Automated Encoding of Clinical Guidelines into Computer-interpretable Guidelines

Yuming Qiu
Chengdu Computer Application Institute of Chinese Academy of Sciences
Chengdu, China
University of Chinese Academy of Sciences
Beijing, China
Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences
Chongqing, China
qiuym@cigit.ac.cn

Peng Tang
Breast and Thyroid Surgery, Southwest Hospital
Chongqing, China
tp1232000@sina.com

Jingzhong Zhang
Chengdu Computer Application Institute of Chinese Academy of Sciences
Chengdu, China
zjz2271@163.com

Jun Jiang
Breast and Thyroid Surgery, Southwest Hospital
Chongqing, China
jcbd@medmail.com.cn

Ju Zhang
Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences
Chongqing, China
Beijing, China
zhangju@cigit.ac.cn

Nanzhi Wang
College of Computer & Information Science, Southwest University
Chongqing, China
wangnz1994@163.com

ABSTRACT

Computer-interpretable guidelines (CIGs) are critical knowledge source for clinical decision support systems (CDSS). However, most of current CIGs are encoded by medical experts and knowledge engineers based on the clinical practice guidelines (CPGs). It is complex, time-consuming and labor-intensive. In this paper, a network structural model had the potential of automated encoding is presented, and a corresponding framework is proposed to encode the guidelines to such a model automatically. The framework consists of three automated sequential pipelines: semi-structural guideline generation, network reduce and validation, CIGs construction. Furthermore. We choose the clinical practice guidelines issued by Notional Comprehensive Cancer Network (NCCN) to carry out the proposed framework, and the results show that automated encoding three breast cancer guidelines save a tremendous amount of time from 25 workdays manual encoding to merely 15 minutes automated encoding plus 5 hours manual validation and correction. This indicates that automated encoding is practicable and the proposed model and framework is feasible, accurate and effective.

CCS Concepts

- **Computing/technology policy→Medical information policy**
- **Information retrieval→Data encoding and canonicalization**

Keywords

Automated Encoding; Computer-interpretable Guidelines; CIGs; Clinical Guidelines

DOI: http://dx.doi.org/10.1145/12345.67890

1. INTRODUCTION

Clinical guidelines constitute an important modality that can reduce the delivery of inappropriate care and support the introduction of new knowledge into clinical practice, and can serve as an up-to-date knowledge source for clinical decision support tools[10; 14; 30]. Unfortunately, it is a tough work to encode them into computer-interpretable guidelines. Many models[3; 11; 17; 18; 28; 32] have been proposed to represent the guideline knowledge, and many efforts[1; 2; 6; 8; 12; 16; 20; 21; 24; 29; 31] have been made to complete the encoding and execution. Nevertheless, most of these endeavors had been made with serious participation with the medical experts and knowledge engineers. As the complexity and amount of medical information keeps increasing, it is more difficult to create new CIGs and maintain the exist CIGs, further harder to give the effective support to clinicians. Therefore, automated encoding methods are becoming increasingly important.

As it is known, the guideline knowledge represents a distillation of research findings combined with the judgement and experience of clinical experts and can be applied into the clinical practice. There are, actually, strong inherent structural and exercisable knowledge in the guidelines, and it can be transform to some format that computer can interpreted. Many models have been proposed to represent the guideline knowledge. The Guideline Interchange Format (GLIF)[3; 4; 17] is a model for representation of sharable computer-interpretable guidelines. GLIF3 enables encoding of a guideline at three levels: a conceptual flowchart, a computable specification that can be verified for logical consistency and completeness, and an implementable specification that is intended to be incorporated into particular institutional information systems. The representation has been tested on a wide variety of guidelines that are typical of the range of guidelines in clinical use.[3] GLIF is a great model to represent the guidelines, but it is meanwhile complex, laborious and tedious to encode a guideline to the GLIF model. The encoding has been experience that a team of two people, an informatician train in GLIF and a clinician, can work collaboratively to encode a guideline to GLIF[3]. Another representative model is the Guideline Elements Model (GEM) proposed by Shiffman, R. N., et al.[26]. The GEM hierarchy includes more than 100 elements, and the GEM model more comprehensive, expressively adequate to represent the heterogeneous information contained in guidelines. Some cases[9] reported the GEM was more specific and covers more clinical situations. Unavoidably, the GEM is complex and difficult to encode a guideline as well. Based on the comprehensive surveys and the practical proposal from the clinician, we develop a simpler but effective model (See the model section) to represent the guideline. The most advantage of the model is that it has the potential of automated encoding, and unnecessariness of complex engine to decode the model. Subsequently, we propose a framework to encode automatically. Lastly, we choose the NCCN guidelines to implement the framework, and explain how to encode the NCCN guidelines issued as PDF format into the model automatically. The results show that it is feasible, effective to encode the semi-structured guidelines to CIGs.

