Communications of the IIMA

Volume 13 | Issue 2 Article 4

2013

Composite Ontology-Based Medical Diagnosis Decision Support System Framework

Hsien-Tseng Wang
The City University of New York

Abdullah Uz Tansel
The City University of New York

Composite Ontology-Based Medical Diagnosis Decision Support System Framework

Hsien-Tseng Wang
The City University of New York, USA
elvinnyc@gmail.com

Abdullah Uz Tansel
The City University of New York, USA
tansel@baruch.cuny.edu

ABSTRACT

Current medical decision support systems have evolved from the automation of medical decision routines to improving the quality of health care services. Knowledge-based systems, compared to conventional data-driven techniques, are promising to support medical decision making. However, knowledge acquisition is usually a bottleneck in the process of developing such systems. One possibility for acquiring medical knowledge, particularly tacit knowledge, is to use data or cases in both syntactic and semantic ways. Case-based Reasoning (CBR) methodology provides a practical way of problem solving with recalled knowledge memory of solved cases. To reduce the difficulty of knowledge acquisition, this paper proposes a design of the system framework that utilizes the simplified medical knowledge: disease-symptom ontology for pre-diagnosis, given patient's symptoms and signs as input. In the first stage, simple pattern matching is used to gather candidate diseases in diagnosis. Following that, case-based reasoning is used to refine diagnostic decision. The case base is structured with ontological knowledge model. The case retrieval process is based on semantic similarity. The diagnostic system uses a composite knowledge base, and will allow automated diagnosis recommendation. The system framework also aims at facilitating semantic explanations to the solution derived.

Keywords: Medical diagnosis, ontology, semantic web, case-based reasoning, pattern matching.

INTRODUCTION

Medical decision making covers important tasks such as diagnosis, therapy planning, interacting with patients, identifying medical errors etc. Medical diagnosis is a process aiming at identifying diseases based on findings, such as symptoms and lab reports. The development of Medical (Clinical) Diagnosis Decision Support Systems (MDDS or CDSS) dates back to 1950s. Such developments, particularly in diagnosis decision support, have high complexity. A limited number of systems are adopted for practical use in the clinical environment. In the diagnosis process, an appropriate representation scheme is necessary for both problem interpretation and knowledge retrieval. From a medical cognition point of view, Long (2001) states that most medical reasoning methods are based on organizing various types of relations that exist in medical domain. These relations were identified by Long (2001) and include: associations, probabilities, causality, functional relationships, temporal relations, locality, similarity, and clinical practice (cases and experiences).

In computer science community, early diagnosis decision support (appeared as expert systems (ES)) research focused on rule-based reasoning (RBR) methods, decision table/tree, and later on Bayesian probabilistic, case-based reasoning (CBR). More recently as the computational power improves, machine learning-based systems emerged. These efforts were novel at the time; however, they did not employ a formal knowledge model (e.g., model of diseases, causation etc.), instead, they relied on the characteristics of data. Further, these approaches mostly operate in syntactic manner. Therefore, it is typically difficult to generate semantic explanation to the decision made by the system. The evolving Semantic Web research has brought a new platform for better knowledge representation, its sharing and semantic reasoning. The ontological model now serves as building blocks for the representation tasks in most knowledge-based applications.

Essentially, a good medical diagnosis system requires a structured knowledge representation component (model) that reflects most of the existing medical relations. It also needs employing efficient reasoning methods that closely follow medical cognition. Organizing these relations obviously needs one or more specific forms of knowledge representation, and computational artifacts that can manipulate them. None of these tasks is easy. Furthermore, physician-like users usually lack knowledge of how the framework works. In the worst case, a common bottleneck in knowledge-based systems is the knowledge acquisition, because of the acquisition mechanism is not transparent to the experts, or tacit knowledge is hardly complete enough to be usable.

We explore the case-based reasoning (CBR) methodology, as CBR not only utilizes the actual data (cases), but also works similarly to how human solves problems by recalling most relevant experiences. We propose a framework for medical diagnosis decision making. This framework incorporates the disease-symptom ontology, and case-based reasoning (CBR) coupling with semantic similarity calculation.

