

TCM Masters Miner for Knowledge Transfer

Xijin Tang, Nan Zhang, and Zheng Wang

Abstract—Traditional Chinese medicine (TCM) has a rich knowledge about human health and disease by its special way evolved along a very long history. As modern medicine is achieving much progress, arguments and disputes toward TCM never ends. To avoid losing lots of precious knowledge of TCM masters, endeavors have been engaged to systematic collection of those knowledge of TCM masters, such as their growth experiences, effective practical cases toward sickness and typical therapeutic principles and methods. Various knowledge mining has been expected to explore some useful or hidden patterns to unveil some mysteries of TCM system. This paper describes a computerized tool, TCM Master Miner, which applies different analytical methods to those collected materials about some living TCM masters in China mainland to show a different way of exposing essential ideas of those TCM masters by correspondence visualization which aims to help people understand TCM holistic views toward disease and body, and facilitate tacit knowledge transfer and sense-making of the TCM essence. This work is one kind of qualitative meta-synthesis of TCM masters' knowledge.

I. INTRODUCTION

Analysis is one of salient features of modern science. The analytical approach is the very foundation of modern medicine. Allied to the notion of analysis are the techniques of quantification and the idea of causality. Analysis is far less important to traditional Chinese medicine (TCM), which views human health and disease in terms of functional entities and disease-causing influence that are observed with the naked senses. "Its sophistication lies in its observation of correspondence between gross phenomena, and its organization of these observations through holistic systems of *yin-yang* and five *phases*"[1]. Qualitativity and holistic correspondence are two principal features in TCM whose basic concepts seem very simple while on the other hand create difficulties in applying them to practical situation. In TCM, the body is conceived as a system, rather than a machine. Then TCM diagnosis requires the identification of subtle variations of the working body and assessment of their significance in relation of each other. This is usually done by qualitative synthesis, rather than analytical reasoning. Ability to synthesize a host of subtle clues into a clear image of human's state, make differentiation of syndromes and then

give corresponding prescriptions is the mark of an experienced TCM physician.

Due to complicated reasons, TCM is confronting difficulties in its own development in comparison to that of modern medicine, which challenges current TCM education mode. Knowledge transmission of TCM meets problems, even lots of precious knowledge of TCM masters are losing. Endeavors have been taken to save those masters' tacit experiences by systematic organizing to collect knowledge of those living TCM masters, such as their growth experiences, effective practical cases toward disease, typical therapeutic methods, principles and prescriptions. On the other hand, a variety of information processing technologies have been applied to different facets of TCM research to explore some patterns or laws. Among those, data mining, text mining and ontology are widely studied [2-9]. That kind of research requires lots of datasets or prerequisites for mining. In this paper, instead of concerning IT applications to TCM research, the delivery of TCM masters' knowledge is firstly addressed as a knowledge conversion process where new insights may be acquired by TCM followers. To facilitate knowledge conversion, a computerized tool, TCM Master Miner, is designed to help find basic concepts or constructs of TCM masters' thoughts by visualizing possible correspondence and applied to the meta-synthetic engineering of TCM knowledge conversion. Firstly, TCM knowledge conversion and meta-synthetic support are addressed.

II. TCM KNOWLEDGE TRANSFER AND COMPUTERIZED SUPPORT

The ability of holistic correspondence and synthesis of determination of therapy by differentiation of syndromes can only be gained through a very long practice. Thus new TCM college graduates still require apprentice training after 5-year institutionalized learning when they need to get familiar with at least 800 herbs and 120 standard traditional formulas of prescriptions. Usually junior physicians copy prescriptions (explicit knowledge) for their mentors during daily practice for rather a period of time to gain the ability of holistic observation. That is one of biggest differences in education between TCM and modern medicine. Through learning and practice under guidance of the experienced TCM physicians, less experienced physicians may gradually sense the insights of their mentors' know-how by careful observation and practice based on their own institutional TCM knowledge, an indication of masters' tacit knowledge transfer to students' own knowledge, which could be regarded as a normal SECI (socialization, externalization, combination and internalization) process of knowledge conversion proposed

Manuscript received April 15, 2007. This work was supported by Natural Science Foundation of China under Grant No. 70571078 & 70221001 and a National Key Technologies R&D Program for TCM Research in China.