In the next section, we review related work on the guideline knowledge representation and guideline encoding. The following sections illustrate the model, explain the proposed framework, and further present our experiments and results.

2. RELATED WORKS

Guideline representation model and the encoding method is long spread topic. In 1995, Purves, I.[23] start to concern the computerized guidelines in primary health care. Increasing research focused on this field in late-1990s and early 2000s[19]. The Computer-interpretable Guidelines(CIGs) representation, integration of CIGs with EHRs, validation and verification, CIGs sharing and CIGs maintenance become hot topic. Many model were proposed in this duration, like GLIF and GEM as mentioned before. Besides the GLIF and GEM, others models represented the guideline through different method and view. Peleg, M., et al. [22] Compared various modes such as GLIF, Guide, Asbru, GASTON, GLARE, HELEN, PROforma, and SAGE. It is concluded that the consensus on several components including plan organization, expression language, conceptual medical record model, medical concept model, and data abstractions, explicit differences existed in the underlying decision models, goal representation, use of scenarios, and structured medical actions.[22]

Tree-like structure was used to introduce to represent guideline because of the tree-like structure is a nature concept hierarchy which can employ the ontology methodology and technologies. Decision tree (DT) [7] were used as the conceptual organization of the guideline knowledge, Each step of the DT was considered as an elementary knowledge, which is to be explicated by elementary messages of prose text. Unlike the decision tree, we treat the guideline structure as an executable network that is made up of Action-Unit s. The network has the superiority that it can encode the loop structure. In our model design, the encoded guideline can be decode to a tree-like structure, but in the encoding phrase, it is allowed to encode the guideline into a network structure.

There are some efforts[5; 6; 9; 25] on the guideline encoding. Georg, G., et al.[9] compared ASTI-based and GEM-based methods with the , and conclude that the GEM-based approach is more specific and covers more clinical situations and led to promising results. Choi, J., et al.[6] encoded a screening guideline using GLIF to transform it into. After identifying all guideline steps from the text format guideline, a flowchart was drawn using the encoded steps in GLIF. The flowchart was validated by an experienced psychiatric nurse clinician. Kim, H. Y., et al.[13] report a case study of pressure-ulcer management with encoding and verification of a computer-interpretable guideline using the SAGE guideline model. This study was conducted using the following procedures: selecting CPGs, extracting rules from the selected CPGs, developing a CIG using the SAGE guideline model, and verifying the obtained CIG with test cases using an execution engine. The CIG for pressure-ulcer management was developed based on 38 rules and three algorithms at the semiformal representation level using MS Excel and MS Visio. The CIG was encoded by two Activity Graphs consisting of 115 instances representing algorithms and rules as knowledge elements in the SAGE guideline model. Two errors were found and corrected. Results of the study demonstrated that a CIG representing knowledge on pressure-ulcer management can be effectively developed using commonly available programs and the SAGE guideline model, and that the obtained CIG can be verified with a locally developed execution engine. The CIG developed in the study could contribute to health information management once it is implemented successfully in a clinical decision support system.

As the complexity and amount of medical information keeps increasing, it is more difficult to create new CIGs and maintain the exist CIGs. Subsequently, Automated encoding is considerable significant.

3. MODEL

Our goal is attempt to develop an automated encoding approach, so a simpler model is a better choose. Unlike the previous works struggle to dig the structural pattern so as to encode the guidelines to it, we treated the guideline from the clinical practice perspective as a simple networks to represent the actions and their transferring relationship. We abstract the clinical practical healthcare process as the logical transmission of different activities based on the decision with clinical data (See Figure1). Moreover, as our investigation of many clinical guidelines, they can be represent with this abstraction as well. The GLIF [3] a conceptual flowchart, a computable specification that can be verified for logical consistency and completeness, and an implementable specification that is intended to be incorporated into particular institutional information systems. The conceptual flowchart in the GLIF is a representation of the clinical practice process to some extends, but it is more comprehensive and complex.

