Ontology-based Semi-automatic Construction of Bayesian Network Models for Diagnosing Diseases in E-health Applications [1]

Beom-Jun Jeon and In-Young Ko
Information and Communications University
119 Munjiro, Yuseong-gu, Daejeon, 305-732, Korea
{shadow, iko}@icu.ac.kr

Abstract

Bayesian network (BN) is one of the popular probabilistic methods of diagnosing diseases in e-health applications. However, it is normally a complex task to construct a BN for diagnosing a specific disease by collecting and analyzing domain knowledge. We propose a semi-automatic way of constructing BNs for diagnosing diseases. Our method automatically generates nodes in a BN out of e-health ontology, and allows developers to easily establish links among nodes based on a meta-model that represents cause-and-effect relationships among ontologies.

1. Introduction

Nowadays ontology is widely used in bioinformatics areas for representing, sharing, and processing various types of bio information for many different application domains [1]. However, because of the non-deterministic nature of the relations among different types of bio information, there must be a way to represent and process non-deterministic information among e-health ontologies [2].

Bayesian network (BN) is one of the popular probabilistic methods of diagnosing diseases in e-health applications [3]. A BN can be used to represent the relationships between symptoms and diseases. However, it is normally a complex task to construct a BN for diagnosing a specific disease by collecting and analyzing domain knowledge. We propose a semi-automatic way of constructing BNs for diagnosing diseases. Our method automatically generates nodes in a BN out of e-health ontology, and allows developers to easily establish links among nodes based on a meta-model that represents the cause-and-effect relationships among ontologies.

There have been some researches to construct a BN automatically from ontology [4] [5]. However, they mainly focus on generating a generic BN out of ontology based on the relationships among ontological terms represented in subsumption relations and properties. The resulting BN is normally too complex to manage for diagnosing diseases because they do not provide any criteria to selectively generate links among many BN nodes.

Section 2 describes requirements of constructing a BN from ontology. Section 3 proposes the semi-automatic BN construction method, and in Section 4, we demonstrate an implementation of this method for diagnosing obesity. Section 5 summarizes our approach and discusses future extensions.

2. Requirements of the ontology-based Bayesian network construction

There are some requirements that we need to consider in generating a BN from ontology. These requirements are needed to produce an ontology-based BN that is manageable and flexible, and has low complexity.

- **Manageability**. Developers need to easily create BNs from large and complex ontologies. There must be a way to help developers construct BN nodes that reflect the classes in ontology, and link the BN nodes based on the relationships among the classes.

- **Flexibility**. Since e-health ontologies normally evolve as we gather more domain knowledge, there must be a mechanism to make BNs be dynamically updated according to the modifications of ontologies.

[1] This research was supported by Samsung electronics research project (Samsung-ICU Mobile u-Health Service)

0-7695-2999-2/07 $25.00 © 2007 IEEE
DOI 10.1109/FBIT.2007.63

▪ **Low Complexity**. As e-health ontology grows larger, it normally becomes harder to produce BN nodes for every class in the ontology. Therefore, there must be a way to produce BNs based on a part of the ontology that is mostly related to an e-health application.

3. A semi-automatic Bayesian network construction from ontology

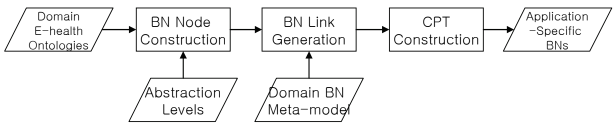

Figure 1. The overall processes of the Bayesian network construction

Figure 1 shows the overall process of constructing BNs semi-automatically from the e-health ontologies. In this process, developers first select abstraction levels in the ontologies to specify the areas that are important for a specific e-health application. Based on these abstraction levels selected, BN nodes are selectively constructed from the ontologies. Then, based on an application-specific meta-model, links among BN nodes are semi-automatically created. Finally, developers need to construct conditional probability tables (CPTs) manually to produce an application-specific BN. We are currently investigating the semi-automatic ways of constructing CPTs.

