A Prototype Patterns Selection Algorithm Based on Semi-supervised Learning

Zhehuang Huang
School of Mathematics Sciences, Huaqiao University, 362021, China
Cognitive Science Department, Xiamen University, Xiamen, 361005, China
Email: hzh98@hqu.edu.cn

Yidong Chen
Cognitive Science Department, Xiamen University, Xiamen, 361005, China
Fujian Key Laboratory of the Brain-like Intelligent Systems, Xiamen,361005, China
Email: ydchen@xmu.edu.cn

***Abstract*—Semantic role labeling (SRL) is a fundamental task in natural language processing to find a sentence-level semantic representation. At present, the mainstream studies of semantic role labeling focus on the use of a variety of statistical machine learning techniques. But it difficult to obtain high quality labeled data. To solve the problem, we proposed a novel prototype patterns selection algorithm based on semi-supervised learning in this paper. There are two main innovations in this article: firstly, order parameter evolution is introduced to expand training data. The strongest order parameter will win by competition and desired pattern will be selected. Secondly, the must-links and cannot-links constraints exist in the train data is used to reduce the noise of extend data. The experiment results show the proposed method has a higher performance for semantic role labeling.**

***Index Terms*—semi-supervised learning, SRL, SNN, Order parameters.**

I. Introduction

Semantic role labeling [1] is a task in natural language processing consisting of the detection of the semantic arguments. Given a sentence and a predicate in it, the task of SRL is to recognize and map all the word sequences in the sentence into their corresponding semantic roles. Semantic role labeling has been widely application in many areas, such as information extraction [2], question answering [3, 4], and machine translation [5].

Nowadays, the main studies of semantic role labeling focus on the use of a variety of machine learning algorithm [6-8]. In traditional supervised learning algorithms, an initial classifier is trained on the large-scale labeled corpus. Then the classifier is used to classify the unlabeled test data. Pradhan and Kadri Hacioglu [9] proposed an algorithm based on Support Vector Machines. Llu´ıs M`arquez [10] presents a semantic role labeling system based on AdaBoost. Maximum entropy classier [11-13] is used in semantic role labeling system, which takes syntactic constituents as the labeling units.

It requires many labeled data according to the method described above. But in many practical applications, it is difficult to obtain high quality labeled data, which involves much human labor and is time-consuming. Semi-supervised learning [14, 15] is a new issue in the field of pattern recognition and machine learning in recent years. It is mainly focus on how to use a small amount of labeled data and unlabeled data to train classifier. Semi-supervised learning can reduce the annotation cost, improve the learning performance, and have very important practical significance.

At present, semantic role labeling method based on semi-supervised learning faces the following problems :(a) how to extend the labeled data? It is essential to propose an effective extend algorithm.(b), how to reduce noise? The noise will increase in the procedure of extending labeled data. So it is important to choose an effective noise edit method. Based on the above ideas, we proposes a prototype patterns selection algorithm based on semi-supervised learning in this paper. Its main idea is as follows: we firstly construct an initial integration classifier based on SNN. Then we extend the train data using order parameter threshold. Finally, reduce the noise of extend data by the must-links and cannot-links constraints exist in the train data.

This paper is organized as follows. Firstly, an integrative Semantic role labeling method based on synergetic theory is introduced. Secondly, a prototype patterns selection algorithm based on semi-supervised learning and a noise filter algorithm based on pair-wise constraints is presented. Finally some experimental tests, results and conclusions are given on the systems.

II. An Integrative Semantic Role Labeling Method Based on Synergetic Theory

SNN model is a top-down network proposed by Haken in the late 1980s [16]. Synergetics is a comprehensive subject which reveals the discipline of the evolution of cooperative system from disorderly to orderly. Collaborative system is composed of many subsystems,

doi:10.4304/jsw.8.8.1984-1990

which can self-organize into open system with ordered structure.

At present, the main research focus on setting of attention parameter [17], the selection of prototype pattern vector [18, 19], and reconstruction algorithm of order parameters [20, 21] and so on.