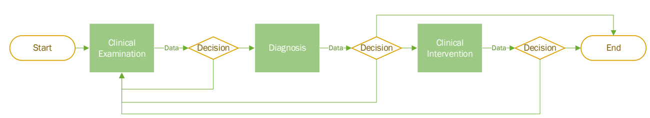

Figure 1 Abstraction of the practical healthcare process.

Therefore, we derived a model from such abstraction. In our designing view, the encoded guideline can be decode to a tree-like structure, but in the encoding phrase, it is allowed to encode the guideline into a network structure. In the model M, so, Action-Unit s are the abstract concept to capsule the clinical actions, clinical data and decision condition. All the Action-Unit s consist a set A can be divide into three subsets: clinical examination A_e, diagnosis A_d, clinical intervention A_i. The logical transmission between units is based on a decision action d, every decision action can be express by a logical expression which driven by the clinical data of one patient. We define our model as follow:

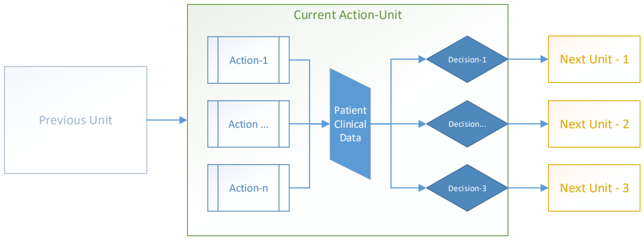

Figure 2 proposed model

$$M = < A, R >$$

$$A = A_e \cup A_d \cup A_i$$

$$a \in A$$

$$a = < \text{actions, patient data, decision condition} >$$

The knowledge contained in the guideline decide the relation between two Action-Unit s but it is constrained by the abstraction paradigm (see Figure 1). If there are some conflicts occurred, it must be resolved by the experts through encoding the conflicting content.

According the defined network model, the guideline can be encoded into a network structure directly. One Action-Unit can be taken as a node in the network, and each node contains three type of components: clinical intervention as actions, patient clinical data and decisional condition expressions. In the runtime phrase, the action output the clinical data, the data drive the condition expression to decide whether trigger the corresponding Action-Unit .

4. ENCODING FRAMEWORK

The upmost advantage of proposed model is the potential of encoding automatically. There are many methods and technology can be employed, but there are some common logical procedures in them. Therefore, we summarized a framework to try extracting the common logical procedures based on the thoughts of Shahar,Y., et al.[25]. They thought that the encoding process is performed gradually using the following representation formats:

1) Semi-structured Text - snippets of text assigned to top-level target-ontology knowledge-roles, such as the eligibility criteria for applying the guideline, or the guideline's objective.
2) Semi-formal representation - further specification of the structured text, adding more explicit procedural control structures, performed jointly by the knowledge engineer and expert physician.
3) Formal representation - final specification performed by the knowledge engineer, resulting with the guideline converted to a machine-comprehensible format, executable by an appropriate runtime execution module specific to the chosen target guideline ontology.

According their opinion, as shown in the Figure3, the framework take free text guidelines as input and the formal CIGs as output. It consists of three steps. Firstly, medical expert makeup the free text guideline according to semantics, and the cooperation between medical expert and knowledge engineer needs to build semi-formal knowledge. Finally, knowledge engineers formalize the semi-formal knowledge to executable code, that is, formal knowledge. This is a complex and cumbersome process involving experts in two fields and their collaboration, and the result depends on the professional level of the experts.

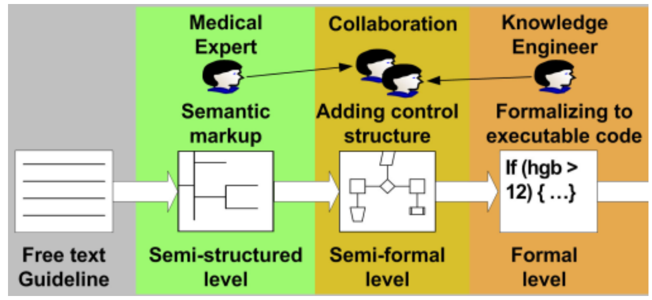

Figure 3 The encoding framework proposed by Shahar,Y., et al.

4.1 Overview of the framework

The framework proposed by Shahar,Y., et al. is an appropriate abstraction about the encoding procedures. However, it was designed for the entire manual participation encoding. Therefore, we design a new framework to reach the automated encoding.

