2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

A New Fuzzy Approach for Multi-Source Decision Fusion

Farnoosh Fatemipour, Mohammad-R. Akbarzadeh-T., Rouhollah Ghasempour
Departments of Computer and Electrical Engineering,
Center of Excellence on Soft Computing and Intelligent Information Processing (SCIIP)
Ferdowsi University of Mashhad
Mashhad, Iran
farnoosh.fatemipour@stu-mail.um.ac.ir, akbarzadeh@ieee.org, rouhollahghasempour@yahoo.com

***Abstract*— Nowadays, we are facing the rapidly growing amount of data being produced in many organizations, social networks and internet. These data are generated in disparate locations and their aggregation into one location is exceedingly time and space consuming. Traditional statistical methods are not sufficient for processing of this massive multi-source data. In this paper, we propose a new fuzzy-based decision fusion approach for classification problems of this kind. The necessity of fuzzy information arises in distributed classification because imprecision, uncertainty and ambiguity can be found at all information sources, from the data itself to the results of the classifiers. In the proposed approach, multiple classifiers are constructed based on different information sources which have different degrees of reliability. Then a fuzzy rule based system is designed for approximating distribution of reliabilities of sources over the input space. The decision fusion of multiple classifiers takes place using the estimated degrees of sources' reliabilities. Comparison results are made between both centralized classification and two other distributed classification methods. One is averaging and the other is discounting each classifier's decision based on its accuracy. Results show the high accuracy of the proposed method in making decisions in distributed environments, without the overhead of aggregating the entire data in one location.**

Keywords*— decision fusion; classifier combination; multi-source classification; Fuzzy logic

I. INTRODUCTION

Classification is one of the most important issues in pattern recognition and data mining tasks. There is a rich collection of research on this topic in the literature and many algorithms have been presented such as decision trees, neural networks and support vector machines. These methods are applicable to problems that all of the training data have the same type, and are available altogether. In today's world, we are facing the rapidly growing amount of data that are being produced in many organizations, social networks, internet, etc. This data is produced in different locations and to maintain data privacy, there is no possibility of aggregating them into one place. It would also be very time and space consuming, if not impossible, to join relevant large data sources for mining patterns consisting of multiple aspects of information [1]. Also, data might consist different data types [2, 3]. Commonly used statistical pattern recognition methods are often not appropriate for the classification of multisource data, because in most cases the data cannot be modeled by a convenient multivariate statistical model [2, 4, 5]. Processing of large volumes of data of different types that are distributed at disparate locations, is still a challenge for data mining methods.

Our concern here is the problem of multi-source classification, where data is split into multiple datasets located in diverse sources, see Fig. 1. The individual data sources may not be equally reliable. One source can be more applicable to describe a specific class, and perhaps another source is more adequate for another class [5]. Consider, for example, data of patients are collected from several hospitals. Each hospital can accommodate a special range of patients according to their position. For example, in a hospital in one place, most visited patients suffer from diseases related to old age, or another one might be children's hospital medical center. In this case, the information collected by each hospital is suitable for certain types of test data. Thus, it is appropriate to weight the different sources during the classification process. On the other hand, conventional statistical techniques do not allow such weighting [5].

Classifier ensemble is considered in many researches. Using multiple classifiers results in increased accuracy of the system. Several methods can be used for combining classifiers such as using multiple classifiers of the same algorithm for several data sets, or several classifiers with different algorithms for the entire data set. Classifiers then make different errors on different instances, and a suitable combination of these classifiers' decisions can hence reduce the total error [6]. In some papers, such as [7-9], when the amount of training data is too large, the data is properly distributed into several parts to deal with the cost of computation and increase the accuracy of the system. Each source is then trained separately and an ensemble is created. How to split the data and how to combine multiple decisions are two primary concerns in these articles.

