A Generic Classifier-Ensemble Approach for Biomedical Named Entity Recognition

Zhihua Liao[1] and Zili Zhang[2,3,⋆]

[1] Modern Foreign-Language Education Technology Center, Foreign Studies College
Hunan Normal University, CS 410081, China
[2] Faculty of Computer and Information Science, Southwest University, CQ 400715, China
[3] School of Information Technology, Deakin University, VIC 3217, Australia
liao.zhihua61@gmail.com, zzhang@deakin.edu.au

Abstract. In named entity recognition (NER) for biomedical literature, approaches based on combined classifiers have demonstrated great performance improvement compared to a single (best) classifier. This is mainly owed to sufficient level of diversity exhibited among classifiers, which is a selective property of classifier set. Given a large number of classifiers, how to select different classifiers to put into a classifier-ensemble is a crucial issue of multiple classifier-ensemble design. With this observation in mind, we proposed a generic genetic classifier-ensemble method for the classifier selection in biomedical NER. Various diversity measures and majority voting are considered, and disjoint feature subsets are selected to construct individual classifiers. A basic type of individual classifier – Support Vector Machine (SVM) classifier is adopted as SVM-classifier committee. A multi-objective Genetic algorithm (GA) is employed as the classifier selector to facilitate the ensemble classifier to improve the overall sample classification accuracy. The proposed approach is tested on the benchmark dataset – GENIA version 3.02 corpus, and compared with both individual best SVM classifier and SVM-classifier ensemble algorithm as well as other machine learning methods such as CRF, HMM and MEMM. The results show that the proposed approach outperforms other classification algorithms and can be a useful method for the biomedical NER problem.

1 Introduction

With the wide applications of information technology in biomedical field, biomedical technology has developed very rapidly. This in turn produces a large amount of biomedical data such as human gene bank. Consequently, biomedical literature available from the Web has experienced unprecedented growth over the past few years. The amount of literature in *MEDLINE* grows by nearly 400,000 citations each year. To mine information from the biomedical databases, a helpful and useful pre-processing step is to extract the valuable biomedical named entity. In other words, this step needs to identify some names from scientific text that is not structured as traditional databases and classify these different names. As a result, biomedical named entity recognition (BioNER) becomes one of the most important issues in automatic text extraction system. Many

⋆ Corresponding author.

P.-N. Tan et al. (Eds.): PAKDD 2012, Part I, LNAI 7301, pp. 86–97, 2012.

popular classification algorithms have been applied to this bioNER problem. These algorithms include Support Vector Machine (SVM) [1,18,19], Conditional Random Fields (CRFs) [3], the Hidden Markov Model (HMM) [5], the Maximum Entropy (ME) [15], decision tree [16], and so on. While successful, each classifier has its own shortcomings and none of them could consistently perform well over all different datasets. To overcome the shortcomings of individual methods, ensemble method has been suggested as a promising alternative.

Ensemble method is more attractive than individual classification algorithm in that it is an effective approach for improving the prediction accuracy of a single classification algorithm. An ensemble of classifiers is a set of classifiers whose individual decisions are combined in some way (typically by weighted or unweighted voting) to classify new examples [8,11]. One of the most active areas of research in supervised learning has been to study methods for constructing good ensembles of classifiers. The most important property of successful ensemble methods is if the individual classifiers have error rate below 0.5 when classifying sample data while these errors are uncorrelated at least in some extent. That is, a necessary and sufficient condition for an ensemble of classifiers over its individual members is that the classifiers are accurate and diverse. Several recent studies indicate that the ensemble learning could improve the performance of a single classifier in many real world text classification [6,7,9,10,12,13,14,23,24].

In this paper, we propose a generic genetic classifier-ensemble approach, which employs multi-objective genetic algorithm and SVM based classifiers to construct an ensemble classifier. Each SVM based classifier is trained on a different feature subset and used as the classification committee. The rest of the paper is organized as follows: Section 2 discusses the generic genetic classifier-ensemble approach in detail. Experimental results and analysis are provided in Section 3. Conclusions and future work are presented in Section 4.