The paper is organized as follows: the next section introduces the readers to knowledge representation. The third section summarizes BioMedical ontologies and the semantic web technologies. The fourth section explains how CBR works. The fifth section presents our composite ontology framework. The diagnosis workflow is set up in the sixth section. This is followed by the conclusion.

KNOWLEDGE REPRESENTATION

In problem solving, knowledge provides the basis for reasoning in either a modal or an ad hoc way. Researches in artificial intelligence early on aimed at using computer power to act as human intelligence for problem solving. A human problem solver either exploits his own or others' experiences (if he understands well). Alternatively, he may visit available formal models (e.g., model of diseases), and search for relevant knowledge coupling with deduction to derive solutions. The latter approach is known to be model-based. Both approaches encounter issues in searching. Using exploit-experiences approach (e.g., machine learning, CBR or rule-based methods) requires certain amount of data or cases. On the other hand, model-based (e.g., knowledge-based) approach usually requires a knowledge base and a reasoning component. The knowledge base is constructed through a sophisticated modeling process. A reasoning component for inferring ground answer to the given problem searches for relevant knowledge from the knowledge base. The knowledge base is present, and the reasoning component is algorithmic. Hence, the answer derived is justifiable.

Knowledge representation (KR) research gained popularity in 1970s and developed main formalisms, such as logic-based and non-logic-based formalisms. Semantic network, production rule and frames are considered as non-logic-based. These were motivated from human cognition, and are high-level in languages and more human-centric. On the other hand, logic-based, such as First Order Logic (FOL) and Description Logic (DL), can unambiguously capture states about the world using logical constructs. The problem solving task becomes checking logic consequences.

The recent development of Semantic Web technology brings ontology as a promising tool in knowledge representation and reasoning. In addition, languages for ontology, such as OWL (McGuinness & van Harmelen, 2004), SWRL (Horrocks et al., 2004), and SQWRL (O'Connor & Das, 2009) were matured. As OWL-family has DL foundation, reasoning services (subsumption and class properties) are built-in as checking logical consequence by deduction. Standard reasoning in OWL-Lite and OWL-DL is decidable, which is a desirable feature at least for inference application developers.

BIOMEDICAL ONTOLOGY AND THE SEMANTIC WEB

Biology and medicine are among the earliest domains that adopted ontological framework for controlled vocabularies and taxonomy, and facilitating semantic interoperability. To develop a universal domain ontology for biology or medicine is difficult, and in fact not possible. Efforts in biomedical ontology such as UMLS (Unified Medical Language System), GO (Gene Ontology) and SNOMED-CT (Systematized nomenclature of medicine clinical terms), emerge as cornerstones supporting better taxonomy for terminologies and semantic interoperability. These systems are based on (Schulz & Hahn, 2005): (1) classification, (2) Multiaxial coding system, (3) Thesaurus-like coding system, and (4) Topography-like system. For instance, SNOMED-CT belongs to multiaxial coding system. In that, each axis has a hierarchical structure and semantic links are allowed for more expressive power. This leads to multiple hierarchies of concepts that may suffer from issues and challenges arising from multiple inheritance.

One special repository, UMLS, was designed to help interpretation and understanding of medical meanings across application systems. It is a huge umbrella system that contains Meta-thesaurus, Semantic Network, Specialist Lexicon and Lexical Tools. As UMLS draws information from mixing sources, it might contain cycles and other similar problems. The main hierarchy is defined by IS_A relation (i.e., specialization—generalization relation), which does not distinguish itself from partonomic relation (part-of relation). For information concerning other ontology-style knowledge repositories, such as OpenGalen, Open Biomedical Ontologies (OBO), readers can refer to the BioPortal homepage by the U. S. National Center for Biomedical Ontology (NCBO, n.d.).

