Exploration of TCM Masters Knowledge Mining

Xijin Tang, Nan Zhang, and Zheng Wang

Academy of Mathematics and Systems Science, Chinese Academy of Sciences
Beijing 100080 P.R.China
{xjtang,zhangnan,wzheng}@amss.ac.cn

Abstract. With a very long history, traditional Chinese medicine (TCM) has a rich knowledge about human health and disease by its special way. To avoid losing lots of precious knowledge of TCM masters, endeavors have been engaged to keep down those knowledge of TCM masters, such as their growth experiences, effective practical cases toward sickness and typical treating methods and principles. In this paper, some computerized methods are applied toward those collected materials about some alive TCM masters in China mainland to show a different way of exposing essential ideas of those TCM masters which aims to help people understand the correspondence of TCM views toward disease and body, and facilitate tacit knowledge transfer. This work is one kind of qualitative meta-synthesis of TCM masters' knowledge.

Keywords: Traditional Chinese medicine, knowledge mining, idea map, meta-synthesis approach.

1 Introduction

Analysis is one of salient features of all 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. 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 synthesis, rather than analytical reasoning. Ability to synthesize a host of subtle clues into a clear image and thus actually visualize a patient's condition 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. The 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 keep down the knowledge of those alive TCM masters, such as their growth experiences, effective practical cases toward disease, typical treating methods, principles and prescriptions. On the other

hand, a variety of information technologies have been applied to different facets of TCM research to find some patterns or laws. Among those, data mining, text mining and ontology are widely studied [2-9]. Those 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 and applied to the meta-synthetic engineering of TCM masters' knowledge conversion. Next TCM knowledge conversion is addressed.

2   TCM Knowledge Transfer and Computerized Support

Due to the holistic correspondence feature, new TCM college graduates still require apprentice training after several year of institutionalized learning. Usually junior physicians write down 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, 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 by Nonaka and his colleague [10]. While the 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 framework or structure of their 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. 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. [11, 12]. GAE has been applied to various conference mining, on-line group discussions of social issues, etc. However, few group activities such as conference exist in TCM practice. It is inappropriate to apply GAE directly to exploring TCM masters' thoughts. A TCM Master Miner is designed with improvements of analytical technologies in GAE.

3   TCM Master Miner for Thoughts Structuring

Current explorations by TCM Master Miner are mainly based on those materials contributed by TCM masters. One piece of thought can be expressed by 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* is manually selected by domain people according the related text. Based on simple representation of thoughts, a 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 concept extraction
– Visualization of idea structure by keyword network
– Comparisons between TCM masters, such as dominance, agreement and discrepancy, etc.

Next mechanisms of two feature functions are explained briefly.

3.1   Visualization of Correspondence Between Masters and Their Academic Thoughts by Exploratory Analysis

This is achieved by correspondence analysis using frequency matrix $F = (a_{ij})$, a_{ij} denotes the frequency of keyword j referred by the master i, $i = 1,2,\ldots,m$; $j = 1,2,\ldots,n$. The keywords are articulated as attributes of the masters.

Given frequency matrix, the mechanism of correspondence analysis is employed to explore the correspondence relations between masters and keywords; the principal components for given relations between keywords and masters could be acquired, then 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.

In TCM Master Miner, "exploratory analysis" fulfills above computing and displays the global structure of masters' collective thoughts. 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 even they have never realized before.

As a matter of fact, the visualized map may be more useful in understanding masters' thought quickly, stimulating further thinking, such as finding interesting or strange ideas which are worth in-depth investigation, etc.

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

3.2   Idea Viewer 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. While above mentioned clustering is not the only way to detect structures of academic thoughts. Here is another way. 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 lth 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 lth record), and if both keyword k_i and keyword k_j occur simultaneously in one record, then an edge exists between two vertexes $e_{ij} = (k_i, k_j), i \neq j$, $e_{ij} \in E_l$ (E_l is the edge set). Each vertex is connected with all others at one keyword graph for one piece of text. Then the aggregation of all complete keyword graphs of one master or a group of selected masters brings forward a topological keyword network, $G = (K, E)$, $K = \cup K_l =$ { $k_1^l, k_2^l, \cdots, k_n^l$ }, $E = \cup E_l = \cup \{e_{ij}\}$, $i, j = 1,2,\ldots,m; i \neq j$. This map is a weighted undirected network where the weight of edge refers to the frequency of co-occurrence of both keywords among all contributed texts of the master(s) and is referred as an idea map of the concerned master(s). Various network analyses can then be undertaken to detect different perspectives of the master's knowledge scope. The basic mechanism had already been discussed in Ref. [13].

