A Framework for Automated Knowledge Graph Construction Towards Traditional Chinese Medicine

Heng Weng[1], Ziqing Liu[2], Shixing Yan[3], Meiyu Fan[1], Aihua Ou[1], Dacan Chen[1(✉)], and Tianyong Hao[4(✉)]

[1] The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, China
ww128@qq.com, 1175383819@qq.com, 4910702@163.com
[2] The Second Clinical Medical College, Guangzhou University of Chinese Medicine, Guangzhou, China
lzq_lby@163.com
[3] Department of Control Science and Engineering, Tongji University, Shanghai, China
yanshixing@jindengtai.cn
[4] School of Information Science and Technology, Guangdong University of Foreign Studies, Guangzhou, China
haoty@gdufs.edu.cn

Abstract. Medical knowledge graph can potentially help knowledge discovery from clinical data, assisting clinical decision making and personalized treatment recommendation. This paper proposes a framework for automated medical knowledge graph construction based on semantic analysis. The framework consists of a number of modules including a medical ontology constructor, a knowledge element generator, a structured knowledge dataset generator, and a graph model constructor. We also present the implementation and application of the constructed knowledge graph with the framework for personalized treatment recommendation. Our experiment dataset contains 886 patient records with hypertension. The result shows that the application of the constructed knowledge graph achieves dramatic accuracy improvements, demonstrating the effectiveness of the framework in automated medical knowledge graph construction and application.

Keywords: Knowledge graph · Semantic analysis · Personalized medical service

1 Introduction

With the fast prevalence and development of precision medicine, more and more patients are seeking personalized medical treatment services. This requires clinicians to continuously pay attention to the rapid development of medical research, and accumulate effective clinical treatment cases based on a large amount of domain knowledge. Clinicians also need to analyze and summarize historical treatment cases time by time. This bring a huge burden on clinicians of new knowledge learning from large amount of medical data in such an information exploration era [1].

S. Siuly et al. (Eds.): HIS 2017, LNCS 10594, pp. 170–181, 2017.
https://doi.org/10.1007/978-3-319-69182-4_18

Medical data has the nature characteristics of big data including volume, variety, velocity, and veracity [2, 3], thus bringing challenges for the storing, transferring, and processing of continuously emerging medical data [4]. On the other hand, the developing data processing techniques provide opportunities for leveraging medical data to assist clinicians in many applications [5], e.g., medical decision support [6–8], medical knowledge mining [9, 10], drug discovery analytics [11, 12], etc. Therefore, considering the limited time of clinicians, extracting knowledge from medical data for personalized treatment are both necessary to assist clinicians and help them improve working efficiency.

Knowledge discovery based on Human-Computer Interaction (HCI) may be a promising approach for such purpose, as addressed by Holzinger [13]. In the knowledge discovery models, Knowledge Graph (KG) obtains increasingly attention in medical domain evidenced by its capability of predicting the cancer clinical treatment via the combination with other patient information such as gene [14]. Moreover, it has been successfully applied on the hyperosmolar byperglycemic state management for ICU adult patients [15]. KG can assist clinicians in retrieval and understanding the clinical practice guidelines and protocols as well. Consequently, KG can be used not only for mining potential hidden knowledge, but also for assisting clinicians in their academic research, clinical decision support, knowledge retrieval, etc.

To assist clinicians in high efficient knowledge learning and retrieval, this paper proposes a framework for Traditional Chinese Medicine (TCM) knowledge graph construction through the information extraction from existing clinical texts. The framework is based on a semantic analysis network containing a large amount of meta knowledge, as the nodes in the network. The constructed knowledge graph can be aggregated into structured vector representations according to different dimensions for the convenience of semantic distance calculation and semantic inference. According to the evaluation of medical dataset containing 866 real patient cases with hypertension, the result shows that the classification performance has been significantly improved by applying the constructed TCM knowledge graph. The experiments indicate that the proposed framework can help data modeling in knowledge graph construction, demonstrating its effectiveness. We also present how the constructed TCM knowledge graph can potentially benefit clinical application such as personalized treatment recommendation.

2 Related Work

Knowledge graph is a symbolic expression of the physical world, which generalizes the world into a logical link among all conceptual entities and attributes. From the perspective of the graph theory, knowledge graph is essentially a conceptual network in which the nodes represent the entities (or concepts) of the physical world, and the edges represent various semantic relations among the entities. The medical concepts are commonly organized in hierarchical structures while the relations among conceptual entities and attributes are intricate.

