Parameter Estimation of Conditional Random Fields Model By Improved Particle Swarm Optimizer

Zengfa Dou
School of computer science, XiDian University, Xi'an, China
Email: zengfa.dou@gmail.com

Lin Gao
School of computer science, XiDian University, Xi'an, China
Email: lgao@mail.xidian.edu.cn

***Abstract*—A new parameter estimation algorithm based on improved particle swarm optimizer is proposed to improve the precision and recall rate of conditional random fields model. Aggregation degree of particle swarm is utilized to control particle swarm optimizer's early local convergence, the relative change ratio of log-likelihood between iterations is employed to end its iterations, and the inertia factor and learning factor are set as linear variables to control the searching scope. We evaluate our method on GENIA, GENETAG and private library. The experiment results prove our method outperforms traditional parameter estimation method on precision and recall.**

***Index Terms*—Conditional Random Fields Model, Particle Swarm Optimizer, Parameter Estimation, Aggregation degree of particle swarm; Relative change ratio of log-likelihood**

I. INTRODUCTION

Conditional random fields [15] (abbreviated as CRFs) is a recently introduced form of conditional model, which allow the strong independence assumptions of HMMs to be relaxed, as well as overcoming the label-bias. Like MEMMs, CRFs are conditional probabilistic sequence models. However, unlike former, CRFs are undirected graphical models.

In Conditional Random Fields, We need to estimate the maximized likelihood parameters, which affect the performance and precision of CRFs in specific application [15]. Traditionally, the nonlinear conjugate gradient algorithm [12,15,20,36] is employed to estimate the parameter of CRFs. Newton's method [7], BFGS [26] are also effective on estimating parameters of CRFs. Recently, Gradient Tree Boosting [32], Virtual Evidence Boosting [18], Piecewise pseudo likelihood [7,10], stochastic gradient methods [30], and minimum divergence beams [9] are also employed.

This paper employs an improved particle swarm optimizer [15,25] (abbreviated as PSO) to estimate the parameters of Conditional Random Fields model. In order to avoid PSO's early local convergence, we use an aggregation degree [35] to control its convergence. In order to prevent PSO from slow convergence near by best position, we employ the relative change ratio of log-likelihood between iterations [27] to end its iteration. Unlike the traditional particle swarm optimizer, we set the inertia factor and learning factor as self-adaptable variables to balance global search and local search. We evaluate CRFs model trained by our method on GENIA library and the result shows our method has better precision and recall than traditional methods.

This paper is organized as follows. Section II describes the conditional random fields model and its parameter estimation. Section III proposes a novel parameter estimation algorithm of CRFs model based on improved PSO. In Section IV, three experiments on bio-entity recognition from GENIA Corpus, GENETAG and private library are given. Section V concludes our work.

II. CONDITIONAL RANDOM FIELDS

Conditional Random Fields are undirected graphical models used to calculate the conditional probability of values on designated output sequence given values assigned to other designated input sequence. Lafferty [15] defined the probability of a particular label sequence y given observation sequence x to be a normalized product of potential functions, each of the form as:

$$p(y|x,\lambda)=\frac{1}{Z(x)}\exp(\sum_j \lambda_j F_j(y,x)) \quad (1)$$

Where $F_j(y,x)$ is either a state function $s(y_{i-1},y_i,x,i)$ or a transition function $t_j(y_{i-1},y_i,x,i)$, λ_j is a weight of indicating the precision of feature f_j, Z (x) is a normalization factor as below:

$$Z(x)=\sum_{x,y} \exp(\sum_k \lambda_k f_k(c,y_c,x)) \quad (2)$$

doi:10.4304/jcp.6.8.1628-1633

The maximum likelihood parameter estimation problem for a CRFs model as equation (2) is the task to estimate the parameters $\lambda=(\lambda_1,\lambda_2,...)$ from a set of training data points $D=\{(y^{(1)},x^{(1)}),...,(y^{(n)},x^{(n)})\}$, which are generated from the empirical distribution $\tilde{p}(x,y)_f$, such that the log-likelihood of the training data is maximized [10]. Definition of the log-likelihood of conditional random fields model is given as below:

$$L(\lambda)=\sum_{i=1}^{n}\log(p(y_i|x_i))-\sum_{j=1}^{n}\frac{\lambda_j^2}{2\sigma^2} \quad (3)$$

where $\lambda=(\lambda_1,\lambda_2,...)$ .From a numerical optimization point of view, the log-likelihood function for a CRF is smooth and concave over the entire parameter space.