X. J. Tang is with Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100080, P.R. China (phone: + 86 10 62553291; fax: + 86 10 62568364; e-mail: xjtang@iss.ac.cn).

N. Zhang and Z. Wang are graduates students of Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100080, P.R. China. (e-mail: {wzheng, zhangnan}@amss.ac.cn).

by Nonaka and his colleague [10], while the spiral process of TCM knowledge conversion lasts longer. Mass production of modern medical doctors is impractical to train genuine TCM physicians.

To enable effective knowledge conversion, ideas of computerized support are naturally adopted to TCM knowledge conversion process to help less experienced physicians or even nonprofessionals to understand those TCM masters' thoughts easier, i.e. to acquire the essential thoughts, especially the mechanism of qualitative correspondence in diagnosis and treatment. Such kind of supporting tools is expected to bring new threads for association and expand human's thinking space. If the disease recognition is an unstructured problem, the particular diagnosing way of TCM is a problem structuring process where to acquire a clear image of the patient with a context of evidences of whole working body. Those computerized aids are expected to visualize the perspectives or structures of those TCM masters' diagnosis. It is actually one kind of qualitative meta-synthesis, i.e. to find assumptions or hypotheses about problems (syndromes) for further actions (treatment).

Among various developed supporting tools, group argumentation environment (GAE) is specifically designed to support divergent group thinking and qualitative meta-synthesis by versatile ways, such as visualization of expert opinion structure, clustering of contributed opinions, various analysis about participation, etc. and has been applied to various conference mining and social problem solving [11, 12]. However, few group activities such as conference exist in normal TCM practice. Even TCM people may prefer applying GAE, it is inappropriate to explore TCM masters' thoughts by GAE which is originally dedicated to grasp the active interaction between participants during group discussing process for procedural rationality of problem-solving process. Then a TCM Master Miner is designed with improvements of analytical technologies in GAE.

III. TCM MASTER MINER FOR THOUGHTS STRUCTURING

Current explorations by TCM Master Miner are mainly based on those materials contributed by living TCM masters. One piece of thought can be expressed with a structure as *<master's name, text of thoughts, keywords set>*, which indicates that a *master* expresses his thoughts by a *text* (one sentence) with a set of *keywords*. The *keywords set* are manually selected by domain people according to the related text. Those keywords can indicate the origin of thoughts (names of ancient TCM masters) and academic background of the master, principles of differentiation of syndromes, therapeutic approaches and prescriptions.

A. Framework of TCM Master Miner Analysis

Fig.1 shows the functional framework of current TCM Master Miner. Based on simple representation of thoughts, a

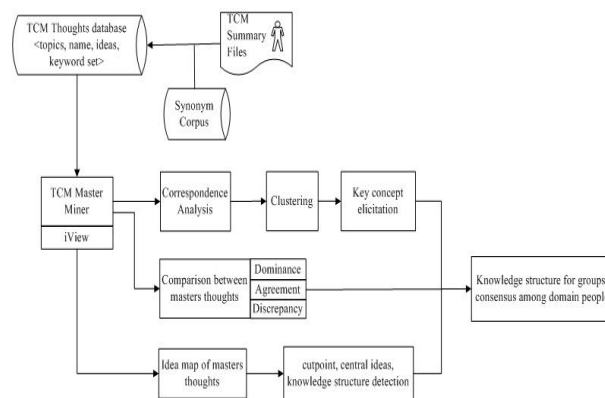


Fig. 1. Framework of TCM Master Miner (v1.0)

variety of explorations toward those masters are provided in TCM Master Miner, such as

- Visualization of correspondence between masters and their academic thoughts by exploratory analysis
- Clustering of masters' academic thoughts and key concept extraction based on correspondence analysis
- Visualization of masters' knowledge structure by idea network
- Comparisons between TCM masters, such as dominance, agreement and discrepancy, etc.

Next, the mechanisms of two feature functions are explained briefly.