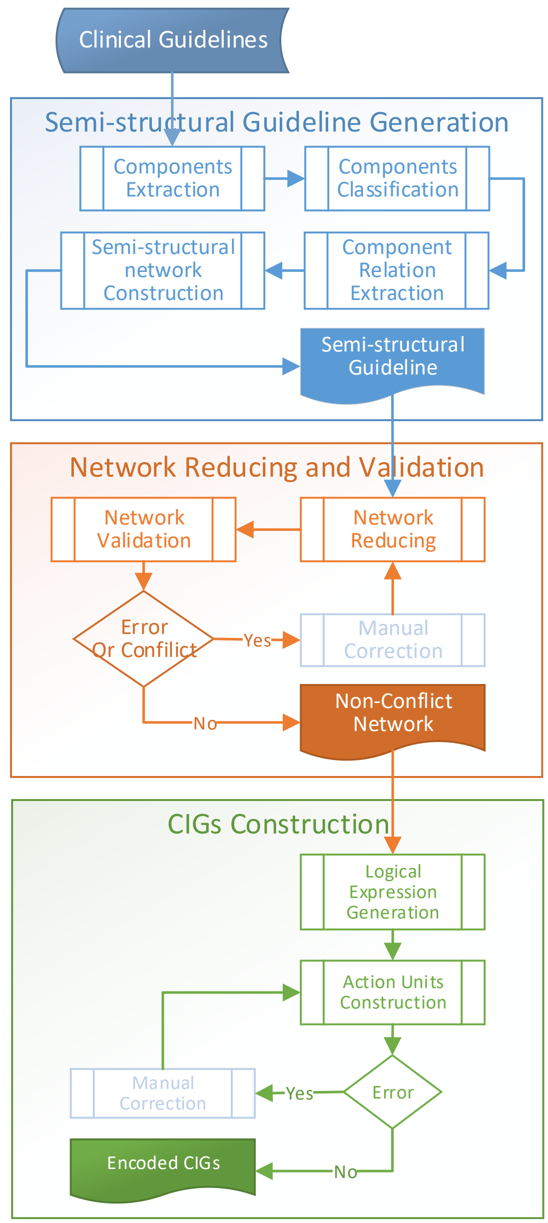

Figure 4 Proposed Automated Framework

Therefore, based on the framework proposed by Shahar,Y., et al, we develop a new framework can automated encoding with few manual participation as shown in Figure4. In our proposed model, the key problem is to construct the Action-Unit node. As showing in Figure2, each Action-Unit node contains three types of components: clinical intervention as actions, patient clinical data and decisional condition expressions. Therefore, how to extract such three types of components is the main object. Besides the patient clinical data is collect from HIS, the actions and condition expression is most challenge.

The framework mainly consists three automated sequential pipelines: semi-structural guideline generation, network reducing and validation, CIGs network construction.

4.2 Pipeline 1: Semi-structural Guideline Generation

In the first pipeline, the target is to build a semi-structural guideline from raw clinical guidelines, so we design four steps: components extraction, components classification, component relation extraction, semi-structural network construction. The term 'component' refers to the basic knowledge points in the raw guidelines, such as a set of clinical examinations, a clinical intervention, and a criteria condition. Actually, the main idea of the framework is scattering the knowledge firstly and then integrating them a new guideline with another format that the computer can processing. The extracting methods may be different while handling various guidelines. For example, methods and technologies of entity extraction, relation extraction, and event extraction in the natural language process (NLP) domain can be used in the free-text guidelines, some image processing algorithms and tricks can be used in some semi-structural guidelines, an implementation of the latter will be demonstrated in the next experiments section.

4.3 Pipeline 2: Network Reducing and Validation

The generated network in the pipeline 1 is semi-structural. It may contains many duplicated knowledge points, error relations and conflict structures. Therefore, the object of pipeline 2 is to get a Non-conflict network. A circuit with few human involvements is designed to meet the challenge. Firstly, network reducing removes the duplicate the knowledge points and retune the network structure. Secondly, network validation check the network whether existing conflict structure and try to repair automatically, if there are some unresolved conflicts, manual correction need to bring in. After the manual correction, enter into a loop of network reducing, network validation, manual correction until a non-conflict network is generated. In this pipeline, some semantic similarity algorithms, formal verification can be employed.