Another interesting development is the disease ontology, introduced in (Schriml et al., 2012), and can be retrieved from (“Disease ontology,” n.d.). The human disease-ontology (DO) contains a comprehensive knowledge base of roughly 8043 inherited, developmental and acquired human diseases. DO also maps itself to common biomedical ontology, such as MeSH (Medical Subject Headings) and SNOMED-CT. The disease ontology browser provides a simple pattern matching mechanism. For instance, when an input *bleeding* keyword is entered, the retrieved list contains 25 candidate diseases. Potentially, this retrieved result can be used as a pre-diagnosis step given an input keyword (symptom).

As the above mentioned biomedical ontology repositories provide re-usability and interoperability, HL7 (High Level Seven) standards (Health Level Seven International, n.d.) provide a protocol-like mechanism for exchanging semantic information in healthcare. While UMLS and HL7 were being developed, the semantic web technology OWL (Ontology Web Language) concurrently emerged as a standard knowledge representation language in semantic-oriented applications. As such, OWL was not incorporated into both UMLS and HL7 yet. However, there is still a possibility of using OWL or its extension as mapping language for UMLS and HL7.

The most apparent use of the semantic web technology in medical domain is to facilitate: (1) querying multiple data sources (database and ontologies) with reduced semantic ambiguity, and (2) semantic preserving integration of distributed medical ontologies. Building on this, deploying a medical decision support system may prove to be very useful, in addition to the possible inclusion of different reasoning strategies.

Since ontology provides a clear semantic and knowledge description of the concepts and interrelationships, it can be adapted to the case description, structuring, storage, and even the knowledge (case) retrieval with semantic operations. Further, more BioMedical ontologies become available, and medical CBR systems can take advantages of these ontologies. Utilizing ontologies in medical CBR systems will enhance the traditional syntactical CBR approach. For instance, semantic similarity measures could be developed for case retrieval in the CBR cycle.

CASE-BASED REASONING

Case-Based Reasoning (CBR) (Kolodner, 1993) is a cyclic problem solving paradigm that emphasizes the reuse of solutions to similar problems in experience memory. The solutions, along with the corresponding problems, are maintained in an indexed case base for faster case retrieval. A case is a problem-solution pair. A typical CBR cycle is shown in Figure 1. Basically each instantiation of a CBR process follows a retrieve, reuse, revise and retain cycle.

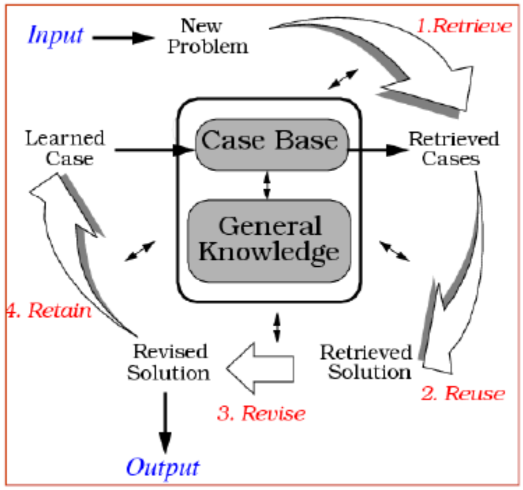

Figure1: The CBR Cycle.

CBR methodology has been applied in many application domains, such as engineering, law, mechanics, medicine etc. It is usually applied to problems in which complex tacit knowledge plays important roles. In a CBR system, expertise knowledge is embodied in a library of past (solved) problems as cases, rather than being encoded in classical rules or formal models (such as ontology). Each case contains a description of the solved problem, and a solution and/or the outcome. The actual knowledge and reasoning process used by an expert to solve the problem is not recorded, but is implicitly embedded in the solution.

A new problem is matched against the cases in the case base, and similar cases are retrieved based on pre-defined similarity measures. For example, in K-nearest neighbor algorithm, the simplest similarity measure could be the Euclidean distance. The retrieved cases are used to suggest a solution. If necessary, the solution is adapted and revised to satisfy the new problem. Finally the current problem and the suggested solution are retained as a newly solved case if necessary. The retaining of cases is one of the most advantageous features of CBR. This enables the learning and growth of problem solving ability. Using CBR is more intuitive to users because it is essentially working with concrete examples, rather than conclusions separated from their context. A case base can also be a powerful resource from a knowledge management point of view.

