

# Sentiment Orientation Identification under Q&A Community based on Two-level Conditions Random Field Improved by Particle Swarm Optimization Algorithm

Wang Caiyin<sup>1,2</sup>, Cui Lin<sup>1,3</sup> and Li Hong<sup>1</sup>

<sup>1</sup>*Intelligent Information Processing Laboratory, Suzhou University, Suzhou 234000, Anhui, China*

<sup>2</sup>*School of Mechanical Electric and Engineering, Suzhou University, Suzhou 234000, Anhui, China*

<sup>3</sup>*College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*

*caiycinwang@163.com; jsjxcuilin@nuaa.edu.cn; lih65@163.com*

## Abstract

*Because the accuracy of traditional sentiment orientation identification algorithm is not high under Q&A community, this paper proposes a new method based on two-level conditional random field improved by particle swarm optimization algorithm for emotion tendency recognition under Q&A community. The proposed method adopts particle swarm optimization algorithm to train two-level conditional random field model, and applies the trained conditional random field model to recognize emotion orientation of question-answer pairs in Q&A community. Experiments were performed on Yahoo! Answers data set and results show that the proposed two-level conditions random field improved by particle swarm optimization algorithm has a higher precision rate, recall rate and F1 value at the micro average and macro average aspects compared with Hidden Markov Model, Max-Entropy Markov Model, Support Vector Machine and traditional condition random domain model, which prove the proposed two-level conditions random field improved by particle swarm optimization algorithm is a more effective method to recognize emotion orientation of question-answer pairs in Q&A community.*

**Keywords:** *Conditional random field model, Particle swarm optimization algorithm, Question-answer pairs, Subjective and objective recognition, Emotional orientation recognition*

## 1. Introduction

With the rapid development of Web 2.0 technology, Q&A community as a new retrieval technology has become the inevitable trend of the development of retrieval system, which can accept natural language questions from users and return exact answers to the users. However, researches on Q&A community mostly focused on objective questions, how to answer subjective emotional questions is rarely involved. Answering subjective emotional questions not only needs various techniques of the traditional question answering system and is also related to sentiment analysis technology. In recent years, sentiment analysis has become a hot research point of text mining and natural processing fields, which has attracted many research scholars. Sentiment analysis usually refers to distinguish the text information with subjective emotion and executes analysis, induction, summary and other series of processing to the text information, until finally mining the opinion of users [1].

At present, most emotion analysis methods are regarded as two independent tasks that are the subjective and objective classification and the sentiment tendency classification respectively. Aiming at the above-mentioned problem, based on the sentence-level, this paper puts sentiment analysis process into the sentiment polarity context information and dependent information between two tasks, constructs a two-level structure conditions random fields model, and particle swarm optimization algorithm is used to estimate the conditional random fields' maximum likelihood parameters, simultaneously completes the subjective and emotional polarity judgment of question-answer pairs. The trained conditional random field model is evaluated on dataset collected from Yahoo! Answers and evaluation results showed that the two-level condition random field model improved by particle swarm optimization algorithm has higher precision rate, recall rate and F1 value at micro average and macro average aspects compared with HMM (Hidden Markov Model), MEMM (Max-Entropy Markov Model), SVM (Support Vector Machine) and traditional CRF model (condition random domain model).

## 2. Related Work

Question answering system is an advanced form of information retrieval system and can use accurate and concise natural language to answer the questions asked by users using natural language, which mainly use natural language parsing, questions classification, named entity recognition technology and so on [2]. Question answering community is also known as the interactive knowledge sharing platform [3], different from the question answering systems, problems in Q&A community are not only provided by users, but also answered by users. Q&A community is a typical application that social network service is in knowledge sharing.

With the growing prosperity of Web 2.0 technology and the explosive growth of data access under the Q&A community, Q&A community has gradually become a hot spot in the field of Web mining research, appearing many theories and algorithms. Agichtein used the classification framework to integrate various types of text information in Q&A community, studied problems about text quality prediction and user satisfaction degree [4]. However, the prediction method of Agichtein required a lot of artificial tagging to data and spent a lot of time. Jeon, *et al.*, adopted a translation model to learn semantic similarity between words and looked for similar problems [5]. Through the translation probability of IBM Model, the semantic similarity between words was estimated, and by using the traditional language model framework, semantic similarity problem was also completed. The best answer of similar questions can be used as the reference to the problem of the answer in order to realize answer recommendation mechanism. Jurczyk, *et al.*, studied the user link structure in the Q&A community, and presented HITS algorithm to predict the user authority [6]. Bian, *et al.*, proposed a semi-supervision mutually reinforcing frame, at the same time calculating content quality and user reputation in the Q&A community [7]. Adamic, *et al.*, made a comprehensive analysis on Yahoo! Answers knowledge sharing behavior, and tested the best answer in the Q&A community to verify the reliability of the best answer selected by the questioner [8]. At the same time, some scholars performed prediction researches on the satisfaction degree of answers obtained by users from Q&A community. After users submitted questions, according to the time of obtaining answers and the quality of answers and other factors, the degree whether users are satisfied with answers was predicted [9].