B. Visualization of Correspondence between Masters and their Academic Thoughts by Exploratory Analysis

This function mainly adopts correspondence analysis which provides a method of factoring categorical variables and displaying them in a property space which maps their association in 2 or more dimensions. This method has been widely used in many disciplines [13]. Here, the singular value decomposition (SVD) is applied. Given an *m-by-n* matrix Z , SVD is defined as $Z = U\Sigma V^T$ where U is an *m-by-m* unitary matrix contains a set of orthonormal vectors called row singular vectors, Σ is *m-by-n* matrix where the diagonal elements are nonnegative singular values sorted in descending order, and V^T denotes the conjugate transpose of V , an *n-by-n* unitary matrix contains a set of orthonormal vectors called column singular vectors.

Given the records of TCM masters' thought, a frequency matrix $A = (a_{ij})$ between masters and their academic thoughts can be acquired where a_{ij} denotes the frequency of keyword j referred by the master i , $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$. The keywords are articulated as attributes of the masters. By performing a series of transformations and SVD towards the transformed matrix, a set of row vectors and column vectors are achieved and then rescaled with the original total frequencies to obtain optimal scores. These optimal scores are weighted by the square root of the singular values and become the coordinates of the points. Given the

coordinates, both masters and keywords can be mapped into 2-dimensional space. As a result, a pair of masters with more shared keywords may locate closer in the 2D space.

Such kind of analysis can be applied to any combination of available TCM masters, and may help to “drill down” into those masters’ thoughts to detect some possible or emerging academic schools among those masters which may have never been realized or admitted before. If applied to an individual master, exploratory analysis may reveal personal thinking structure.

Moreover, a variety of clustering methods such as k -means clustering can then be applied to ideas clustering and concept extraction for qualitative meta-synthesis based on the spatial relations.

The goal of such visualization is for holistic thinking. For TCM students or nonprofessionals, it is easier to grasp the fundamental concepts of TCM from the visualized map of masters’ academic thoughts, find interesting or strange ideas which may lead to some in-depth investigation for curiosity.

It should be indicated that the main goal of applying correspondence map in TCM Master Miner is to provide a basic or possible association between masters and their academic thoughts represented by keywords. As correspondence analysis is only a method for exploratory analysis, the visualized association is not a confirmatory one, even 2 dimensions may not visualize more than 75% of the association between masters and keywords.

C. Idea Map by Keyword Network

The clustering of the thoughts of the concerned masters by spatial correspondence provides perspectives of those masters, which is easier for novices to understand the major ideas of those masters. TCM Master Miner provides another way to detect structures of academic thoughts. Each text record of the masters’ thought has a group of keywords, which actually explain the basic constructs or ideas applied to the specific problem solving by the masters. Then a keyword graph $G_l = (K_l, E_l)$ of the l th record of the thoughts can be constructed where the vertex refers to a keyword $k_i \in K_l$ (K_l is the keyword set of the l th record), and if both keyword k_i and keyword k_j exist simultaneously in one record, then an edge exists between two vertices $e_{ij} = (k_i, k_j), i \neq j, e_{ij} \in E_l$ (E_l is the edge set). Each vertex is connected with all others at one keyword graph for one piece of text. The aggregation of all keyword graphs of one master or a group of selected masters brings forward a topological keyword network, $G = (K, E), K = \cup K_l = \{k_1^l, k_2^l, \dots, k_n^l\}, E = \cup E_l = \cup \{e_{ij}\}, i, j = 1, 2, \dots, m; i \neq j$. This map is a weighted undirected network where the weight of edge refers to the frequency of co-occurrence of keywords among all contributed texts of the master(s) and is referred as an idea map of the concerned master(s). Given such a network, more senses may be acquired by a variety of network analysis in detecting some features of the idea map, such as cutpoints, centrality of

keywords, clustering of keywords, etc. which expose different perspectives of the master’s knowledge scope. For example, a cutpoint (articulation point) of a graph is a vertex whose removal increases the number of connected component [14]; then the cutpoint keyword in the idea network may reveal the important concept for holistic correspondence. So does the centrality analysis, such as degree and betweenness of the keyword vertex.

Next some trials are given.