3.1 Selecting abstraction levels

E-health ontologies normally contain a lot of classes which are hierachically structured. Therefore, it is usually a complex task to convert all ontology classes to BN nodes.

We identified that only the classes on a specific abstraction level of ontology are mostly useful in diagnosing diseases for an e-health application. We provide a way to specify the areas of ontology classes by selecting abstraction levels of ontology. Then, each class in those selected areas is converted into a BN node.

For examaple, when there are classes in a subsumption hierarchy in ontology, *Disease – Mental_Disorders – Stress,* application developers select *Stress* as an abstraction level if the application is specialized for diagnosing stress. By adjusting the level of abstraction for selecting ontology classes, we can limit the number of BN nodes generated for the application.

3.2 Converting ontology classes to Bayesian network nodes

After selecting abstraction levels, the system that we propose automatically constructs BN nodes from ontology. In a BN, the states of a node represent probabilistic conditions that the node can have. When a class does not have any subclasses, the system constructs BN nodes with the true and false states to represent absence or presence of the situation denoted by the ontology class. In addition, when the node has subclasses, the system generates another BN, a sub-BN, which only contains nodes constructed by the subclasses. However, when the subclasses of the node are disjoint with each other (i.e., do not intersect in meaning) the system uses subclasses as states of the node. Algorithm 1 describes a detailed procedure of automatically converting classes in ontology to BN nodes.

```
foreach class, c, in ontology do
    if c is selected as an abstraction level then
        if c doesn't have any subclasses then
            Make a node of c for the main BN, M, with true/false
            states;
        else
            if all subclasses of c are disjoint with each other then
                Make a node of c for M with subclasses of c as states;
            end
            else
                Make a node of c for M with true/false states;
                foreach subclass, s, of c do
                    if all subclasses of s are disjoint then
                        For a sub-BN of c, make a node of s with
                        subclasses of s as states;
                    end
                    else
                        For a sub-BN of c, make a node of s with
                        true/false states;
                    end
                end
            end
        end
    end
end
```

Algorithm 1. Constructing Bayesian network nodes from ontology classes

Figure 2 is an example of constructing BN nodes by using Algorithm 1 when the developers select *Thickness_of_Subcutaneous_Fat* as an abstraction level in the Bio_Data ontology (the ontology shown in the figure is partial).

The main BN in this example includes nodes generated from ontology classes that are directly marked as abstraction levels. Sub-BNs are smaller BNs containing nodes that are generated from the subclasses of the selected classes. Each node in the main BN can have only one sub-BN, i.e., a sub-BN is mapped to a node in the main BN. A sub-BN is accessed when a user requests more detail diagnosis information for the node. The use of sub-BNs makes it possible to provide more accurate diagnosis functions with low complexity.

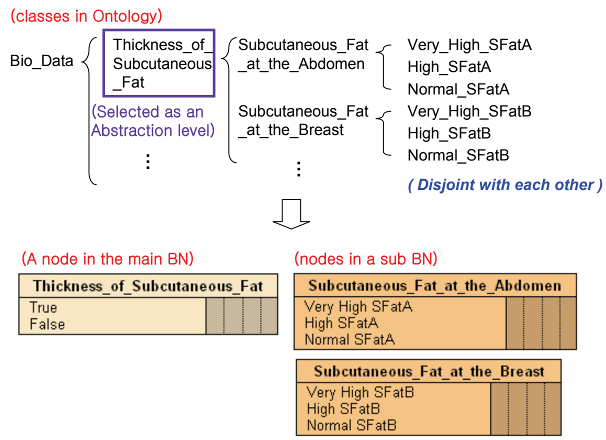

Figure 2. Abstraction level selection from the *Bio_Data* ontology and BN nodes constructed from the selected classes

3.3 Generating Bayesian network links

As we discussed at the previous sub-section, we can generate nodes to construct a BN from ontologies in a semi-automatic way. However, to produce links among the BN nodes, we need to analyze dependencies and orders among many BN nodes, which is normally a complex task. Our approach provides a way to generate links between nodes semi-automatically by using a relational model, a meta-model, among the ontologies.