A dynamic equation can be given for a unrecognized pattern q:

$$\dot{q}=\sum_{k=1}^{M}\lambda_k v_k(v_k^{+}q)-B\sum_{k'\neq k}(v_{k'}^{+}q)^2(v_k^{+}q)v_k-C(q^{+}q)q+F$$

, Where q is the status vector of input pattern with initial value q_0 , λ_k is attention parameter, v_k is prototype pattern vector. v_k^{+} is the adjoint vector of v_k that satisfies

$(v_k^{+}, v_{k'}^{T}) = v_k^{+}.v_{k'}^{T} = \delta_{kk'}$.

(1) Obtain feature vectors from the train corpus and test corpus, construct test pattern q_l $(l=1,2\cdots)$ and possible roles R_k $(k=1,2\cdots)$.

(2) Calculate the order parameter of q_l.

(3) Find the N-best candidate role $(R_{l1}, R_{l2}, \cdots R_{lNbest})$ of q_l $(l=1,2\cdots)$

(4) Combination of all possible roles of q_l $(l=1,2\cdots)$, obtain all the possible roles chains, and calculate the corresponding probability.

(5)Set the attention parameters B, C and non- attention parameter λ_k .

(6) Obtain the best roles chain through the evaluating of order parameter equation.

Algorithm 1.The semantic role labeling based on Synergetic Neural Network

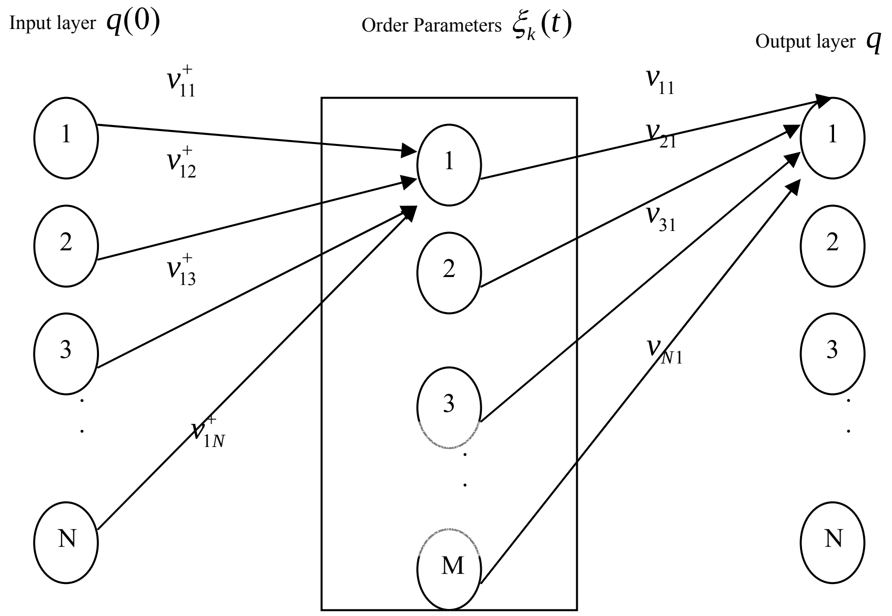

Fig .1. Network structure of synergetic neural network

Corresponding dynamic equation of order parameters is

$$\dot{\xi}_k=\lambda_k\xi_k-B\sum_{k'\neq k}\xi_{k'}^2\xi_k-C\left|\sum_{k'=1}^{M}\xi_{k'}^2\right|\xi_k$$

Figure 1 shows the network structure of synergetic neural network.

The role labeling procedure can be viewed as the competition progress of many order parameters. A method for semantic role labeling using synergetic neural network (SNN) technique is presented as algorithm 1.

III. Prototype Patterns Selection Algorithm

The basic principle of synergetic neural network is that the pattern recognition procedure can be viewed as the competition progress of many order parameters. The Semantic role labeling architecture based on SNN is shown in Figure 2.

At present, most of supervised learning methods can achieve better results in large-scale labeled training data application. But it is difficult to obtain labeled data in practical applications. Semi-supervised learning is a new machine learning method which make use of both labeled and unlabeled data[22,23] , such as a small amount of labeled data with a large amount of unlabeled data. In this paper, we proposed a semi-supervised method based on SNN.