2 The Generic Genetic Classifier-Ensemble Approach

Classifier-ensemble is a popular technique in pattern recognition domain. It reflects the generalization accuracy if an ensemble depends not only on the performances of the individual classifier but also on the diversity among the classifiers [6,8,10,7,12,22]. Therefore, a classifier-ensemble system is usually made up of two major components: the classifiers forming the ensemble members and the combination scheme. In order to achieve this goal, we develop a generic genetic classifier-ensemble algorithm. In the proposed approach, SVM is used as the basic classifier and the genetic algorithm was used to search the optimal solution of weighted classifier combination.

2.1 Feature Set and SVM Based Classifier

Since the main issue using machine learning method for BioNER task is to design a proper feature set, choosing the suitable feature is very important for improving the performance of the system. Here various types of features have been considered for bioNER task in different combinations (see Table 1).

- Word: All words appearing in the training data.
- Orthography: Table 2 shows the orthographic features. If the token has more than one feature, then we used the feature list of Table 2 from left to right and from up to down orderly.
- Prefix: Uni-,bi-, and tri-grams(in letters) of the starting letters of the current token.
- Suffix: Uni-,bi-, and tri-grams(in letters) of the ending letters of the current token.
- Lexical: POS tags, base phrase classes, and base noun phrase chunks. POS tags are generated by Geniatagger[1].
- Preceding class: The prediction of the classifier for the preceding tokens are computed dynamically and used as feature.
- Surface word: Surface words forming a list of tokens that are tagged as an entity in the training data. In our system, the surface word includes simple surface word lists, name aliases and trigger words [17,21].

Table 1. The features in our generic genetic classifier-ensemble system

Feature	Value
words	all words in the training data
orthographic	capital, symbol, etc.(see Table 2)
prefix	1,2, and 3 gram of starting letters of word
suffix	1,2, and 3 gram of ending letters of word
lexical	POS tags, base phrase classes, and base noun phrase chunks
preceding class	-4,-3, -2, -1
surface word	simple surface word lists, name aliases and trigger words

Table 2. Orthographic features

Feature	Example	Feature	Example
DigitNumber	15	Greek	alpha
SingleCap	M	CapsAndDigits	I2
TwoCaps	RalGDS	LettersAndDigits	p52
InitCaps	Interleukin	LowCaps	kappaB
Lowercase	kinases	Hyphen	-
Backslash	/	OpenSquare	[
CloseSquare	]	Colon	:
SemiColon	;	Pecent	%
OpenParen	(	CloseParen	)
Comma	,	FullStop	.
Determiner	the	Conjunction	and
Other	* @		

[1] http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/tagger/

Table 3. The parameters of Yamcha

Parameter	Value
kernel	polynomial
degree of kernel	1,2,3
direction of parsing	forward, backward
windows position	9 words(position -4, -3,-2,-1,0,+1,+2,+3,+4)
multi-class	pair-wise

Next, due to the fact that support vector machines(SVMs) are powerful methods for learning a classifier and have been applied successfully to many NLP tasks, SVMs construct the base classifier in BioNER. The general-purpose text chunker named Yet Another Multipurpose Chunk Annotator-Yamcha[2] uses TinySVM[3] for learning the classifiers. Yamcha is utilized to transform the input data into feature vectors usable by TinySVM [18,19]. Table 3 shows the Yamcha parameters. Accordingly, each classifier is unique in at least one of the following properties: window size, degree of the polynomial kernel, parsing direction as well as feature set. Consequently, this constructs 46 individual SVM classifier committees [17,20,21].

2.2 Generic Genetic Classifier-Ensemble Algorithm

The genetic algorithm (GA) was developed in the 1970s by Holland as an effective evolutionary optimization method [25]. In GA the two core elements are chromosome and fitness. Chromosome is used to encode representation of the optimal solution to the classifier-ensemble problem. Fitness is designed to measure the chromosome's performance.

Genetic Classifier-Ensemble-I. The basic idea behind the genetic classifier-ensemble-I is that different classes in each classifier differ with contributing degrees of prediction classes. In other words, each class in each classifier has been assigned a weight which corresponds with the contributing degree of prediction class. To use genetic algorithm, we first need to represent the problem domain as a chromosome. Here, we want to find an optimal set of weight for classifier ensemble scheme shown in Figure 1. Assume that there are totally N tags (classes) corresponding to the named entities considered in the BioNER task. Set the total number of available classifiers denoted by M. The optimal weight solution of the classifier ensemble scheme is encoded in the form of a weight chromosome,which has N*M genes. First N genes belong to the first classifier and the next N genes the second classifier and so on. The encoding of a chromosome is illustrated in Figure 1. Each value of gene in the chromosome is initialized to a small random number, said within the range[0,1]. Thus, we obtain a chromosome.

