Sensors & Transducers, Vol. 157, Issue 10, October 2013, pp. 1-5

www.sensorsportal.com

Sensors & Transducers

http://www.sensorsportal.com

A Semi-Supervised User Group Identification Based on Synergetic Neural Network and Information Entropy

[1, 2] Zhehuang HUANG, [2, 3] Yidong CHEN

[1] School of Mathematics Sciences, Huaqiao University, 362021, China

[2] Cognitive Science Department, Xiamen University, Xiamen, 361005, China

[3] Fujian Key Laboratory of the Brain-like Intelligent Systems, Xiamen, 361005, China

[1] E-mail: hzh98@hqu.edu.cn

Received: 30 July 2013 /Accepted: 25 October 2013 /Published: 31 October 2013

Abstract: User group identification is an important task in intelligent personalized information service. A key problem of intelligence user server model is how to classify and indentify the user groups. Only when the user group can be effectively identified, the desired service can be offered. At present, it is difficult to obtain a large number of labeled corpuses which takes a certain amount of human and material resources. How to improve the comprehensive utilization of a small amount of labeled sample and a large number of unlabeled samples is an important task. To solve the problem, we propose novel semi-supervised user group identification based on SNN and information entropy in this paper. This paper has two main works. Firstly, a user group identification using synergetic neural network (SNN) is presented, which can effectively identify user groups; Secondly, we propose a noise filter based on information entropy to reduce the noise of expand data. The experiment results show the proposed model in this paper has a higher performance for user group identification, and provide a good practicability and a promising future for other tasks. *Copyright © 2013 IFSA.*

Keywords: User group identification, SNN, Information entropy, PSO, Semi-supervised learning.

1. Introduction

At present, network information services everywhere. But the existing resource service model is simply assigned the resources to users which can not effectively meet the practical needs of users. How to better share network resources, so as to provide a personalized service is a key problem to be solved [1]. Only when the information of user's interests, preferences, and access mode are understood better, it is possible to assign suitable information to users based on the characteristics of the users, and achieve the desired service .User modeling method should be automatically constructed according to the user's characteristics [2-4].

Synergetic neural network is proposed by Haken [5] to explain the phase transition and self-organization in non-equilibrium system. One advantage of synergetic neural network method is robust against noise and occluded, and the use of this method to realize the user group identification will be able to better handle fuzzy matching problem [6-8]. User group identification can also be considered as a problem of pattern recognition, so it is also entirely

possible to use this method to solve user group identification.

In traditional supervised learning algorithms [9], an initial classifier is trained on the large-scale labeled corpus. But it is difficult to obtain a large number labeled corpus. Semi-supervised learning [10] has become one of the important areas of research in machine learning and pattern recognition.

Entropy refers to the disarray extent of a system. The core idea of entropy is that it is the expected average information content associated with a random variable. Entropy [11] is widely used in control theory, probability theory, number theory, astrophysics, life sciences and other fields. Shannon first introduced the concept of entropy into information theory. Information Entropy [12] is a probabilistic measure of uncertainty or ignorance.

Take into full consideration of user's emotional information, preference information and Interest information, a new model was put forwarded based on user's information in this paper from a viewpoint of global which provide a unified model of high-efficient, consistent and personalized characters in the network environment to share knowledge resources.

This paper is organized as follows. Firstly, a user group identification using synergetic neural network is presented. Secondly, a semi-supervised model based on SNN and a noise filter based on information entropy is introduced. Finally some experimental tests, results and conclusions are given on the systems.

2. A Brief Introduction to Synergetic Neural Network

In synergetic, high dimension and nonlinear problem can be described as a set of low-dimension nonlinear equations, which focuses on the research of the qualitative change of macro-feature in complex system. A pattern remained to be recognized, q, is constructed by a dynamic process which translates q into one of prototype pattern vectors v_k through status $q(t)$, namely, this prototype pattern is closest to $q(0)$.The process is described as following equation: $q \to q(t) \to v_k$.

A dynamic equation can be given for an unrecognized pattern *q*.

$$\dot{q} = \sum_{k=1}^{M} \lambda_k v_k (v_k^+ q) - B \sum_{k' \neq k} (v_{k'}^+ q)^2 (v_k^+ q) v_k - C(q^+ q) q + F(t),$$

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

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

Corresponding dynamic equation of order parameters is

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

where ξ_k satisfies initial condition: $\xi_k = v_k^+ q(0)$.