4.4 Pipeline 3: CIGs Network Construction

The non-conflict network obtained in the pipeline 2 does not have the executable ability. In our model design (see Figure2), the Action-Units have the decision component inside. Therefore, the next step is to generate the logical expression. As described in section 4.2, there are many criteria conditions in the non-conflict network, we can take them out of the network and retune the network again. The network construction need the human involvements as well until obtaining the final encoded CIGs.

5. EXPERIMENTS

In order to verify competence of our proposed model and the feasibility of the automated framework, we design some experiments. In the first step, we select NCCN to carry out experiments after consulting serval kinds of guidelines published by CCN, AHA, NICE, WHO, ACOG, ESMO. The main reason to choose the NCCN guidelines base on the following considerations: 1)the guidelines published by NCCN is very authoritative; 2)there are a mount of semi-structure knowledge in their guidelines which make the first pipeline of the framework become easier and relatively simpler, as is well known that extraction the structural from the free text is more difficult; 3)it can be easily obtained on their official website.

In the NCCN Clinical Practice Guidelines (which are available at https://www.nccn.org/professionals/physician_gls/default.aspx), a typical page need to be processed is shown in Figure5

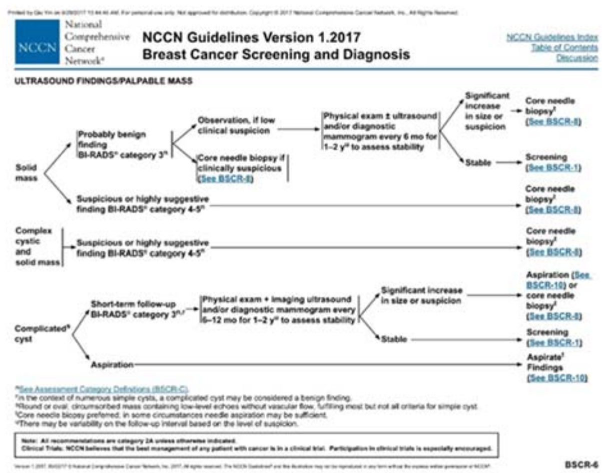

Figure 5 A sample page of the NCCN guideline

As shown in the Figure.5, the logical structure diagram is just a part of the whole page. In the logical structure diagrams, nodes point to their child nodes from left to right. Based on the proposed model, we divide node types into 3 classes: action, condition and link. It worth noting that the right-most nodes on this page may point to other pages. Therefore, identifying the relationships across pages is a challenge to handle the NCCN guidelines.

5.1 Semi-structural Guideline Generation

Although the NCCN guideline contain the semi-structural knowledge, it is very tough to convert them into a computer-interpretable format, because the NCCN guidelines are issued as PDF format. There are a few methods to complete the semi-structural generation, such as text extraction, HTML conversion, image processing. We choose the image processing method because this method reserve more positional information. In order to meet the challenge, an integrated approach was proposed, and it combines computer vision, morphology, machine learning, and neural networks.

5.1.1 Preprocessing

Firstly, we converted the pdf-format guidelines into JPG format images. So, the key problem changes to how to locate the logical structure diagram in the discrete images. A simple but effective morphological method was employed to solve this problem. Erosion operation and Dilation operation (see Figure6) are the main taken function.

Erosion: this function erodes the source image using the specified structuring element that determines the shape of a pixel neighborhood over which the minimum is taken:

$$\mathrm{dst}(x, y) = \min_{(x',y'):element(x',y')\neq 0} src(x+x', y+y')$$

Dilation: this function dilates the source image using the specified structuring element that determines the shape of a pixel neighborhood over which the maximum is taken:

$$\mathrm{dst}(\mathrm{x},\mathrm{y}) = \max_{(x',y'):element(x',y')\neq 0} src(x+x', y+y')$$

Figure 6 The results after morphological operation. Left:Erosion Right:Dilation

After filling the hollow by using flood-fill algorithm, the region of the logical structure diagram on the page can be located.

5.1.2 Blocks detection

The term ‘Block’ specify the text snippet with complete semantics and positional information. A block is consistent with the action or conditional statement in the guidelines, and is same as the ‘component’ term in the automated framework in Figure4. The detection have been done with the following four steps:

1) Detecting all the locations and other relevant information of arrows and lines on the logical structure diagram by analyzing the pixel density.

2) Discarding the regions of arrows and lines to simplify the next processing.

3) Extracting the block position information by comprehensive analyzing the position and size of each candidate regions.

4) Recognizing the text contents of the nodes by OCR technology[27] to generate the blocks set.

