

Insult Detection in Social Network Comments Using Possibilistic Based Fusion Approach

Mohamed Maher Ben Ismail and Ouiem Bchir

Abstract This paper aims to propose a novel approach to automatically detect verbal offense in social network comments. It relies on a local approach that adapts the fusion method to different regions of the feature space in order to classify comments from social networks as insult or not. The proposed algorithm is formulated mathematically through the minimization of some objective function. It combines context identification and multi-algorithm fusion criteria into a joint objective function. This optimization is intended to produce contexts as compact clusters in subspaces of the high-dimensional feature space via possibilistic unsupervised learning and feature weighting. Our initial experiments have indicated that the proposed fusion approach outperforms individual classifiers and the global fusion method. Also, in order to validate the obtained results, we compared the performance of the proposed approach with related fusion methods.

Keywords Supervised learning · Fusion · Social networks · Insult detection

1 Introduction

The widespread of smart devices and broadband internet connections yield an exponential growth of social networks. These networks are hosted and managed by very big companies which are employing thousands of people, and investing millions of dollars in order to improve their services, features, and performance. Also, millions of users gathered within these virtual societies, and formed several communities sharing the same interest and skills. Thus, blogs and social networks have become very active spaces where people express, comment, and share their opinions. However, the cultural heterogeneity of some users yields some misunderstanding of

M.M. Ben Ismail (✉) · O. Bchir
College of Computer and Information Sciences, King Saud University,
Riyadh, Kingdom of Saudi Arabia
e-mail: mbenismail@ksu.edu.sa

O. Bchir
e-mail: obchir1@ksu.edu.sa

each other comment meanings. In fact, a user may consider someone else comment inappropriate even though it was not meant to offend him. Moreover, sometimes verbal offenses, insults and other defamatory language shared by certain users cause hurt feelings, especially when they are addressed to “conservative” people. A natural solution to overcome this problem was to appoint human moderators monitoring online conversation. However, this solution can be expensive and labor intensive task for the moderator. Moreover, comments can be very frequent, which makes the process not efficient enough. The earliest efforts in this area were directed towards matching comments with a vocabulary of “prohibited” words. In other words, if the comment contains one or more keywords from the banned word list, then the comment is denied. These efforts posed the problem of insult detection as a string matching problem. In [1], the authors outlined a system which relies on a static dictionary and some patterns based on socio-linguistic. However, the obtained results proved that the approach suffers from high false positive rates and low coverage. The authors in [2] proposed an approach which consists in differentiating between insult and factive statements by parsing the sentences and using the semantic rules. The main drawback of this approach is its inability to discriminate efficiently between insults directed to non-participant and participant of conversation. The system proposed in [3] relies on a dictionary abusing language dictionary which is coupled with a bag-of-words features. However, these state-of-the-arts show two main drawbacks. The first one consists of the use of seed words and naive matching approach. The second drawback is that these solutions are not able to distinguish between insults directed towards people participating in blog/forum conversation and non-participants such as celebrities, public figures etc. In other words, comments which contain racial slurs or profanity may not necessarily be insulting to other users. Using machine learning techniques was a natural alternative to automatically detect verbal offenses in social network comments. More specifically, the problem has been perceived as text classification problem.

During the last decade, several researches was proposed to overcome the challenge of automatic insult detection in social network comments. In [4], topical features and lexicon features and various machine learning techniques have been used in order to detect offensive tweets. In [5], the authors proposed an approach that exploits linguistic regularities in profane language via statistical topic modeling on comments corpus. Another flame detection software was proposed in [3]. This tool applies models from multi-level classifiers, boosted by a dictionary of Insulting and Abusing Language. One can notice that these state-of-the-art solutions rely on efficient supervised learning algorithms [6]. These algorithms learn models from known examples (labeled comments) and use them in order to automatically classify new samples (unlabeled comments).