In Traditional Chinese Medicine (TCM) domain, there are some existing research works on TCM knowledge graph construction. Zhang et al. [16] addressed that the basic

structure of TCM knowledge graph consisted of concept hierarchical relations and entity relations. They defined semantic inferences between the nodes according to general TCM knowledge. They regarded knowledge graph as a mapping between the relational tree of concepts and the relational graph of entities. However, the research only provided the application direction of TCM knowledge graph without offering practical application cases. Moreover, the semantic references still relied on the manual work of domain experts.

Yu et al. [17] focused on the concept organization of TCM and integrated the structured knowledge resource into a large-scale knowledge graph, which embedded with literature search, knowledge retrieval and other functions to provide knowledge navigation, integration and visualization services, etc. Based on an ontology, the knowledge graph was further divided into concept semantic network and thesaurus. The former defined the correlation among TCM concepts and knowledge resources, while the latter structured concepts and terms. The research reported some promising applications in KG visualization and ontology retrieval. However, the method still needed tedious manual work on semantic inference definition.

Shi et al. [18] claimed that a computation framework for Textual Medical Knowledge (TMK) is necessary to construct a TCM knowledge graph. They emphasized that the usage of framework needed to meet three requirements: (1) able to organize heterogeneous TMK and integrate with HIS data to transfer data; (2) should have reasonable knowledge element expressions supporting both human and machine interpretation to realize efficient retrieval; (3) should have a retrieval function to facilitate the promotion of latest knowledge to users. They constructed a healthcare organization model that contained three parts: Medical Knowledge Model (MKM), Health Data Model (HDM), and Terminology Glossary (TG), for organizing TMK into concept maps to define normalize Electronic Health Records (EHRs) and to provide the meta-thesaurus of TMK and HDM cases. It applied First-order Predicate Logic for semantic inference and adopted text categorization algorithms to rectified semantic inference errors. Yet, the application still has limitations on practical applications such as clinical prescription patterns summarization.

The existing works focused on the content-aware natural language processing. It was feasible for acquiring knowledge with explicit description. However, they seldom deal with hidden knowledge with implicit descriptions in medical texts, e.g., main syndrome and concurrent syndrome, prescription based on syndrome differentiation, etc. To that end, we propose a new automated extraction method for TCM knowledge graph construction. The purpose of the TCM knowledge graph is to realize automatic extraction of semantic inference, discovering hidden knowledge in accumulated treatment cases of experienced physicians and finding diagnose, treatment and prescription patterns, etc. The knowledge graph includes two kinds of the visualization of complex knowledge element associations. This research also applies deep learning technology to annotate each knowledge unit with individual coordinate mapping and distance information to express the correlation among knowledge elements, which can not only be used in data description of current TMK to bring clinic physicians convenience in understanding general ideas of data set, but also be applied in relevant research work such as couplet medicine retrieval, core prescription, single substance drug, etc.

3 The Framework

A knowledge graph construction framework based on the ontology model and deep learning technique is proposed. The framework aims to automate the meta knowledge extraction and conversion processes which transfer meta knowledge to vector representation for semantic distance calculation and semantic inference. The vector representation is used to regenerate structured datasets according to clinical scenario differences. The generated datasets can be stored into meta knowledge warehouse for further usage. Each sample of the dataset exists in a sparse matrix and is assigned with a list of labels, where the labels correspond to meta knowledge. The generated datasets are further used to train a Recurrent Neural Network (RNN) [19] model for calculating the semantic distance and relation paths of given meta knowledge to discover the potential hidden knowledge so as to construct a domain-specific knowledge graph. As shown in Fig. 1, the whole framework consist of four main modules including: (1) a medical