IV. PRACTICAL ANALYSIS OF SOME MASTERS’ THOUGHTS USING TCM MASTER MINER

A national key technologies R&D program for TCM research was initiated to build a platform (<http://tcm.wfcms.org/>) and collect those renowned TCM masters’ materials in China since 2005. It is a time-consuming process to upload personalized academic summaries at a standard format. With only very limited finished materials contributed by active TCM masters, original design of a variety of data mining could not be carried out. A different attempt is tried to analyze those small-size datasets of academic thoughts of a few masters using TCM Master Miner (v1.0).

A. Data Preprocessing

Before applying TCM Master Miner, data preprocessing is undertaken.

- 1) Select representative texts from each TCM master’s contributing file, which is fulfilled by domain people.
- 2) Convert each selected text into the structure: $\langle \text{master's name, text of thoughts, keywords set} \rangle$ where each keyword denotes only one idea, syndrome, disease, diagnosis or treatment.
- 3) Put all structured records of the concerned master(s) together into a data base file;
- 4) Converge keywords by their synonyms based on a synonym corpus of TCM masters’ thoughts. The corpus is not a comprehensive one but growing with increasing TCM masters’ materials. For example, TCM masters prefer to cite an ancient book; while sometimes they refer its author. For keyword convergence, if the keyword is the book name, then it is replaced by its author name.

With 8 TCM masters materials, some testing is undertaken here to show basic features of TCM Master Miner in exposing different perspectives about those masters’ thoughts, and help to experience the holistic view in TCM thinking.

B. Exploratory Analysis

Fig.2 shows a global correspondence structure of 8 selected TCM masters. It is easily to find that at the center of the map lies the keyword 内经 (actually denotes the famous TCM book *Yellow Emperor’s Inner Canon*) which is surrounded by names of some famous ancient TCM masters (keywords). This reveals a basic fact that those living TCM masters mainly got basic ideas set forth in the Inner Canon written in the beginning of the first millennium and from other ancient

masters. Moreover, the specialty of some masters could also be speculated, such as both **郑新** and **胡建华** are experts on stomach and spleen disease according to their surrounding keywords. Here 4 experts, **郑新**(in the center), **胡建华**(below the center), **张珍玉**(close to the bottom) and **何任**(close to left border), are selected and their group structure is as shown in Fig.3.

The absolute location of each expert is changed in Fig.4 while the relative location of each expert still maintains, which may infer a somewhat stable joint knowledge structure of those 4 experts. Further observation indicates those 4 experts all treat stomach and spleen disease. Moreover, it could be noticed that those keywords at the center area of Fig.4 are 行气, 益气 and 理气, all related with 气(qi, the dynamic product of the orchestration of muscle action, or the invisible but observable force that carries food downward or upward in the digestive tract), which also exposes the treatment principles applied by those TCM masters to stomach disease. With simple materials, basic principles of those TCM



Fig. 2. Visualization of 8 TCM masters' thought structure



Fig. 3. Visualization of the selected 4 TCM masters' thought structure

experts' thoughts are easily acquired.

With the spatial relations as shown in Fig.2, a centroid-based *k*-means clustering of keywords is undertaken. Here as *k* = 7, seven clusters are acquired as shown in Fig. 4.



Fig. 4. Clustering of keywords of the 8 TCM masters' thought

The keyword (whose label is of bigger size of fonts) which is closest to the centroid of the affiliated cluster could be regarded as the label of the cluster. For example, the keyword “脾胃虚寒” in Cluster No.4 of 63 keywords located in the central area of Fig.4 is denoted as the representative of that cluster, which actually reflects one kind of concept extraction. Observers can check details of that cluster and define a more appropriate label.

C. Masters' Knowledge Structure

The idea maps of both individual master and a group can be generated. From available 14 records about one selected TCM masters **张珍玉** (“Zhang Zhenyu”) in the TCM master database, 57 keywords are extracted. Fig. 5 is the keyword network of “Zhang Zhenyu” generated by TCM Master iView.