### 3. The Proposed Two-level CRF Model Improved by Particle Swarm Optimization Algorithm

#### 3.1. Particle Swarm Optimization Algorithm

In 1995, inspired by social behavior of birds and fish, Eberhart and Kennedy proposed particle swarm optimization algorithm [10]. In the PSO algorithm, Suppose  $N$  particles in  $D$  dimensions motion space,  $v_i = (v_{i,1}, v_{i,2}, \dots, v_{i,D})$  denotes the velocity vector of particle  $i$ ,  $X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,D})$  represents space position vector of particles  $i$ , according to  $v_i$  and  $X_i$ , particle  $i$  continues to regulate its speed and space position, until it reaches the global optimal solution [11]. The optimal solution of the particle is denoted as  $P_{best}$ ,  $P_{best} = (p_{i,1}, p_{i,2}, \dots, p_{i,D})$ , the optimal solution found by the population is represented as  $g_{best}$ ,  $g_{best} = (g_{i,1}, g_{i,2}, \dots, g_{i,D})$ .  $v_i^{(t)} = (v_{i,1}^t, v_{i,2}^t, \dots, v_{i,D}^t)$  denotes the velocity of particle  $i$  in the  $t$ th iteration, then the iteration velocity of  $i$  in the  $t+1$  times is:

$$v_i^{(t+1)} = wv_i^t + c_1 rand() (P_{bi} - X_i^{(t)}) + c_2 rand() (g_{bi} - X_i^{(t)}) \quad (1)$$

Where,  $w$  is regarded as the inertia factor and greater than 0,  $c_1$  is a cognitive learning factor,  $c_2$  is a social learning factor,  $c_1$  and  $c_2$  are both the positive constants,  $rand()$  is a randomized function that can produce the range between 0 and 1. Velocity of each dimension of particle is in  $[v_{\min}, v_{\max}]$ , if the velocity of some dimension of a particle is not in  $[v_{\min}, v_{\max}]$ , then this velocity of some dimension of a particle is set to the corresponding threshold value. When performing  $t+1$ th iteration, the location of  $i$  is as shown in the formula (2):

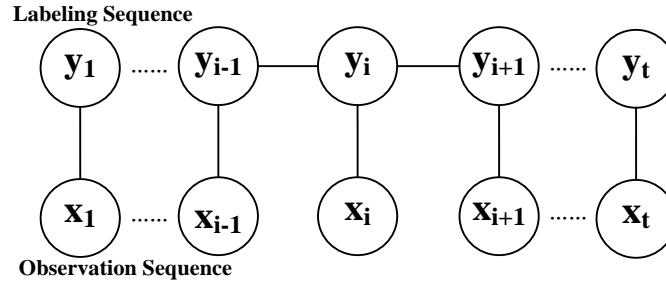
$$X_i^{(t+1)} = X_i^t + v_i^{(t+1)} \quad (2)$$

Applying PSO algorithm in fact, when the number of iterations reaches the maximum or particles finally find the optimal position with minimum adaptation degree threshold, iteration is terminated.

#### 3.2. Introduction to the Proposed Two-level CRF Model

##### 3.2.1. The Traditional CRF Model

CRF (Conditional Random Field) model is a typical discrimination model that was firstly proposed by Lafferty, *et al.*, in 2001, which models the target sequence based on observation series and focuses on solving the sequential labeling problem [12]. Related technologies about CRF have been widely applied into the different fields of natural language processing, such as word segmentation, named entity recognition, syntactic analysis and so on [13]. CRF is an undirected graph model that calculates output node conditional probability under given input node conditions. Under the given observation sequence, conditions need to be labeled and the joint probability of the whole label sequence is calculated. For the input value of some specific nodes, the conditional probability of output values of specified nodes can be calculate out; the training goal is to make the conditional probability maximization. Defining  $X = x_1, x_2, \dots, x_T$  is a given input observation value sequence, and there are  $T$  input node values in the undirected graph model;  $Y$  is defined as a finite state machine state set,  $Y = y_1, y_2, \dots, y_T$  is a state sequence that its length is equal to  $T$ , *i.e.*,  $T$  output node value in the undirected graph model. CRF model is shown in Figure 1:



