, i.e.:

$$A(i,j) = T[I^1(i), I^2(j)]. \quad (2)$$

The *Extraction* operator recovers the un-normalized output in vector form:

$$\text{Extraction}[A] = [I'(1), I'(2), \dots, I'(N+1)] \quad (3)$$

$$\text{where : } I'(i) = A(i, i) + A(i, N+1) + A(N+1, i) \text{ for } i = 1, N \quad (4)$$

$$\text{and } I'(N+1) = A(N+1, N+1) \quad (5)$$

The extraction operator leverages the fact that all classes are disjoint. A justification for equations (4) and (5) can be found in the case analysis illustrated in eq. (7).

There is an infinite number of T-norms. Therefore, to cover their space we used the family proposed by Schweizer and Sklar [15], in which each T-norm is uniquely defined by a real valued parameter p . We selected the six most representative ones for some practical values of information granularity, as listed in Table 1.

Table 1. Example of Outer Product using as operator the function $T(x,y)$

T-norm	Value of p	Correlation Type
$T_1(x, y) = \max(0, x + y - 1)$	$p = -1$	Extreme case of negative correlation
$T_2 = \text{Max}(0, x^{0.5} + y^{0.5} - 1)^2$	$p = -0.5$	Partial case of negative correlation
$T_3 = x * y$	$p \rightarrow 0$	No correlation
$T_4(x, y) = (x^{-0.5} + y^{-0.5} - 1)^{-0.5}$	$p = 0.5$	Mildly Positive Correlation
$T_5(x, y) = (x^{-1} + y^{-1} - 1)^{-1}$	$p = 1$	Partial case of positive correlation
$T_6 = \min(x, y)$	$p \rightarrow \infty$	Extreme case of positive correlation

The selection of the best T-norm to be used as intersection operation in the fusion of the classifiers depends on the potential correlation among their errors. T_6 (the minimum operator) should be used when one classifier subsumes the other one (**extreme case of positive correlation**). T_3 should be selected when the classifiers errors are **uncorrelated** (similar to the evidential independence in Dempster-Shafer). T_1 should be used if the classifiers are mutually exclusive (**extreme case of negative correlation**). The operators T_2 , T_4 , and T_5 should be selected when the classifiers errors show intermediate stages of negative (T_2) or positive correlation (T_4 and T_5). Of course, other T-norms could also be used. These five T-norms are simply good representatives of the infinite number of functions that satisfy the T-norm properties.

From the associativity of the T-norms we can derive the associativity of the fusion

$$F(I^1, F(I^2, I^3)) = F(F(I^1, I^2), I^3) \quad (6)$$

providing that equation (4) only contains one term. This is the case when: a) there is *no lack of commitment* in the classifiers [$A(i, N+1) = A(N+1, i) = 0$ for $i = 1, \dots, N$]; b) the *lack of commitment is complete* in one classifiers [$A(i, i) = 0$ for $i = 1, \dots, N$ and either $A(i, N+1) = I^1(i)$ & $A(N+1, i) = 0$ or $A(N+1, i) = I^2(i)$ & $A(i, N+1) = 0$]; c) the *lack of commitment is complete* in both classifiers [$eq(4) = 0$ & $eq(5) = 1$]. In any other case, the preservation of associativity requires the distributivity of the T-norm over the addition operator of equation (4). This distributivity is satisfied only by T-norm T_2 (scalar product). For any other T-norm the lack of associativity means that the outer-product A defined by equation (2) must be computed on all dimensions at once, using as operator a dimensionally extended Schweizer & Sklar T-norm:

$$T_p(x_1, \dots, x_k) = \left[\sum_{i=1}^k x_i^{-p} - k + 1 \right]^{-1/p} \quad \text{if } (p > 0) \text{ or } (p < 0) \text{ and } \left[\sum_{i=1}^k x_i^{-p} \right] \geq (k - 1)$$

Given our premise that the classes are disjoint, we have four possible situations:

- (a) when $i=j$ and $i < (N+1)$ then $r_i \cap r_j = r_j \cap r_i = r_i$ (7a)
- (b) when $i=j$ and $i = (N+1)$ then $r_i \cap r_j = U$ (the universe of classes) (7b)
- (c) when $i \neq j$ and $i < (N+1)$ and $j < (N+1)$ then $r_i \cap r_j = \phi$ (the empty set) (7c)
- (d) when $i \neq j$ and either $i = (N+1)$ or $j = (N+1)$ then $r_i \cap r_j = r_j$ or r_i (7d)

Figure 1 illustrates the result of the intersections of the classes and the universe U.