Fig. 1 shows the network structure of synergetic neural network.

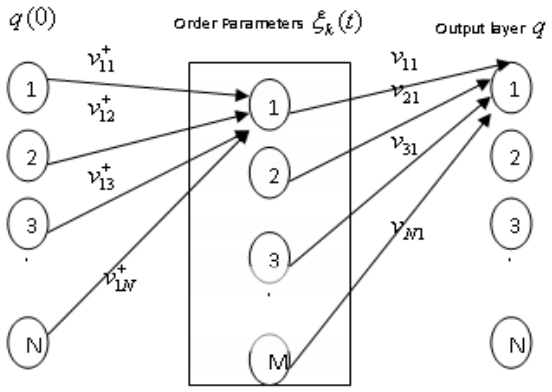

Fig 1. Network structure of synergetic neural network.

3. User Group Identification Based on SNN Model

The user group recognition procedure can be viewed as the competition progress of user group order parameters. The strongest will win by competition and desired user group will be recognized. A method for user group identification using synergetic neural network is presented in Fig. 2.

3.1. The Reconstruction of User Order Parameters

User order parameters reflect the degree of similarity between user groups. In order to make recognition process reasonably reflect the relation among user groups, it is necessary to propose a new order parameters reconstruction algorithm.

Supposed user i is represented as a vector $u_i(w_{i1}, w_{i2} \cdots w_{in})$, $i = 1, 2 \cdots$. where w_{ij} $(j = 1, 2 \cdots n)$ represent user's personalized

information, such as emotional information, preference information and interest information. The user groups are represented as $g_l\ (l=1,2\cdots)$.

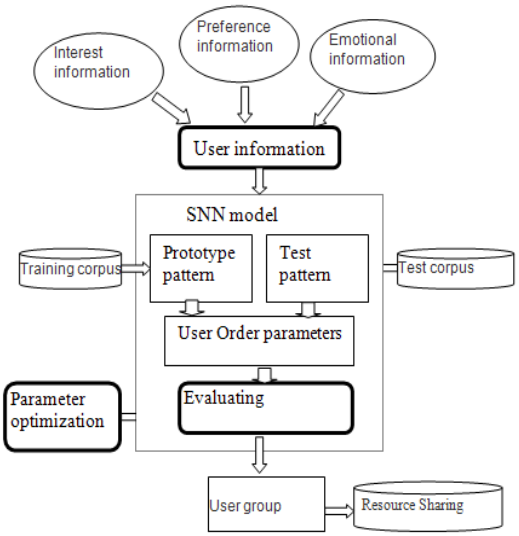

Fig. 2. User group identification based on SNN.

How to reflect the similarity between two users is a core issue, the similarity between two users u_i and u_j can be calculated using the following formula:

$$sim(u_i,u_j)=\sum_{k=1}^{n}\beta_k sim(w_{ik},w_{jk}) \qquad (1)$$

where β_k is the weight of feature,

$$sim(w_{ik},w_{jk})=\frac{\alpha}{dis(w_{ik},w_{jk})+\alpha}.$$

where $dis(w_{ik},w_{jk})$ Indicates semantic distance in Semantic dictionary TongYiCiYu Lin, α is an adjustable constant.

Then the order parameters between user i and user group $g_l(u_{l1},u_{l2},\cdots u_{lm})$ is:

$$\xi_k=\frac{\sum_{j=l1}^{lm}sim(u_i,u_j)}{m}=\frac{\sum_{j=l1}^{lm}\sum_{k=1}^{n}\beta_k sim(w_{ik},w_{jk})}{m} \qquad (2)$$

3.2. Parameter Optimization Algorithm

Corresponding dynamic equation of user order parameters is

$$\dot{\xi}_k=\lambda_k\xi_k-B\sum_{k'\neq k}\xi_{k'}^2\xi_k-C\left|\sum_{k'=1}^{M}\xi_{k'}^2\right|\xi_k \qquad (3)$$

where ξ_k can be calculated as formula (2).