**Figure 1. Model Structure of Linear-Chain CRF**

As Figure 1 shows, giving an observation sequence  $X$ , the conditional probability of tokens sequence  $Y$  based on the conditional random fields model is:

$$P(Y | X) = \frac{1}{Z_k} \exp \left\{ \sum_k l_k f_k(y_{i-1}, y_i, X, i) \right\} \quad (3)$$

Where  $Z(X) = \sum_{y \in Y} \exp \left\{ \sum_k l_k f_k(y_{i-1}, y_i, X, i) \right\}$  is the normalization

factor,  $f_k(y_{i-1}, y_i, X, i)$  is the general feature definition form under conditional random fields, which can be decomposed into two specific feature definitions:

(a) Edge features (transferring characteristics) are

$$t_{y'y}(y_{i-1}, y_i, x_i) = \begin{cases} 1 & \text{if } y_{i-1} = y', y_i = y \text{ and } x_i = x; \\ 0 & \text{else} \end{cases} \quad (4)$$

(b) Vertex feature (state feature) is

$$s_{y,x}(y_i, x_i) = \begin{cases} 1 & \text{if } y_i = y \text{ and } x_i = x; \\ 0 & \text{else} \end{cases} \quad (5)$$

According to the formula (4) and (5), the formula (3) can be divided into:

$$P(Y | X) = \frac{1}{Z(X)} \exp \left\{ \sum_k \mu_k t_k(y_{i-1}, y_i, X, i) + \zeta_k s_k(y_i, X, i) \right\} \quad (6)$$

Where  $\mu_k$  and  $\zeta_k$  are weight parameters for transferring features and state features. Based on the maximum likelihood estimate principle and L-BFGS algorithm, the model executes parameter training and makes  $P(Y | X)$  be log likelihood maximization as follows:

$$L_\Lambda = \sum_{i=1}^N \log \left( P_\Lambda \left( \frac{y_i}{x_i} \right) \right) - \sum_{k=1}^K \frac{l_k^2}{2\sigma_k^2} \quad (7)$$

Where the second part is Gauss's transcendental value of characteristic parameters providing smoothing processing;  $\sigma_k^2$  is deviation of  $K$  dimension feature. Since the L-BFGS algorithm only requires a derivative of log likelihood, then derivation of the formula (7) is:

$$\frac{\partial L_\Lambda}{\partial l_k} = \sum_{i=1}^N f_k(y_{i-1}, y_i, X, i) - \sum_{i=1}^N P_\Lambda \left( \frac{y_i}{x_i} \right) f_k(y_{i-1}, y_i, X, i) - \frac{l_k}{\sigma_k^2} \quad (8)$$

Finally, the optimal sequence  $Y^*$  solved by using dynamic programming algorithm is

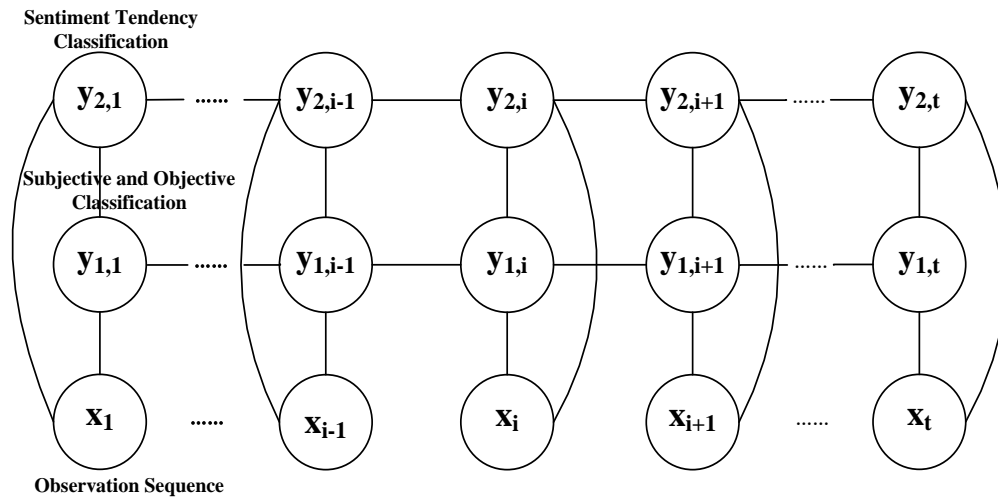
$$Y^* = \arg \max_Y \left\{ P_t \left( \frac{Y}{X} \right) \right\} \quad (9)$$

For the most possible path,  $\hat{Y}$  corresponding to given input sequence  $X$  can be gained by Viterbi algorithm as follows:

$$\hat{Y} = \arg \max_{Y \in Y(x)} P(Y | X) = \arg \max_Y \lambda \cdot F(Y, X) \quad (10)$$