Intersection	r_1	r_2	...	r_{N-1}	r_N	U
r_1	r_1	ϕ	ϕ	ϕ	ϕ	r_1
r_2	ϕ	r_2	ϕ	ϕ	ϕ	r_2
...	ϕ	ϕ	...	ϕ	ϕ	ϕ
r_{N-1}	ϕ	ϕ	ϕ	r_{N-1}	ϕ	r_{N-1}
r_N	ϕ	ϕ	ϕ	ϕ	r_N	r_N
U	r_1	r_2	...	r_{N-1}	r_N	U

Figure 1. Intersection of disjoint classes and the universe U

Note that when DS theory is applied to subsets containing only singletons or the entire universe of classes U , its fusion rule becomes a special case of our proposed methodology.

2.3 Measure of Confidence

In cases where classes are ordered, we can compute the confidence in the fusion by defining a measure of the scattering around the main diagonal of the confusion matrix. The more weights are assigned to elements outside the main diagonal, the less is the measure of the consensus among the classifiers. We can represent this concept by defining a penalty matrix $\mathbf{P}=[P(i,j)]$, of the form:

$$P(i, j) = \begin{cases} \max(0, (1 - W^{|i-j|})^d) & \text{for } 1 \leq i \leq N \text{ and } 1 \leq j \leq N \\ 1 & \text{for } i = (N+1) \text{ or } j = (N+1) \end{cases} \quad (8)$$

This function rewards the presence of weights on the main diagonal, indicating agreement between the two classifiers, and penalizes the presence of elements off the main diagonal, indicating conflict. The conflict increases in magnitude as the distance from the main diagonal increases. For example, for $W=0.2$ and $d=5$ we have the following penalty matrix:

\mathbf{P}	r_1	r_2			r_N	U
r_1	1	0.3	0.1	0	0	1
r_2	0.3	1	0.3	0.1	0	1
	0.1	0.3	1	0.3	0.1	1
	0	0.1	0.3	1	0.3	1
r_N	0	0	0.1	0.3	1	1
U	1	1	1	1	1	1

Figure 2. Penalty matrix \mathbf{P} for $W=0.2$ and $d=5$

Of course any other function penalizing elements off the main diagonal (any suitable non-linear function of the distance from the main diagonal, i.e. the absolute value $|i-j|$) could also be used. The reason for this penalty function is that **conflict is gradual**, since the classes have an ordering. Therefore, we want to capture the fact the discrepancy between classes r_1 and r_2 is smaller than the discrepancy between r_1 and r_3 . The shape of the penalty matrix \mathbf{P} captures this concept, as \mathbf{P} shows that the confidence decreases non-linearly with the distance from the main diagonal.

A measure of the normalized confidence \hat{C} is the sum of element-wise products between \hat{A} and \mathbf{P} , e.g.:

$$\hat{C} = \text{Normalized Confidence}(\hat{A}, \mathbf{P}) = \sum_{i=1}^{N+1} \sum_{j=1}^{N+1} \hat{A}(i, j) * P(i, j) \quad (9)$$

where \hat{A} is the normalized fusion matrix.

We could interpret the confidence factor \hat{C} as the weighted cardinality of the normalized assignments around the main diagonal, after all the classifiers have been fused.