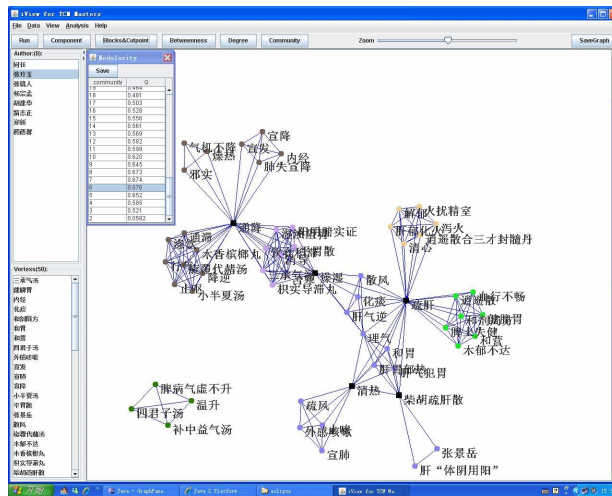


Fig. 5. One TCM masters' thought map via keyword network (cutpoint: non circle vertex)

The idea map of “Zhang Zhenyu” has 2 components and 5 cutpoints exist in the biggest component, 柴胡疏肝散, 清热, 疏肝, 燥湿 and 通降 (from bottom right to up left) which connect 5 clusters detected with the highest modularity (Q=0.676) by

Newman's fast algorithm [15]. The centrality of vertex is measured, too. For degree centrality of vertex, the top 5 vertexes are 通降 (24.00), 疏肝 (24.00), 燥湿 (14.00), 理气 (11.00) and 清热 (10.00). Here only 理气 is not the cutpoint. For centrality of betweenness, all 5 cutpoints ranks top 5. Except the prescription 柴胡疏肝散, all other keywords belong to treatment principles. By synthesis of those analyses, we may say those 6 keywords denote the constructs of knowledge system of the master "Zhang Zhenyu".

Fig. 6 is the collective keyword network of the selected 4

TCM masters whose knowledge correspondence is as shown in Fig.3. There are 35 clusters detected by Newman's fast algorithm ($Q=0.766$), where the biggest component contains 6 clusters. 理气, 益气 and 疏肝 are three cutpoints, which may reflect their principal roles among those 4 masters' knowledge. Also they locate in the central area of Fig. 3. Thus together with Fig.3, more senses could be acquired about the major treating principles applied during long practice by those 4 experts.