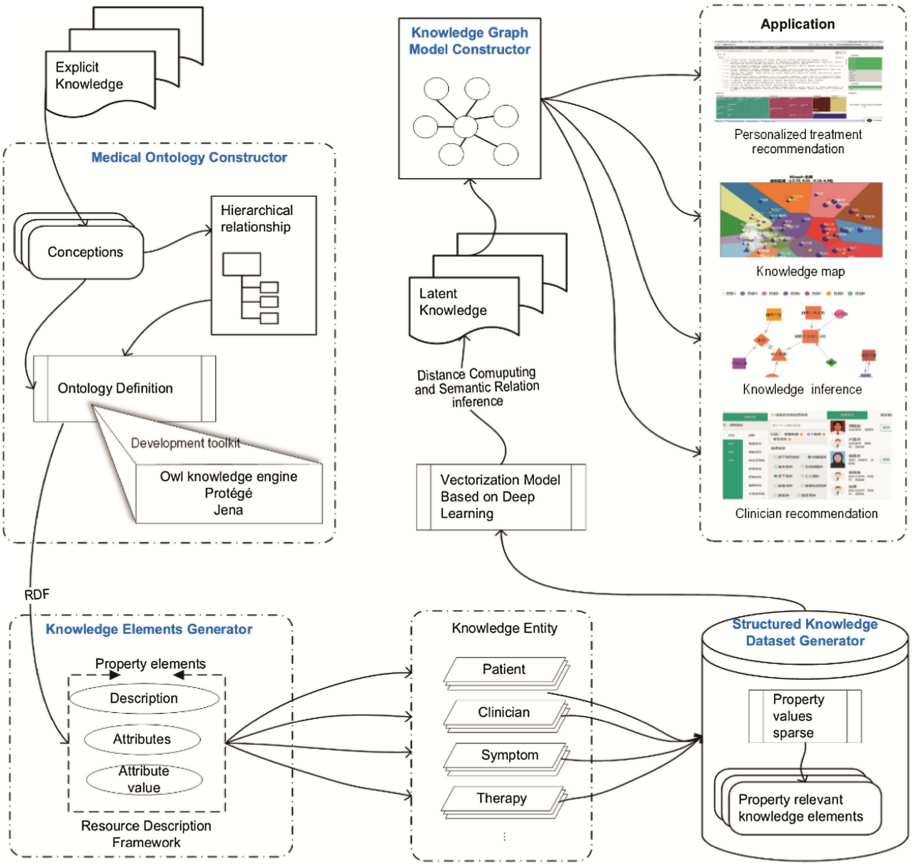

Fig. 1. The framework of TCM Knowledge Graph construction

ontology constructor, (2) a knowledge element generator, (3) a structured knowledge dataset generator, and (4) a graph model constructor.

The Medical ontology constructor is the module to construct medical domain ontology using explicit knowledge. Utilizing Natural Language Processing (NLP) technique, e.g., named entity recognition and text classification, we extract meta data from unstructured clinical texts. After that, The explicit knowledge including expert-defined traditional Chinese medicines, modern medical knowledge from clinical protocol guidelines and medical textbooks are acquired. According to the Chinese medicine terminology standards published by Chinese government, we generate a hierarchical structure as the base of the ontology by following the Resource Description Framework (RDF) and Ontology Web Language (OWL-Lite). The process is under the supervision of domain experts and assisted with an ontology edition tool Protégé[1].

The knowledge element generator is a module to generate knowledge triples containing meta knowledge attributes and relations. Here "meta knowledge" is an extensive notion including all concepts and their relations defined in RDF. For example, "inspection" associates with specific scope (related to human body parts) including "head", "thoracoabdominal", "limb", "sprit", "urination & defecation", etc. "head" further associates with "face", "eye", "lip", "tongue", etc. Every meta knowledge has attributes with attribute values. For example, "tongue body" has the attribute values "tough", "tender", "enlarged", "thin", "luxuriant", "withered", etc. Therefore, a specific disease ontology has rich information in terms of concepts, relations, attributes, attribute values. All concepts in the same ontology have semantic similarity calculated through their locations, the depths, and nearby densities in the ontology structure. The relevant concepts are closer, e.g., "floating pulse"-"sunken pulse" and "limb"-"foot". The generated meta knowledge triples can be used for semantic inference in the knowledge graph construction procedure.