5.1.3 Blocks classification

Detected blocks have rich knowledge of the guideline, but they are chaotic. In the proposed model (see Figure1 and Figure2), we finally want to get all action-units containing clinical actions classified three categories and logical decisions.

So, we firstly mapped each word to a high-dimensional vector by utilizing word2vec tool[15] with training more than 100 official guidelines. Word vector is a distributed representation of the word and can capture more contextual information [15].

Then, we classified the blocks into four categories. Among of them, three categories are same as the proposed model including clinical examination, diagnosis, clinical intervention, and the other category is condition. The blocks with condition category will be used to generate the logical expression in the next pipelines.

Concretely, we manual tagged more than 600 nodes as corpus, then used the corpus to train a classification model by employing the deep forward neural networks, and lastly used the classification model to predict new blocks.

5.1.4 Relation construction

Relations among blocks are most important knowledge in the guidelines. In NCCN guidelines, the relations mainly present as many kinds of arrow line. So, we use the following strategy to construct the relation of blocks.

1) Recognizing all arrows and lines from the morphological-operated binary image.

2) Finding the correspondence of arrow and block, then build the relation of them logically. As our priori observation, there is a strong rule that: the arrow tails used to point the nearest block. According to this rule, we almost find all correct relations among blocks.

3) Validating the built relations through finding the isolated blocks and correct the relations through some manual rules. Actually, there are a small number of isolated blocks and incorrect relations.

5.1.5 Semi-structural network generation

From above extracted blocks and relations, we constructed a semi-structural network to represent knowledge in the guidelines.

5.2 Network Reducing and Validation

The semi-structural network contains many duplicate similar nodes. Network reducing is to deduplicate them and rebuild the broken relation. Firstly, we used the cosine law to calculate the similarity of each pair in all of the nodes, and delete one of them when they are in a same branch. Secondly, we define a metric to measure one node containing percentage in another node. Thirdly, we validate the network through calculate the network connectivity, if the network is valid, manual correction is introduced to correct the error or conflicts. Loop the three step until get a well-structure network.

5.3 CIGs Construction

In the CIGs construction pipeline, the utmost difficult challenge is to generate the logical expression based on the natural language free-text. We employ the Finite State Transducer method to analyze the lexicon of natural described condition, and convert them into a computer-executable logical expression. Then, we reduce and integrate the nodes into a higher level Action-Unit s, and construction the Action-Unit s network.

5.4 Experiment results

We have converted 116 NCCN guidelines in serval hours. After automated recognition and construction, we get 56 condition-blocks, 226 action-blocks, and 35 overlapped duplicate blocks on average per one guideline. Specially, we contrast the time-consumption between manual manner and automated manner in encoding three breast cancer related guidelines. We develop a visualized system to manually encode, validate and correct the encoded CIGs. We record the time consumption of encoding. the results show that automated encoding three breast cancer guidelines save a tremendous amount of time from 25 workdays manual encoding to merely 15 minutes automated encoding plus 5 hours manual validation and correction. Additionally, the network reduce strategy is very effective, 1358 relations can be reduce to 505 relations rapidly.

6. CONCLUSION

The recent advances of computer version, natural language processing and other technologies make it possible to encode the CPGs into CIGs. In this paper, we present a guideline-representational model, proposed a framework to encode the CPGs into CIGs automatically. At last, based on the proposed model and framework, we perform an experiment and successfully encode the NCCN guideline into CIGs automatically. The result demonstrated that automated encoding the guidelines to the proposed CIG model is feasible, accurate and effective. Honestly, there are still some problems that we have not solved such as the missing extracted terms, incomplete extracting conditions in natural language.

In the future work, we will integrate more methods to improve the efficiency and accuracy of the CPGs encoding. Furthermore, we will try to extract the semi-structural network from the free-text guideline without visual diagram.

7. ACKNOWLEDGMENT

This work was supported by Chongqing Science & Technology Commission under grant cstc2015shmszx0624 and grant cstc2015shms-ztzx10005.

8. REFERENCES

[1] BACKERE, F.D., MOENS, H., STEURBAUT, K., COLPAERT, K., DECRUYENAERE, J., and TURCK, F.D., 2012. Towards automated generation and execution of clinical guidelines: Engine design and implementation through the ICU Modified Schofield use case. *Computers in Biology & Medicine 42*, 8, 793-805.