### 3.2.2. Introduction to the Two-level CRF Model

At present, when the traditional CRF model performs sentence-level sentiment analysis, sentiment analysis is generally divided into two independent tasks which are firstly judging the sentence is subjective or objective, namely whether the sentence contains emotion; then aiming at the sentences that are judged to be subjective, the judgment of emotional tendency is further executed, which is to judge whether sentiment included by the sentence is positive or negative. Two independent CRF models are often required to achieve the above judgments. To some extent, this research idea ignored the contact between the judgment of subjective and objective sentences and the judgment of emotional tendency. In order to consider the connection information between the subjective and objective classification and the emotional tendency classification, based on the traditional CRF mode, a kind of conditional random field model with two-level structure is built and its structure as shown in Figure 2:



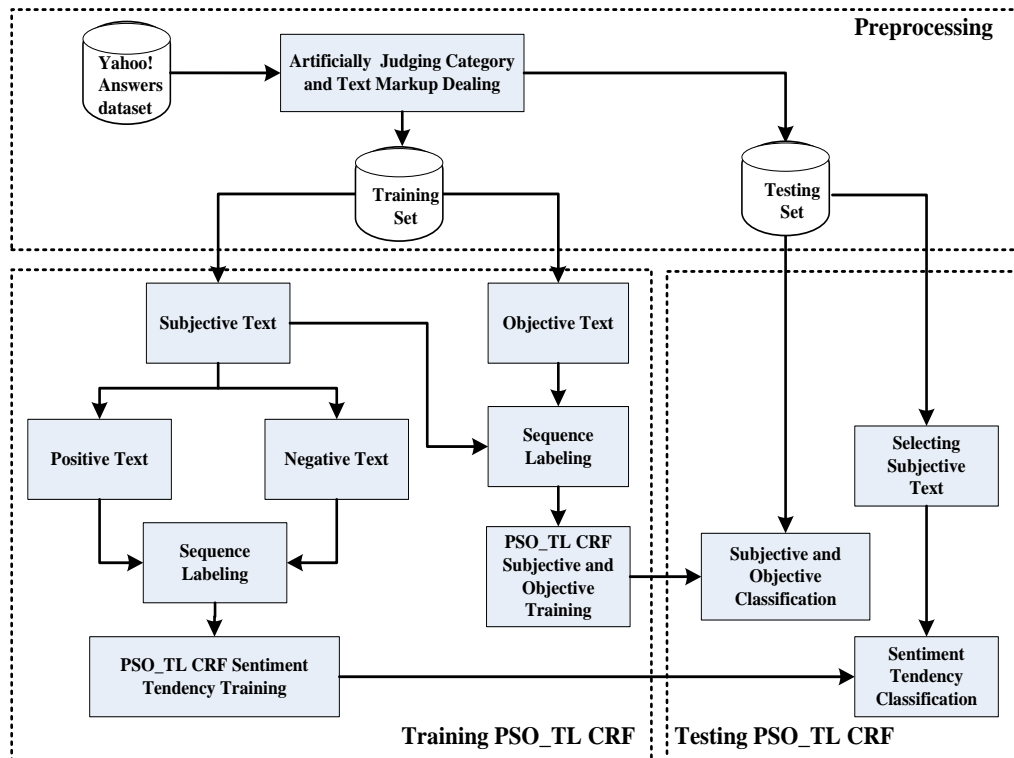
**Figure 2. Model Structure of Two-level CRF**

As Figure 2 shows, two-level conditional random field model can be seen as the combination of two linear-chain CRF model. It has a two-level tag chain structure, the first level chain structure of the model models the subjective and objective classification process, the second level chain structure models the sentiment orientation classification process, simultaneously connected to each other in the same location nodes at different levels. Random variables  $X$  denote the marked observed sequence, and random variables  $Y$  represent the corresponding marker sequences. All elements  $L$  in tag sequence  $Y$  is a finite state set.

### 3.3. Two-level CRF Model Improved by Particle Swarm Optimization Algorithm

In training the two-level conditional random fields model, the log likelihood function of conditional random fields model is regarded as the fitness function, the parameter vector  $K$  is regarded as the particle swarm, and allowing them to search the optimal value in a  $D$  dimensional space. Question answering system sentiment analysis framework based on two-level conditional random fields model improved by particle swarm optimization algorithm that is abbreviated as PSO\_TL CRF is shown in Figure 3, which

shows that the sentiment tendency analysis of the whole question answering system is divided into many modules which are artificially judging category and text markup processing, the classification of positive text and subjective text, sequence labeling, PSO\_TL CRF subjective and objective training, PSO\_TL CRF sentiment tendency training and sentiment tendency classification, some main steps are introduced as follows:



**Figure 3. Flowchart of Subjective and Objective Classification and Sentiment Tendency Classification based on PSO\_TL CRF**

### (1) Preprocessing

Firstly getting question-answer pairs from Yahoo! answers, as files that are downloaded are mostly HTML format, so what the first step does is to execute clean work to remove HTML tags. Similar with traditional text classification process, when using PSO\_TL CRF classification, firstly Q&A community text corpus were preprocessed, and the corpus is divided into training set and the testing set.