Since the 1990's, CBR has grown into a field of widespread interest, both in academic and commercial domains. Mature tools are available, and application-focused conferences exist. For instance, in (Bichindaritz & Montani, 2011), advances and efforts of CBR in health science are briefly ranked by dates from the 1980s. Case-based reasoning is often used as a generic term to describe techniques including, but not limited to, CBR as described above.

In traditional CBR-based problem solving, more syntactical manipulation is applied to cases. For instance, to retrieve the most relevant cases in the case base, sometimes only the calculation of *distance* is needed, followed by the actual retrieval process. The usability of distance calculation is assumed and based on a vector-style representation of cases. In this mode, the distances derived may not have any contextual meaning. To improve the performance and usability of

CBR-based systems, one can consider applying the semantic web technology and ontological knowledge model, in addition to the *syntactical* CBR.

In (Bichindaritz, 2006), formal semantics and CBR in biology and medicine are discussed. The conclusion shows that adding semantic support to traditional CBR systems is common, and the ontological support improves the performance of CBR systems. In addition, traditional CBR case representation and contents do not usually contain enough details required by a CBR system to act effectively in all CBR cyclic tasks. Examples seen in (Bichindaritz, 2006), show that most medical CBR systems use ontology for case representation and interpretation. Therefore, simple *distance calculation* can possibly be adapted to semantic distance calculation. In addition, some type of hybrid methodologies emerges in medical CBR-based systems for decision support or problem solving. Examples are "CBR combined with Electronic Patient Record" in (van den Branden, Wiratunga, Burton, & Craw, 2011), and "CBR combined with RBR (Rule-Based Reasoning)" in (Berka, 2011).

SYSTEM ONTOLOGY FRAMEWORK

The proposed composite ontology framework is motivated from the usefulness of disease-ontology (DO) and hybrid CBR approaches. The composite ontology contains the following ontology parts:

Medical Scenario Ontology

This is the episode driving ontology that specifies the context, domain, situation, and the structure of case library.

Medical Knowledge Ontology

This part specifies general categories of medical knowledge as templates of fundamental medical concepts, generic concepts, anatomy concepts, and disease concepts in the domain that is referenced by Medical Scenario ontology.

Electronic Medical Record Ontology (EMR)

EMR ontology contains clinical data created for health professionals (including their notes) in the course of providing care in a hospital or clinician's office. A piece of typical EMR may contain administrative data, diagnosis and treatment history of the patient. The EMR ontology could support the construction of medical cases.

Disease - Symptom Ontology

Disease-Symptom ontology contains a generalized disease ontology derived from DO mentioned in (“Disease Ontology,” n.d.). In DO, no explicit symptoms are defined, other than referenced textual definition. So applying Natural Language Processing techniques to derive the symptoms for corresponding diseases is necessary. The disease domain could be confined based on the specification in Medical Scenario ontology. In addition, this ontology part is expected to expand as new supporting conceptualization model, such as disease causal model or prototypical model is developed.

Medical Case Store

Medical case store is a case ontology that contains necessary descriptive constructs for constructing and maintaining a case base. In addition, a concept of context forming is included for case retrieval and case matching purpose.

COMPOSITE MEDICAL ONTOLOGY SUPPORTING PATTERN MATCHING AND CASE-BASED REASONING

With the composite ontology mentioned above, a work flow for a diagnosis problem solving is outlined (assuming input of Symptoms) in Figure 2. The main stages in the workflow are as follows:

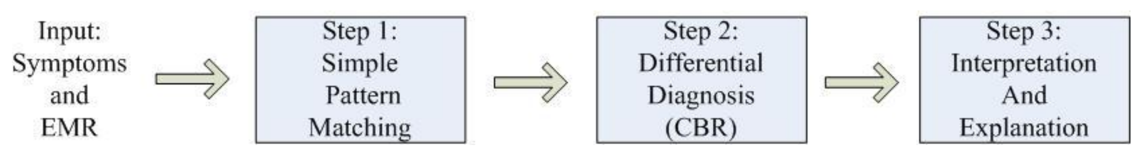

Figure 2: System Workflow.