The second step is to define a fitness function for evaluating the chromosome's performance. This function must estimate the performance of a given classifier-ensemble

[2] http://cl.aist-nara.ac.jp/∼taku-ku/software/yamcha/

[3] http://chasen.org/∼taku/software/TinySVM/

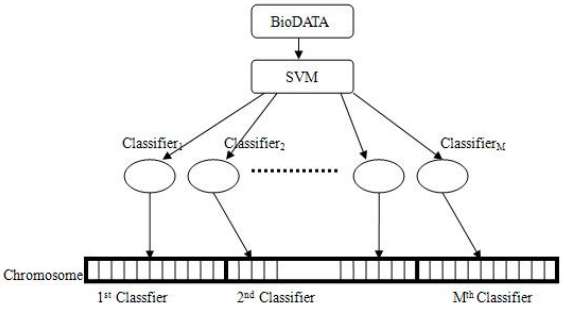

Fig. 1. Genetic Classifier-Ensemble-I

problem with weights. We define the fitness of a chromosome as the full object F-score provided by the weighted majority voting type decision combination rule [12,17,22]. In this rule, the class receiving the maximum combined score is selected as the joint decision. By the definition of the combined score of a particular class,

$$f(c_i) = \sum_{m=1}^{M} F_m \cdot w(m,i)$$

we obtain the fitness as follows:

$$f_n(c_l) = max(f(c_1), f(c_2), \cdots, f(c_n))$$

where M denotes the total number of classifiers and F_m denotes the full object F-score of *m*th classifier. *w(m,i)*is assigned to a weight value in the gene of *i*th class of *m*th classifier in the chromosome.

The third step is to choose the genetic operators-crossover and mutation. A crossover operator takes two parent chromosomes and creates two children with genetic material from both parents. In the proposed approach, either uniform or two point crossover method is randomly selected with equal probability. The selected operator is applied with a probability p_{cross} to generate two offspring. A mutation operator randomly selects a gene in offspring chromosomes with a probability p_{mut} and adds a small random number within the range[0,1] to each weight in the gene. In addition, we still need to specify the tournament size,elitism, population size and the number of generations. Tournament size is used in tournament selection during the reproduction. Elitism is applied at the end of each iteration where the best *elit_size%* of the original population are used to replace those in the offspring producing the lowest fitness.

Genetic Classifier-Ensemble-II. The basic principle behind the genetic classifier-ensemble-II is that different classifiers have different contributing degrees of prediction of classes. In other words, each classifier can be assigned a weight which corresponds with the contributing degree of prediction of class. Suppose each chromosome is encoded as a weight string having M genes, one for each classifier(see Figure 2). If the

value of a gene is w_m, this means that the contributing degree of the *m*th classifier in this ensemble is w_m. Accordingly, the combined score of a given class can be redefined as:

$$f(c_i) = \sum_{m=1}^{M} F_m \cdot w_m$$

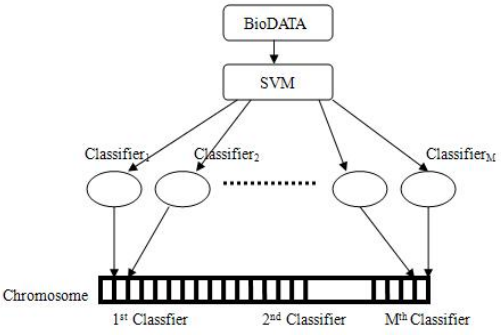

Fig. 2. Genetic Classifier-Ensemble-II

At the same time, all parameters of this algorithm described above including population size, the number of generations, crossover and mutation rate etc. are kept the same.

Genetic Classifier-Ensemble-III. Based on the above consideration in both subsections 2.2.1 and 2.2.2, not only contributing degrees of prediction classes among different classes in the same classifier are different, but also contributing degrees of prediction classes among different classifiers differ. Thus, the chromosome is made up of the chromosome in genetic classifier-ensemble-I and the chromosome in genetic classifier-ensemble-II, and has (N+1)*M genes (see Figure 3). Therefore, the combined score of a given class is determined as:

$$f(c_i) = \sum_{m=1}^{M} F_m \cdot w(m,i) \cdot w_m$$

Similarly, all the other parameters are kept the same.