### (2) Keywords Recognition of the Question-answer Pairs

After obtaining the information of question-answer pairs and getting the cleaning results, then the next work is to recognize answer topics and answer keywords. For these two types of words, different strategies were used. As answer evaluation words are usually verb, adjective or adverb and their number generally remain unchanged and are finite, so WordNet ontology is adopted to judge the sentiment orientation of the evaluated words, for the text to be processed, the maximum matching principle is used to match these words. However, because topics of the answered questions mostly belong to proper nouns, and new nouns constantly arise, so using machine learning methods to identify the topics of the answered questions.

### (3) Grain Size of Features

The traditional text classification algorithms such as SVM, KNN and so on usually adopt the phrase as feature unit. Because the question-answer pairs are generally not too long, and segmentation algorithms have some errors, so we use the word as a feature unit when using PSO\_TL CRF training.

### (4) Sequence Labeling

This paper adopts the word labeling method to annotate sequence. For each sentence in the training, each word is regarded as the first column; the obtained sentence classification is regarded as the second labeling columns. The second column category is empty and waits for prediction. And the following feature template is used to extract feature:

```
#Unigram
U01:%x[-1,0]
U02:%x[0,0]
U03:%x[-1,0]/ %x[0,0]
#Bigram
```

## 4. Experimental Analyses

### 4.1. Experimental Data Set

Experimental data set adopts question-answer pairs capturing from the website of Yahoo! Answers. For each question-answer pair, firstly we conducted subjective and objective classification and then perform emotional tendency artificial classification. In the experiment, 70% data are randomly selected as training set; the remaining 30% data are regarded as testing set. Subjective and objective question-answer pair text data and emotional tendency question-answer pair text data employed in experiments are respectively as follows:

**Table 1. Subjective and Objective Experimental Data of Question-answer Pairs**

| Category  | Subjective | Objective |
|---|------------|-----------|
| Question-answer pairs text number from Yahoo! Answers | 12985      | 4697      |

**Table 2. Emotional Tendency Experimental Data of Question-answer Pairs**

| Category  | Positive | Negative |
|---|----------|----------|
| question-answer pairs text number of Yahoo! Answers | 1755     | 2366     |

### 4.2. Experimental Tools and Software

In this paper, experimental codes were implemented using C# language in Visual Studio 2010 platform, including the models which are text label processing, text category extraction, text sequence labeling, partitioning the training set and test set, judging the classification result, word segmentation and feature words extraction respectively. The comparative CRF model adopts open source tool CRF++-0.57, which uses the fast L-BFGS training method. SVM classification algorithm is also selected as comparison experiment and SVM method adopts LIBSVM tool.

### 4.3. Experimental Evaluation Indexes

When evaluating classification task's classification performance, parameters indexes accurate rate (Precision), recall rate (Recall) and F1 measure (F1 value) are usually adopted. In this experiment, the general accurate rate Precision, Recall and F1 value are defined as follows:

$$Precision = \frac{m_c}{m_{all}} \quad (10)$$

$$Recall = \frac{m_i}{m_{all}} \quad (11)$$

$$F1\ value = \frac{2 * Precision * Recall}{Precision + Recall} \quad (12)$$

Where  $m_c$  denotes the number of correct annotation question-answer pairs,  $m_i$  represent the number of question-answer pairs that can be identified,  $m_{all}$  represents the total number of question-answer pairs involved in the experiments.

Since the classification system contains more than one category, so the micro-average method and macro-average method are adopted to calculate accuracy, recall and F1 values. The detailed definitions are as follows:

Micro-average accuracy rate is shown as:

$$MicroP = \frac{\sum_{j=1}^N correct(polarity = POS, NEG or Classification = SUB, OBJ)}{\sum_{j=1}^N proposed(polarity = POS, NEG or Classification = SUB, OBJ)} \quad (13)$$

Micro-average recall rate is shown as:

$$MicroR = \frac{\sum_{j=1}^N correct(polarity = POS, NEG or Classification = SUB, OBJ)}{\sum_{j=1}^N annotated(polarity = POS, NEG or Classification = SUB, OBJ)} \quad (14)$$

Micro-average F1-measure value is shown as:

$$MicroF1 = \frac{MicroP * MicroR * 2}{MicroP + MicroR} \quad (15)$$

Macro-average accuracy rate is shown as:

$$MacroP = \frac{1}{N} \sum_{i=1}^N Precision_i \quad (16)$$

Macro-average recall rate is shown as:

$$MacroR = \frac{1}{N} \sum_{i=1}^N Recall_i \quad (17)$$

Macro-average F1-measure value is shown as:

$$MacroF1 = \frac{MacroP * MacroR * 2}{MacroP + MacroR} \quad (18)$$

### 4.4. Experiment Results Analysis

Experimental results are shown in table 1 and table 2 respectively.