III. Parameter Estimation of Conditional Random Fields by Improved Particle Swarm Optimizer

A.Improved Particle Swarm Optimizer

Kennedy and Eberhart [16,25] originally designed particle swarm optimizer (abbreviated as PSO), which simulates the behaviors of bird flocking and uses it to solve the optimization problems. PSO is initialized with a group of random particles (x_i) and then searches for optima by updating generations. In every generation, each particle is updated by two 'best' values, the local best value and global best value. The local best value is the best solution (fitness) it has achieved in its dimension so far. This value is named p_{id} .The global best value that is tracked by the particle swarm optimizer is the best value obtained so far by any particle in the population. This value is named p_g . The velocity and positions of each particle are updated according to their local best position and the global best position meet by any particle according to equation (5) and equation (6).

$$\begin{aligned} v_{id}(t+1) &= w*v_{id}(t)+ \\ & c_1*rand()*(p_{id}-x_{id})+ \\ & c_2*rand()*(p_g-x_{id}) \end{aligned} \quad (5)$$

$$x_{id}(t+1)=x_{id}(t)+v_{id}(t+1) \quad (6)$$

In equation (5) and (6), $v_{id}(t)$ is the current particle velocity in dimension d, $v_{id}(t+1)$ is the next particle velocity in dimension d, $x_{id}(t)$ is the current particle position in dimension d, $x_{id}(t+1)$ is the next position, w is the inertia weight, c_1 and c_2 are learning factors, p_{id} and p_g are defined as local best position and global best position, and $rand()$ is a random function in the range [0,1]. In order to keep the particles in the search space, usually a velocity limit $V_{\max}$ is given. If the velocity is higher than $V_{\max}$, reset the new velocity to $V_{\max}$.

In order to overcome the shortage of early local convergence, we employ the aggregation degree of the particle swarm [35] during searching. The aggregation degree [35] describes the discrete degree of the swarm, namely diversity. It is defined as its biggest value of the absolute difference values of each dimensional coordinate to denote the distance:

$$d(t)=\max\{|x_{id}-x_{jd}|,i,j=1,2,...,m;i\neq j;d=1,2,...,N\} \quad (7)$$

Where m is the size of swarm, N is the dimensionality of search space, x_{id} is the ith particle and x_{jd} is the jth particle on dimension d.

In PSO, the search speed is fast at the beginning run and slow at the near best position, even run into infinite iteration. In order to avoid infinite iteration and speed up convergence near by best position, we employ relative change ratio of log-likelihood, which Malouf [27] has successfully used to compare parameter estimation algorithms for conditional maximum entropy models, as stop criterion of PSO. The relative change ratio of log-likelihood between iterations is defined as:

$$L(x_{id})=\frac{Log(x_{id}(t+1))-Log(x_{id}(t))}{Log(x_{id}(t+1))} \quad (8)$$

When it falls below a predetermined threshold of 10^{-7}, all iterations stop.

Inertia weight decides the search space. A larger inertia weight facilitates a global search while a small inertia weight facilitates a local search. In order to make PSO has good global search ability at the beginning and good local search ability by the end, we make the inertia weight decrease as following:

$$w(t)=w_{start}-\frac{w_{start}-w_{end}}{M}\times t \quad (9)$$

The learning weight c_1 is the local learning weight, and learning weight c_2 is the global learning weight. In standard PSO, they are constant values. In order to make PSO has good global learning ability at the beginning and good local learning ability by the end, we set c_1 and c_2 as 2 at beginning, and then they are varying with iterations as below:

$$c_1=2+\frac{t}{M},c_2=2-\frac{t}{M} \quad (t=0\sim M-1) \quad (10)$$

In above equations, $w(t)$ is the current inertia weight, c_1 and c_2 are the learning factors, M is the maximum of epoch, and t is the current epoch.