Episode Driving Engine

The Episode Driving Engine is the central controlling unit that facilitates the workflows. It first initiates a diagnosis episode and organizes a workspace and decides the context of the diagnosis problem. More importantly, the Episode Driving Engine serves as the common denominator of all system components.

Simple Disease-Symptom Matching

In this stage, an iterative pattern matching is performed by the pattern matching engine, similar to algorithm found in (Carvalho, Isola, & Tripathy, 2011). Basically, an input symptom fed to the system, then it presents a list of all possible disease ranked based on the number of matched symptoms. This process goes iteratively until all input symptoms are consumed by the system. The end product of this stage is a rank list of candidate diseases. If the list contains only one disease, the one is the primary solution for the diagnosis problem. Otherwise, differential diagnosis is needed.

Differential Diagnosis With CBR and Semantic Similarity Calculation

Following the result of the first stage, a CBR cycle is performed. The CBR inference engine (reasoner) first retrieves the most similar cases based on the input case (i.e. symptoms and EMR). The retrieval process is time consuming, as it has to search through the specified case ontology. To reduce the searching complexity, candidate diseases derived from previous stage can be used to confine the searching space. As for similarity measure, a recent developed method called DOPCA presented in (Gan, Dou, Wang, & Jiang, 2011) is adopted. This similarity measure will guide the retrieval of similar cases.

After the retrieval, the CBR cycle encounters candidate cases that may not be exact match to the problem in consideration. Before the adaptation process starts, the CBR module invokes the Disease-Symptom Ontology and determine if any supporting model is available and uses it. Otherwise, the adaptation (revise) process is activated. The adaptation is based on pre-defined rules. Once the adaptation is finished, the solution is derived.

Explanation and Interpretation

The Episode Driving Engine keeps track of all tasks performed and is able to provide a complete episode for the diagnosis with integrated formal semantics.

A template system diagram is shown in Figure 3.

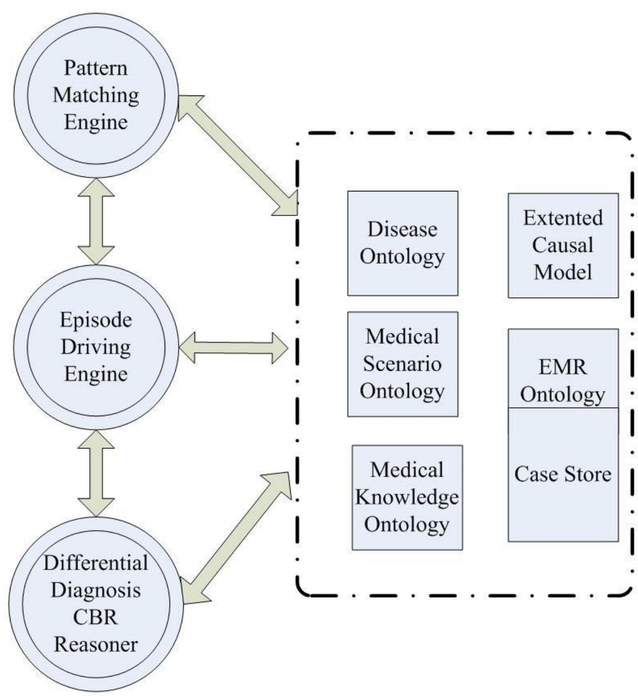

Figure 3: System Diagram.

CONCLUSION

CBR methodology not only utilizes the actual data (cases), but also works similarly to how human solves problems by recalling most relevant experiences. We propose a composite ontology for diagnosis that encompasses both simple pattern matching and CBR for differential diagnosis. The composite ontology is constructed using semantic web standard languages such as OWL and its variants. In the iterative simple pattern matching step, candidate diseases are obtained. Potentially, differential diagnosis step continues the diagnosis process. The search space for the CBR is confined to that suggested by the candidate diseases. The case retrieval accuracy is augmented by semantic similarity calculation. The candidate disease set can be further refined by invoking the extended causal inference model.