Classifier selection and classifier fusion are two main approaches discussed in the literature for combining classifiers. The presumption in classifier selection is that each classifier has expertise in some local area of the feature space [10]. On the other hand, classifier fusion assumes that all classifiers are equally "experienced" in the whole feature space [10]. When a new pattern x is submitted for classification, the classifier selection approaches use the most expert classifier in the vicinity of x [11]. While classifier fusion approaches take into account the decisions of all classifiers by using some weighted average of many classifiers' outputs [6, 12, 13].

Information fusion deals with the integration of information from several different sources. The idea of decision fusion approaches is to let each local classifier make a (local) decision based only on its own data set and forward that decision to the central classifier which finalizes a decision based on the set of local decisions and any available prior knowledge, such as the reliabilities of the respective local decisions [14]. The necessity of fuzzy information arises in distributed classification because imprecision, uncertainty and ambiguity can be found at all information sources, from the data itself to the results of classifiers. In addition, individual classifiers are usually unreliable and misleading. Thus integration of classifiers' results is necessary to obtain a reliable classification.

In this paper, we propose a new approach for decision fusion using fuzzy IF-THEN rules. Some classifiers are constructed based on different information sources which have different degrees of reliability. The main idea behind our proposed approach is that each system is trained with only a part of the whole data that has been collected independently, so its reliability varies over input space. If we model these reliabilities suitably, we can significantly increase classification accuracy. Therefore we are going to estimate the distribution of reliabilities of sources in the entire input space. To this end, we train a fuzzy rule based system to learn these weights and set appropriate weights to classifiers based on the input data that is submitted for classification. For each unknown pattern, classifiers' results will be combined using these weights. This means that those classifiers that are more reliable in the vicinity of x will have more effect on the final decision and vice versa. Evaluation results on five datasets shows the superiority of our method for classification in distributed data environments.

The rest of this paper is organized as follows. Section II includes a brief review of the preliminary concepts. In section III our proposed decision fusion method is explained in details. Evaluation results are included in section IV, and section V finalizes the paper with the conclusion.

II. PRELIMINARIES

In this section we review some preliminary concepts briefly. In the following, fuzzy rule based systems and fuzzy approximation are described which we are going to use in the next sections.

Fuzzy logic was first introduced by Zadeh in 1965 [15]. It is a complementary to the classical set theory that elements either belong to a set or not. In fuzzy set theory elements can belong to a set to some degree between 0 and 1.

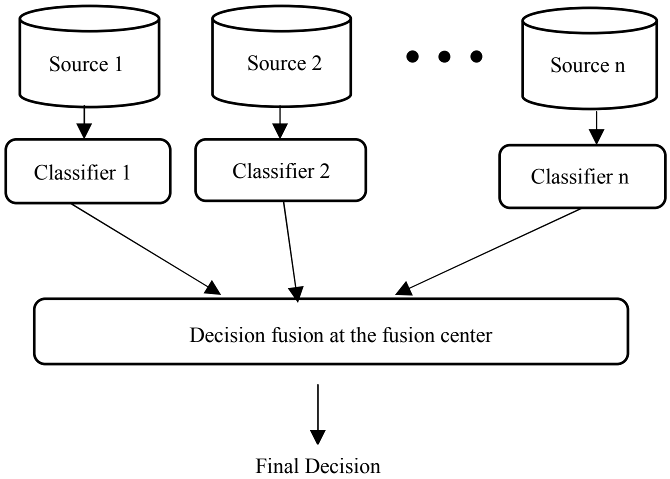

Fig. 1. Decision fusion with multiple classifiers

A fuzzy rule is expressed in the form of (1). Each A_i and B_i is a fuzzy set. Any type of membership functions such as triangular, exponential, etc. can be used to describe these fuzzy sets.

$$\begin{aligned}&\text{IF } x_1 \text{ is } A_1 \text{ and } \dots \text{ and } x_n \text{ is } A_n\\&\text{THEN} y_1 \text{ is } B_1 \text{ and } \dots \text{ and } y_m \text{ is } A_m\end{aligned} \quad (1)$$

A fuzzy rule based is the main part of a fuzzy system. A fuzzy rule based is a set of fuzzy rules that each rule fires with a degree of possibility depending on the input feature vector. The matching degree between the current value of x and the antecedent of each rule defines its firing degree.