B. Parameter Estimation of CRF model based on Improved Particel Swarm Optimizer

In parameter estimation of CRF model, we assume the weight vector λ as particle swarm and let them fly in a d-dimension search space. The details are as follows:

Step 1. Set current iteration generation $t = 0, v_{id}(0) = 0$. Set search space as d-dimension. Initialize a population X including m particles x.

$$x_{id} = \{\lambda_{1d}, \lambda_{2d}, ..., \lambda_{nd}\}$$
$$= \{random_1, random_2, ..., random_n\}$$
$$(i = 0 \sim m-1)$$

Set local best position and global best position as: $p_{id} = x_{id}, p_g = 0$. Store the local best fitness in an array $F[i][d]$, and global best fitness in a variable F_g. An empirical value is inertia weight starting with a value close to 1 and linearly decreasing to 0.4 through the course of the run [25], so we set $w_{start} = 0.9, w_{end} = 0.4$.

Step 2. Evaluate the fitness for each particle.

$$fitness_{id} = \sum_{i=1}^{n} \log(p(y_i|x_i)) - \sum_{j=1}^{n} \frac{\lambda_j^2}{2\sigma^2}$$

Step 3. Compare the evaluated fitness value of each particle with its $F[i][d]$. If current value is larger than $F[i][d]$, then set the current position as the local best position and set current fitness as $F[i][d]$. Furthermore, if current value is larger than F_g, then reset global best position to the current position in particle array and reset F_g to the current fitness.

Step 4. Change the velocity and position of the particle:

$$v_{id}(t+1) = (w_{start} - \frac{w_{start} - w_{end}}{M} \times t) * v_{id}(t)$$
$$+(2 + \frac{t}{M}) * rand() * (p_{id} - x_{id})$$
$$+(2 - \frac{t}{M}) * rand() * (p_g - x_{id})$$

If the next velocity $v_{id}(t+1)$ is larger than $V_{\max}$, then set $v_{id}(t+1) = V_{\max}$.

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1)$$

If $p_{id} > P_{\max}$, set $P_{\max} = p_{id}$.

Step 5. If ($t\%m = 0$), calculate aggregation degree $d(t)$:

$$d(t) = \max\{| x_{id} - x_{jd} |, i, j = 1, 2, ..., m; i \neq j\}$$
$$= \max\{\sqrt{(\lambda_{di1} - \lambda_{dj1})^2 + ... + (\lambda_{din} - \lambda_{djn})^2}, i, j = 1, 2, ..., m; i \neq j\}$$
$$= \max\{\sqrt{\sum_{k=1}^{n} (\lambda_{dik} - \lambda_{djk})^2}, i, j = 1, 2, ..., m; i \neq j\}$$

If $d(t)$ is less than given threshold value e, reinitialize velocities and position of particle on d-dimension.

Step 6. $t = t + 1$. If $L(x_{id}) < 10^{-7}$ or $t = M - 1$ is met, end algorithm, else turn to step 2.

IV. EXPERIMENTAL STUDIES

We evaluated our method on bio-entity recognition based on GENIA, GENETAG and a private library. Compared with L-BFGS, stochastic gradient (abbreviated as SG-CRF), Virtual Evidence Boosting (abbreviated as VEB-CRF), Gradient Tree Boosting (abbreviated as Tree-CRF), the CRF trained by our method got a better precision and recall.

We used about 6213 semantic features during training CRFs model, most of which follow the features used by Settles [5]. We extract most features from training data, such as, "peri-kappa_B_site", "human_immunodeficiency_virus_type_2_enhancer", "monocyte", As well as "T_cell", etc.

In addition to semantic features, we used orthographic features [5] as table I.

TABLE I. ORTHOGRAPHIC FEATURES USED IN THE EXPERIMENT

Orthographic Feature	Regular Expression
Init Caps	[A-Z].*
Init Caps Alpha	[A-Z][a-z]*
All Caps	[A-Z]+
Caps Mix	[A-Za-z]+
Has Digit	. *[0-9]. *
Single Digit	[0-9]
Double Digit	[0-9][0-9]
Natural Number	[0-9]+
Real Number	[-\+][[0-9]+[\.,]+[0-9].,]+
Alpha-Numeric	[A-Za-z0-9]+
Roman	[ivxdlcm]+|[IVXDLCM]+
Has Dash	.*-.*
Init Dash	-.*
End Dash	.*-
Punctuation	[,\.;:\?!-\+"]
Greek	(alpha|beta|…|omega)
Has Greek	.*\b(alpha|beta|…|omega)\b.*
Mutation Pattern	\w*\d+-*\D+