After given the definition of chromosome and fitness as well as all parameters, the complete genetic classifier-ensemble algorithm can be described in the following steps:

1. Generate randomly an initial chromosome population of size **MAX_POPULATION**
2. For each chromosome in the population
   2.1 Apply weighted majority to all classifiers vector
   2.2 Compute full object **F-score** as fitness of the chromosome

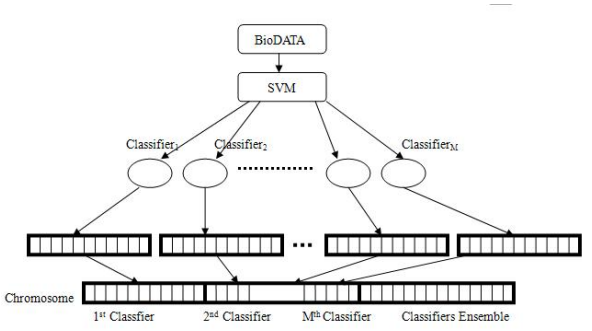

Fig. 3. Genetic Classifier-Ensemble-III

3. For generation_index in 1 ...**MAX_GENERATION**
   3.1 For chromosome_index in 1 ...**MAX_POPULATION**
   - Select two parents from the old population
   - Crossover the two parents to produce two offspring with probability p_{cross}
   - Mutate each gene of each offspring with probability p_{mut}
   - Apply weighted majority to each of the offspring
   - Compute full object **F-score** as fitness of each offspring

   3.2 Replace the worst **ELIT_SIZE%** of the offspring with the best chromosomes from the original population to form the new population
4. Select the best chromosome as the resultant ensemble

Figure 4 presents the flow of the proposed generic genetic classifier-ensemble algorithm.

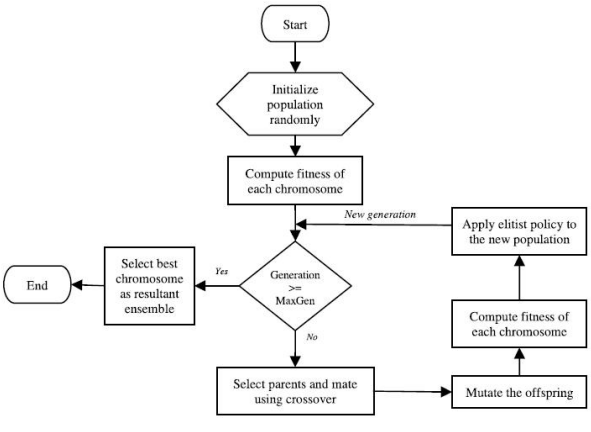

Fig. 4. The flow of the proposed generic genetic classifier-ensemble algorithm

The overall system architecture is illustrated in Figure 5. The best-fitting solution of weighted classifier-ensemble is obtained by using the classifier outputs generated

through three-fold cross-validation on the training data. In our proposed algorithm, the training data is initially partitioned into three parts. Each classifier is trained using two parts and then tested with the remaining part. This procedure is repeated three times and the whole set of training data is used for computing the best-fitting solution. Multi-class SVM is used for all individual classifier. The major differences among the individual classifiers are in their modeling parameter values and feature sets. Each classifier is different from the rest in at least one modeling parameter or the feature set. During testing, the outputs of the individual classifiers are combined by using the computed best-fitting solution of weight classifier-ensemble.