The value of composite ontology is two folds: the ontology parts are all constructed by using semantic web standard languages. This enables the interoperability of the composite ontology store to agents outside of the system. On the other hand, the composite ontology is applicable to multiple reasoning strategies. Since reasoning strategies applied in the system are algorithmic, explanation to the diagnosis derived can be constructed with the Episode Driving Engine.

This paper only proposes a design sketch. In the future, we anticipate research on core components of a system prototype based on the proposed design. Further, we shall design the Episode Driving Engine, and explore efficient ways of constructing cases from EMR.

REFERENCES

Berka, P. (2011). NEST: A compositional approach to rule-based and case-based reasoning. *Advances in Articial Intelligence, 2011*. doi: 10.1155/2011/374250

Bichindaritz, I. (2006). Mémoire: A framework for semantic interoperability of case-based reasoning systems in biology and medicine. *Artificial Intelligence in Medicine, 36*(2), 177-192.

Bichindaritz, I., & Montani, S. (2011). Guest editorial: Advances in case-based reasoning in the health sciences. *Artificial Intelligence in Medicine, 51*(2), 75-79. doi: 10.1016/j.artmed. 2011.01.001

Carvalho, R., Isola, R., & Tripathy, A. K. (2011). MediQuery: An automated decision support system. *Proceedings of the 24th IEEE International Symposium on Computer-Based Medical Systems (CBMS),* Bristol, UK, 1-6. doi: 10.1109/CBMS.2011.5999050

Disease ontology. (n.d.). Retrieved April 20, 2013, from http://disease-ontology.org

Gan, M., Dou, X., Wang, D., & Jiang, R. (2011). DOPCA: A new method for calculating ontology-based semantic similarity. *2011 IEEE/ACIS 10th International Conference on Computer and Information Science* (pp. 110-115). doi: 10.1109/ICIS.2011.25

Health Level Seven International (HL7). (n.d.). *Introduction to HL7 standards*. Retrieved on April 25, 2013, from http://www.hl7.org/implement/standards/index.cfm?ref=nav

Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosof, B., & Dean, M. (2004). *SWRL: A semantic web rule language combining OWL and RuleML*. Retrieved from http://www. w3.org/Submission/SWRL/

Kolodner, J. (1993). *Case-based reasoning*. San Mateo, CA: Morgan Kaufmann.

Long, W. J. (2001). Medical informatics: Reasoning methods. *Artificial Intelligence in Medicine, 23*(1), 71-87. doi: 10.1016/S0933-3657(01)00076-8

McGuinness, D. L., & van Harmelen, F. (2004). *OWL web ontology language overview*. Retrieved from http://www.w3.org/TR/owl-features/

National Center for Biomedical Ontology (NCBO). (n.d.). *BioPortal*. Retrieved on April 25, 2013, from http://www. bioontology.org/

O'Connor, M. J., & Das, A. K. (2009). SQWRL: A query language for OWL. In R. Hoekstra & P. F. Patel-Schneider (Eds.), *Proceedings of OWL, 529*. Retrieved from http://ceur-ws.org/Vol-529/owled2009_submission_42.pdf

Schriml, L. M., Arze, C., Nadendla, S., Chang, Y.-W. W., Mazaitis, M., Felix, V., Feng, G., & Kibbe, W. A. (2012). Disease ontology: A backbone for disease semantic integration. *Nucleic Acids Research, 40*(D1), D940-D946. Doi: 10.1093/nar/gkr972

Schulz, S., & Hahn, U. (2005). Part-whole representation and reasoning in formal biomedical ontologies. *Artificial Intelligence in Medicine, 34*(3), 179-200.

van den Branden, M., Wiratunga, N., Burton, D., & Craw, S. (2011). Integrating case-based reasoning with an electronic patient record system. *Artificial Intelligence in Medicine, 51*(2), 117-123.