#### (1) Performance Comparison of Different Size and Dimension of PSO Algorithm

In order to validate whether size and dimension of PSO algorithm particle swarm is appropriate, suppose a particle swarm size is 20, 40, 80 respectively, for each group of



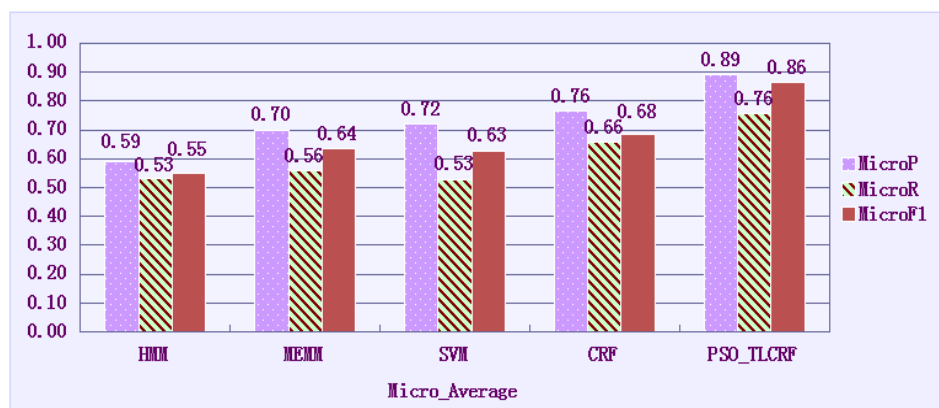
population, respectively using different dimensions 10, 20 and 30 to test. In order to obtain more reasonable likelihood parameters, every group of dimension-particle number is operated 45 times, and then takes the average value of parameters to offset the randomization of evolutionary algorithm itself. As shown in table 3, Precision, Recall and F1 value of POS-TLCRF are different with different size and dimension, when the size of particle swarm is 80 and its dimension is 30, the performance of POS-TLCRF is the best, Precision, Recall and F1 value are 0.879, 0.771 and 0.821 respectively.

**Table 3. Performance Comparison of Different Size and Dimension of PSO Algorithm on Yahoo! Answers Data Set**

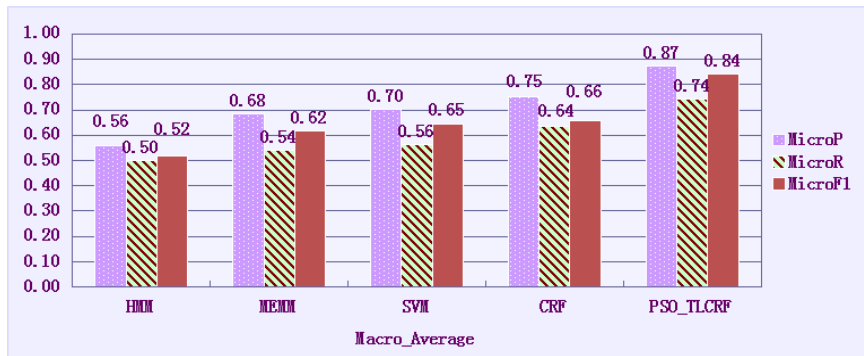
| Size | Dimension | Precision    | Recall       | F1 value     |
|------|-----------|--------------|--------------|--------------|
| 20   | 10        | 0.816        | 0.753        | 0.783        |
|      | 20        | 0.823        | 0.756        | 0.788        |
|      | 30        | 0.831        | 0.759        | 0.793        |
| 40   | 10        | 0.845        | 0.761        | 0.801        |
|      | 20        | 0.852        | 0.764        | 0.806        |
|      | 30        | 0.865        | 0.765        | 0.812        |
| 80   | 10        | 0.871        | 0.768        | 0.816        |
|      | 20        | 0.873        | 0.769        | 0.818        |
|      | 30        | <b>0.879</b> | <b>0.771</b> | <b>0.821</b> |

**(2) Subjective and Objective Classification Experiment of Question-answer Pairs**

Adopting subjective and objective experimental data of question-answer pairs in table 1, from the above experimental data, it can be found that the number of subjective text is more than the number of objective text, the comparative experimental results are shown in Figure 4(a) and Figure 4(b), among which, Figure 4(a) denotes subjective and objective micro average classification comparative results and shows the proposed POS-TLCRF algorithm outperforms HMM, MEMM, SVM and CRF in *MicroP*, *MicroR* and *MicroF1*, and *MicroP*, *MicroR* and *MicroF1* are 0.89,0.76,0.86 respectively. Figure 4(b) denotes the subjective and objective macro average classification comparative results and also shows the proposed POS-TLCRF algorithm outperforms HMM, MEMM, SVM and CRF in *MacroP*, *MacroR* and *MacroF1*, *MacroP*, *MacroR* and *MacroF1* are 0.87,0.74,0.84 respectively.