We decode CRFs using VITER algorithm. We evaluate its performance with precision, recall and f-score. Precision is measured by the fraction of predicted gene mentions that are correct. Recall is measured by the fraction of actual gene mentions that were identified. The

F-score is a value decided by precision and recall. The three indicators are defined as bellows:

$$precision = \frac{m_c}{m_{all}} \quad (11)$$

$$recall = \frac{m_i}{m_{all}} \quad (12)$$

$$F - score = \frac{2 \times precision \times recall}{precision + recall} \quad (13)$$

In above equations, m_c means the number of bio-entity that has been recognized correctly, m_i means the number of bio-entity that has been identified, and m_{all} means total number of bio-entity included in experiment.

A. Experiment Result on GENIA

GENIA builds a corpus of annotated abstracts taken from National Library of Medicine's MEDLINE database. In GENIA Corpus, a subset of the substances and the biological locations involved in reactions of proteins is annotated, based on a data model of the biological domain, in XML format (GPML). GENIA Corpus Version 3.02 consists of 2000 abstracts. The base abstracts are selected from the search results with keywords (MeSH terms) Human, Blood Cells, and Transcription Factors.

In our experiment, we selected 1500 abstracts from GENIA Corpus Version 3.02 as training data of CRFs model and select the left 500 abstracts for testing CRFs model. Table II shows the experiment result based on GENIA.

TABLE II. A COMPARISON AMONG L-BFGS CRF, SG-CRF, VEB-CRF, TREE-CRF, AND PSO-CRF ON GENIA

Algorithm	Size	Dimension	P	R	F-Score
L-BFGS CRF	--	--	0.816	0.759	0.797
SG -CRF	--	--	0.802	0.713	0.766
VEB-CRF	--	--	0.782	0.711	0.756
Tree-CRF	--	--	0.817	0.756	0.796
PSO-CRF	20	10	0.825	0.760	0.791
		20	0.826	0.766	0.794
		30	0.820	0.766	0.795
	40	10	0.828	0.77	0.798
		20	0.829	0.769	0.798
		30	0.830	0.769	0.798
	80	10	0.833	0.771	0.800
		20	0.832	0.772	0.801
		30	0.836	0.773	0.803
	160	10	0.839	0.774	0.805
		20	0.840	0.777	0.807
		30	0.843	0.78	0.813

B Experiment result on GENETAG

GENETAG is a gene and protein corpus built by Lorraine Tababe. The data resource of GENETAG is the abstracts of MEDLINE. GENETAG is comprised of 20000 sentences selected randomly and the gene and protein names of each sentence have been annotated by some specific rules. Some justification has been made by manually. The latest version of GENETAG is GENETAG-05. We trained and compared CRFs based on GENETAG and the result is shown as table III.

TABLE III. A COMPARISON AMONG L-BFGS CRF, SG-CRF, VEB-CRF, TREE-CRF, AND PSO-CRF ON GENETAG

Algorithm	Size	Dimension	P	R	F-Score
L-BFGS CRF	--	--	0.811	0.769	0.800
SG -CRF	--	--	0.799	0.712	0.764
VEB-CRF	--	--	0.770	0.713	0.751
Tree-CRF	--	--	0.802	0.755	0.789
PSO-CRF	20	10	0.824	0.77	0.796
		20	0.823	0.771	0.796
		30	0.824	0.772	0.797
	40	10	0.833	0.773	0.802
		20	0.828	0.769	0.797
		30	0.83	0.772	0.780
	80	10	0.834	0.774	0.802
		20	0.835	0.776	0.804
		30	0.836	0.778	0.806
	160	10	0.84	0.78	0.809
		20	0.842	0.78	0.810
		30	0.849	0.787	0.817

C Experiment Resutl on Private Library

We built a private library by selecting 200 abstracts from PubMed and annotating the names of disease, gene, and protein. We use 100 abstracts as training data and other 100 abstracts as testing data. We compare L-BFGS CRF, SG-CRF, VEB-CRF, Tree-CRF, and PSO-CRF based on private library and the result is shown as table IV.