In the special case of Dempster Shafer, the measure of confidence \hat{C} is the complement (to one) of the measure of conflict K, i.e.: $\hat{C} = 1 - K$, where K is the sum of weights assigned to the empty set. We would like to point out that the fusion rule based on Dempster-Shafer corresponds to the selection of:

- (a) T-norm operator $T(x,y) = x * y$
- (b) Penalty function using $d = \infty$

Constraint b) implies that the penalty matrix P will be:

P	r_1	r_2			r_N	U
r_1	1	0	0	0	0	1
r_2	0	1	0	0	0	1
	0	0	1	0	0	1
	0	0	0	1	0	1
r_N	0	0	0	0	1	1
U	1	1	1	1	1	1

Figure 3. Penalty matrix P for the special case of DS confidence computation: $d=\infty$

Therefore the two additional constraints a) and b) required by Dempster-Shafer theory imply that

- (a) The classifiers to be fused must be uncorrelated (evidentially independent) and
- (b) There is no ordering over the classes, and any kind of disagreement (weights assigned to elements off the main diagonal) can only contribute to a measure of conflict and not (at least to a partial degree) to a measure of confidence. In DS, the measure of conflict K is the sum of weights assigned to the empty set. This corresponds to the elements with a 0 in the penalty matrix P illustrated in Figure 3.

Clearly, our proposed method is more general, since it does not require these additional constraints, and includes the DS fusion rule as a special case.

2.4 Decisions from the Fusion Process

The decisions for each class can be gathered by adding up *all* the weights assigned to them, as indicated in equations (4) and (5).

3 Examples

3.1 Example (1): Agreement between the classifiers without discounting.

Let $I^1 = [0.8, 0.15, 0.05, 0, 0, 0]$ and $I^2 = [0.9, 0.05, 0.05, 0, 0, 0]$

This indicates that both classifiers are showing a strong preference for the first class as they are assigned 0.8 and 0.9, respectively. We fuse these classifiers using each of the T-norm operators defined above and normalize the results so that *the sum of the entries is equal to one*. Note that during the process we will use the un-normalized matrices A to preserve the associative property. Using the expressions for weights of a class, we can compute the final weights for the N classes and the universe. Table 2 shows the results of the fusion of classifiers S1 and S2, using each of the five T-norms, with the associated normalized confidence measure.

Table 2. Fusion of two classifiers and associated confidence measure with agreement

	C1	C2	C3	C4	C5	U	Norm. Conf.
$T_6 = \min(x, y)$	0.615	0.038	0.038	0	0	0	0.774
$T_5(x, y) = (x^{-1} + y^{-1} - 1)^{-1}$	0.633	0.034	0.022	0	0	0	0.770
$T_3 = x * y$	0.720	0.008	0.003	0	0	0	0.797
$T_2 = \text{Max}(0, x^{0.5} + y^{0.5} - 1)^2$	0.807	0.000	0.000	0	0	0	0.858
$T_1 = \max(0, x + y - 1)$	0.933	0.000	0.000	0	0	0	0.955

3.2 Example (2): Disagreement between the classifiers, without ignorance

In a situation in which there is a *discrepancy* between the two classifiers, this fact will be captured by the confidence measure. For instance, let's assume that the two classifiers are showing strong preferences for two *different* classes, e.g.

$$I^1 = [0.15, 0.85, 0.05, 0, 0, 0] \quad \text{and} \quad I^2 = [0.9, 0.05, 0.05, 0, 0, 0]$$

The results of their fusion are summarized in Table 3. None of the classes has a high weight and the normalized confidence has dropped, with respect to Table 2.

Table 3. Fusion of two classifiers and related confidence measure with disagreement

	C1	C2	C3	C4	C5	U	Norm. Conf.
$T_6 = \min(x, y)$	0.115	0.038	0.038	0	0	0	0.438
$T_5(x, y) = (x^{-1} + y^{-1} - 1)^{-1}$	0.127	0.043	0.022	0	0	0	0.438
$T_3 = x * y$	0.135	0.040	0.003	0	0	0	0.434
$T_2 = \text{Max}(0, x^{0.5} + y^{0.5} - 1)^2$	0.128	0.016	0.000	0	0	0	0.416
$T_1 = \max(0, x + y - 1)$	0.067	0.000	0.000	0	0	0	0.373