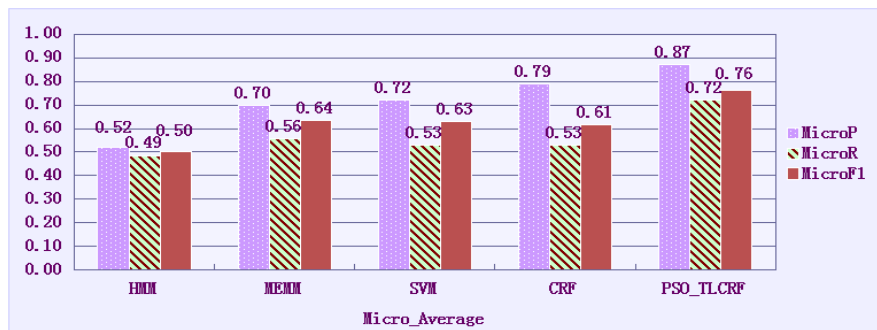
**Figure 4(a). Micro\_average Comparative Results based on Subjective and Objective Classification**



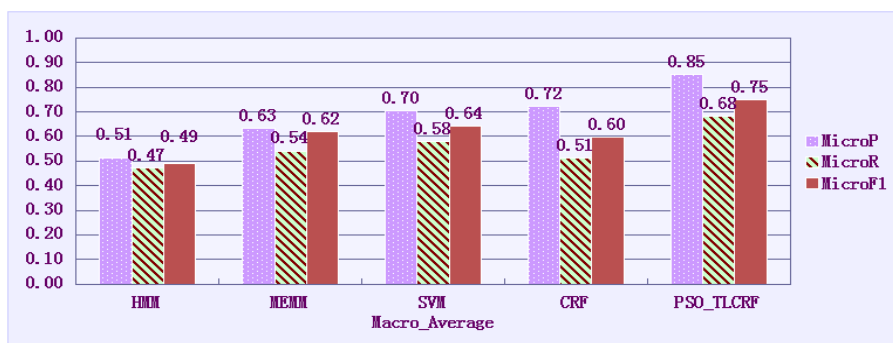
**Figure 4(b). Macro\_average Comparative Results based on Subjective and Objective Classification**

## (2) Question-answer Pairs Sentiment Orientation Classification Experiments

Emotional tendency experimental data of question-answer pairs in table 2 were adopted, which include 1755 positive question-answer pairs and 2366 positive question-answer pairs. The comparative experimental results are shown in Figure 5(a) and Figure 5(b), among which, Figure 5(a) denotes positive and negative classification micro average comparative results and shows the proposed POS-TLCRF algorithm outperforms HMM, MEMM, SVM and CRF in *MicroP*, *MicroR* and *MicroF1*, and *MicroP*, *MicroR* and *MicroF1* are 0.87,0.68,0.73 respectively. Figure 5(b) denotes positive and negative macro average classification comparative results and also shows the proposed POS-TLCRF algorithm outperforms HMM, MEMM, SVM and CRF in *MacroP*, *MacroR* and *MacroF1*, and *MacroP*, *MacroR* and *MacroF1* are 0.85,0.74,0.78 respectively.



**Figure 5(a). Micro\_average Comparative Results based on Sentiment Tendency Judgment**



**Figure 5(b). Macro\_average Comparative Results based on Sentiment Tendency Judgment**

## 5. Conclusion and Future Work

This paper adopts particle swarm optimization algorithm to estimate the maximum likelihood parameters of conditional random field model. Dimensions of population scale and search space in particle swarm optimization algorithm are obtained through the experiments, when the size of particle swarm is 80, dimensions are 30, and the obtained parameter estimation value is the best. Through comparative experiments on Yahoo! Answers, it can be proved that the proposed POS-TL-CRF has higher precision rate, recall rate and F1 value than HMM, MEMM, SVM and traditional CRF model at micro average aspect and macro average aspect. However, in practical engineering applications, the scale and dimension of particle swarm would depend on the quantity of training data. Therefore, the actual values would influence the accuracy and computational complexity of the proposed algorithm, how to find the best suitable experimental data is one of the directions of our future researches. In the future, we would like to extract more interesting and useful features to improve sentiment identification under Q&A community and we can also incorporate a topic model to refine the sentence sentiment with respect to the main topic or topics of the document.