TABLE IV. A COMPARISON AMONG L-BFGS CRF, SG-CRF, VEB-CRF, TREE-CRF, AND PSO-CRF ON PRIVATE LIBRARY

Algorithm	Size	Dimension	P	R	F-Score
L-BFGS CRF	--	--	0.700	0.658	0.689
SG -CRF	--	--	0.692	0.603	0.655
VEB-CRF	--	--	0.660	0.605	0.642
Tree-CRF	--	--	0.693	0.644	0.679
PSO-CRF	20	10	0.714	0.661	0.686
		20	0.712	0.661	0.686
		30	0.714	0.663	0.688
	40	10	0.722	0.662	0.691
		20	0.719	0.659	0.689
		30	0.721	0.663	0.691
	80	10	0.724	0.664	0.693
		20	0.725	0.667	0.695
		30	0.726	0.668	0.696
	160	10	0.731	0.673	0.701
		20	0.735	0.679	0.706
		30	0.739	0.702	0.720

Figure 1 shows the performance of PSO-CRF on GENIA, GENETAG and Private library. We find PSO-CRF gets the best F-Score when the dimension size is 30 and the particle swarm size is 160.

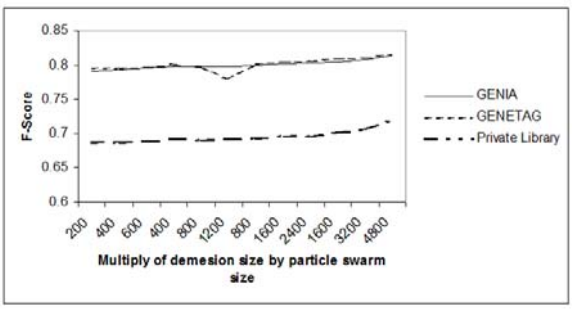

Figure 1.F-Score of PSO-CRF on GENIA, GENETAG, and Private Library

Figure 2 is an overall comparison of F-Score among L-BFGS CRF, SG-CRF, VEB-CRF, Tree-CRF, and PSO-CRF on GENIA, GENETAG and Private library.

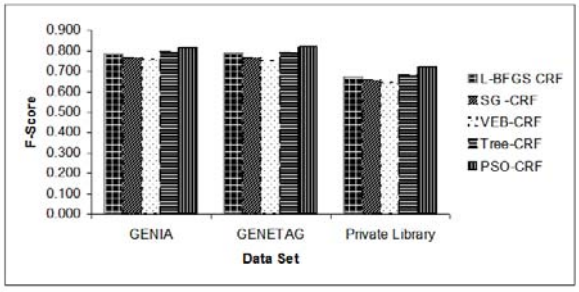

Figure 2. Comparison of F-Score among all parameter estimation algorithm on GENIA, GENETAG, AND Private Library

V.Conclusion

Conditional Random Fields model have shown to be competitive in a few domains, such as biomedical entity recognition, part-of-speech tagging, natural language processing, etc. However, Its performance is highly dependent on the parameter estimation method. This paper employs improved particle swarm optimizer to estimate the maximum likelihood parameters of CRFs model. Compared with L-BFGS CRF, SG-CRF, VEB-CRF, Tree-CRF, our method has a better precision and recall in generally.

In order to avoid PSO's early local convergence, we utilize aggregation degree of particle swarm [35] to control convergence. In order to avoid infinite iteration of PSO near by best position, we employed new stop criterion, the relative change ratio of log-likelihood between iteration, to end all iteration. In order to make particle swarm optimizer search reasonably, we make the inertia weights and learning factors change by self-adaptation. The particle size and dimension size of particle swarm optimizer are determined by experiments. In this paper, we set the particle size as 20,40 80 and 160, and set dimension size as 10,20 and 30. The experiment results prove the best values of particle size and dimension size are 160 and 30 separately.

ACKNOWLEDGEMENT

This work was supported by a grant from the National Key Natural Science Foundation of China (Grant No. 60933009), Specialized Research Fund for the Doctoral Program of Higher Education (Grant No. 200807010013) and the National Natural Science Foundation of China (Grant No.60970065).

