

Causal Knowledge Modeling for Traditional Chinese Medicine using OWL 2

Peiqin Gu

College of Computer Science, Zhejiang University, P.R.China
gupeiqin@zju.edu.cn

Abstract. Unlike Western Medicine, those in Traditional Chinese Medicine (TCM) are based on inherent rules or patterns, which can be considered as causal links. Existing approaches tend to apply computational methods on semantic ontology to do knowledge mining, but it cannot perfectly make use of internal principles in TCM. When it comes to knowledge representation, we can transform this inherent knowledge into causal graphs. In this paper, we present an approach to build a TCM knowledge model with the capability of rule reasoning using OWL 2. In particular, we focused on the causal relations among syndrome and symptoms, changes between syndromes. We evaluated our approach by giving two typical use cases and implemented them using Jena, a Java framework supporting RDF, OWL, and including a rule-based inference engine. The evaluation results suggested that our approach clearly displayed the causal relations in TCM and shows a great potential in TCM knowledge mining.

Keywords: Causal knowledge modeling, TCM, Rule reasoning, OWL 2

1 Introduction

The difference between Western Medicine (WM) and TCM is that WM focuses on the anatomy of human body, but TCM is based on an entirely different system of inherent rules. Because of the complicated philosophy, TCM has not had a proper understanding using current computer technologies.

The primary goal of Semantic Web [1] is to use URIs as a universal space to name anything, expanding from using URIs for web pages to URIs for “real objects and imaginary concepts”, as phrased by Berners-Lee. Among those efforts done by W3C group, Resource Description Framework (RDF) [2] provides data model specifications and XML-based serialization syntax, Web Ontology Language (OWL) [3] enables the definition of domain ontologies and sharing of domain vocabularies. The OWL 2 Web Ontology Language, informally OWL 2, provides a possible solution for rule reasoning using property chains.

For what we concern, we wish to apply the Semantic Technologies to represent TCM knowledge, which is mainly focused on rule reasoning derived from TCM philosophy. For example, the Five Phase Theory [4] of TCM divides all the things in the whole world to five types, which are Wood, Fire, Earth, Metal, and Water. The system of five phases was used for describing interactions and relationships between phenomena.

A lot of work has also been done to study the causal relationship in TCM. A stepwise causal adjacent relationship discovery algorithm [5] has been developed to study correlation between composition and bioactivity of herbal medicine and identify active components from the complex mixture of TCM. Chinese Medical Diagnostic System (CMDS) [6] contains an integrated medical ontology and the prototype of it can diagnose about 50 types of diseases by using over 500 rules and 600 images for various diseases. Wang et al. [7] developed a self-learning expert system for diagnosis in TCM using a hybrid Bayesian network learning algorithm, Naïve-Bayes classifiers with a novel score-based strategy for feature selection and a method for mining constrained association rules. However, these researches focused on applying mathematical methods on TCM knowledge mining and learning, so domain ontology only acts as knowledge base without self-learning capability, although building inherent knowledge links can be applicable using current ontology language.

TCM is rather a static theory model than an ever-evolving statistical one, so an expressive causal knowledge model with built-in rules can reveal the nature of TCM better. In this paper, we present an approach to build a TCM knowledge model with the capability of rule reasoning based on property chains using OWL 2.

2 Causal TCM Knowledge Modeling

We implement the causal TCM knowledge model with five functional layers. Ontology layer gives the basic terminology and assertions represented using OWL 2. Association rules are gathered in rules layer. Rule engine layer is recognized as an engine to deal with rules. Method layer do knowledge mining based on the defined ontology and rules. The utility of causal TCM knowledge model is ought to be the natural representation of TCM knowledge; it is also a reference model for TCM knowledge reasoning and knowledge mining, more importantly.

Our TCM knowledge model is designed based on the basic principle of the Five Phase Theory and the aim is to enable causal reasoning of TCM. According to the theory, all the things in the universe can be mapped to one of the five elements known as Wood, Fire, Earth, Metal, and Water. We define seven top ontology classes and corresponding sub classes to enable later causal reasoning.