III. MULTI SOURCE DECISION FUSION APPROACH

Bayesian classifiers are designed to classify an unknown pattern into the most probable class [16]. Our concern here is to find the most probable class when there are multiple classifiers, each trained on a part of the whole data. In this section we propose a new approach for decision fusion in combining outputs of these classifiers.

Assume, we have N number of data sources $S_i, i = 1 \dots \mathrm{N}$ such that each source is trained with training data d_i. An unknown pattern which is represented by a feature vector x, is supposed to be classified into one of the M classes $C_j, j = 1 \dots M$. Each source i's outputs for data x are in the form of $P_i(C_j|x)$, which is the posterior probability that data x belongs to class C_j, according to source i. Let y_d be the desired value for label of x. After the training phase in the fusion center and learning sources' weights, each unknown pattern is classified to one of the M possible classes. In the following, the training phase of the fusion center is described in subsection A. Subsection B explains the process of assigning a label to an unlabeled data.

A. Training the fusion center

As mentioned before, each classifier is trained with only a part of the entire data. Different data sets have different reliabilities [5]. Each source has collected data independently.

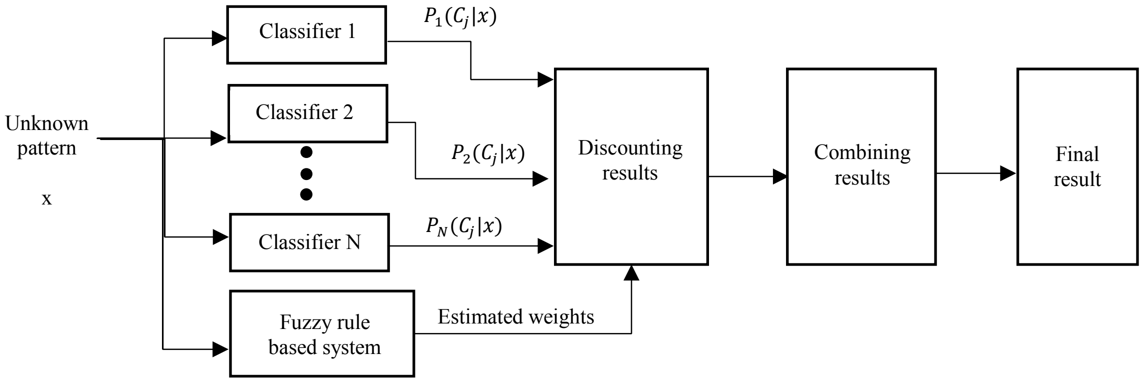

Fig. 2. Overview of assigning a label to an unlabeled data. $P_i(C_j|x)$ indicates the posterior probability that data x belongs to class $C_j, j = 1, \dots, M$, according to source i.

Data might not be balanced between sources and some sources might lack or not have enough data for some classes. The different sources' data types also might be different.

Instead of setting only one static weight for each source, we first learn each source's distribution of reliability over input space. This is because each classifier's reliability is not a single number, but varies over data space. In other words, the classifier's reliability is determined according to the data that is supposed to be classified. After this phase, we will be able to discount different classifiers' posterior probabilities based on their reliability in the vicinity of each input data, and gain higher accuracies. Because for each input data, the most suitable classifiers with higher accuracy will have more effect on the final result and vice versa.