## ACKNOWLEDGEMENTS

This work was supported by the key project of Anhui Province Colleges and Universities Natural Science Foundation of China (No.KJ2014A250), the teaching research project at Suzhou University of China (No.szxjyxm201413), the open project of Intelligent Information Processing Laboratory at Suzhou University of China (No.2013YKF14, No.2014YKF40, No.2014YKF43, No.2011YKF10), and Innovation and entrepreneurship training program of university students in Anhui Province (No.201310379081).

## References

- [1] P. Melville, W. Gryc and R. D. Lawrence, "Sentiment analysis of blogs by combining lexical knowledge with text classification", Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (2009) June 1275–1284, Paris, France.
- [2] P. Moreda, H. Llorens, E. Saquete and M. Palomar. "Combining semantic information in question answering systems", Information Processing and Management, vol. 6, no. 47, (2011).
- [3] A. Frank, H. Krieger and F. Xu, "Question answering from structured knowledge sources", Journal of Applied Logic, vol. 1, no. 5, (2007).
- [4] E. Agichtein, C. Castillo and D. Donato, "Finding high quality content in social media", Proceedings of the First ACM international Conference on Web Search and Web Data Mining, (2008) February 183-194, Stanford, USA.
- [5] J. Jeon, W. B. Croft and J. H. Lee, "A framework to predict the quality of answers with no textual features", Proceedings of the 29th International ACM SIGIR Conference on Research and Development in Information Retrieval, (2006) August 228-235, Washington, USA.
- [6] P. Jurczyk and E. Agichtein, "Discovering authorities in question answer communities by using link analysis", Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, (2007) November 919-922, Lisbon, Portugal.
- [7] J. Bian, Y. Liu and E. Agichtein, "Finding the right facts in the crowd: Factoid question answering over social media", Proceedings of the 17th International Conference on World Wide Web, (2008) April 467-476, Beijing, China.
- [8] L. A. Adamic, J. Zhang and E. Bakshy, "Knowledge Sharing and Yahoo Answers: Everyone Knows Something", Proceedings of the 17th International Conference on World Wide Web, (2008) April 467-476, Beijing, China.
- [9] D.-R. Liu, Y.-H. Chen and C.-K. Huang, "QA document recommendations for communities of question-answering websites", Knowledge-Based Systems, vol. 57, (2014).
- [10] J. Kennedy and R. Eberhart, "Particle swarm optimization", Proceeding of IEEE international conference on Neural Networks, (1995) March 1942-1948, Perth, Australia.
- [11] G. Yasser and S. Chen, "Identifying and exploiting the scale of a search space in particle swarm optimization", Proceedings of the 2014 conference on Genetic and evolutionary computation, (2014) July 17-24, New York, USA.

- [12] K. Zhang, Y. Xie and Y. Yang, "Incorporating conditional random fields and active learning to improve sentiment identification", *Neural Networks*, vol. 58, (2014).
- [13] A. Delaye and C.-L. Liu, "Contextual text/non-text stroke classification in online handwritten notes with conditional random fields", *Pattern Recognition*, vol. 3, no. 47, (2014).
- [14] Y.-M. Li and T.-Y. Li, "Deriving market intelligence from micro blogs", *Decision Support Systems*, vol. 1, no. 55, (2013).

## Copyright Forms

You must include your fully-completed, signed SERSC copyright release form when you submit your paper. WE MUST HAVE THIS FORM BEFORE YOUR PAPER CAN BE PUBLISHED IN THE PROCEEDINGS. The copyright form is available from journal home page. Authors should send their copyright forms to FAX. +82-42-624-2205 or E-mail [sersc@sersc.org](mailto:sersc@sersc.org)

## Authors

**Caiyin Wang**, received the master degree in computer science and technology from Hefei University of Technology, China in 2010. Currently, he is an associate professor at Intelligent Information Processing Lab, Suzhou University, China. His research interests include Social Network and Semantic P2P.

**Lin Cui**, received the master degree in computer science and technology from Hefei University of Technology, China in 2008. Currently, she is studying for doctoral degree in Nanjing University of Aeronautics and Astronautics of China and she is also a researcher at Intelligent Information Processing Lab, Suzhou University, China. Her research interests include Web text mining and Social network.

**Hong Li**, received the master degree in computer science and technology from Hefei University of Technology, China. Currently, he is studying for doctoral degree in China University of Mining and Technology. He is also a professor at Intelligent Information Processing Lab Suzhou University, China. His research interests include Granule computing and Social Network.