Supervised learning algorithms have been extensively applied to several challenges in real world applications. For instance, in [7], the authors outlined how classification is used to perform opinion mining. The authors in [8–10] proposed several approaches in order to either automatically classify emails using the subject or detect junk email [11]. Usually, it is admitted that there is no one best way to solve the challenges and it may be useless to argue which type of classification technique is

best [12]. Therefore, many approaches to combine the outputs of several classifiers were proposed in order to enhance the effectiveness of standard single classifier systems. Classifier fusion or ensemble classifier has become a very active research field, and promising results have been obtained with several applications [13]. Namely, fusion has been applied to pattern recognition, including character recognition [14], speech recognition [15], and text categorization [16], and have outperformed single classifier systems both theoretically and experimentally. Motivated by the classifiers' complementary characteristics, ensemble of classifiers can outperform individual algorithms by exploiting the advantages of the individual classifiers and limiting the effect of their disadvantages. Nonetheless, a necessary and sufficient conditions for a fusion classifier to be more accurate than single classifier are diversity and accuracy [17]. Classifier fusion methods rely on an effective combination of the classifiers outputs. This process considers all experts competitive and equally trained on the whole feature space. For unlabeled point, single experts are launched simultaneously. Then, the obtained outputs are combined in a way to take a group decision. The classifier combinations methods can be grouped based on how they assign weights to the single experts. Namely, global methods assign an average degree of worthiness over the feature space to each expert. On the other hand, local methods formulate the classifiers' worthiness with respect to different feature subspaces. These data-dependent weights, when learned properly more accurate classifiers. In [18], the authors outlined a method where they estimate the accuracy of each expert in local regions of the feature space neighboring an unlabeled test point. Then, the most accurate classifier in that specific local region is used for the final decision. However, the need to estimate the accuracy for each test sample makes the approach time-consuming. In [19], the clustering-and-selection method was proposed. Basically, this method selects statistically the best classifier. First, the training samples are clustered to form the decision regions. Then, the classifier that performed the best in terms of accuracy on this local region is chosen. However, the method was not generic enough to consider more than one classifiers for one region. The authors of [20] extended the clustering-and-selection approach, and exploited the class labels. In other words, they divided the training set into correctly and incorrectly classified samples. Then, they categorized them in order to form a partition of the feature space. For testing, they pick the most effective classifier based its accuracy in the vicinity of the input point in order to make the final decision. Thus, each classifier should maintain its own partition. This makes the decision process computationally expensive. Lately, in [21, 22] the authors presented a local fusion technique which partitions the feature space into homogeneous regions based on their features, and adopts the obtained feature space structure when performing the fusion. On the other hand, the fusion component assigns an aggregation weight to each detector in each context based on its relative performance within the context. However, the adopted fuzzy approach makes the fusion stage sensitive to outliers. In fact, outliers may affect the obtained partition and reduce the accuracy of the final decision.

To overcome this limitation, we propose a possibilistic based optimization to partition the feature space and the fusion of the classifiers. The partitioning of the feature space is based on the standard sum of within cluster distances. However, for

complex classification problems, the data is usually noisy which yields inaccurate partitions of the feature space. To alleviate this drawback, we propose a possibilistic based local approach that adapts the fusion method to different regions of the feature space. The aggregation weights are then estimated by the fusion component to each detector. These weights assignment is based on the relative performance of each detector within the context. Categorizing the input samples into regions during the training phase is a main requirement of the fusion component. Then, this approach appoints an expert for each region. These experts represent the best classifiers for the corresponding region.

2 Fusion Based on Possibilistic Context Extraction

Let N training observations with desired output $T = \{t_j | j = 1, \dots, N\}$. These outputs were obtained using K classifiers. Each classifier k extracts its own feature set $X_k = \{x_j^k | j = 1, \dots, N\}$ and generates confidence values, $Y^k = \{y_{kj} | j = 1, \dots, N\}$. The K feature sets are then concatenated to generate one global descriptor, $\chi = \bigcup_{k=1}^K \chi^k = \{x_j = [x_j^1, \dots, x_j^K | j = 1, \dots, N]\}$. The original Context Extraction for Local Fusion algorithm [9] minimizes