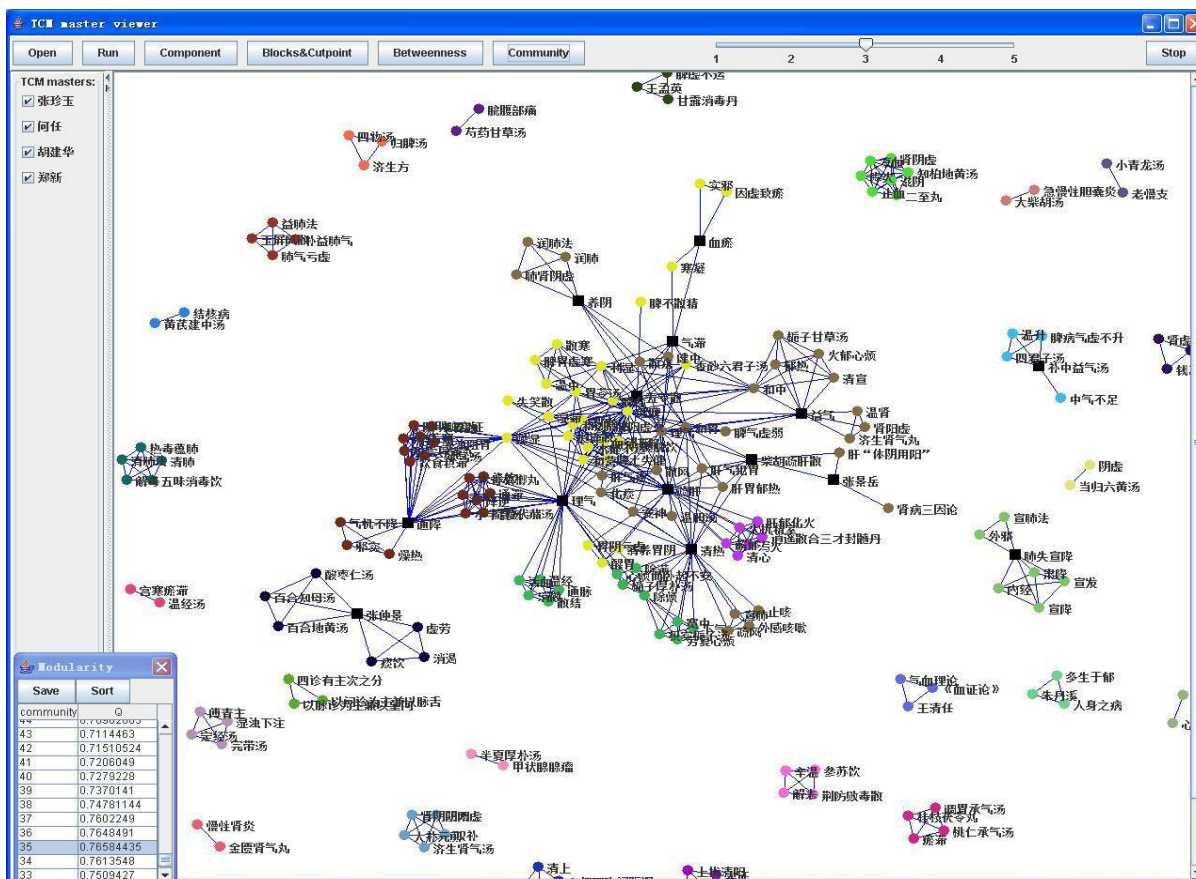


Fig. 6. Four TCM masters' thought map via keyword network (cutpoint: non circle vertex)

V. CONCLUDING REMARKS

With long history, traditional Chinese medicine confronts a lot of difficulties, such as the dissemination of its thoughts with its salient features in holistic thinking toward the working body and determination of treatment based on differentiation of syndromes. Many current TCM physicians even do not know how to apply the fundamental TCM approaches in their practice due to inappropriate education mode of TCM. This paper focus on conceptual modeling of the TCM knowledge conversion and proposes a computerized tool, TCM Master Miner, which aims to facilitate TCM knowledge conversion and qualitative meta-synthesis during

structuring process of masters' thought.

By adopting various methods, such as correspondence analysis, graph theory and complex networks analysis, TCM Master Miner provides

- perspective analysis of TCM masters' thoughts, which help people to acquire TCM basic scenario about the working body of human beings easier;
- exploratory detection of possible academic schools of current TCM masters;
- extraction of essential TCM masters' ideas;
- awareness of unknown correspondence between different masters, between syndromes, diagnosis and treatment, etc.

Many disputes are held towards the possibility of survival

of TCM, while some western philosophers and system scientists are even engaged in disseminating TCM thoughts [16], which may bring stimuli to unveil mysteries of TCM.

TCM masters belong to TCM expert system and a variety of IT supports is regarded as machine system for quantitative computing and analysis. Both systems contribute TCM knowledge to the TCM knowledge system which is being increasing and validating. Then those three systems (human expert system, machine system and knowledge system) construct a meta-synthetic system of TCM masters for TCM knowledge production. TCM Master Miner belongs to machine system and undertakes somewhat knowledge mining by exposing hidden structure of those TCM masters' thought and characteristics of basic TCM thinking, which may even reflect basic situations of current TCM diagnosis and treatment, and then help understand the situation of TCM in China in a right way.

Our current work is still at a very initial stage at both research and practice. Here only shows very basic analysis provided by TCM Master Miner in exploring living TCM famous experts' thoughts. Lots of further work will be under exploration, such as expert group detection by considering the working location of TCM masters, consideration of more semantic meanings of keywords, such as origins or background of the thoughts, concerned patterns or syndromes, treating methods and principles or prescriptions, in correspondence analysis, etc. It is also worth adopting or comparing with others' ideas [17]. Exploring a new linkage may be more difficult to be validated at TCM field. Besides, with more TCM masters' materials provided, more analysis will be undertaken for verification and validation of TCM Master Miner in practice.