[2] BARAHONA, P., AZEVEDO, F., VELOSO, M., ESTÊVÃO, N., and GALLEGO, R., 2001. Computerising a guideline for the management of diabetes. *International Journal of Medical Informatics 64*, 2-3, 275-284.

[3] BOXWALA, A.A., PELEG, M., TU, S., OGUNYEMI, O., ZENG, Q.T., WANG, D., PATEL, V.L., GREENES, R.A., and SHORTLIFFE, E.H., 2004. GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines. *Journal of Biomedical Informatics 37*, 3, 147.

[4] BOXWALA, A.A., TU, S., PELEG, M., ZENG, Q., OGUNYEMI, O., GREENES, R.A., SHORTLIFFE, E.H., and PATEL, V.L., 2001. Toward a representation format for sharable clinical guidelines. *Computers & Biomedical Research 34*, 3, 157-169.

[5] CHOI, J., CURRIE, L.M., WANG, D., and BAKKEN, S., 2007. Encoding a clinical practice guideline using guideline interchange format: a case study of a depression screening and management guideline. *International Journal of Medical Informatics 76*, 76 Suppl 2, S302-S307.

[6] CHOI, J., SAPP, J., and BAKKEN, S., 2006. Encoding a depression screening guideline using GLIF. *Studies in Health Technology & Informatics 122*, 905.

[7] COLOMBET, I., AGUIRREJUNCO, A.R., ZUNINO, S., JAULENT, M.C., LENEVEUT, L., and CHATELLIER, G., 2005. Electronic implementation of guidelines in the EsPeR system: a knowledge specification method. *International Journal of Medical Informatics 74*, 7–8, 597-604.

[8] DOWNS, S.M., BIONDICH, P.G., ANAND, V., ZORE, M., and CARROLL, A.E., 2006. Using Arden Syntax and adaptive turnaround documents to evaluate clinical guidelines. *AMIA Annu Symp Proc*, 214-218.

[9] GEORG, G., SÉROUSSI, B., and BOUAUD, J., 2003. Does GEM-encoding clinical practice guidelines improve the quality of knowledge bases? A study with the rule-based formalism. *AMIA Annu Symp Proc 2003*, 254-258.

[10] GUIDELINES, I.O.M.C.O.S.F.D.T.C.P., 2011. Clinical Practice Guidelines We Can Trust. *Port J Nephrol Hypert*.

[11] ISERN, D., SANCHEZ, D., and MORENO, A., 2012. Ontology-driven execution of clinical guidelines. *Comput Methods Programs Biomed 107*, 2 (Aug), 122-139. DOI= http://dx.doi.org/10.1016/j.cmpb.2011.06.006.

[12] KAISER, K. and MIKSCH, S., 2009. Versioning computer-interpretable guidelines: semi-automatic modeling of 'Living Guidelines' using an information extraction method. *Artificial Intelligence in Medicine 46*, 1, 55-66.

[13] KIM, H.Y., PARK, H.A., CHO, I., KIM, J., and LEE, J., 2015. Encoding and verification of a computer-interpretable guideline: a case study of pressure-ulcer management. *Him j 44*, 1, 39-48.

[14] MERRITT, T.A., PALMER, D., BERGMAN, D.A., and SHIONO, P.H., 1997. Clinical practice guidelines in pediatric and newborn medicine: implications for their use in practice. *Pediatrics 99*, 1, 100.

[15] MIKOLOV, T., SUTSKEVER, I., CHEN, K., CORRADO, G., and DEAN, J., 2013. Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems 26*, 3111-3119.

[16] MOSKOVITCH, R. and COHEN, S., 2006. Multiple hierarchical classification of free-text clinical guidelines. *Artificial Intelligence in Medicine 37*, 3, 177-190.

[17] OHNOMACHADO, L., GENNARI, J.H., MURPHY, S.N., JAIN, N.L., TU, S.W., OLIVER, D.E., PATTISONGORDON, E., GREENES, R.A., SHORTLIFFE, E.H., and BARNETT, G.O., 1998. The GuideLine Interchange Format: A Model for Representing Guidelines. *Journal of the American Medical Informatics Association 5*, 4, 357.

[18] PATEL, V.L., ALLEN, V.G., AROCHA, J.F., and SHORTLIFFE, E.H., 1998. Representing clinical guidelines in GLIF: individual and collaborative expertise. *Journal of the American Medical Informatics Association Jamia 5*, 5, 467.

[19] PELEG, M., 2013. Computer-interpretable clinical guidelines: a methodological review. *Journal of Biomedical Informatics 46*, 4, 744–763.