Firstly, an initial classifier with a certain accuracy rate was constructed based on small-scale labeled data to predict some new candidate instances from unlabeled data. Secondly, order parameters[24,25] was applied by setting different value to expand training data .New instances with higher credibility from candidate instances were selected to add to the training data. Finally, the training classifier was re-iteration with the expanded training data until classifier performance tended to stable, and iteration termination.

IV. Noise Filter of Prototype Patterns

Probabilistic latent semantic analysis (PLSA) [26] which evolved from latent semantic analysis is a statistical technique for the analysis of co-occurrence data. PLSA has widely applications [27-29] in information retrieval, natural language processing, machine learning and Image Classification [30-32]. Must-link and cannot-link constraints are used in PLSA and supervised Clustering.

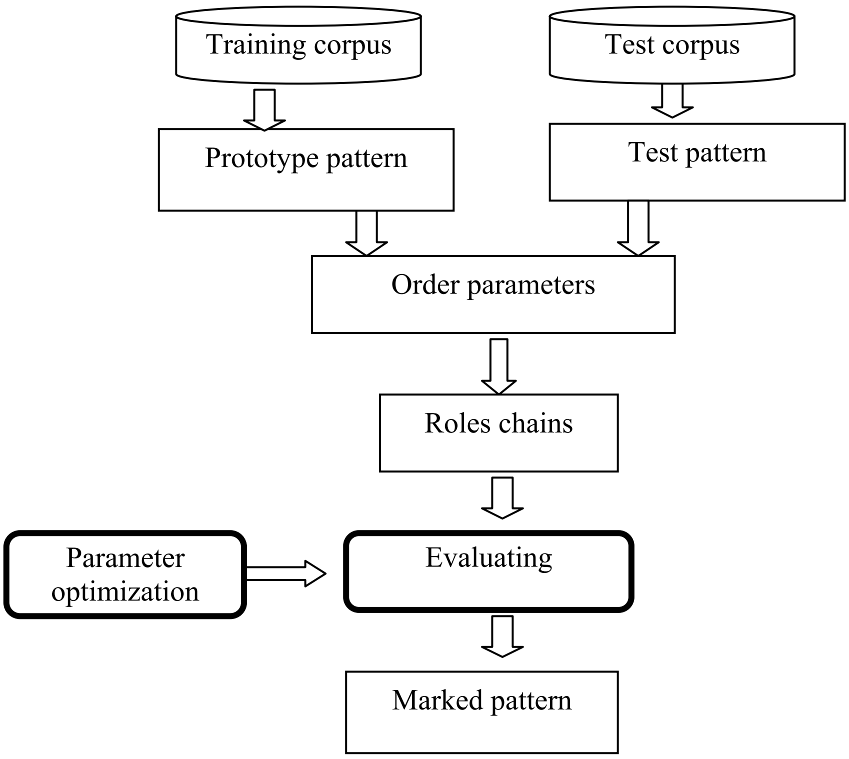

Fig .2. Semantic role labeling architecture based on SNN

The semi-supervised method based on SNN is shown as follow:

---

(1). Select the initial prototype pattern, build the initial classifier with a certain accuracy rate.

(2). Expand prototype pattern based on SNN.

a) Calculate the order parameter;

b) Find out the N-best candidate role;

c) Obtain the best roles chain through the evaluating of order parameter equation.

(3). Reiteration with the expanded train data until classifier performance tend to stable.

Algorithm 2. Semi-supervised prototype pattern selection based on SNN

A. Problem Definition

Supposed $D=\{d_1,d_2\cdots d_m\}$ is prototype pattern and $T=\{t_1,t_2\cdots t_n\}$ is test pattern from the train corpus and test corpus. R_k $(k=1,2\cdots)$ is possible roles. We encode these two kinds of constraints as follows:

$$f_1=\sum_r p(d_l^i\mid r)p(d_l^j\mid r)\ldots\ldots\ldots\ldots\ldots(1)$$

Note that $p(d_l^i\mid r)$ $p(d_l^j\mid r)$ represents the probability that two documents d_l^i and d_l^j generated by same role r , and f_1 denotes the probability that two prototype pattern d_l^i and d_l^j are on the same role.

In the same manner, the cannot-link constraint as follows:

$$f_2 = \sum_{r_i \neq r_j} p(d_l^i \mid r_i) p(d_l^j \mid r_j) ....(2)$$

Where f_2 denotes the probability that two prototype pattern d_l^i and d_l^j are on the different roles.