The structured knowledge dataset generator is a module to map real word data to meta knowledge for structuring medical text data to adapt different application scenarios. The medical texts contain ancient literature, Electrical Medical Record (EMR), public health textbox, scientific articles for health education, etc. The original texts are used to establish mapping relations with generated knowledge elements. Due to the differences of practical applications, the dataset organization method may also alters accordingly to form knowledge entities, namely, dimensional aggregation (e.g., from clinician, patient and disease dimensions) of knowledge element nodes according to different perspectives. Each category of entities contains related knowledge element nodes, e.g., the clinician dimension contains symptom, treatment, etc., while the patient dimension contains disease history, symptoms, inspection indexes, etc. Using the module, each data sample is automatically structuralized into a sparse matrix, which is the +collection of involved knowledge elements with corresponding attributes structured values. The structured datasets are internally related to the medical ontology repository.

The graph model constructor is a module to construct knowledge graphs based on the structured knowledge datasets and to generate knowledge maps and knowledge element networks. Each involved knowledge element is transformed into a vector

[1] http://protege.stanford.edu/.

representation after the structured datasets goes through a vectorization model based on deep learning algorithms. To calculate the semantic distance and the inference of semantic relations, an unsupervised learning is applied to generate a knowledge map by calculating the distance among knowledge element vectors according to preset categories. The semantic inference refers to the prediction of correlations of knowledge elements based on the graph model, which returns the weighted directed complex network according to relation weights. The knowledge map reflects the latent correlation among knowledge elements, and the directed knowledge element complex network reflects the latent logical relation among knowledge elements, while the weight reflects the popularity degree of the logical rules. The entire construction process of the knowledge graph can be regarded as a process of discovering latent knowledge.

4 Experiments and Results

To evaluate the effectiveness of the framework in Traditional Chinese Medicine (TCM) knowledge graph construction, we use a publically available "Levis hypertension" Chinese clinical dataset [20], which contains 908 hypertension TCM cases. The dataset has rich case information and each case has 129 dimensions of diagnosis and symptoms including "inspection diagnosis" (*望诊*), "inquiry diagnosis" (*问诊*), "tongue diagnosis and palpation diagnosis" (*舌脉*) etc. After removing 22 cases because of diagnosis information missing, we obtain 886 cases eventually for the evaluation with ten-folder cross-validation. The summary of the dataset is shown in Table 1.

Table 1. The summary of the hypertension TCM dataset.

Dataset	Feature dimensions	# of cases	Data type
Training set	129	797	*Boolean*
Testing set	129	89	*Boolean*
Total	129	886	*Boolean*

According to the standards of the syndrome of TCM (*中医证候*) [21], we manually extract major characteristics of TCM syndrome for each case and use them as gold reference labels, in which each case has 2 to 5 labels. The experiment on the dataset thus is converted to a multi-label classification problem. Part of the characteristics of TCM syndrome elements (*证候要素*) is listed in Table 2.

In order to optimize the iteration parameter β in the learning process, we use the ML-KNN algorithm [22] and RAKEL-SMO algorithm [23] on the training dataset. Using evaluation metrics including hamming loss, average precision, micro-averaged precision, micro-averaged F-measure, macro-averaged precision, macro-averaged F-measure, and micro-averaged AUC, the performances are presented in Fig. 2. The first of Fig. 2 shows that the ML-KNN algorithm ($k = 12$, $V = 0.1$) tends to be more stable when iteration β is greater than or equal to 75, while the second presents that the RAKEL-SMO algorithm ($S = 6$, $V = 0.1$) becomes stable when β is greater than or equal to 100. We therefore select the best iteration parameter β as 100.

Table 2. The summarized characteristics of the symptoms of Traditional Chinese Medicine

Categorized inquiry features	Clinical symptoms
Face	Facial flush, Red face, Dark pale complexion, Flushing, Pale white, A pale complexion, Pale complexion
Head	Empty pain of head, Dizziness, Head and eye distending pain, Headache, Fullness in head, Heavy sensation of head, Head with binding sensation, Vertigo
Eye	Red eye, Dry eye, Dizzy, Fullness in eye, Hypopsia, Blurred vision
Lip	Purple lips, Lip colorless, Dark lips
Ear	Eared, Deaf, Tinnitus
Thoracoabdominal & limb	Oppression in chest, Pectoralgia, Distention and fullness, Hypochondriac pain, Soreness of waist, Limb numbness, Leg soft, Ventosity, Soreness and weakness of knees