In the top class design, “Five Phases” refers to the basic five elements in Five Phase Theory. “Environment” defines all kinds of natural elements that we use to diagnose diseases. “Body Elements” include all the physiological components. “Physiology” describes the related descriptions which describe human body conditions. “Pathology” describes pathologic states of human body. “Treatment” includes Chinese medicine, recipe, rules of treatment and therapy. “Symptoms” contains abnormal human body states.

In OWL 2, applications need to model interactions that are referred as one property “propagating” or being “transitive across” another. For now, we define about 30 object properties, which will help us define the property chains in the reasoning stage. For instance, we have :ObjectPropertyAssertion(:creates :Wood :Fire) to state Wood generates Fire in the theory. These internal properties among them give a basic causal foundation of TCM knowledge model.

As stated above, ontology classes establish a terminology structure for TCM domain knowledge; refined properties connect the terminology nodes tightly to form a knowledge model.

3 Causal Reasoning

The clinical diagnosis in TCM is mainly based on the internal rules. In this paper, we transform the diagnosis process into a layered causal graph.

We define the layered Causal Graph based on the fact that TCM diagnosis mainly focuses on syndrome. In a causal graph, there are nodes as terminology base, and causal links. Thus, formally a Causal Graph $G=\{V, E\}$, is defined as:

- $V \subseteq R$, where $R=\{Symptom, Syndrome, Treatment Rule, Therapy, Prescription, Herbal Medicine\}$.
- $E \subseteq U$, where $U=\{\rightarrow, \rightsquigarrow, \rightsquigarrow\}$.
 - $X \rightarrow Y$ represents that X has a *Direct Causation* relationship with Y meaning that X is a direct cause of Y independently.
 - $X \rightsquigarrow Y$ represents that X has a *Logic Causation* relationship with Y as combined causation with logical operators, which exists from symptoms to syndrome.
 - $X \rightsquigarrow Y$ represents that X has a *Weighted Causation* relationship with Y that X plays a partial causation in Y , which exists from previous layers to prescriptions.
- A *Rule Pattern* is defined as a direct path from X to Z or a logical path from a set of X to Z , which means nodes connected by links can be defined as a rule.

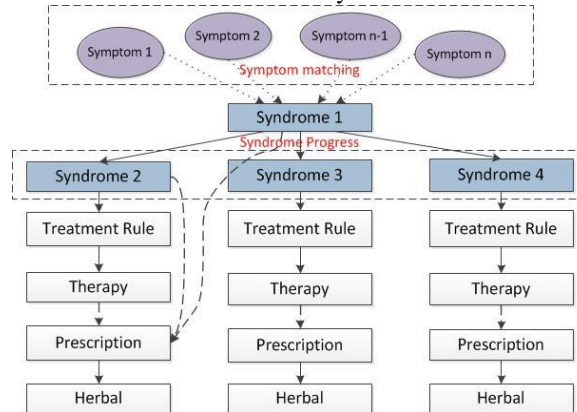


Fig. 1. The causal reasoning graph in TCM knowledge model.

The basic idea of Fig.1 is that certain set of symptoms identifies certain syndrome or disease (logic causation), and some syndrome or disease leads to the corresponding rule of treatment and therapy (direct causation), however the final prescription will be decided from the symptoms, the developing syndromes and the therapy together (weighted causation).

Generally, causal reasoning makes the TCM diagnosis process into a structural graph, and provides different layers of the graph with suitable algorithms.

4 Evaluation

In this section we discuss the experimental evaluation of our model. In total, we designed 821 classes, 26 object properties, 134 class assertion axioms and 78 object property axioms in our ontology. Since our causal knowledge model is derived from TCM diagnostic principles, we evaluated the process using two typical medical use cases in TCM.

Use case 1 (as shown in Fig 2)

Input: Angry (single symptom)

Output: A causal graph generated by user input as follow.

Description: When our system gets the input, it will search the RDF graph starting with the input node “angry”, that means searching reachable nodes through property chains and pre-defined rules in the generated graph. All the white nodes are the final displayed results. The blue ones are the latent causal links upon each edge, which can be displayed when user clicks the edge.