Considering the two constraint f_1 and f_2 at the same time, the objective function can be expressed as follows.

$$L = \alpha_1 \sum_{d_l} \log \sum_{r} p(d_l^i \mid r) p(d_l^i \mid r) +$$

$$\alpha_2 \sum_{d_l} \log \sum_{r_i \neq r_j} p(d_l^i \mid r_i) p(d_l^j \mid r_j) ....(3)$$

The Semi-supervised prototype pattern selection is shown in Figure 4.

With the introduction of a large amount noise of unlabeled corpus, the performance of the classifier will significantly decrease.

So the noise filter is necessary. We find there are some constraint conditions in the unlabeled corpus, such as must-link and cannot-link constraints, which provided the possibility for out proposed noise filter.

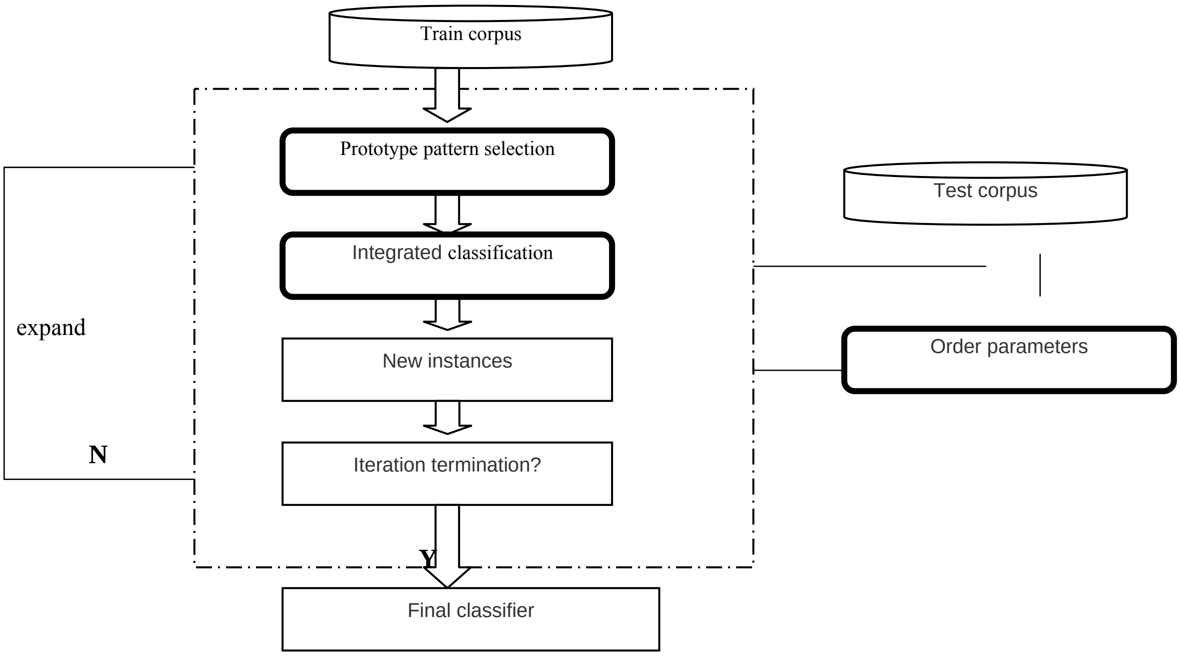

Fig .4. Semi-supervised prototype pattern selection

Here α_1 and α_2 are the parameters to set the weights for the must-link and cannot-link constraints.

$$P(d_i \mid r) = \frac{p(d_i) p(r \mid d_i)}{p(r)} ....(4)$$

The Overall constraint is shown in Figure 3.

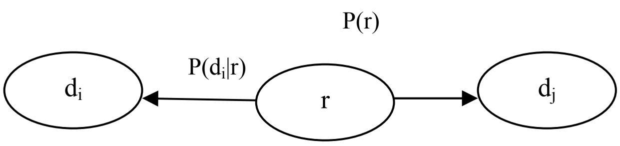

Fig .3. The Overall constraint

B. Noise Filter

The noise filter algorithm based on the must-links and cannot-links constraint is shown as follow.