The domain BN meta-model describes the cause-and-effect relationships between two ontologies in a specific domain. For example, if we consider a cause-and-effect relationship between the *Obesity* ontology and the *Symptom* ontology, the *Obesity* ontology becomes 'cause', and the *Symptom* ontology beomes 'effect' in the meta-model. Based on the constructed meta-model, for each node in the main BN and sub-BNs, the system constructs links from the node to other relevant BN nodes.

Algorithm 2 explains the detail procedure of generating links among BN nodes based on a meta-model. By using this algorithm, our system identifies a list of candidate nodes that can be connected to a specific node that is selected by a developer. The core idea of this algorithm is to consider the cause-and-effect relationship represented in the meta-model, and to identify relevant BN nodes among which there are cause-and-effect relationships. In the process of making links between nodes, if a node has a sub-BN, the system creates a clone copy of the node to which a link needs to be established, and puts the clone copy into the sub-BN of the source node. If the target node also has a sub-BN, the system allows the developers to selectively make clones of the nodes in the sub-BN, and put cloned nodes into the sub-BN of the source node. Links among the nodes in a sub-BN are established by repeating the same algorithm.

```
Developers, D, request linking for a specific node, n, in a BN
    Cause nodes list of n, C := empty;
    Effect nodes list of n, E := empty ;
    foreach node, o, in the same level of BN with n do
        if o has a Cause relation with n at the meta-model then
            Add o at C;
        end
        if o has a Effect relation with n at the meta-model then
            Add o at E;
        end
    end
    foreach nodes, o', in C and E do
        Ask D whether add link between o' and n;
        if D answer yes then
            mark o' to be link node with n;
        end
    end
    foreach marked nodes, mo, in C and E
        if mo has a sub-BN then
            foreach n or n's lineal sub-BN nodes, sn, that are directly
            converted from subclasses of n do
                Ask D whether add n or sn to mo's subBN;
                if D answer yes then
                    Create and add a clone of the node to the sub-BN
                    of mo;
                end
            end
        end
        if n has a sub-BN then
            foreach mo or mo's lineal sub-BN nodes, smo, that are
            directly converted from subclasses of mo do
                Ask D whether add mo or each smo to n's subBN;
                if D answer yes then
                    Create and add a clone of the node to the sub-BN
                    of n;
                end
            end
        end
        if mo is the node in C then
            Generate the link from mo to n;
        end
        if mo is the node in E then
            Generate the link from n to mo;
        end
end
```

Algorithm 2. Generating Bayesian network links by using a meta-model

4. Implementation

The overall architecture of our system for supporting semi-automatic construction of BNs is shown in Figure 3.

In this architecture, the Ontology Manager is the core element that implements the entire process of constructing BNs from ontologies. The Ontology Manager is composed of Data Pre-processor, Ontology Processor, and BN Processor. The role of Data Pre-processor is to aggregate and map various e-health data into the e-health ontologies. Ontology Processor enables other elements in the Ontology Manager to access and manage ontology data. BN Processor manages BN nodes and their sub-BNs to be consistent with the e-heath ontology. It is also in charge of creating and accessing sub-BNs while performing inferences based on a BN.

Jena [6], a Java API for Semantic Web applications, provides the Ontology Processor with a set of APIs by which in-memory models of ontology can be created and managed. The Protégé ontology editor [7] provides an environment in which developers can edit e-health ontology. BN Processor is implemented by using Netica [8], which provides a GUI and APIs to manage BNs and to perform inferencing.