The network parameters B,C,λ_k and $(\beta_1,\beta_2,\cdots\beta_n)$ are very important for better recognition performance. There is no effective way to control parameters. A novel parameters optimization algorithm based on particle swarm optimization algorithm is proposed in this paper.

Particle swarm optimization (PSO) [13] is a new evolutionary algorithm proposed by Kennedy and Eberhart in 1995. Because of simple in principle and robust, PSO has been applied successfully to all kinds of optimization problems [14, 15].

Each particle moves over the search space according to the historical behaviors of the particle and its companions. Suppose that the location of the i[th] particle is represented as $X_i=(x_{i1},x_{i2,}\cdots x_{id})$. The best previous position of the ith particle is represented as $P_i=(p_{i1},p_{i2,}\cdots p_{id})$, which is also called pbest. The location p_g is also called gbest. The velocity for the ith particle is represented as $V_i=(v_{i1},v_{i2,}\cdots v_{id})$

Then the speed and position of the particle is updated based on the following formula:

$$v_{ij}(t+1)=wv_{ij}(t)+c_1r_1[p_{ij}-x_{ij}(t)]+ +c_2r_2[p_{gj}-x_{ij}(t)]$$

$$x_{ij}(t+1)=x_{ij}(t)+v_{ij}(t+1),\ j=1,2,\cdots,d.$$

The parameter optimization algorithm based on PSO can be described as follow.

Algorithm 1. User group recognition based on SNN.

1) Construct user groups vector space from the train corpus;

2) Calculate user order parameter ξ_k according to formula (2);

3) Evolution of the order parameter according to formula (3);

4) PSO algorithm is used to search the global optimum parameters B,C, λ_k and $(\beta_1,\beta_2,\cdots\beta_n)$ in the corresponding parameter space.

5) Get best users group.

4. User Group Identification Based on Semi-Supervised Learning

Because less of labeled corpus, we present a semi-supervised user group identification model

based on SNN and a noise filter based on information entropy in this section.

4.1. Expand Algorithm Based on User Order Parameters

Firstly, an initial SNN classifier with a certain accuracy rate was constructed based on small-scale labeled data to predict some new candidate instances from unlabeled data. Secondly, user order parameters 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.

A Semi-supervised user group identification based on SNN is presented in Fig. 3.

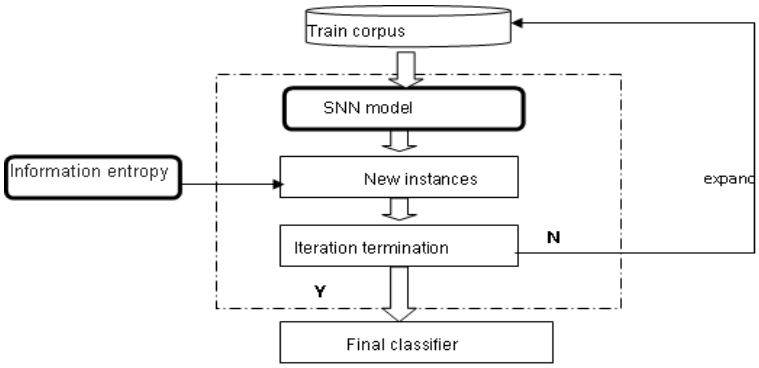

Fig. 3. Semi-supervised user group identification.

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

Algorithm 2. Semi-supervised user group identification based on SNN.

1) Construct the initial classifier with a certain accuracy rate.

2) Expand training data based on SNN.

a) Calculate user order parameter;

b) Evolution of user order parameter;

c) New instances with higher credibility were selected to expand the training data.

3) Reiteration with the expanded data until classifier performance tends to stable.

4.2. Noise Filter Based on Information Entropy

With the introduction of a large amount noise of unlabeled corpus, the noise will be significantly introduced. So the noise filter is necessary. Information entropy is a probabilistic measure of uncertainty or ignorance which provided the possibility for noise filter.

Shannon denoted the entropy H of a discrete random variable X with possible values $\{x_1, x_2 \cdots x_n\}$ and possibility $p(x_i)\,(i = 1, 2 \cdots n)$ for X taking value x_i as

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i).$$

The noise filter algorithm based on information entropy is shown as follow.