Input: Prototype pattern $D = \{d_1, d_2 \cdots d_m\}$, test pattern $T = \{t_1, t_2 \cdots t_n\}$ and possible roles $R_k \ (k = 1, 2 \cdots)$;

1) An initial classifier C with a certain accuracy rate was constructed based on small-scale labeled corpus.

2) Calculates $L(0)$ using Equation (3) ,and Calculates $p(r \mid d)$ using Equation (4);

3) Loop performs the following steps, until meet certain degree of accuracy:

For each $t_i \in T$ Do

Step1: Combination of all possible roles of $t_l\ (l = 1,2\cdots n)$, calculate the corresponding order parameter ξ_l . Obtain the best roles chain through the evaluating of order parameter equation.

Step2: If ξ_l > threshold, do the following steps：

a) Expand training data： $L = L + \{t_i\}, i = 1,2\cdots n$;

Noise filter：

Calculates $L(m+1)$ using Equation （3）;

b) If $L(m+1) < L(m)$,

Then $L = L - \{t_i\}, i = 1,2\cdots n$

Otherwise：m=m+1, Go to step1;

c) The training classifier was re-iteration with the expanded training data.

4) Output the final classification results.

Algorithm 3. Noise filter based on the must-links and cannot-links constraint

V. EXPERIMENT

A. Data Description

In this section we empirically evaluate our algorithm for semantic role labeling and compare it with other state-of-the-art algorithms.

In our experiments, we use the OntoNotes Release 2.0 corpus [33] which is a large, multilingual richly-annotated corpus .The corpus has been annotated with multiple levels of annotation, including constituency trees, predicate argument structure, word senses, co-reference, and named entities.

The corpus has English, Chinese, and Arabic portions, and we just use the English portion, which has been split into four sections: broadcast conversation (bc), broadcast news (bn), magazine (mz), andnewswire (nw).

The corpus used is shown as Table 1.

TABLE 1
ONTONOTES CORPUS

ABC	0001-0059
CNN	0001-0437
MNB	0001-0025
NBC	0001-0039
PRI	0001-0032
VOA	0001-0265

The features are listed in Table 2

TABLE 2
THE FEATURES

ID	Feature name	Description
1	ID	Token counter
2	FORM	Word form or punctuation symbol
3	LEMMA	Predicted lemma of FORM
4	GPOS	Gold part-of-speech
5	PPOS	Predicted POS tag
6	SPLIT_FORM	Tokens split at hyphens and slashes
7	SPLIT_LEMMA	Predicted lemma of SPLIT_FORM
8	PPOSS	Predicted POS tags of the split forms
9	HEAD	Syntactic head of the current token
10	DEPREL	Syntactic dependency relation to the HEAD
11	PRED	Rolesets of the semantic predicates in this sentence
12+	ARG	Roles

The roles set are listed in Table 3

TABLE 3
ROLES SET

Field name	Description
ARG0	Agent
ARG1	The action of the recipient
ARG2	Role associated with the predicate
ARG3	Role associated with the predicate
ARG4	Role associated with the predicate
ARGM- EXT	extent
ARGM-DIR	direction
ARGM-LOC	location
ARGM-TMP	temporal
ARGM-REC	reciprocal
ARGM-PRD	predication
ARGM-NEG	negation
ARGM-MOD	modal
ARGM ADV	adverbial
ARGM-MNR	manner
ARGM-CAU	cause
ARGM-PNC	purpose not cause
ARGM-DIS	Discourse

B. Experimental Results and Analysis

We used precision, recall and F1 values as evaluation indicators.

$$F1 = \frac{2 \times precision \times recall}{precision + recall}.$$

For comparison, we use three strategies:

① Baseline: Integrative semantic role labeling method based on synergetic theory;

② SNN expansion algorithm: Prototype patterns selection algorithm based on SNN without noise filter;

③ Expansion with noise filter: Prototype patterns selection algorithm based on SNN with noise filter.

Algorithm	Threshold		
	0.9	0.7	0.6
①	65%	65%	65%
②	65.2%	66.4%	65.9%
③	66.1%	66.9%	67.2%

Table 4 The performances of the semantic role labeling system based on different method.