Next some trials are given.

4   Practical Analysis of Some Masters' Thoughts Using TCM Master Miner

Before applying TCM Master Miner, data preprocessing is undertaken.

1) Select representative texts from each TCM master's contributing file;
2) Convert each selected text into the structure: <*master's name, text of thoughts, keywords set*> 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 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.

Fig.1 shows a global correspondence structure of the 8 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 the basic fact that those alive 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,

Fig. 1. Visualization of the 8 TCM masters' thought structure

Fig. 2. Visualization of the selected 4 TCM masters' thought structure

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.2.

The absolute location of each expert is changed in Fig.2 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 of Fig. 2 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
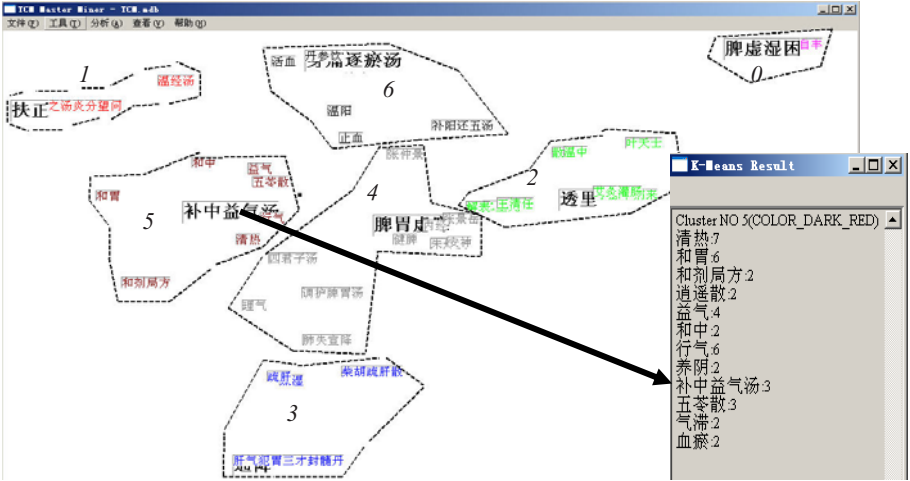
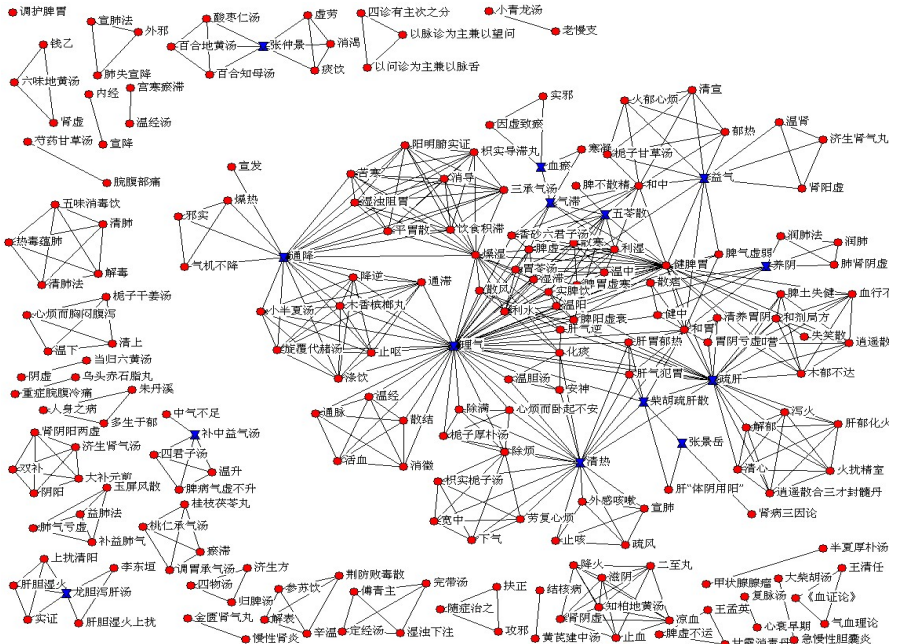**Fig. 3.** Clustering of keywords of the 8 TCM masters' thought

Fig. 4. Four TCM masters' thought map via keyword network (cutpoint: non circle node)

those TCM masters to stomach disease. With simple materials, basic principles of those TCM experts' thoughts are easily acquired.