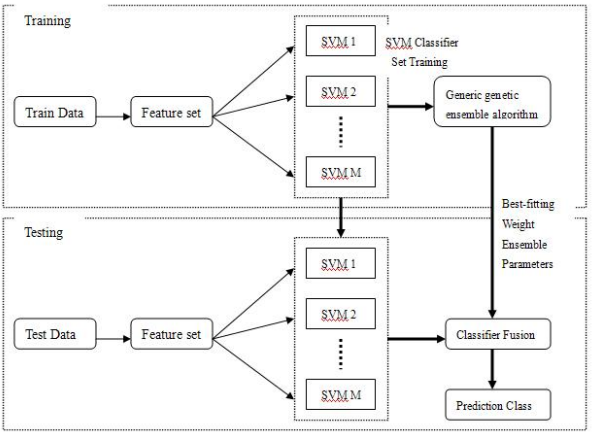

Fig. 5. Overall system architecture

3 Experiments and Results

To conduct the experiment, we use the latest GENIA[4] version 3.02 corpus provided by the shared task in COLING 2004 JNLPBA. The corpus includes the training dataset and the testing dataset. The training dataset consists of 2000 MEDLINE abstracts of the GENIA corpus with named entities in IOB2 format. The testing dataset consists of 404 abstracts. There are 18546 sentences and 492551 words in the training dataset and 3856 sentences and 101039 words in the testing dataset. Each word is tagged with "B-X", "I-X", or "O" to indicate that the word is at the "beginning"(B) or "inside"(I) of a named entity of type X, or "outside"(O) of a named entity. For BioNER task, the named entity types are DNA, RNA, cell_line, cell_type, and protein. Table 4 shows the number of 5 different biomedical named entities in this corpus. For each entity, two different tags(classes) result in 10 tags for the named entities and one additional tag for all non-named entities called class. Accordingly, this translate to a total of N=11 classes. Besides, we present M=46 single SVM base classifier committees on the basis

[4] http://www-tsujiii.is.s.u-tokyo.ac.jp/GENIA/

of different combination within feature set and Yamcha parameter. The experimental performance is evaluated by the standard measures, namely precision, recall and F-score which is the harmonic mean of precision and recall.

Table 4. Number of different biomedical named entities in GENIA 3.02 corpus

Types	Train data	Test data
DNA	9,534	1,056
RNA	951	118
Cell_line	3,830	500
Cell_type	6,718	1,921
Protein	30,269	5,067
Total	51,302	8,662

In the simulation experiments, The tournament size, crossover probability, mutation probability and elitism ratio are empirically computed as 40, 0.7, 0.02, and 20%, respectively. The population size of the generic genetic classifier-ensemble algorithm is set to 100. This means that one hundred different ensemble candidates evolve simultaneously. The algorithm is run for 10000 iterations. The weight classifier-ensemble corresponding to the chromosome with the highest fitness value in the last generation is selected as the optimal solution. We perform simulation experiments repeatedly by changing the weight values of these chromosomes and selected the weight genes of the chromosome providing the best performance of BioNER on the training data. In the testing, the test data is measured by using the optimal solution. This solution provides the best-fitting ensemble parameter with weights in the simulation experiments.

Table 5 shows the performance of the proposed three genetic classifier-ensemble scheme on precision, recall, and Fscore for BioNER. In this table, the genetic classifier-ensemble-III gets the better results compared with the genetic classifier-ensemble-I and genetic classifier-ensemble-II, where the performance of precision, recall and Fsore reach 75.65%, 78.52%, and 77.85% respectively.

It can be seen that in Table 6 we compare our best result with those of the recent work that employ support vector machines as classifier. The individual best SVM-classifier has the full feature set and optimal setting parameters[20,21]. Dimililer et al. used a vote-based classifier selection approach to construct a classifier ensemble and effective post-processing techniques for biomedical named entity recognition task[17,20,21]. Compared with the individual best SVM-classifier and SVM-classifier ensemble, our method outperforms them. It means that our generic genetic classifier-ensemble approach which searched the best-fitting ensemble parameter with weights can be powerful and efficient to combine orderly individual SVM base classifier with their strengths through giving the corresponding weights and to avoid individual classifier's weakness.

Table 7 shows that the best result of our experiment outperforms that of other individual classifier algorithms [26]. Their approaches include the Hidden Markov Model (HMM) [5], the Maximum Entropy Markov Model (MEMM) [4] and the Conditional Random Field (CRF) [3], which use deep knowledge resources with extra costs in

Table 5. The performances of different biomedical named entities on three genetic classifier-ensemble schemes