The results showed, without noise filter, the noise increase when extending data. When the threshold was 0.9 and 0.7, the accuracy rate is improved. But when the threshold is 0.6, the introduction of noise is too much, the accurate rate declines. With noise filter, the accuracy rate increases obviously. So the method proposed in this paper is effective.

Experimental evaluation shows that our proposed algorithm can improve the performance of semantic role labeling significantly.

VI. CONCLUSION

In this paper, we proposed a prototype patterns selection algorithm based on semi-supervised learning and a noise filter algorithm based on pair-wise constraints. The experiment results show the integrative method have a higher performance for semantic role labeling , and provide a good practicability and a promising future for other natural language processing tasks. We will do further investigation to inspect the inherent constraint among train data, and apply it to other areas.

ACKNOWLEDGMENTS

This work was supported by the science and technology project of Quanzhou (Grant No.2012Z91), the National Natural Science Foundation of China (Grant No. 61005052), the Fundamental Research Funds for the Central Universities (Grant No. 2010121068) and the Natural Science Foundation of Fujian Province of China (Grant No.2011J01369).

REFERENCES

[1] Punyakanok, Vasin, Dan Roth, and Wen tau Yih.The importance of syntactic parsing and inference in semantic role labeling. Computational Linguistics,2008,34(2),pp.258-285.

[2] Surdeanu Mihai,Sanda Harabagiu,John Williams and Paul Aarseth.Using Predicate-argument Structures for Information Extraction.In Proceedings of ACL 2003,pp.8-15.

[3] Narayanan Srini and Sanda Harabagiu.Question Answering based on Semantic Structures.In Proceedings of COLING 2004,pp.693-701.

[4] Shen Dan and Mirella Lapata.Using Semantic Roles to Improve Question Answering.In Proceedings of EMNLP-CoNLL 2007,pp.12-21.

[5] Wu Dekai and Pascale Fung.Can Semantic Role Labeling Improve SMT?. In Proceedings of EAMT 2009,pp.218-225.