Fuzzy approximation is a promising framework for dealing with uncertainty. Here we use fuzzy function approximation for estimating sources' reliabilities. The rule based system is shown in (3). A_is are fuzzy sets. x is an n dimensional input vector, and y is the output vector representing estimated weights for each source i. These weights in the consequent part of rules are defined as the classifier's accuracy in the corresponding partition. In other words, the ratio of correctly classified patterns to the number of all patterns for each classifier in the corresponding partition. P, the number of rules, is defined as the multiplication of the number of partitions for each input dimension. Assuming partitions(d) is the number of partitions in dimension d, the formula of calculating P is shown in (2).

$$P = \prod_{d=1}^{n} partitions(d) \tag{2}$$

rule 1 IF x_1 is A_1^1 and … and x_n is A_n^1 THEN y is $[w_1^1, \dots, w_N^1]$

rule 2 IF x_1 is A_1^2 and … and x_n is A_n^2 THEN y is $[w_1^2, \dots, w_N^2]$ (3)

…

rule P IF x_1 is A_1^p and … and x_n is A_n^p THEN y is $[w_1^P, \dots, w_N^P]$

B. Decision fusion at the fusion center

Assigning a label to an unknown pattern x is performed in several stages: (see Fig. 2)

- *Stage 1*: Obtaining sources' outputs

For every information source $i, i = 1, \dots, N$, local decision for label of x in the form of posterior probabilities, $P_i(C_j|x)$, is obtained for each class $j, j = 1, \dots, M$.

- *Stage 2*: Calculating membership degree of x to each rule: μ_p

μ_p, the matching degree of x to each rule is calculated.

- Stage 3: Decision fusion

The input pattern might match with more than one rule. So the rule based system approximates multiple sets of weights with different firing degrees. Equation (3) is the defuzzification formula named "center of gravity" [17]. In this equation, x is the input and P is the total number of rules in the rule based system. y_p is the output of rule number p.

$$f(x) = \frac{\sum_{p=1}^{P} \mu_p y_p}{\sum_{p=1}^{P} \mu_p} \tag{3}$$

In the proposed method, the approximated weights are used to discount the sources' results before defuzzification. All sources are assumed equally reliable after the discounting process and posterior probability $P(C_i|x)$ for each class $C_i, i = 1, \dots, M$ is calculated using (4). Data will be classified into the most probable class (class with maximum estimated probability).

$$P(C_j|x) = \sum_{p=1}^{P} \mu_p \sum_{i=1}^{N} w_i^p * P_i(C_j|x) \tag{4}$$

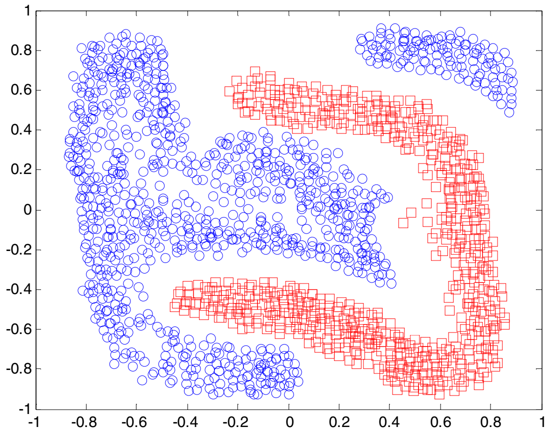

Fig. 4. Generated non-linear 2-dimensional 2-class data

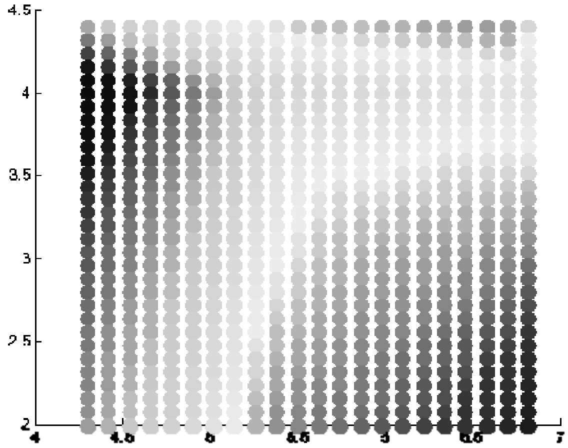

Fig. 3. A sample estimate of distribution of reliability of one source for first two dimensions of banknote dataset. Darker circles indicate higher reliability