4 Application and Results

Insurance underwriting is a classification problem. The underwriting (UW) process consists of assigning a given insurance application, described by its medical and demographic records, to one of the risk categories (referred to as rate classes). Traditionally, highly trained human individuals perform insurance underwriting. The granularization of risk into rate classes is dictated by industry regulation and by the impossibility of human underwriters to achieve consistency of decisions over real-valued risk granularity. A given application for insurance is compared against standards put forward by the insurance company and classified into one of the risk categories available for the type of insurance requested by the applicant. The risk categories then affect the premium paid by the applicant; the higher the risk category, the higher the premium. Such manual underwriting, however, is not only time-consuming, but also often inadequate in consistency and reliability. The inadequacy becomes more apparent as the complexity of insurance applications increases. The automation of this decision-making process has strong accuracy, coverage and transparency requirements. There are two main tradeoffs to be considered in designing a classifier for the insurance underwriting problem: 1) *Accuracy versus coverage* - requiring low misclassification rates for high volume of applications; 2) *Accuracy versus interpretability* - requiring a transparent, traceable decision-making process. Given the highly non-linear boundaries of the rate classes, continuous approaches such as multiple regressions were not successful and a discrete classifier was deployed for production.

Once this process is automated by a production engine, we need to perform *off-line* audits to monitor and assure the quality of the production engine decisions. Auditing automated insurance underwriting could be accomplished by fusing the outputs of multiple classifiers. The purpose of this fusion, then, is to assess the quality of the decisions taken by a production engine. In addition, this fusion will identify the best cases, which could be used to tune the production engine in future releases, as well as controversial or unusual cases that should be highlighted for manual audit by senior underwriters, as part of the Quality Assurance process.

We tested the approach against a case base containing a total of 1,875 cases, which could be assigned to one of five rate-classes. The fusion was performed using different T-norms. The classifiers to be fused were a classification and regression tree (MARS), a bank of binary MVP neural (NN) nets using back-propagation as the learning algorithm, a case-based reasoning engine (CBE), and a bank of binary support vector machines (SVM). Table 4 shows the performance of the individual classifiers as expressed by the true positive rate.

Table 4: Classifier Performance

Classifier	True Positive Rate
MARS	94.72%
NN	94.08%
CBE	91.52%
SVM	94.51%

4.2 Experiments Results using different T-Norms

Table 5 shows the results of using three *Tnorms* as the outer-product operators (using the same set of 1,875 non-nicotine users). When using T3 (*scalar product*) combined with a strict determination of the conflict (e.g. for $d=\infty$ in the computation of penalty matrix P), we have an aggregation analogous to that of Dempster-Shafer’s rule of combination. This aggregation relies on the *evidential uncorrelation* of the sources, a constraint that is not easy to satisfy in real situations. In our case, we used a pre-processing stage (*Tagging*) to annotate the input data with domain knowledge. Since this pre-processing stage is applied to a common set of inputs for *all* models, it is possible to introduce a common bias; hence the models could exhibit partial positive correlation. We sampled the space of T-norms corresponding to positive correlation and we found that *T4* is the *Pareto-best* to account for this positive correlation, according to the metrics in Table 5. We also compared the T-norm fusion against a baseline fusion using an averaging scheme. The strict or liberal errors used in Table 5 are extensions of type 1 and type 2 errors in a classification matrix of the rate classes.

Table 5: Performance of Average and Fusion (using three T-norms)

Metrics	Average	Fusion using T3 ($p \rightarrow 0$)	Fusion using T4 ($p = 0.5$)	Fusion using T5 ($p = 1$)
True Positive Rate	94.24%	94.08%	94.72%	94.08%
Incorrect Rate (too strict)	3.47%	3.47%	3.09%	3.47%
Incorrect Rate (too liberal)	2.29%	2.35%	2.19%	2.40%
No Decisions	0%	0.11 %	0%	0.05 %

5 Conclusions

We introduced a fusion mechanism based on an outer-product of the outputs, using a Triangular norm as the operator. The output of the fusion was a class distribution and a measure of conflict among all the classifiers. We normalized the final output, identified the strongest selection of the fusion, and qualified the decision with an associated confidence measure. We considered the output of each classifier as a weight assignment, representing the (un-normalized) degree to which a given class was selected by the classifier. As in DS theory, the assignment of weights to the universe all rate-classes represented the lack of commitment to a specific decision.