	Genetic Classifier-ensemble-I			Genetic Classifier-ensemble-II			Genetic Classifier-ensemble-III		
Types	**Precision**	**Recall**	**F-score**	**Precision**	**Recall**	**F-score**	**Precision**	**Recall**	**F-score**
DNA	73.54	74.25	72.92	70.68	70.98	70.76	74.65	75.59	75.21
RNA	74.33	75.85	75.22	71.15	70.25	70.46	75.88	76.79	76.42
Cell_line	72.50	71.56	72.12	68.25	67.20	67.82	74.60	73.82	74.36
Cell_type	73.15	72.87	72.04	69.62	72.58	70.37	74.85	75.39	75.06
Protein	83.36	76.58	79.65	80.56	71.25	75.86	84.58	80.06	83.57
Total	74.33	73.52	73.86	71.28	71.02	71.16	75.65	78.52	77.85

Table 6. The comparison with individual best SVM classifier and Vote-based SVM-classifier selection for bioNER task

Approaches	**Precision**	**Recall**	**F-score**
Single best SVM-classifier[20,21]	69.40	70.60	69.99
Vote-based SVM-classifier selection[20,21]	71.74	73.76	72.74
Genetic classifier-ensemble-III	75.65	78.52	77.85

Table 7. The comparison with other different individual classifier algorithms on bioNER task

Approaches	**Precision**	**Recall**	**F-score**
Zhou and Su[1]	69.42	75.99	72.55
Finkel et al.[2]	68.56	71.62	70.06
Settles[3]	69.30	70.30	69.80
Song et al.[4]	64.80	67.80	66.30
Zhao[5]	61.00	69.10	64.80
Genetic classifier-ensemble-III	75.65	78.52	77.85

pre-processing and post-processing. For instance, Zhou and Su [1] used name alias resolution, cascaded entity name resolution, abbreviation resolution and an open dictionary (around 700,000 entries). Finkel et al. used gazetteers and web-querying [2]. Settles used 17 lexicons that include Greek letters, amino acids, and so forth [3]. In contrast, our system did not include these similar processing.

4 Conclusion and Future Work

We proposed a generic genetic classifier-ensemble approach to recognizing the biomedical named entities. The contributions of this paper are that a novel genetic classifier-ensemble algorithm with weights is provided to deal with bioNER task and improve the BioNER performance compared with both of SVM-based classifiers as well as other individual machine learning algorithms. In the future, we will incorporate much more

effective features and more classifiers using different machine learning algorithms in our ensemble approach, and include some post-processing techniques and comparison of computational cost.

References

1. Zhou, G., Su, J.: Exploring Deep Knowledge Resources in Biomedical Name Recognition. In: Proceedings of the Joint Workshop on Natural Language Processing in Biomedicine and its Applications (JNLPBA 2004), pp. 70–75 (2004)
2. Finkel, J., Dingare, S., Nguyen, H., Nissim, M., Sinclair, G., Manning, C.: Exploiting Context for Biomedical Entity Recognition: From Syntax to the Web. In: Proceedings of the Joint Workshop on Natural Language Processing in Biomedicine and its Applications, JNLPBA 2004 (2004)
3. Settles, B.: Biomedical Named Entity Recognition Using Conditional Random Fields and Novel Feature Sets. In: Proceedings of the Joint Workshop on Natural Language Processing in Biomedicine and its Applications (JNLPBA 2004), pp. 104–107 (2004)
4. Song, Y., Kim, E., Lee, G.-G., Yi, B.-K.: POSBIOTM-NER in the shared task of BioNLP/NLPBA 2004. In: Proceedings of the Joint Workshop on Natural Language Processing in Biomedicine and its Applications, JNLPBA 2004 (2004)
5. Zhao, S.: Name Entity Recognition in Biomedical Text using a HMM model. In: Proceedings of the Joint Workshop on Natural Language Processing in Biomedicine and its Applications (JNLPBA 2004), pp. 84–87 (2004)
6. Zhang, Z., Yang, P.: An ensemble of classifiers with genetic algorithm based feature selection. IEEE Intelligent Informatics Bulletin 9, 18–24 (2008)
7. Yang, P., Zhang, Z., Zhou, B.B., Zomaya, A.Y.: Sample Subset Optimization for Classifying Imbalanced Biological Data. In: Huang, J.Z., Cao, L., Srivastava, J. (eds.) PAKDD 2011, Part II. LNCS(LNAI), vol. 6635, pp. 333–344. Springer, Heidelberg (2011)
8. Yang, P., Yang, Y.-H., Zhou, B.B., Zomaya, A.Y.: A review of ensemble methods in bioinformatics. Current Bioinformatics 5, 296–308 (2010)
9. Yang, P., Ho, J.W.K., Zomaya, A.Y., Zhou, B.B.: A genetic ensemble approach for gene-gene interaction identification. BMC Bioinformatics 11, 524 (2010)
10. Kuncheva, L.I., Jain, L.C.: Designing classifier fusion systems by genetic algorithms. IEEE Transaction on Evolutionary Computation 4(4) (September 2000)
11. Dietterich, T.G.: Ensemble Methods in Machine Learning. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, pp. 1–5. Springer, Heidelberg (2000)
12. Ruta, D., Gabrys, B.: Application of the Evolutionary Algorithms for Classifier Selection in Multiple Classifier Systems with Majority Voting. In: Kittler, J., Roli, F. (eds.) MCS 2001. LNCS, vol. 2096, pp. 399–408. Springer, Heidelberg (2001)
13. Larkey, L.S., Croft, W.B.: Combining classifier in text categorization. In: SIGIR 1996, pp. 289–297 (1996)
14. Patrick, J., Wang, Y.: Biomedical Named Entity Recognition System. In: Proceedings of the 10th Australasian Document Computing Symposium (2005)
15. Tsai, T.-H., Wu, C.-W., Hsu, W.-L.: Using Maximum Entropy to Extract Biomedical Named Entities without Dictionaries. In: JNLPBA 2006, pp. 268–273 (2006)
16. Chan, S.-K., Lam, W., Yu, X.: A Cascaded Approach to Biomedical Named Entity Recognition Using a Unified Model. In: The 7th IEEE International Conference on Data Mining, pp. 93–102