REFERENCES

[1] AR Kinjo, F Rossello, G Valiente, Profile Conditional Random Fields for Modeling Protein Families with Structural Information, BIOPHYSICS, Vol 5,2009, pp.37-44.

[2] Brill E, Transformation-based errordriven learning and natural language processing: A case study in part-of-speech tagging, Comput. Linguistics, Vol. 21(4), pp.543–565.

[3] Brill E, Processing natural language without natural language processing, in Computational Linguistics and Intelligent Text Processing, Proceedings, Lecture Notes in Computer Science, Vol. 2588, Gelbukh, A. F., Ed, Springer, pp. 360–369.

[4] Brill E and Mooney RJ, An overview of empirical natural language processing, AI Magazine, Vol. 18(4), pp. 13–24.

[5] B Settles, 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.

[6] Chang JT, Schutze H and Altman RB, GAPSCORE: Finding gene and protein names one word at a time, Bioinformatics, Vol. 20(2), 2004, pp. 216–225.

[7] Charles A Sutton, Efficient Training Methods For Conditional Random Fields, PhD thesis, University of Massachusetts Amherst, February 2008.

[8] Chen L and Friedman C,Extracting phenotypic information from the literature via natural language processing, in Proceedings of the 11th World Congress on Medical Informatics, IMIA, San Francisco, CA, 2004,pp. 758–762.

[9] Chris Pal, Charles Sutton, and Andrew McCallum, Sparse forward backward using minimum divergence beams for fast training of conditional random Fields, In International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2006,pp.v-v.

[10] C Sutton, A McCallum, Piecewise pseudolikelihood for efficient training of conditional random fields, Proceedings of the 24th international conference on Machine learning, 2007,pp. 863-870.

[11] Erik F, Tjong Kim Sang and Fien De Meulder, Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition, In Walter Daelemans and Miles Osborne, editors, Proceedings of CoNLL-2003, Edmonton, Canada, 2003.

[12] Fletcher R and C M Reeves, Function Minimization by Conjugate Gradients, Comp.J, 7,1964,pp. 149-154.

[13] Hanisch D, Fluck J, Mevissen HT and Zimmer R, Playing biology's name game: Identifying protein names in scientific text, in Proceedings of the 8th Pacific Symposium on Biocomputing, 3rd–7th January,2003, Hawaii, pp. 403–414.

[14] Han-Shen Huang, Yu-Ming Chang, Chun-Nan Hsu, Training Conditional Random Fields by Periodic Step Size Adaptation for Large-Scale Text Mining, Seventh IEEE International Conference on Data Mining, 2007,pp. 511-516.

[15] J Lafferty, A McCallum, F Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, Proc. 18th International Conf On Machine Learning, pp. 282-289.

[16] J Kennedy and RC Eberhart, Particle Swarm Optimization, Proc. on feedback mechanism IEEE Int'l. Conf. on Neural Networks, vol. VI, IEEE Service Center, 1995,pp. 1942-1948.

[17] Kim JD, Ohta T, Tateisi Y and Tsujii J, GENIA corpus – a semantically annotated corpus for bio-textmining, Bioinformatics, Vol. 19, Suppl. 1, 2003, pp. i180–182.

[18] Lin Liao, Tanzeem Choudhury, Dieter Fox, Henry Kautz, Training Conditional Random Fields using Virtual Evidence Boosting, In Proc. of the International Joint Conference on Artificial Intelligence (IJCAI), 2007,pp.2530-2535.

[19] Lovbjerg M, Rasmussen T K, Krink T, Hybrid particle swarm optimization with breeding and subpopulations,