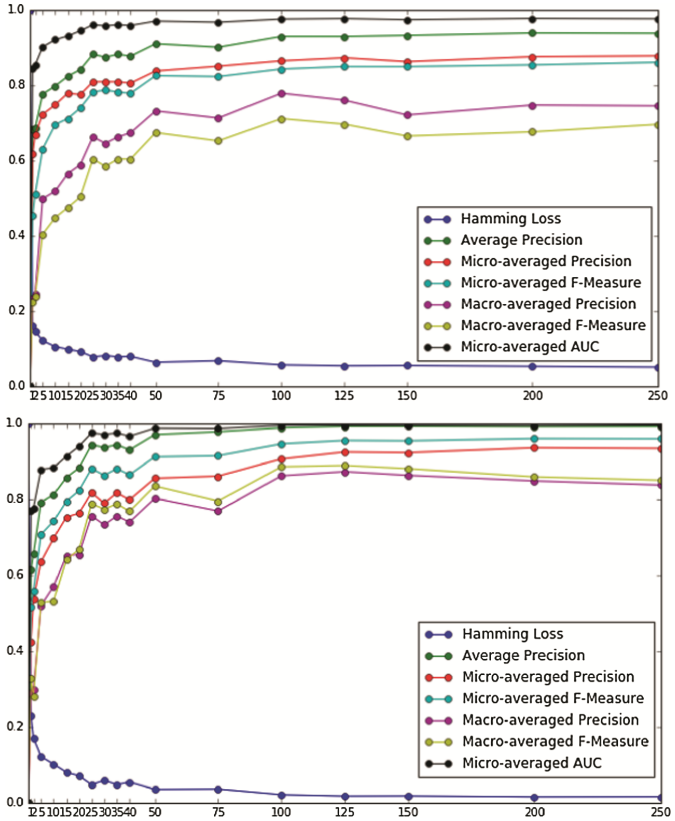

Fig. 2. The performance of ML-KNN and RAKEL-SMO algorithms with the increasing number of learning iterations

Due to difficulties to acquire entity relations corresponding to knowledge graph from unstructured texts directly, the annotation on texts to build a gold standard for the evaluation relation prediction by knowledge graph thus is infeasible. Therefore, we test the effectiveness of the constructed TCM knowledge graph by comparing the classification performance differences of machine learning algorithms with and without the knowledge graph. Using the exact same ML-KNN and RAKEL-SMO algorithms with the optimized iteration β, we take the converted vectors of meta knowledge from the TCM knowledge graph as features as "KG" regarding to "conventional" features using commonly used algorithms. In each experiment, ten-folder cross-validation evaluation was used on the testing dataset. The comparison result is reported in Table 3.

Table 3. The performance comparison with and without knowledge graph vectors as features.

Methods/Metrics	ML-KNN ($k = 12$, $V = 0.1$)			RAKEL-SMO ($S = 6$, $V = 0.1$)		
	Conventional	+KG	Change	Conventional	+KG	Change
Micro-averaged precision	0.717 ± 0.064	0.888 ± 0.022	23.8%	0.686 ± 0.057	0.945 ± 0.019	37.8%
Micro-averaged recall	0.489 ± 0.037	0.848 ± 0.019	73.4%	0.608 ± 0.051	0.988 ± 0.005	62.5%
Micro-averaged F-measure	0.581 ± 0.043	0.867 ± 0.017	49.2%	0.644 ± 0.051	0.966 ± 0.010	50.0%
Macro-averaged precision	0.376 ± 0.064	0.760 ± 0.032	102.1%	0.340 ± 0.050	0.861 ± 0.071	153.2%
Macro-averaged recall	0.289 ± 0.038	0.671 ± 0.041	132.2%	0.526 ± 0.054	0.881 ± 0.060	67.5%
Macro-averaged F-measure	0.303 ± 0.040	0.697 ± 0.034	130.0%	0.404 ± 0.052	0.868 ± 0.063	114.9%
Average precision	0.745 ± 0.048	0.946 ± 0.011	27.0%	0.755 ± 0.044	0.980 ± 0.007	29.8%
Mean average precision	0.385 ± 0.038	0.813 ± 0.047	111.2%	0.418 ± 0.045	0.908 ± 0.044	117.2%
Ranking loss	0.119 ± 0.021	0.022 ± 0.005	−81.5%	0.132 ± 0.026	0.006 ± 0.004	−95.5%
Logarithmic soss	4.626 ± 0.344	1.937 ± 0.227	−58.1%	14.06 ± 2.341	0.732 ± 0.275	−94.8%