$$J = \sum_{j=1}^N \sum_{i=1}^C u_{ji}^m \sum_{s=1}^K v_{ik}^q d_{ijk}^2 + \sum_{j=1}^N \sum_{i=1}^C \beta_i u_{ji}^m \left(\sum_{k=1}^K \omega_{ik} y_{kj} - t_j \right)^2, \quad (1)$$

subject to $\sum_{i=1}^C u_{ji} = 1 \forall j$, $u_{ji} \in [0, 1] \forall i, j$, $\sum_{k=1}^K v_{ik} = 1 \forall i$, $v_{ik} \in [0, 1] \forall i, k$, and $\sum_{k=1}^K \omega_{ik} = 1 \forall i$.

The first term in (1) corresponds to the objective function of the Fuzzy C-Means (FCM) algorithm [23]. It is intended to categorize the N points into C clusters centered in c_i . Each data point x_j will be assigned to all clusters with fuzzy membership degrees. When a partition of C compact clusters with minimum sum of intra-cluster distances is discovered this FCM term is minimized. The second term in (1) attempts to learn cluster-dependent aggregation weights of the K algorithm outputs. ω_{ik} is the aggregation weight assigned to classifier k within cluster i . This term is minimized when the aggregated partial output values match the desired output. When both terms are combined and β is chosen properly, the algorithm seeks to partition the data into compact and homogeneous clusters while learning optimal aggregation weights for each algorithm within each cluster.

For real world classification problems, multiple sources of information and multiple classifiers for each source may be needed to obtain satisfactory results. In this case, the resulting feature space can be noisy and high dimensional. This complicates the clustering task, and the true partition of the data cannot be generated. This is due to the influence of the noisy points on the obtained clusters. To alleviate this drawback, we propose a possibilistic version of the algorithm. The proposed algorithm generates possibilistic memberships in order to represent the degree of typicality of each data point within every category, and reduce the influence of noise

points on the learning process. We extend the objective function (1), and formulate the context extraction for local fusion using the following objective function

$$J = \sum_{j=1}^N \sum_{i=1}^C u_{ji}^m \sum_{k=1}^K v_{ik}^q d_{ijk}^2 + \sum_{j=1}^N \sum_{i=1}^C \beta_i u_{ji}^m \left(\sum_{k=1}^K \omega_{ik} y_{kj} - t_j \right)^2 + \sum_{i=1}^C \eta_i \sum_{j=1}^N (1 - u_{ji})^m, \quad (2)$$

In (2), u_{ji} represents the possibilistic membership of \mathbf{X}_j in cluster i . The $M \times N$ matrix, $U = [u_{ji}]$ is called a possibilistic partition if it satisfies:

$$\begin{cases} u_{ji} \in [0, 1], & \forall j \\ 0 < \sum_{i=1}^C u_{ji} < N \quad \forall i, j \end{cases} \quad (3)$$

On the other hand the $M \times d$ matrix of feature subset weight, $V = [v_{ik}]$ satisfies

$$\begin{cases} v_{ik} \in [0, 1] \quad \forall i, k \\ \sum_{k=1}^K v_{ik} = 1 \quad \forall i \end{cases} \quad (4)$$

In (2), $m \in [1, \infty)$ is called the fuzzier, and η_i are positive constants that controls the importance of the second term with respect to the first one. This term is minimized when u_{ji} are close to 1, thus, avoiding the trivial solution of the first term (where $u_{ji} = 0$). Note that $\sum_{i=1}^C u_{ji}$ is not constrained to sum to 1. In fact, points that are not representative of any cluster will have $\sum_{i=1}^C u_{ji}$ close to zero and will be considered as noise. This constraint relaxation overcomes the disadvantage of the constrained fuzzy membership approach which is the high sensitivity to noise and outliers. The parameter η_i is related to the resolution parameter in the potential function and the deterministic annealing approaches. It is also related to the idea of “scale” in robust statistics. In any case, the value of 0.7 determines the distance at which the membership becomes 0.5. The value of η_i determines the “zone of influence” of a point. A point \mathbf{X}_j will have little influence on the estimates of the model parameters of a cluster if $\sum_{k=1}^K v_{ik}^q (d_{ijk})^2$ is large when compared with η_i . On the other hand, the “fuzzier” m determines the rate of decay of the membership value. When $m = 1$, the memberships are crisp. When $m \rightarrow \infty$, the membership function does not decay to zero at all. In this possibilistic approach, increasing values of m represent increased possibility of all points in the data set completely belonging to a given cluster.