Algorithm 3. Noise filter based on information entropy.

Input: training data $U = \{u_1, u_2 \cdots u_m\}$, test pattern $T = \{t_1, t_2 \cdots t_n\}$;

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

2) Calculate the corresponding user order parameter ξ_l. Obtain the best user group through the evaluating of order parameter equation.

3) If ξ_l > threshold T_1, do the following steps:

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

b) Noise filter:

If information entropy H >threshold T_2, Then $U = U - \{t_i\}, i = 1, 2 \cdots n$

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

4) Output the final classification results.

5. Experiment

At present, it is difficult to obtain a large number of marked corpus with user information which takes a certain amount of human and material resources. In this paper, we only do some small-scale simulation experiments. We used precisions as evaluation indicators.

The performance on test set is shown in Table 1.

Table 1. The performance of semi-supervised learning.

Algorithm	Total	Correct	Precision
The improved model	85	56	0.66

The experiments show that the proposed model can well handle user group identification, so as to achieve the purpose of the shared resource for user groups.

6. Conclusions

In this paper, we present a user group identification algorithm base on SNN and information entropy. This is the first effort to introduce SNN model and information entropy to user group identification. The simulation results show the proposed model provides a good practicability and a promising future for pattern classification. We will do further investigation to inspect the inherent similarity among pattern, and apply it to other areas.

Acknowledgements

This work was supported by 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), the Science and Technology Project of Quanzhou (Grant No.2012Z91).

References

[1]. L. Kerschberg, W. Kim, A. Scime, A semantic taxonomy_based personalizable meta-search agent, in the *Proceedings of the 2th International Conference on Web Information System Engineering*, Kyoto, Japan, 2001, pp. 53-62.

[2]. M. A. Ramiz, A new sentence similarity measure and sentence based extractive technique for automatic text summarization, *Expert Systems with Applications*, Vol. 36, No. 4, 2009, pp. 7764-7772.

[3]. Diana McCarthy, Rob Koeling, Julie Weeds, and John Carroll, Finding predominant senses in untagged text, in the *Proceedings of the 42th Annual Meeting of the Association for Computational Linguistics*, Barcelona, Spain, 2004, pp. 280-287.

[4]. Huang Cheng Hui, Yin Jian, Hou Fang, A text similarity measurement combining word semantic information with TF/IDF method, *Chinese Journal of Computers*, Vol. 34, No. 5, 2011, pp. 856-864.

[5]. H. Haken, Synergetic computers and cognition – A top-down approach to neural nets, *Springer-Verlag*, Berlin, 1991.

[6]. Weizhi Wang, Lei Pan and Binghan Liu, Synergetic method of traffic state recognition based on manifold learning, in the *Proceedings of the IEEE International Conference on Automation and Logistics*, Shenyang, China, 2009, pp. 587-591.

[7]. 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*, Vol. 19, Issue 4, 2004, pp. 1140-1150.

[8]. 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, 3497, 2005, pp.26-31.

[9]. O. Bousquet, New approaches to statistical learning theory, *Ann Inst Statist Math*, Vol. 55, No. 2, 2003, pp. 371–389.

[10]. X. Zhu, Semi-supervised learning literature survey, CS TR 1530, *University of Wisconsin-Madison*, 2006.

[11]. T. Liu, W. Che, S. Li, et al., Semantic role labeling system using maximum entropy classifier, in *Proceedings of the CoNLL'05*, 2005, pp. 189–192.

[12]. Ren He, Zeng Junfang, A Chinese word extraction algorithm based on information entropy, *Journal of Chinese Information Processing*, Vol. 20, No. 5, 2006, pp. 40-44.

[13]. J. Kennedy, R. C. Eberhart, Particle swarm optimization, in *Proceedings of the IEEE International Conference on Neural Networks*, 1995, pp. 1942-1948.

[14]. R. C. Eberhard and J. Kennedy, A new optimizer using particle swarm theory, in *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, Japan, 1995, pp. 39-43.

[15]. M. Clerc and J. Kennedy, The particle swarm – Explosion, stability, and convergence in a multi-dimensional complex space, *IEEE Trans. Evol. Comput.*, Vol. 6, 2002, pp. 58–73.