From the results, the ML-KNN and RAkEL-SMO algorithms with conventional feature extraction strategy obtain an average precision of 0.745 ± 0.048 and 0.755 ± 0.044, respectively. By combining with the knowledge graph (+KG), the average precision is increased to 0.946 ± 0.011 and 0.980 ± 0.007 with an improvement of 27.0% and 29.8%, respectively. Similarly, the micro-averaged F-measure performance is increased from 0.581 ± 0.043 and 0.644 ± 0.051 to 0.867 ± 0.017 and 0.966 ± 0.010 with an improvement of 49.2% and 50.0%, while the macro-averaged F-Measure performance is increased from 0.303 ± 0.040 and 0.404 ± 0.052 to 0.697 ± 0.034 and 0.868 ± 0.063 with an improvement of 130.0% and 114.9%, respectively. The results on ranking loss and logarithmic metrics also show the usage of TCM knowledge graph significantly outperforming the conventional feature extraction, demonstrating that the constructed TCM knowledge graph can benefit the performance of machine learning algorithms on multi-label classification tasks.

5 Discussions

The constructed TCM knowledge graph can be visualized into dynamic map for clinicians to interactively observe the connections between concepts, for providing the references for the purpose of disease syndrome type summarization. On the other hand, the knowledge graph can generate complex network for reflecting the inferences between meta knowledge in the network. As shown in Fig. 3, the inferences are the edges representing the concept relations among concept nodes. All the edges have directions and weights, where the directions denote sequential relations, e.g., *X medicine treat Y disease*, *X disease has Y symptom*, etc. The weight values of the edges denotes the strengthness and weakness of the relations. The weights can be learned and adjusted for helping clinicians in observing and filtering the relations to obtain relation patterns.

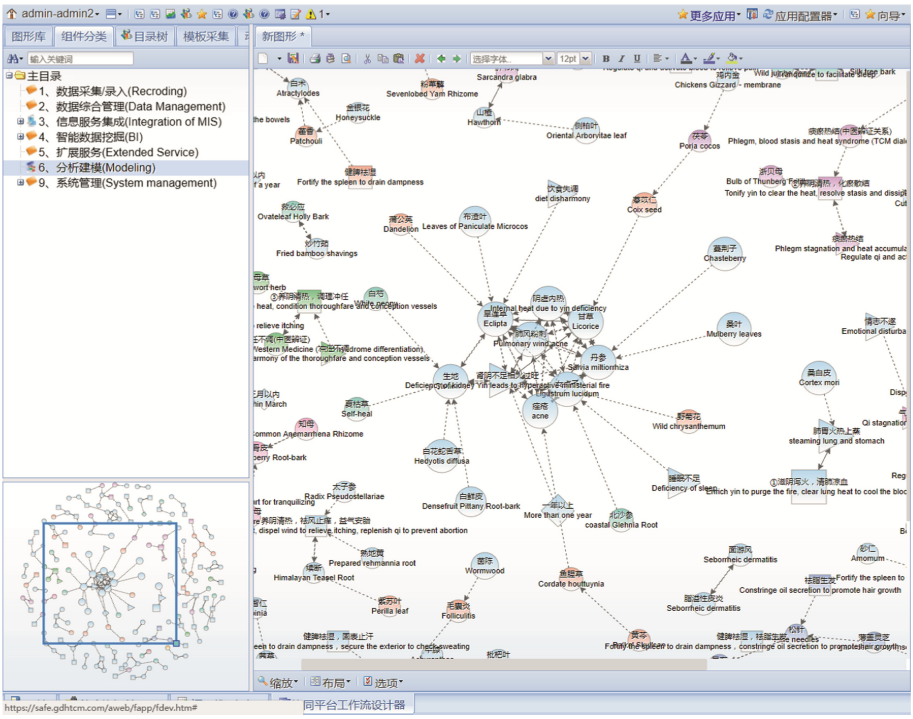

Fig. 3. The visualization of the TCM knowledge graph for clinicians to operate interactively

The constructed TCM knowledge graph can also be potentially utilized for decision making assistant. We develop a system named as "Intelligent profile analysis and recommendation based on TCM knowledge graph", as shown in Fig. 4. The decision making assistant in the system mainly has the following steps: (1) retrieve and filter the meta knowledge in the knowledge graph according to the symptoms, observational data, lab test, etc.; (2) observe the distributions of the attributes of diagnosis, medicine, prognosis in the knowledge graph and import the meta nodes into the vectorization model

as described in the framework for structuring the data; (3) generate a knowledge map network according to the vector representations; (4) observe the results of network clustering and analyze the references for initial patient diagnosis and treatment strategies; (5) obtain assisted decision making references of patient treatment strategies according to the semantic inference among the meta knowledge such as the symptoms and lab test values of the patients.