IV. EXPERIMENTAL RESULTS

In this section we explain the details of experimental results. To evaluate the performance of our proposed method, we chose four datasets from [18] and [19] and a nonlinear dataset that we generated as shown in Fig. 4. The datasets are Blood Transfusion, Banknote and magic Gamma Telescope from [18] and svmguide1 from [19]. Table 1 shows a detailed description of the data sets. The Non-Linear dataset is generated for further expressing of the ability of the proposed method in classifying non-linear distributed data.

TABLE 1. A BRIEF DESCRIPTION OF THE DATA SETS USED

Data set	Attributes type	Examples	Features	Classes
Blood Transfusion	Real	748	5	2
Bank note	Real	1372	5	2
Non-linear data	Real	1601	2	2
magic	Real	19020	10	2
svmguide	Real	3089	4	2

We used 5-fold cross validation for this purpose. Cross validation helps to ensure that the performance is data-independent. In each iteration, we chose 3 folds of data for dividing between sources and the remaining 2 folds of data is used for training the fusion center and testing. The sources' data is first divided to 2 sources, then we increase the number of sources two by two up to 12 sources.

In our experiment, each dimension of the input space is divided into ten partitions with triangular membership functions. We use each sources' accuracy for determining weights in the consequent part of each rule. Matching degree between the input data and the membership functions in the rule is computed using the product operator. As mentioned before, we estimate distribution of the sources' reliabilities using fuzzy approximation. An example of the estimated reliability of one source first two dimensions of Banknote dataset is depicted in Fig. 3.

We used two scenarios for splitting data into multiple sources: (as in [20]).

- First case: We split data randomly between sources. In this case, sources will have roughly the same distribution [20].
- Second case: data are split equally into N segments based on the value of the first feature [20]. In this case, different sources will have different distributions.

Comparisons are made between the proposed method and the non-distributed naïve Bayes classifier where the entire sources' data is available at once for training. Table 2 shows the mean accuracy results of 5-fold cross validation training and testing. As it is shown in Table 2, simulated results on five datasets show that our method can outperform the accuracy of the non-distributed case where all data are available for training at once. This means that the proposed method can make decisions in distributed data environments with high accuracy without overhead of aggregating the entire data in one place.

The proposed approach for weighting sources is also compared to two other methods, averaging and static weighting. In averaging, classifiers' outputs are combined using mean operator, without being weighted. In static weighting, one weight is assigned to each classifier, and decision fusion is done using weighted mean of classifiers' outputs, for every unknown pattern. This weight is determined by the ratio of correctly classified patterns to the number of all patterns for each classifier in the corresponding training data set. Results of this comparison are included in Table 3. As the results indicate, our method outperforms these weighting methods and increases the accuracy of the classification significantly. This is simply because defining weights based on input pattern which is supposed to be classified, lets the classifiers that are more reliable in the vicinity of x, have more effect on the final decision and vice versa. These result in

TABLE 2. MEAN RESULTS OF 5-FOLD CROSS VALIDATION. COMPARISONS ARE MADE BETWEEN THE PROPOSED DISTRIBUTED METHOD, AND NON-DISTRIBUTED NAÏVE BAYES CLASSIFIER