Proc of the third Genetic and Evolutionary computation conference, San Francisco, USA, 200l,pp. 469-476.
[20] M Al-Baali, Descent property and global convergence of the Fletcher-Reeves method with inexact line search, IMA Journal of Numerical Analysis, Vol 5,1985, pp.121-124.
[21] M Bundschus, M Dejori, M Stetter, V Tresp, HP Kriegel, Extraction of semantic biomedical relations from text using conditional random fields, BMC Bioinformatics 9 (2008): 207.
[22] Mika S and Rost B, Protein names precisely peeled off free text, Bioinformatics, Vol. 20, Suppl. 1, 2004, pp. I241–247.
[23] M Mahdaviani, T Choudhury, Fast and Scalable Training of Semi-Supervised CRFs with Application to Activity Recognition, Advances in Neural Information Processing Systems, 2007
[24] Narayanaswamy M, Ravikumar KE and Vijay-Shanker K, A biological named entity recognizer, in Proceedings of the 8th Pacific Symposium on Biocomputing, 3rd–7th January 2003, Hawaii, pp. 427–438.
[25] R Eberhart, J Kennedy, A new optimizer using particle swarm theory, Proc. 6th Int. Symposium on Micro Machine and Human Science, IEEE, 1995,pp. 39-43.
[26] Richard H Byrd, Peihuang Lu, Jorge Nocedal and Ciyou Zhu, A Limited Memory Algorithm For Bound Constrained Optimization, SIAM Journal on Scientific Computing, Volume 16, 1995, pp. 1190-1208.
[27] R Malouf, A comparison of algorithms for maximum entropy parameter estimation, In Proceedings of The Sixth Conference on Natural Language Learning (CoNLL-2002), 2002,pp. 49–55.
[28] Settles B, Biomedical named entity recognition using conditional random fields and rich feature sets, in Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications (NLPBA), Geneva, Switzerland, 2004.
[29] SS Gross, O Russakovsky, CB Do, S Batzoglou, Training Conditional Random Fields for Maximum Labelwise Accuracy, Advances in Neural Information Processing Systems, 2007,pp. 529-536.
[30] SVN Vishwanathan, NN Schraudolph, MW Schmidt, KP Murphy, Accelerated training of conditional random fields with stochastic gradient methods, Proceedings of the 23 rd International Conference on Machine Learning, Pittsburgh, PA, 2006,pp. 969-976.
[31] Tanabe L and Wilbur WJ, Tagging gene and protein names in biomedical text, Bioinformatics, Vol. 18(8), 2002, pp. 1124–1132.
[32] Thomas G Dietterich, Guohua Hao, Adam Ashenfelter, Gradient Tree Boosting for Training Conditional Random Fields, Journal of Machine Learning Research 9, 2008, pp.2113-2139.
[33] Van den Bergh F Engelbrecht AP, Training product unit networks using cooperative particle swarm optimizers, Proc of the third Genetic and Evolutionary computation conference, San Francisco, USA, 200l,pp. 126-131.
[34] Xiaoxu Ma, W Eric, L Grimson, Learning Coupled Conditional Random Field for Image Decomposition with Application on Object Categorization, IEEE Conference on Computer Vision and Pattern Recognition, June 2008,pp. 1-8.
[35] Y Guangyou. A modifed particle swarm optimizer algorithm. 8th International Conference on Electronic Measurement and Instruments, ICEMI '07,2007, pp.2-675-2-679
[36] YH Dai and Y Yuan, Convergence properties of the Fletcher-Reeves method, IMA Journal of Numerical Analysis 16, 1996, pp. 155-164.
[37] Y Hifny, S Renals, Speech Recognition using Augmented Conditional Random Fields, IEEE Transactions on Audio, Speech, and Language Processing, Volume17, 2009,pp. 354-365.
[38] Y Xiong, J Zhu, H Huang, H Xu, Minimum tag error for discriminative training of conditional random fields, Information Sciences, 2009,pp. 169-179.
[39] Zhou G, Zhang J, Su J et al, Recognizing names in biomedical texts: A machine learning approach, Bioinformatics, Vol. 20(7), 2004,pp. 1178–1190.

Zengfa.Dou received his Master degree in Computer Science from XiDian University in 2007.He obtained his BE from the Northwestern Normal University in 2003. Now, He is working for PhD degree in computer science at XiDian University. His research interest is biomedical text mining.

Lin.Gao is a professor of XiDian University. She received her PhD degree in circuit and system from XiDian university in 2003, obtained her Master degree in computing mathematics from Northwestern university in 1990 and got BE in computing mathematics from Xian Jiao Tong university in 1987. She was a visiting scholar at University of Guelph, Canada from 2004 to 2005. Her research interests include bioinformatics, data mining in biological data, graph theory and intelligence computation.