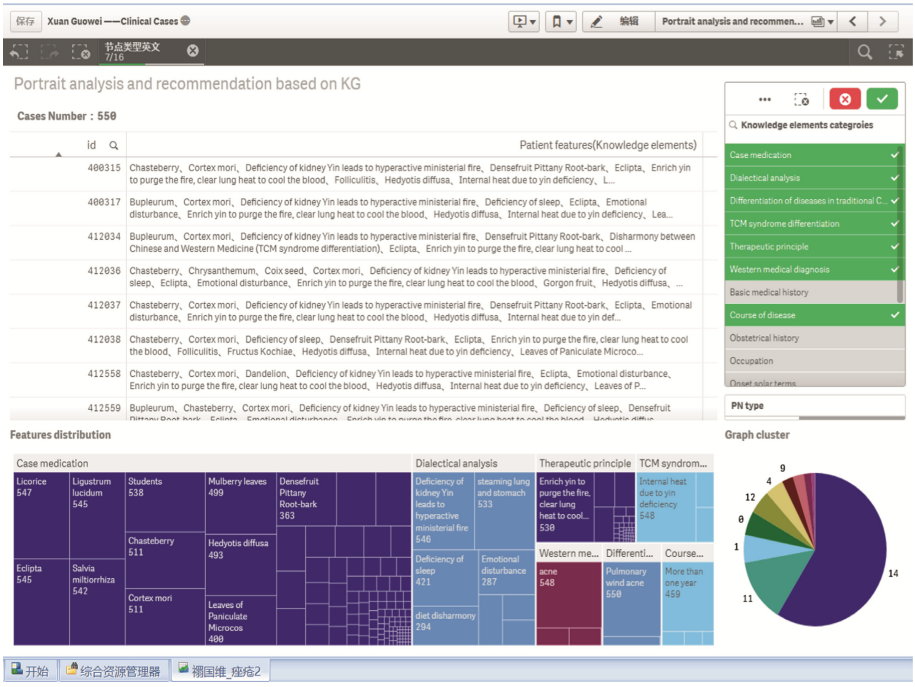

Fig. 4. The user interface of a developed system named as "Intelligent profile analysis and recommendation based on TCM knowledge graph" for decision making assistant

Until now, the system based on the TCM knowledge graph has been applied to the analysis of more than 1000 ancient Chinese medicine books, and the information extraction from the medical records for more than ten TCM departments in provincial hospitals. Particularly, the system has been used to serve for 5 national/provincial level famous TCM experts in the summarization of their clinical cases. In short, the system not only implements the TCM knowledge retrieval and network analysis but also provides the summarization and visualization of famous TCM experts through the knowledge discovery from their related EMR text data. We believe the system could further benefit the interactions among TCM clinicians and even the knowledge accumulation for public health knowledge spread.

6 Conclusions

Targeting at medical knowledge graph construction, this paper proposes a framework for automated Traditional Chinese Medicine knowledge graph construction from existing clinical texts. The framework consists of four major modules. Based on a standard dataset containing 886 patient cases, the evaluation results present that the usage of the knowledge graph can significantly improve the classification performances, demonstrating the effectiveness of the proposed framework in medical knowledge graph construction.

Acknowledgements. This work was supported by Frontier and Key Technology Innovation Special Grant of Guangdong Province (No. 2014B010118005), Public Interest Research and Capability Building Grant of Guangdong Province (No. 2014A020221039), and National Natural Science Foundation of China (No. 61772146 & 61403088).