Setting the gradient of J with respect to u_{ji} to zero yields the following necessary condition to update the possibilistic membership degrees [24]:

$$u_{ji} = \left[1 - \left(\frac{D_{ij}^2}{\eta_j} \right)^{\frac{1}{m-1}} \right]^{-1}. \quad (5)$$

where $D_{ij} = \sum_{k=1}^K v_{ik}^q d_{ijk}^2 + \beta \sum_{k=1}^K v_{ik}^q \left(\sum_{l=1}^K \omega_{il} y_{lj} - t_j \right)^2$. D_{ij} represents the total cost when considering point x_j in cluster i . As it can be seen, this cost depends on the distance between point x_j and the cluster's centroid c_i , and the deviation of the combined algorithms' decision from the desired output (weighted by β). More specifically, points to be assigned to the same cluster: (i) are close to each other in the feature space, and (ii) their confidence values could be combined linearly with the same coefficients to match the desired output.

Minimizing J with respect to the feature weights yields

$$v_{ik} = \sum_{l=1}^K \left[(D_{ik}^2 / D_{il})^{\frac{1}{q-1}} \right] \quad (6)$$

where $D_{il} = \sum_{j=1}^N u_{ij}^m d_{ijl}^2$.

Minimization of J with respect to the prototype parameters, and the aggregation weights yields

$$c_{jk} = \frac{\sum_{j=1}^N u_{ij}^m \mathbf{X}_{jk}}{\sum_{j=1}^N u_{ij}^m}. \quad (7)$$

and

$$w_{ik} = \frac{\sum_{j=1}^N u_{ij}^m y_{kj} \left(t_j - \sum_{l \neq k}^K \omega_{il} y_{lj} \right) - \zeta_i}{\sum_{j=1}^N u_{ij}^m y_{kj}^2}. \quad (8)$$

where ζ_i is a Lagrange multiplier that assures that the constraint in (3) is satisfied, and is defined as

$$\zeta_i = \frac{\sum_{l=1}^K \frac{\sum_{j=1}^N u_{ij}^m y_{lj} \left(t_j - \sum_{k=1}^K \omega_{ik} y_{kj} \right)}{\sum_{j=1}^N u_{ij}^m y_{lj}^2}}{\sum_{l=1}^K \frac{1}{\sum_{j=1}^N u_{ij}^m y_{lj}^2}}. \quad (9)$$

The behavior of this algorithm depends on the value of β . Over estimating it yields the multi-algorithm fusion criteria to be dominant which results in non-compact clusters. On the other hand, a small value of β reduces the influence of the multi-algorithm fusion criteria and categorizes the data based mainly on the distances in the feature space.

The obtained algorithm is an iterative algorithm that starts with an initial partition and alternates between the update equations of u_{ji} , v_{ik} , and c_{ik} . It is summarized below.

Algorithm 1 Fusion algorithm based on Possibilistic Context Extraction

```

Begin
  Fix the number of clusters  $C$ ;
  Fix  $m$ ,  $q$  and  $\beta$ .
  Initialize the centers and the possibilistic  $M$  partition matrix  $U$ ;
  Initialize the relevance weights to  $1/K$ ;
  Repeat
    Compute  $d_{ijk}^2$ , for  $1 \leq i \leq C$  and  $1 \leq j \leq N$  and  $1 \leq k \leq K$  ;
    Update the relevance weights  $v_{ik}$  using equation (6);
    Compute  $D_{ij}^2$ 
    Update the partition matrix  $U$  using equation (5);
    Update the partition matrix  $W$  using equation (8);
    Update the centers using equation (7);
  Until (centers stabilize)
End

```

3 Experiments

A range of experiments were performed to assess the strengths and weaknesses of the proposed approach. We used the KAGGLE data [25]. This collection of social commentary consists of two subsets. The first one represents the training set with 3948 comments. The second subset is the testing collection, and it consists of 2235 comments.