ACKNOWLEDGMENT

The authors are grateful to Professor Weiliang WENG and Mr. Fei WU who provide the basic datasets and help in segmentation during this study.

REFERENCES

- [1] N. Wiseman and K. Boss, *Introduction to Glossary of Chinese Medical Terms and Acupuncture Points*. Library of Congress number 89-2982, Paradigm, Brookline, 1995.
- [2] Z.-G. Xiang, "A 3-Stage Voting Algorithm for Mining Optimal Ingredient Pattern of Traditional Chinese Medicine," *Journal of Software*, vol.14, no.11, pp. 1882-1890, 2003.
- [3] Y. N. Sun, et al, "OLAP and Data Mining Technology in Decision Supporting System for Chinese Traditional Medical Diagnosis," *Computer Engineering*, vol. 32, no.9, pp. 251-252, 255, 2006. (in Chinese)
- [4] H. Y. Liu, Y. F. Cao and L. N. Qin, "Knowledge Acquisition Method of Traditional Chinese Medical Expert by Case Base on Ontology," *Computers System and Applications*, no. 3, pp80-83, 2005. (in Chinese)
- [5] J.-F Yan and W.-F. Zhu, "Apply Rough Sets Theory in TCM Syndrome Factor Diagnosis Research," *Chinese Journal of Basic Medicine In Traditional Chinese Medicine*, vol.12, no.2, pp90-93, 2006. (in Chinese)
- [6] Y.-H. Zhu and W.-F. Zhu, "Syndrome Differentiation System of Traditional Chinese Medicine Based on Bayesian Network," *Journal*

- of Hunan University (Natural Sciences part)*, vol.33, no.4, 123-125, 2006. (in Chinese)
- [7] C.-G. Cao, "Extracting and Sharing Medical Knowledge," *Journal of Computer Science and Technology*, vol.17, no.3, 295-303, 2002.
- [8] C. Li, et al, "TCMiner: A High Performance Data Mining System for Multi-dimensional Data Analysis of Traditional Chinese Medicine Prescriptions," in *Proc. ER Workshops 2004* (S. Wang, et al. eds), LNCS 3289, Springer-Verlag, Berlin Heidelberg, pp246-257.
- [9] Z. H. Wu, et al, "Text Mining for Finding Functional Community of Related Genes Using TCM Knowledge," in *Proc. PKDD 2004* (J.-F. Boulicaut, et al. eds.), LNAI 3202. Springer-Verlag, Berlin Heidelberg pp.459-470.
- [10] I. Nonaka and Y. Takeuchi, *Knowledge Creating Company*. New York: Oxford University Press, 1995.
- [11] X. J. Tang and Y. J. Liu, "Exploring computerized support for group argumentation for idea generation," in *Proc. the 5th International Symposium on Knowledge and Systems Sciences* (KSS'2004, Y. Nakamori, et al. eds.), Japan, pp296-302.
- [12] X. J. Tang and Y. J. Liu, "Computerized Support for Qualitative Meta-synthesis as Perspective Development for Complex Problem Solving," in *Creativity and Innovation in Decision Making and Decision Support* (proceedings of IFIP WG 8.3 International Conference, CIDMDS'2006, F. Adam, et al. eds.), Vol.1. London: Decision Support Press, pp432-448.
- [13] E. J. Beh, "Simple Correspondence Analysis: a Bibliographic Review", *International Statistical Review*, vol.72, no.2, pp. 257-284, 2004
- [14] F. Harary, *Graph Theory*, Addison-Wesley, 1970. Chapter 3.
- [15] M E J. Newman, "Fast Algorithm for Detecting Community Structure in Networks", *Phys Rev E*, vol.69, no.6, 2004, 066133.
- [16] F.G. Wallner. *What Practitioners of TCM Should Know*, Berlin: Peter Lang Pub Inc, 2006.
- [17] Y. Matsuo, Y. Ohsawa and M. Ishizuka, "KeyWorld: Extracting Keywords from a Document as a Small World", in *Proc. DS 2001* (K.P.Jantke and A. Shinohara eds.), LNAI 2226, Springer-Verlag, Berlin Heidelberg, pp271-281.