References

1. Jameson, J.L., Longo, D.L.: Precision medicine-personalized, problematic, and promising. Obstetr. Gynecol. Surv. **70**(10), 612–614 (2015)
2. Raghupathi, W., Raghupathi, V.: Big data analytics in healthcare: promise and potential. Health Inf. Sci. Syst. **2**(1), 3 (2014)
3. IBM-The FOUR V's of Big Data. http://www-01.ibm.com/software/data/bigdata/. Accessed 2017
4. Sagiroglu, S., Sinanc, D.: Big data: a review. In: Proceedings of International Conference on Collaboration Technologies and Systems, pp. 42–47 (2013)
5. Belle, A., Thiagarajan, R., Soroushmehr, S., et al.: Big Data Analytics in Healthcare. BioMed Research International (2015)
6. Alickovic, E., Subasi, A.: Medical decision support system for diagnosis of heart arrhythmia using DWT and random forests classifier. J. Med. Syst. **40**(4), 1–12 (2016)
7. Constantinou, A.C., Fenton, N., Marsh, W., et al.: From complex questionnaire and interviewing data to intelligent Bayesian network models for medical decision support. Artif. Intell. Med. **67**, 75–93 (2016)
8. Woosley, R., Whyte, J., Mohamadi, A., et al.: Medical decision support systems and therapeutics: the role of autopilots. Clin. Pharmacol. Ther. **99**(2), 161–164 (2016)
9. Cambria, E., Olsher, D., Rajagopal, D.: SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In: Proceedings of Twenty-Eighth AAAI Conference on Artificial Intelligence (2014)
10. Mirzaa, G.M., Millen, K.J., Barkovich, A.J., et al.: The developmental brain disorders database (DBDB): a curated neurogenetics knowledge base with clinical and research applications. Am. J. Med. Genet. Part A **164**(6), 1503–1511 (2014)
11. Taglang, G.D., Jackson, B.: Use of "big data" in drug discovery and clinical trials. Gynecol. Oncol. **141**(1), 17–23 (2016)
12. Vicini, P., Fields, O., Lai, E., et al.: Precision medicine in the age of big data: the present and future role of large-scale unbiased sequencing in drug discovery and development. Clin. Pharmacol. Ther. **99**(2), 198–207 (2016)
13. Holzinger, A.: Trends in interactive knowledge discovery for personalized medicine: cognitive science meets machine learning. IEEE Intell. Inform. Bull. **15**(1), 6–14 (2014)

14. Kim, D., Joung, J.G., Sohn, K.A., et al.: Knowledge boosting: a graph-based integration approach with multi-omics data and genomic knowledge for cancer clinical outcome prediction. J. Am. Med. Inform. Assoc. **22**(1), 109–120 (2015)
15. Kamsu-Foguem, B., Tchuenté-Foguem, G., Foguem, C.: Using conceptual graphs for clinical guidelines representation and knowledge visualization. Inf. Syst. Front. **16**(4), 571–589 (2014)
16. Zhang, D., Xie, Y., Li, M., et al.: Construction of knowledge graph of traditional Chinese medicine based on the ontology. Technol. Intell. Eng. **3**(1), 8 (2017)
17. Yu, T., Li, J., Yu, Q., et al.: Knowledge graph for TCM health preservation: design, construction, and applications. Artif. Intell. Med. **77**, 48–52 (2017)
18. Shi, L., Li, S., Yang, X., et al.: Semantic Health Knowledge Graph: Semantic Integration of Heterogeneous Medical Knowledge and Services. BioMed Research International (2017)
19. Mikolov, T., Kombrink, S., Deoras, A., et al.: RNNLM-recurrent neural network language modeling toolkit. In: Proceedings of the 2011 ASRU Workshop, pp. 196–201 (2011)
20. Ou, A., Lin, X., Li, G., et al.: LEVIS: a hypertension dataset in traditional Chinese medicine. In: Proceedings of Bioinformatics and Biomedicine (BIBM), pp. 192–197 (2013)
21. State Administration of Traditional Chinese Medicine of People's Republic of China: Clinic terminology of traditional Chinese medical diagnosis and treatment–Syndromes. Standards Press of China, Beijing, GB/T 16751.2–1997 (1997)
22. Zhang, M.L., Zhou, Z.H.: ML-KNN: a lazy learning approach to multi-label learning. Pattern Recogn. **40**(7), 2038–2048 (2007)
23. Sorower, M.S.: A Literature Survey on Algorithms for Multi-label Learning. Oregon State University, Corvallis (2010)