First, we preprocessed the comments collection in order to discard some encoding parts that may affect the results, gather similar words with stemming and discard the less frequent words. For instance, a raw comments looks like “\ \xc\ \xa0If you take out the fags and booze...”. After preprocessing it, we obtain “If you take out the fags and booze...”. Also, we deleted words starting with “@”. Then, substituted words like “u” to “you”, and “da” to “the” etc.

For feature extraction, we used standard TFIDF [26] technique in order to map comments x_j into a compact representation of its content. Thus, each comment x_j was represented using one 800-dimensional feature vector $x_j = \langle w_{1,j}, \dots, w_{|\tau|,j} \rangle$. Where τ is the vocabulary of words that occur at least once in at least one comment, and $0 < w_{k,j} < 1$ represents how much the k th word contributes to the semantics of comment x_j .

Due to the space limitation, the effect of these parameters cannot be illustrated in this work. To adapt this data to our application, we assume that we have 3 sets of features and that we have one classifier for each set. These sets are extracted as subsets from the original features. Specifically, The first subset includes features from 1 to 400. The second one includes features from 200 to 600, and the third subset includes features 400–800. For each set, we use a simple K-NN classifier to generate confidence values. We classify the training data using the 3 K-NN classifiers with their appropriate feature subsets. Then, we use the proposed local fusion to partition the training data into 3 clusters. For each cluster, the algorithm learns the optimal aggregation weights. The testing phase starts by classifying the test point using the

Table 1 Confusion matrices obtained using three single classifiers, the Method in [9], and the proposed method, respectively

		Predicted as not insult	Predicted as insult
Classifier 1			
	Not insult	1449	505
	Insult	538	155
Classifier 2			
	Not insult	1352	602
	Insult	545	148
Classifier 3			
	Not insult	1448	506
	Insult	535	158
Method in [9]			
	Not insult	1320	591
	Insult	495	198
Proposed method			
	Not insult	1309	645
	Insult	430	290

four classifiers and generating the corresponding partial confidence values. Then, we assign it to the closest cluster. Finally, the final decision is obtained by combining the four partial confidence values using the aggregation weights of the closest cluster.

Table 1 shows the confusion matrix obtained using the three single classifiers, the Method in [9], and the proposed method. One can notice that the proposed method outperforms the other approaches in terms of Specificity. More specifically, our approach detected about 50 % more insult comments than single classifiers. On the other hand, our approach classifies less accurately non-insult comments which yields lower sensitivity value. Despite this discrepancy between sensitivity and specificity, we consider these results promising because for the problem of automatic detection of insults in social network comets, we assume that the True Negative predictions are not equally relevant to True Positive ones. In other words, we do not consider misclassifying an insult comment as serious as misclassifying a non-insult one. Moreover, since the testing data contains 720 insulting comments only out of 2674 comments, the accuracy cannot be an appropriate performance measure for this application. In Table 1, we show the aggregation weights learned by the proposed algorithms. As it can be seen, the relative performances of the individual classifiers varies significantly from one cluster to another. For instance for cluster 1, classifier 3 outperforms the other classifiers. Consequently, this classifier is considered the most reliable one for this cluster and is assigned the highest aggregation weight as shown in Table 2. Similarly, for cluster 3, classifier 2 is assigned the highest weight.