[6] Che, Wanxiang, Ting Liu, and Yongqiang Li. Improving semantic role labeling with word sense.In NAACL 2010, pp. 246–249.
[7] Hajiˇc, Jan, Massimiliano Ciaramita, Richard Johansson,Daisuke Kawahara, Maria Ant`onia Mart´ı, Llu´ıs M`arquez, Adam Meyers, Joakim Nivre, Sebastian Pad´o, Jan ˇ Stˇep´anek, Pavel Straˇn´ak, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. The conll-2009 shared task: Syntactic and semantic dependencies in multiple languages. In Proceedings of CoNLL 2009: Shared Task,pp. 1–18.
[8] X. Carreras, L. M`arquez, G. Chrupała. Hierarchical Recognition of Propositional Arguments with Perceptrons. H. T. Ng, E. Riloff, (Editors) HLTNAACL 2004 Workshop: Eighth Conference on Computational Natural Language Learning (CoNLL-2004). Boston, Massachusetts, USA, 2004.
[9] S. Pradhan, K. Hacioglu, V. Krugler, et al. Support Vector Learning for Semantic Argument Classification. Machine Learning Journal. 2005, 60(3), pp. 11–39.
[10] L.M`arquez, P. Comas, J. Gim´enez, et al. Semantic Role Labeling as Sequential Tagging. Proceedings of CoNLL-2005. Ann Arbor, Michigan, 2005.
[11] T. Liu, W. Che, S. Li, et al. Semantic Role Labeling System Using Maximum Entropy Classifier. Proceedings of CoNLL-2005, pp.189–192.
[12] N. Kwon, M. Fleischman, E. Hovy. Senseval Automatic Labeling of Semantic Roles Using Maximum Entropy Models. R. Mihalcea, P. Edmonds, (Editors) Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text. Barcelona, Spain, 2004.
[13] T.-H. Tsai, C.-W. Wu, Y.-C. Lin, et al. Exploiting Full Parsing Information to Label Semantic Roles Using an Ensemble of ME and SVM Via Integer Linear Programming. Proceedings of CoNLL-2005. Ann Arbor, Michigan, 2005.
[14] Basu, S., Banerjee, A., and Mooney, R. J. Semi-Supervised Clustering by Seeding. In ICML, 2002.
[15] Cohn, D., Caruana, R., and McCallum, A. Semi-Supervised Clustering with User Feedback.Technical Report TR2003-1892, Cornell University, 2003.
[16] Haken. Synergetic Computers and Cognition–A Top-Down Approach to Neural Nets. Springer-Verlag, Berlin ,1991.
[17] WangGang,Gao Yang,xia Jie. Quick train of artificial neural network based on diffierential evolution.Chinese Journal of Management,2005,2(4),pp.450-454.
[18] Jing Shao, Jun Gao and Xuezhi Yang. Synergetic Face Recognition Algorithm Based on ICA. In: Proceedings of the International Conference on Neural Networks and Brain, Beijing, China, 2005, pp. 249-253.
[19] Zhenhua Jiang and Roger A. Dougal. Synergetic Control of Power Converters for Pulse Current Charging of Advanced Batteries From a Fuel Cell Power Source. IEEE Transactions on Power Electronics, 2004, 19(4),pp.1140-1150.
[20] Xiuli Ma, Shuang Wang, Licheng Jiao. Robust Classification of Immunity Clonal Synergetic Network Inspired by Fuzzy Integral. In: Jun Wang, Xiaofeng Liao, Yi Zhang (Eds.), Lecture Notes in Computer Science, Springer, Berlin. 2005, 3497,pp.26-31.
[21] Gao Jun, Dong Huo-Ming, Shao Jing. Parameters optimization of synergetic recognition approach. Chinese Journal of Electronics,2005,14(2), pp.192-197.
[22] Xinmeng Zhang, Shengyi Jiang. A Splitting Criteria Based on Similarity in Decision Tree Learning. Journal of software , 2012, 7(8),pp.1775-1782.
[23] Lin Yao, Chengjie Sun, Xiaolong Wang, Xuan Wang. Combining Self Learning and Active Learning for Chinese Named Entity Recognition. Journal of software,2010,5(5), pp.530-537.
[24]] Wang, H.L. The Research of Application of Image Recognition Using Synergetic Neural Network. Ph.D. Dissertation. Shanghai Jiao Tong University, China 2000.
[25] Xiuli Ma, Licheng Jiao. Reconstruction of Order Parameters Based on Immunity Clonal Strategy for Image Classification. In: Aurélio Campilho, Mohamed Kamel (Eds.), Lecture Notes in Computer Science, Springer, Berlin. 2004, 3211, pp.455-462.
[26] Cohn, D., and Hofmann, T. The Missing Link - a Probabilistic Model of Document Content and Hypertext Connectivity. In NIPS, 2001.
[27] Dempster, A., Laird, N., and Rubin, D. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of Royal Statistical Society, Series B, 1977,39(1) , pp.1–38.
[28] Nigam, K., McCallum, A. K., Thrun, S., and Mitchell, T. Text Classification from Labeled and Unlabeled Documents using EM. Machine Learning , 2000,39(2-3) , pp.103–134.
[29] Hofmann T . Unsupervised Learning by Probabilistic Latent Semantic Analysis . Machine Learning,2001,42(2) , pp.177– 196
[30] Sivic J,Russell B C,Efros A A . Discovering Object Categories in Image Collections . Proc of the 10th IEEE International Conference on Computer Vision. Rio de Janeiro,Brazil,2005, pp.1–13
[31] Liu D,Chen T. Semantic-Shift for Unsupervised Object Detectionv. Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington,USA,2006, pp.16– 23.
[32] Liu D,Chen T. Unsupervised Image Categorization and ObjectLocalization Using Topic Models and Correspondences between Images. Proc of the 11th International Conference on Computer Vision. Rio de Janeiro,Brazil,2007, pp.1– 7.
[33] E. Hovy, M. Marcus, M. Palmer, L. Ramshaw, and R. Weischedel, “Ontonotes: The 90% solution,” in Proceedings of NAACL 2006, June 2006, pp. 57–60.

Zhehuang Huang is a lecturer in the School of Mathematics Sciences Huaqiao University. His research interest covers intelligent algorithm, machine learning and data fusion.

Yidong Chen, Dr., associate professor of Cognitive Science Department, Xiamen University , His research interest covers intelligent algorithm, machine learning and machine translation.