dataset		**Proposed Method**						**Naive Bayes**
		# of sources						
		2	***4***	***6***	***8***	***10***	***12***	
Transfusion	Random distribution	75.3354	75.6973	76.3665	76.6916	77.0433	77.0874	75.8632
	Split by first feature	76.9961	77.3597	76.8137	76.9378	77.6579	75.5426	
Banknote	Random distribution	84.6392	86.0092	86.8223	87.3899	87.7853	89.4655	84.1910
	Split by first feature	97.7906	98.1911	97.9048	97.8224	98.3047	97.1321	
Nonlinear	Random distribution	76.1570	77.8455	78.5646	79.8871	80.9271	81.2849	75.6329
	Split by first feature	78.6858	85.0376	91.1037	87.9055	87.5311	87.8116	
Magic	Random distribution	76.6365	76.6325	76.6322	76.6100	76.5792	76.1684	76.61
	Split by first feature	76.6478	76.6856	76.6613	76.6175	76.5991	76.1743	
SVMGuide	Random distribution	92.8858	93.1081	93.3885	93.4371	93.6935	93.7656	92.8376
	Split by first feature	90.2879	90.1823	93.3478	89.0873	93.1595	90.8792	

TABLE 3. MEAN RESULTS OF 5-FOLD CROSS VALIDATION. COMPARISONS ARE MADE BETWEEN THE PROPOSED METHOD AND TWO OTHER METHODS, MEAN AND STATIC WEIGHTING.

dataset	**Algorithm**	**Random distribution**						**Split by first feature**					
		# of sources											
		2	***4***	***6***	***8***	***10***	***12***	***2***	***4***	***6***	***8***	***10***	***12***
Transfusion	Proposed approach	**75.33**	**75.69**	**76.36**	**76.69**	**77.04**	**77.08**	**76.99**	**77.35**	**76.81**	**76.93**	**77.65**	75.54
	Averaging	75.28	75.19	75.86	76.23	76.45	76.66	72.14	69.51	72.92	76.31	76.28	**76.63**
	Static weighting	75.12	75.40	75.86	76.36	76.20	76.56	73.35	71.81	74.65	76.51	76.48	76.23
Banknote	Proposed approach	**84.63**	**86.00**	**86.82**	**87.38**	**87.78**	**89.46**	**97.79**	**98.19**	**97.90**	**97.82**	**98.30**	**97.13**
	Averaging	84.18	84.21	84.22	84.25	84.35	84.29	59.81	67.52	72.97	78.07	71.99	70.35
	Static weighting	84.18	84.19	84.22	84.32	84.37	84.36	67.59	69.03	76.86	76.00	73.02	66.80
Nonlinear	Proposed approach	**76.15**	**77.84**	**78.56**	**79.88**	**80.92**	**81.28**	**78.68**	**85.03**	**91.10**	**87.90**	**87.53**	**87.81**
	Averaging	75.53	75.65	75.66	75.79	75.70	75.54	60.10	61.95	66.76	74.77	72.78	72.88
	Static weighting	75.56	75.65	75.68	75.85	75.71	75.49	73.20	62.50	73.03	74.70	73.68	71.89
Magic	Proposed approach	**76.63**	**76.63**	**76.63**	76.61	**76.58**	**76.16**	**76.65**	**76.69**	**76.66**	**76.62**	**76.60**	**76.17**
	Averaging	76.60	74.61	75.62	**76.62**	75.57	76.13	74.26	73.28	73.56	72.38	74.10	72.284
	Static weighting	76.59	76.53	76.52	76.60	76.54	76.15	74.45	74.67	74.98	75.69	74.76	75.72
SVMGuide	Proposed approach	**92.88**	**93.10**	93.38	**93.45**	**93.71**	**93.76**	**90.27**	**92.18**	**94.34**	**91.08**	**93.15**	**90.87**
	Averaging	92.87	92.10	93.38	93.43	93.69	93.74	90.26	90.18	93.34	89.08	**93.15**	89.87
	Static weighting	92.87	91.10	**93.40**	93.43	93.69	93.73	91.12	92.03	93.56	90.87	93.14	90.31

great decrement of the multi-source classification error.

V. CONCLUSION AND FUTURE WORK

In this paper, we propose a new method for decision fusion when the data sources are disparate. Instead of assigning one single weight to each classifier, we estimate the distribution of reliability of each source over the input space. This is done using fuzzy function approximation which is a promising framework for dealing with uncertain and incomplete knowledge. Simulated results prove the superiority of the proposed method in making decisions with disparate information sources.