17. Dimililer, N., Varoğlu, E.: Recognizing Biomedical Named Entities Using SVMs: Improving Recognition Performance with a Minimal Set of Features. In: Bremer, E.G., Hakenberg, J., Han, E.-H(S.), Berrar, D., Dubitzky, W. (eds.) KDLL 2006. LNCS (LNBI), vol. 3886, pp. 53–67. Springer, Heidelberg (2006)
18. Kazamay, J.-I., Makinoz, T., Ohta, Y., Tsujiiy, J.-I.: Tuning Support Vector Machines for Biomedical Named Entity Recognition. In: ACL NLP, pp. 1–8 (2002)
19. Mitsumori, T., Fation, S., Murata, M., Doi, K., Doi, H.: Gene/protein name recognition based on support vector machine using dictionary as features. BMC Bioinformatics 6(suppl. 1) (2005)
20. Dimililer, N., Varoğlu, E., Altınçay, H.: Vote-Based Classifier Selection for Biomedical NER Using Genetic Algorithms. In: Martí, J., Benedí, J.M., Mendonça, A.M., Serrat, J. (eds.) IbPRIA 2007, Part II. LNCS, vol. 4478, pp. 202–209. Springer, Heidelberg (2007)
21. Dimililer, N., Varoglu, E., Altmcay, H.: Classifier subset selection for biomedical named entity recognition. Appl. Intell., 267–282 (2009)
22. Ruta, D., Gabrys, B.: Classifier selection for majority voting. Inf. Fusion 1, 63–81 (2005)
23. Yang, T., Kecman, V., Cao, L., Zhang, C., Huang, J.Z.: Margin-based ensemble classifier for protein fold recognition. Expert Syst. Appl. 38(10), 12348–12355 (2011)
24. Zhang, P., Zhu, X., Shi, Y., Wu, X.: An Aggregate Ensemble for Mining Concept Drifting Data Streams with Noise. In: Theeramunkong, T., Kijsirikul, B., Cercone, N., Ho, T.-B. (eds.) PAKDD 2009. LNCS, vol. 5476, pp. 1021–1029. Springer, Heidelberg (2009)
25. John, H.: Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to biology, control and artificial intelligence. MIT Press (1998) ISBN 0-262-58111-6
26. Kim, J.-D., Ohta, T., Tsuruoka, Y., Tateisi, Y., Collier, N.: Introduction to the Bio-Entity Recognition Task at JNLPBA. In: Proceedings of the International Workshop on Natural Language Processing in Biomedicine and its Applications (JNLPBA 2004), pp. 70–75 (2004)