We applied the fusion to the quality assurance (QA) problem for automated underwriting. When the degree of confidence of a fused decision was below a lower confidence bound, that case became a candidate for *auditing*. When the degree of confidence of a fused decision was above an upper confidence bound, that case became a candidate for *augmenting the Standard Reference Decision set (SRD)*. We tested the fusion module with data sets for nicotine and non-nicotine users. In an analysis of 1,875 non-nicotine applications, we generated a distribution of the quality of the production engine. By focusing on the two tails of such distribution, we identified **~9%** of the most reliable decisions and **4.9%** of the least reliable ones.

The proposed fusion module plays a key role in supporting the quality assurance of the knowledge-based classifier, by selecting the most questionable cases for auditing. At the same time, the fusion module supports the standard reference decision lifecycle by identifying the most reliable cases used for updating it [4]. Our next step will be to use backward search, guided by a diversity measure [12,14], to select the smallest subset of classifiers to be deployed in the production version of this QA process.

References

- [1] Al-Ani, A. and Deriche, M. A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence, *J. Artificial Intelligence Research*, 17:333-361, 2002.
- [2] Bonissone, P.P. and Decker, K.S. Selecting Uncertainty Calculi and Granularity: An Experiment in Trading-off Precision and Complexity, in Kanal and Lemmer (editors) *Uncertainty in Artificial Intelligence*, pages 217-247, North-Holland, 1986.
- [3] Bonissone, P.P. Summarizing and Propagating Uncertain Information with Triangular Norms, *International Journal of Approximate Reasoning*, 1(1):71-101, January 1987.
- [4] Bonissone, P.P. Automating the Quality Assurance of an On-line Knowledge-Based Classifier By Fusing Multiple Off-line Classifiers, *Proc. IPMU 2004*, Perugia (Italy), July 2004.
- [5] Cho, S. and Kim, J., Combining multiple neural networks by fuzzy integral for robust classification, *IEEE Trans. on Systems, Man, and Cybernetics*, 25(2):380-384, 1995
- [6] Dempster, A.P. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325-339, 1967.
- [7] Ho, T.K., Hull, J.J., and Srihari, S.N. Decision Combination in Multiple Classifier Systems, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 16, No.1, pp.66-75, 1994.
- [8] Kim, J., Seo, K., and Chung, K. A Systematic approach to Classifier selection on combining multiple classifiers for handwritten digit recognition, *Proceedings of the 4th International Conference on Document Analysis and Recognition*, 2:459-462, 1997.
- [9] Kuncheva L.I. Switching between selection and fusion in combining classifiers: An experiment, *IEEE Transactions on SMC, Part B*, 32 (2), 2002, 146-156.
- [10] Lam L. and Suen, C. Y. Increasing experts for majority vote in OCR: Theoretical considerations and strategies. *Proc. 4th Int. Workshop on Frontiers in Handwriting Recognition*, Taipei, Taiwan, 245-254, 1994.
- [11] Niyogi, P Pierrot, J-B, and Siohan.O. On decorrelating classifiers and combining them, MIT AI Lab., September 2001
- [12] Partridge, D., Yates, W.B. Engineering multiversion neural-net systems. *Neural Computation*, 8:869-893, 1996.
- [13] Petrakos, M., Kannelopoulos, I., Benediktsson, J., and Pesaresi, M. The Effect of Correlation on the Accuracy of the Combined Classifier in Decision Level Fusion, *Proceedings of IEEE 2000 International Geo-science and Remote Sensing Symposium*, Vol. 6, 2000.
- [14] Roli, F., Giacinto, G., and Vernazza, G. Methods for Designing Multiple Classifier Systems, *MCS 200, LNCS 2096*, 78-87, 2001
- [15] Schweizer, B., and Sklar, A. Associative Functions and Abstract Semi-Groups., *Publicationes Mathematicae Debrecen*, 10:69-81, 1963.
- [16] Shafer, G. *A Mathematical Theory of Evidence*, Princeton University Press, NJ, 1976.
- [17] Tumer, K.& Ghosh, J. Error correlation and error reduction in ensemble classifiers, *Connection Science*, 8:385-404, 1996.
- [18] Xu, L., Krzyzak, A., & Suen, C.Y. Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition, *IEEE Syst. Man and Cybernetics*, 22(3), 1992.