The proposed method can be useful for many applications such as when data sources are geographically distributed and it is not possible to merge them into a single source. Another application of the proposed method is when we have different types of training data, so we cannot use a single type of classifier for all of them. It is also applicable for the case which we have a massive amount of data. By dividing the data randomly into some smaller sources, the proposed method can increase the speed of processing without a considerable loss of accuracy.

For the future work we are looking forward to using our method in real applications and show its performance in real world distributed decision making problems.

REFERENCES

[1] L. Cao, H. Zhang, Y. Zhao, D. Luo, and C. Zhang, "Combined mining: discovering informative knowledge in complex data," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol. 41, pp. 699-712, 2011.

[2] J. Benediktsson, P. H. Swain, and O. K. Ersoy, "Neural network approaches versus statistical methods in classification of multisource remote sensing data," IEEE Transactions on geoscience and remote sensing, vol. 28, pp. 540-552, 1990.

[3] R. R. Yager, "Set measure directed multi-source information fusion," Fuzzy Systems, IEEE Transactions on, vol. 19, pp. 1031-1039, 2011.

[4] T. Lee, J. A. Richards, and P. H. Swain, "Probabilistic and evidential approaches for multisource data analysis," Geoscience and Remote Sensing, IEEE Transactions on, pp. 283-293, 1987.

[5] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," Geoscience and Remote Sensing, IEEE Transactions on, vol. 45, pp. 3858-3866, 2007.

[6] J. Kittler, M. Hatef, R. P. Duin, and J. Matas, "On combining classifiers," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 20, pp. 226-239, 1998.

[7] H. Cevikalp and R. Polikar, "Local classifier weighting by quadratic programming," Neural Networks, IEEE Transactions on, vol. 19, pp. 1832-1838, 2008.

[8] J. A. Benediktsson and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," Geoscience and Remote Sensing, IEEE Transactions on, vol. 37, pp. 1367-1377, 1999.

[9] X. Ceamanos, B. Waske, J. A. Benediktsson, J. Chanussot, M. Fauvel, and J. R. Sveinsson, "A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data," International Journal of Image and Data Fusion, vol. 1, pp. 293-307, 2010.

[10] L. I. Kuncheva, "Switching between selection and fusion in combining classifiers: An experiment," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol. 32, pp. 146-156, 2002.

[11] K. Woods, W. P. Kegelmeyer Jr, and K. Bowyer, "Combination of multiple classifiers using local accuracy estimates," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 19, pp. 405-410, 1997.

[12] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 16, pp. 66-75, 1994.

[13] J. Grim, J. Kittler, P. Pudil, and P. Somol, "Multiple classifier fusion in probabilistic neural networks," Pattern Analysis & Applications, vol. 5, pp. 221-233, 2002.

[14] B. Jeon and D. A. Landgrebe, "Decision fusion with reliabilities in multisource data classification," in Systems, Man and Cybernetics, 1992., IEEE International Conference on, 1992, pp. 617-622.

[15] L. A. Zadeh, "Fuzzy sets," Information and control, vol. 8, pp. 338-353, 1965.

[16] S. Theodoridis and K. Koutroumbas, Pattern recognition: Academic Press, 1999.

[17] L.-X. Wang, A Course in Fuzzy Systems: Prentice-Hall press, USA, 1999.

[18] D. Newman, S. Hettich, C. Blake, C. Merz, and D. Aha, "UCI repository of machine learning databases. Department of Information and Computer Science, University of California, Irvine, CA," in 1998 of Conference, http://archive. ics. uci. edu/ml/datasets. html, 1998.

[19] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," ed, 2003.

[20] P. Phoungphol, "A Classification Framework for Imbalanced Data," Computer Science Dissertations. Paper 78, 2013.