[20] PELEG, M., BOXWALA, A.A., BERNSTAM, E., TU, S., GREENES, R.A., and SHORTLIFFE, E.H., 2001. Sharable Representation of Clinical Guidelines in GLIF: Relationship to the Arden Syntax. *Journal of Biomedical Informatics 34*, 3, 170.

[21] PELEG, M., KEREN, S., and DENEKAMP, Y., 2008. Mapping computerized clinical guidelines to electronic medical records: knowledge-data ontological mapper (KDOM). *Journal of Biomedical Informatics 41*, 1, 180-201.

[22] PELEG, M., TU, S., BURY, J., CICCARESE, P., FOX, J., GREENES, R.A., HALL, R., JOHNSON, P.D., JONES, N., and KUMAR, A., 2002. Comparing computer-interpretable guideline models: a case-study approach. *Journal of the American Medical Informatics Association 10*, 1, 52.

[23] PURVES, I., 1995. Computerised guidelines in primary health care: reflections and implications. *Stud Health Technol Inform 16*, 57-74.

[24] SHAHAR, Y., YOUNG, O., SHALOM, E., GALPERIN, M., MAYAFFIT, A., MOSKOVITCH, R., and HESSING, A., 2004. A framework for a distributed, hybrid, multiple-ontology clinical-guideline library, and automated guideline-support tools. *J Biomed Inform 37*, 5 (Oct), 325-344. DOI= http://dx.doi.org/10.1016/j.jbi.2004.07.001.

[25] SHAHAR, Y., YOUNG, O., SHALOM, E., MAYAFFIT, A., MOSKOVITCH, R., HESSING, A., and GALPERIN, M., 2003. DEGEL: A Hybrid, Multiple-Ontology Framework for Specification and Retrieval of Clinical Guidelines. In *Artificial Intelligence in Medicine, Conference on Artificial Intelligence in Medicine in Europe, Aime 2003, Protaras, Cyprus, October 18-22, 2003, Proceedings*, 122-131.

[26] SHIFFMAN, R.N., KARRAS, B.T., AGRAWAL, A., CHEN, R., MARENCO, L., and NATH, S., 2000. GEM: A Proposal for a More Comprehensive Guideline Document Model Using XML. *Journal of the American Medical Informatics Association Jamia 7*, 5, 488.

[27] SMITH, R., 2007. An Overview of the Tesseract OCR Engine. In *International Conference on Document Analysis and Recognition*, 629-633.

[28] SORDO, M., OGUNYEMI, O., BOXWALA, A.A., and GREENES, R.A., 2003. GELLO: An Object-Oriented Query and Expression Language for Clinical Decision Support: AMIA 2003 Open Source Expo. *Amia Annual Symposium Proceedings 2003*, 1012.

[29] TERENZIANI, P., MOLINO, G., and TORCHIO, M., 2001. A modular approach for representing and executing clinical guidelines. *Artificial Intelligence in Medicine 23*, 3, 249-276.

[30] TRAN, N., MICHEL, G., KRAUTHAMMER, M., SHIFFMAN, and N., R., 2008. Embedding the guideline elements model in web ontology language. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium 2009*, 2009, 645.

[31] TU, S.W., MUSEN, M.A., and M., D.P.D., 2000. Representation Formalisms and Computational Methods for Modeling Guideline-Based Patient Care *83*.

[32] WANG, D., PELEG, M., TU, S.W., SHORTLIFFE, E.H., and GREENES, R.A., 2001. Representation of clinical practice guidelines for computer-based implementations. *Studies in Health Technology & Informatics 84*, 1, 285-289.

Authors' background

Your Name	Title*	Research Field	Personal website
Qiu Yumin	PhD candidate, assistant professor	Natural Language Processing, Biomedical Image Processing	
Tang Peng	associate professor	Breast Surgery, Clinical Medicine	
Zhang Jingzhong	full professor	Automated Reasoning, Distance Geometry, Dynamical System	
Jiang Jun	full professor	Breast Surgery, Clinical Medicine	
Zhang Ju	full professor	Intelligent Systems, Artificial Intelligence	
Wang Nanzhi	master student	Knowledge Engineer, Machine Learning	

***This form helps us to understand your paper better, the form itself will not be published.**

***Title can be chosen from: master student, Phd candidate, assistant professor, lecture, senior lecture, associate professor, full professor**