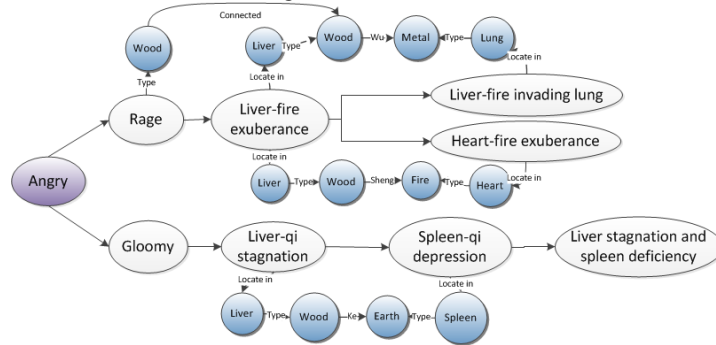


Fig. 2.Use case 1

Use cases 2 (as shown in Fig 3)

Input: A set of symptoms

Output: One or multiple possible syndromes diagnosed from the input symptoms

Description: When user submits a set of symptom, our approach searches possible corresponding syndromes in our knowledge base and output it to the user.

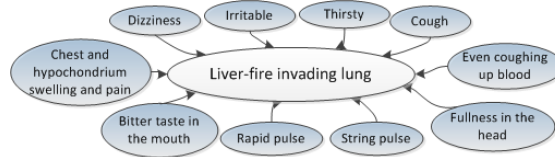


Fig. 3.Use case 2

The evaluation results are represented in Chinese in our system, so we depicted them in use cases above using alternative graphs. As we described, use case 1 shows the important principles in TCM philosophy, that’s the action cycles between the basic five elements, based on which we conduct knowledge learning and knowledge

mining. Our TCM knowledge model gets a satisfying result in TCM knowledge representation, it also has a great potential in knowledge mining.

5 Conclusion

In this paper, we present a causal knowledge modeling method that can be applied to TCM diagnosis process. The principle objective of TCM knowledge modeling is to figure out a formal method to represent Chinese medicine knowledge.

We build a TCM causal knowledge model based on the belief that the underlying causal relations inside TCM ontology can be represented using OWL 2. We defined seven ontology classes and corresponding sub classes based on Five Phase Theory. All the properties are defined according to the key activities between key concepts in diagnosis. The relationship between symptoms and diseases or syndromes is the focus in determining the disease. However, the relationship is not pure one-to-one relation, but many to uncertain. As for this reason, we viewed TCM knowledge as a layered causal graph, particularly viewed sets of symptoms as a whole, and our algorithms upon causal graph involves symptom matching and syndrome progress. Our causal knowledge reasoning, which is a method of integrating defined ontology with pre-defined rules to compose a causal graph, can clearly demonstrate the process of TCM diagnosis and shows a potential to do knowledge mining.

6 Acknowledgments

Thanks to the Grid Computation Group in CCNT Lab of Zhejiang University. This paper is supported by NSFC61070156, China 863 program with No.2009AA011903.

References

1. T.Berners-Lee, J.Hendler, and O.Lassila, J: The Semantic Web. Scientific American. (2001).
2. Frank Manola, Eric Miller, S: RDF Primer, <http://www.w3.org/TR/rdf-primer/>, (2004)
3. F.van Harmelen, S.Bechhofer, J.Hendler,et.al. S: OWL Web Ontology Language Reference. <http://www.w3.org/TR/owl-ref/>, (2002).
4. EB/OL: Wikipedia-Wu Xing. http://en.wikipedia.org/wiki/Wu_Xing
5. Cheng Y, Wang Y, Wang X. J: A Causal Relationship Discovery-based Approach to Identifying Active Components of Herbal Medicine. *Comput.Biol.Chem*, Vol.30, pp.148--154(2006).
6. M. Huang, M. Chen. J: Integrated Design of the Intelligent Web-based Chinese Medical Diagnostic system (CMDS)-Systematic Development for Digestive Health. *Expert Syst. Appl*. Vol.32, pp.658--673(2007).
7. X. Wang, H. Qu, P. Liu, et al. J: A Self-learning Expert System for Diagnosis in Traditional Chinese Medicine. *Expert Syst. Appl*, Vol.26. pp.557--566(2006).