Table 2 Learned weights for each classifier in each cluster

Cluster #	1	2	3
Classifier 1	0.4295	−0.0475	0.7301
Classifier 2	0.2198	0.6888	−0.0019
Classifier 3	0.3536	0.3297	0.2811

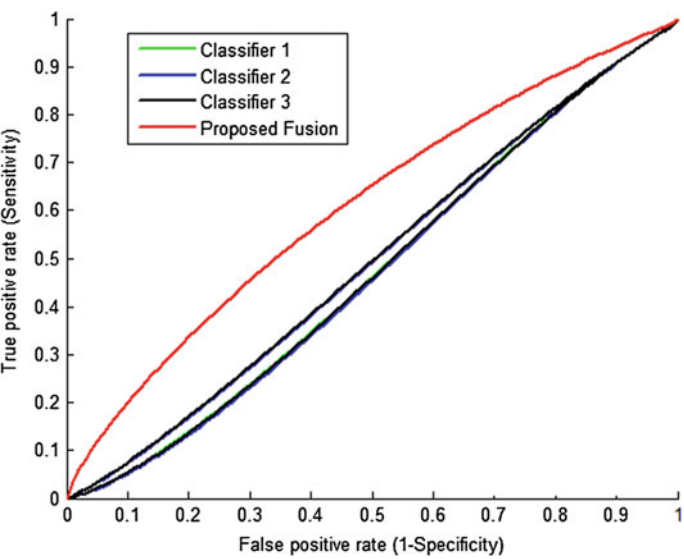


Fig. 1 Comparison of the three individual classifiers that use subsets of the features with the proposed fusion

Figure 1 displays Receiver Operating Characteristic (ROC) curve that compare the performance of the individual classifiers and the proposed fusion performances. As one can see, the proposed fusion outperforms the three individual classifiers.

To illustrate the local fusion ability of the proposed approach, we display the accuracy of the three individual classifiers (K - NN) for the 3 clusters. These accuracy values are obtained based on the crisp partition generated by the proposed algorithm, and testing the samples within each cluster independently. These results are displayed in Table 3 As it can be seen, the relative performances of the individual K - NN varies from one cluster to another. For instance in cluster 1, classifier 1 overcomes the classifier 2 and classifier 3. Consequently, for cluster 1 the most relevant classifier

Table 3 Per-cluster accuracy within each cluster obtained using the proposed Fusion

Cluster #	Cluster 1	Cluster 2	Cluster 3
Classifier 1	0.9066	0.7923	0.9637
Classifier 2	0.8934	0.8239	0.9592
Classifier 3	0.8574	0.7878	0.9589

is classifier 1. Thus, the highest aggregation weight is assigned to this classifier as shown in Table 2. Similarly, in cluster 2, the highest weights is assigned to classifier 2.

4 Conclusion

In this paper we have proposed a novel approach of automatic insult detection in social network comments. This approach relies on a local multi-classifier fusion method. Specifically, it categorizes the feature space into homogeneous clusters where a linear aggregation of the different classifier outputs yields a more accurate decision. The clustering process generates a possibilistic membership degree that represents the typicality, and is used to identify and discard noise frames. Moreover, the proposed algorithm provides optimal fusion parameters for each context. The initial experiments have shown that the fusion approach outperforms the individual classifiers performance. In order to overcome the need to specify the number of clusters a priori, a nice property of the possibilistic approach to generate duplicated clusters can be investigated in order to find the optimal number of clusters in an unsupervised manner.

Acknowledgments This work was supported by the Research Center of College of Computer and Information Sciences, King Saud University (Project RC131013). The authors are grateful for this support.

References

1. Spertus, E., Smokey: Automatic recognition of hostile messages. In: Proceedings of the Ninth Conference on Innovative Applications of Artificial Intelligence, pp. 1058–1065 (1997)
2. Mahmud, A., Ahmed, K.Z., Khan, M., Detecting flames and insults in text. In: Proceedings of the Sixth International Conference on Natural Language Processing (2008)
3. Razavi, A.H., Inkpen, D., Uritsky, S., Matwin, S., Offensive language detection using multi-level classification. In: Proceedings of the 23rd Canadian Conference on Artificial Intelligence, pp. 16–27 (2010)
4. Xiang, G., Hong, J., & Rosé, C. P., Detecting Offensive Tweets via Topical Feature Discovery over a Large Scale Twitter Corpus, Proceedings of The 21st ACM Conference on Information and Knowledge Management, Sheraton, Maui Hawaii, October 29–November 2, (2012)
5. Xiang, G., Fan, B., Wang, L., Jason I., Carolyn, H., Rose, P., Detecting Offensive Tweets via Topical Feature Discovery over a Large Scale Twitter Corpus, Proceeding of the 21st ACM international conference on Information and knowledge management (CIKM '12), pp. 1980–1984 (2012)
6. Namburu, S.M., Tu, H., Luo, J., Pattipati, K.R., Experiments on Supervised Learning Algorithms for Text Categorization. International Conference, IEEE Computer Society, pp.1–8 (2005)
7. Liu, B., Zhang, L.: A survey of opinion mining and sentiment analysis. In: Aggarwal, C., Zhai, C. (eds.) Mining Text Data. Springer, Berlin (2011)
8. Lewis, D., Knowles, K.: Threading electronic mail: a preliminary study. Inf. Process. Manag. 33(2), 209–217 (1997)