With the spatial relations as shown in Fig.1, a centroid-based *k*-means clustering of keywords is undertaken. Here as *k* = 7, seven clusters are acquired as shown in Fig. 3.

The keyword (whose label is of bigger size of fonts) which is closest to the centroid of the affiliated cluster could be regarded as label of the cluster. For example, Cluster No.5 includes 12 keywords and the keyword "补中益气汤" 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.

Fig. 4 is the keyword network of those 4 selected TCM masters whose knowledge correspondence is as shown in Fig.2. 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, keyword structure of the network, etc. For example, 行气 and 益气 are two cutpoints, which may reflect their principal roles among those 4 masters. Together with Fig.2, more senses could be acquired about the major treating principles applied by those 4 experts.

5   Concluding Remarks

With long history, traditional Chinese medicine still confronts a lot of difficulties, such as the dissemination of its thoughts. This paper initially focus on modeling the TCM knowledge conversion and then 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.

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 increasing and validating TCM knowledge system. Then those three systems 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 a right way.

Our current work is still at a very initial stage at both research and practice. Here shows very basic analysis provided by TCM Master Miner in exploring alive TCM famous experts' thoughts due to space limits. 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. Besides, with more TCM masters' materials provided, more analysis will be undertaken for verification and validation of TCM Master Miner in practice.

References

1. Wiseman, N., Boss, K.: Introduction to Glossary of Chinese Medical Terms and Acupuncture Points. Library of Congress number 89-2982, Paradigm, Brookline (1995)
2. Xiang, Z.-G.: A 3-Stage Voting Algorithm for Mining Optimal Ingredient Pattern of Traditional Chinese Medicine. Journal of Software. 14(11) (2003)1882-1890
3. Sun, Y.N., et al. OLAP and Data Mining Technology in Decision Supporting System for Chinese Traditional Medical Diagnosis. Computer Engineering. 32(9) (2006) 251-252, 255 (in Chinese)
4. Liu, H.Y., Cao, Y.F., Qin, L.N.: Knowledge Acquisition Method of Traditional Chinese Medical Expert by Case Base on Ontology. Computers System and Applications. 3 (2005) 80-83 (in Chinese)
5. Yan, J.-F, Zhu, W.-F.: Apply Rough Sets Theory in TCM Syndrome Factor Diagnosis Research. Chinese Journal of Basic Medicine In Traditional Chinese Medicine.12(2) (2006) 90-93 (in Chinese)
6. Zhu, Y.-H., Zhu, W.-F.: Syndrome Differentiation System of Traditional Chinese Medicine Based on Bayesian Network. Journal of Hunan University (Natural Sciences). 33(4) (2006) 123-125 (in Chinese)
7. Cao, C.-G.: Extracting and Sharing Medical Knowledge. Journal of Computer Science and Technology. 17(3) (2002) 295-303
8. Li, C., et al.: TCMiner: A High Performance Data Mining System for Multi-dimensional Data Analysis of Traditional Chinese Medicine Prescriptions. In: Wang, S., et al. eds.: ER Workshops 2004, LNCS 3289. Springer-Verlag, Berlin Heidelberg (2004) 246-257
9. Wu, Z. H., et al.: Text Mining for Finding Functional Community of Related Genes Using TCM Knowledge. In: Boulicaut, J.-F., et al. (eds.): PKDD 2004, LNAI 3202. Springer-Verlag, Berlin Heidelberg (2004) 459-470
10. Nonaka, I., Takeuchi, Y.: Knowledge Creating Company. Oxford University Press, New York (1995)
11. Tang, X. J., Liu, Y. J.: Computerized Support for Idea Generation during Knowledge Creating Process. In: Cao, C.G., Sui, Y.F. (eds.): Knowledge Economy Meets Science and Technology (KEST'2004). Tsinghua University Press, Beijing, (2004) 81-88
12. Tang, X. J., Liu, Y. J.: Exploring computerized support for group argumentation for idea generation. In: Nakamori, Y. et al. (eds.): Proceedings of the 5th International Symposium on Knowledge and Systems Sciences (KSS'2004), Japan (2004) 296-302
13. Tang, X. J., Liu, Y. J.: Computerized Support for Qualitative Meta-synthesis as Perspective Development for Complex Problem Solving. In: Adam, F. et al. (eds.): Creativity and Innovation in Decision Making and Decision Support (proceedings of IFIP WG 8.3 International Conference, CIDMDS'2006), Vol.1. Decision Support Press, London (2006) 432-448