Expert Systems with Applications 38 (2011) 15253–15261

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Medicine expert system dynamic Bayesian Network and ontology based

Octavian Arsene *, Ioan Dumitrache, Ioana Mihu

Laboratory of Intelligent Systems, "Politehnica" University of Bucharest, SPl. Independentei 313, 060042 Bucharest, Romania

A R T I C L E   I N F O

Keywords:
Expert system
Ontology
Bayesian Network

A B S T R A C T

The paper proposes an application framework to be used for medicine assisted diagnosis based on ontology and Bayesian Network (DBNO). There are two goals: (1) to separate the domain knowledge from the probabilistic information and (2) to create an intuitive user interface. The framework architecture has three layers: knowledge, uncertainty model and user interface. The contributions of the domain experts are decoupled, the ontology builder will create the domain concepts and relationships focusing on the domain knowledge only. The uncertainty model is Bayesian Network and the probabilities of the variables states are stored in a profile repository. The diagnostician will use the user interface feeded with the domain ontology and one uncertainty profile. The application was tested on a sample medicine model for the diagnose of heart disease.

1. Introduction

The diagnosis can be defined as the process of identifying a set of hypotheses that model the problem domain and finding that one with highest probability of matching the real world state. In medical diagnosis, the uncertainty arises from the inability to evaluate the degree of truth of a hypothesis due to unreliable and incomplete information or inconsistent knowledge.

The ontology and Bayesian Network (BN) methodologies have been chosen to address knowledge management and uncertainty. The ontology enables the representation of a domain knowledge in a machine understandable form. It can represent the organizational structure of a large complex domains, but the inability to deal with the uncertainty can be a drawback for its application. One disadvantage of the BN is representation of complex structured domains point of view. The ontology and BN can complement themselves in order to overcome the each other disadvantages, thus an ontology-driven uncertainty model can be created.

The main goal of the paper is to propose an application framework as a collaborative expert system for creating, developing and maintaining a general model for medicine assisted diagnosis (Fig. 1). From user point of view there will be three roles based on their competencies: concepts and relations definition, connect probabilities to states of the concepts, setting evidences in order to assist the diagnosis. Each role is assigned to one of the triangle' sides and thus depicting three operational layers. The automated connection between all three layers increases the efficiency of the entire process. The proposed model is implemented as a software using PROTEGE as an ontology framework, NETICA API (Application Programming Interface) as a Bayesian API and Java technology as a development platform.

The mapping between domain knowledge and uncertainty model is based on the fact that each concept defined into ontology is part of the BN as a variable.

The diagnostician will use an intuitive graphical user interface for changing evidences of the BN variables and based on a threshold the application will depict a chart having the most significant nodes. The final chart will assist the medicine diagnostician to identify the significant factors for a particular case. In order to offer more flexibility the domain knowledge and probabilities are already build and ready to be used as a medicine ontology and respectively uncertainty profile. There is no need to be re-created each time during diagnosis phase. The diagnostician can choose between more than one uncertainty profiles for a medicine ontology. The relation between ontology and uncertainty profile is one to many (in case there are several sources for probabilities tables for the same ontology). This approach speed up the entire diagnosis process and allows the asynchronous update of the ontology and uncertainty profiles by the domain experts in order to increase the degree of accuracy of the information.

A similar model, OntoBayes, was proposed in Yi (2007). The major difference between OntoBayes and this proposed model resides in separation between domain knowledge and quantitative component of BN in order to decouple the ontology from the uncertainty probabilities. BayesOWL (Zhongli, 2005) and PR-OWL (Cesar, Costa, Laskey, & Laskey, 2003) are others probabilistic ontology approaches facilitating ontology mapping in the semantic web. Some of their limitations refer to: two-valued variables only

* Corresponding author.
E-mail address: octavianarsene@gmail.com (O. Arsene).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.05.074

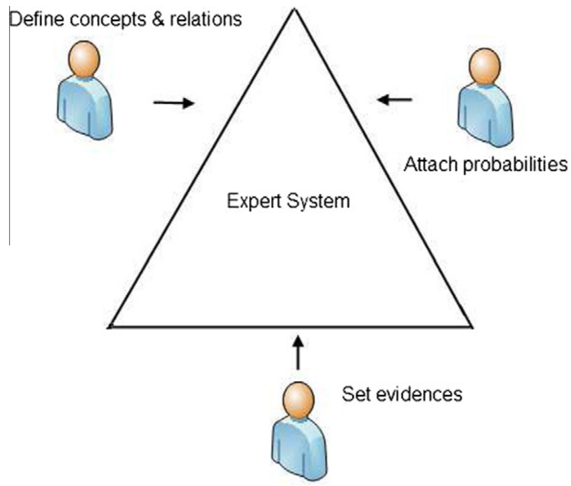

Fig. 1. Architecture of the proposed concept.

allowed in BayesOWL, the ontology model is too complex in PR-OWL. An automated BN construction based on an ontology model was proposed in Devitt, Danev, and Matusikova (2006) to be applied in the telecommunication network management domain. It's more an algorithm model than an application. Similar approach was proposed in Mihu and Arsene (2009) for monitoring and diagnosis of an IT infrastructure, all layers (gathering evidences, ontology model and presentation layer) were implemented as software agents.