9. Cohen, W., Learning rules that classify e-mail. AAAI Conference (1996)
10. de Carvalho, V.R., Cohen, W., On the collective classification of email “speech acts”, ACM SIGIR Conference (2005)
11. Sahami, M., Dumais, S., Heckerman, D., Horvitz, E., A Bayesian approach to filtering junk e-mail. AAAI Workshop on Learning for Text Categorization. Technical Representation WS-98-05, AAAI Press. <http://robotics.stanford.edu/users/sahami/papers.html>
12. Bi, Y., Bell, D., Wang, H., Guo, G., Guan, J.: Combining multiple classifiers using dempster’s rule for text categorization. *Appl. Artif. Intell.* **21**(3), 211–239 (2007)
13. Kuncheva, L.I.: *Combining Pattern Classifiers*. Wiley, New York (2004)
14. Sirlantzis, K., Hoque, S., Fairhurst, M. C., Trainable multiple classifier schemes for handwritten character recognition. In: *Proceedings of the 3rd International Workshop on Multiple Classifier Systems*, pp. 319–322, Cagliari, Italy (2002)
15. Huenupan, F., Yoma, N.B., Molina, C., Garreton, C.: Confidence based multiple classifier fusion in speaker verification. *Pattern Recognit. Lett.* **29**(7), 957–966 (2008)
16. Minsky, M.: Logical versus analogical or symbolic versus connectionist or neat versus scruffy. *AI Mag.* **12**(2), 34–51 (1991)
17. Hansen, L.K., Salamon, P.: Neural network ensembles. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(10), 993–1001 (1990)
18. Woods, K., Kegelmeyer Jr, W.P., Bowyer, K.: Combination of multiple classifiers using local accuracy estimates. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(4), 405–410 (1997)
19. Kuncheva, L., Clustering-and-selection model for classifier combination. In: *Proceedings of Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies*, vol. 1, pp. 185–188 (2000)
20. Liu, R., Yuan, B.: *Multiple classifiers combination by clustering and selection*. Information Fusion, pp. 163–168. Elsevier, New York (2001)
21. Frigui, H., Zhang, L., Gader, P.D., Ho, D., Context-dependent fusion for landmine detection with ground penetrating radar. In: *Proceedings of the SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets*, Orlando, FL, USA, 2007
22. Abdallah, A.C.B., Frigui, H., Gader, P.D.: Adaptive Local Fusion With Fuzzy Integrals. *IEEE T. Fuzzy Syst.* **20**(5), 849–864 (2012)
23. Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York (1981)
24. Krishnapuram, R., Keller, J.: A possibilistic approach to clustering. *IEEE Trans. Fuzzy Syst.* **1**, 98–110 (1993)
25. <http://www.kaggle.com/c/detecting-insults-in-social-commentary/prospector#169> (2013)
26. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. *Inform. Process. Man.* **24**(5), 513–523 (1988). Also reprinted in Sparck Jones and Willett [1997], pp. 323–328



<http://www.springer.com/978-3-319-10508-6>

Computer and Information Science

Lee, R. (Ed.)

2015, XIII, 209 p. 107 illus., Hardcover

ISBN: 978-3-319-10508-6