Section 2 briefly introduces the ontology concept, the knowledge representation and the implementation of the medicine ontology using PROTEGE application (Horridge, Knublauch, Rector, Stevens, & Wroe, 2004). Section 3 describes Bayesian Network concept, the uncertainty representation and the implementation module within the proposed software application. Section 4 covers the description of proposed software application. Section 5 presents the results of the tests using heart disease data presented in Ghosh and Valtorta (1999) and Ghosh and Valtorta (2000).

2. Knowledge representation: ontology

The ontology may take a variety of forms, but necessarily it will include a vocabulary of terms and some specification of their meaning. The ontology is virtually always the manifestation of a shared understanding of a domain that is agreed between a numbers of parties. Such agreement facilitates accurate and effective communication of meaning, which in turn leads to other benefits such as inter-operability, reuse and sharing (Uschold, 1996). The main steps of the methodology for developing ontologies are the following: (1) identify purpose and scope, (2) building the ontology.

2.1. Purpose and scope

The proposed ontology addresses the medicine domain. It will be built and used by doctors. The purpose on the proposed ontology is to be knowledge representation of the medicine domain integrated into an application. The doctors will use it in order to define an accurate diagnosis based on the cumulated experience captured into ontology and probabilities tables.

2.2. Building the ontology

The first step in building ontology is the identification of the key concepts. In the Fig. 2 are depicted the key concepts defined within the medicine ontology based on the medical documents and

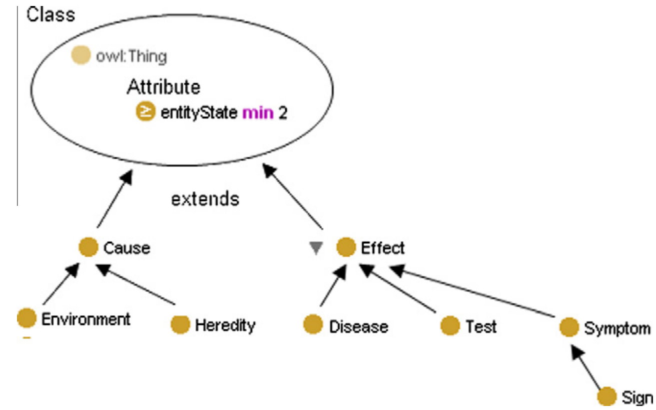

Fig. 2. Hierarchy of the medicine ontology classes defined in PROTEGE.

discussions with medical experts. The class is the formal way to represent a concept, e.g. the *MedicineNode* is the root concept of the medicine classes. The two main classes are proposed here: *Cause* and *Effect*. The Cause subclasses are considered: *Environment* and *Heredity*. The Effect subclasses are: *Disease*, *Symptom*, *Sign* and *Test*.

The second step is the identification of the relationships between key concepts that were defined before. We distinguished two kinds of relations between proposed classes (see Fig. 3):

- *any Cause will have at least one Effect*, implemented by *Cause has Effect* (the inverse relation is: *Effect of Cause*);
- *any Effect can generate another Effect*, implemented by *Effect generates Effect* (the inverse relation is: *Effect generatedBy Effect*). This seems to be odd but let's take these examples: *Disease generates Symptom* or *Sign generates Test*.

In the Fig. 4 is presented the implementation of the relations between concepts in PROTEGE.

Both relations are inverse type based. This kind of relation will be used in bidirectional inference necessary to graphical representation of the Bayesian Network.

Different ontology languages provide different facilities. The most recent development in standard ontology languages is OWL, endorsed by the World Wide Web Consortium (W3C) to promote the Semantic Web vision. The PROTEGE framework is going to be used for creating, maintaining the medicine ontology. PROTEGE is a free, open source ontology editor and knowledge-base framework. The PROTEGE-OWL editor is an extension of PROTEGE that supports the Web Ontology Language (OWL). The OWL ontology may include descriptions of classes, properties and their instances. The OWL-DL language is used due to its expressivity and computational efficiency. The real life entities from the medicine domain are encoded as classes instantiations within an ontology (called individuals).

Each individuals has a clear identity that makes them different than others (even they have common attributes). The individuals used in the proposed application are based on the heart disease model presented in Ghosh and Valtorta (2000), they are depicted in the Fig. 5. The domain expert (a senior doctor) will create the individuals according to his expertise and experience. In the Fig. 6 HighBloodPresure individual is defined:

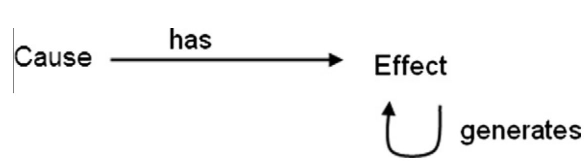

Fig. 3. Medicine ontology – relations between classes.