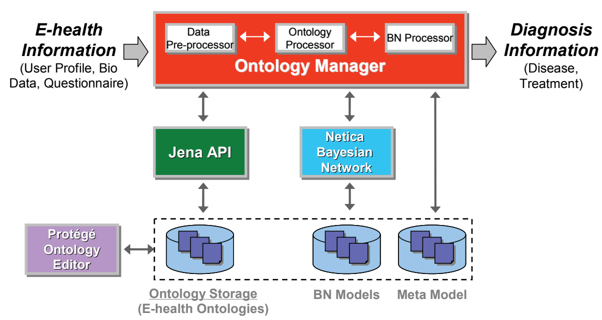

Figure 3. The system architecture

We adopt our system to an e-health application for managing obesity. We implemented a BN for diagnosing obesity. To do that, we first constructed application specific-ontologies: *Disease, Symptom, Obesity, Obesity_Causes,* and *Bio_Data* ontologies. We constructed our *Disease* and *Symptom* ontologies based on the Disease Ontology from the NuGene project [9].

To support more sophisticated diagnosis about obesity, we constructed the *Obesity* ontology categorized by body parts, causes, and personal characteristics. To combine and represent various factors that cause obesity, we define the *Obesity_Causes* ontology. The *Obesity_Causes* ontology maps between obesity causes to specific obesity types. The *Bio_Data* ontology represents data which can be acquired from bio sensors.

In our example shown in Figure 4, we set the abstraction levels such that classes can be mapped to BN nodes to represent specific obesity types, relevant causes, and their effects.

5. Conclusion

In this paper, we explained the semi-automatic BN construction system based on e-health ontologies. The system allows developers to select abstraction levels in e-health ontologies to specify the areas that are important for a specific e-health application. BN nodes are selectively constructed from the selected ontology areas to reduce complexity of the BN. Finally, by using an application-specific meta-model, BN nodes can be linked with each other semi-automatically.

Our approach satisfies the major requirements of generating a BN from ontology. Our system enables the production of a BN that is manageable and flexible, and has low complexity.

Our framework enables probabilistic inferencing capability for various E-health applications by allowing developers to construct domain-specific BNs in a semi-automatic way. This approach contributes to reduce the complexity of BNs, and BN-based inference, which is critical to practically support E-health diagnosis systems.

Currently, we are working on developing a mechanism to help developers construct CPT more efficiently. For example, a CPT can be automatically constructed based on a probabilistic distribution across multiple factors in a BN node. In addition, we are conducting a research on how to produce personalized BNs by reconfiguring existing BNs based on users' feedback information.

6. References

[1] C. Golbreich, O. Dameron, B. Gibaud, and A. Burgun, "Web ontology language requirements w.r.t. expressiveness of taxonomy and. axioms in medicine", International Semantic Web Conference, 2003.

[2] P. C. G. da Costa, K. B. Laskey, K. J. Laskey, "PR-OWL: A Bayesian Ontology Language for the Semantic Web", ISWC 2005 Workshop Uncertainty Reasoning for the Semantic Web, 2005, pp.23-33.

[3] P. J.F. Lucas, L. C. van der Gaag, and A. Abu-Hanna, "Bayesian networks in biomedicine and health-care", Artificial Intelligence in Medicine 30, 2004, pp. 201–214.

[4] Yi Yang and Jacques Calmet, "OntoBayes: An Ontology-Driven Uncertainty Model", In Proceeding of the International Conference on Intelligent Agents, Web Technologies and Internet Commerce (IAWTIC'05), 2005, pp. 457-463.

[5] Z. Ding, Y. Peng, and R. Pan, "BayesOWL: Uncertainty Modeling in Semantic Web Ontologies", In Soft Computing

in Ontologies and Semantic Web, Springer-Verlag, December 2005.

[6] Jena – A semantic web Framework for java, http://jena.sourceforge.net/

[7] Protégé, ontology editor, http://protege.stanford.edu/

[8] Netica, a Bayesian network development software, http://www.norsys.com/

[9] Disease Ontology, http://diseaseontology.sourceforge.net/

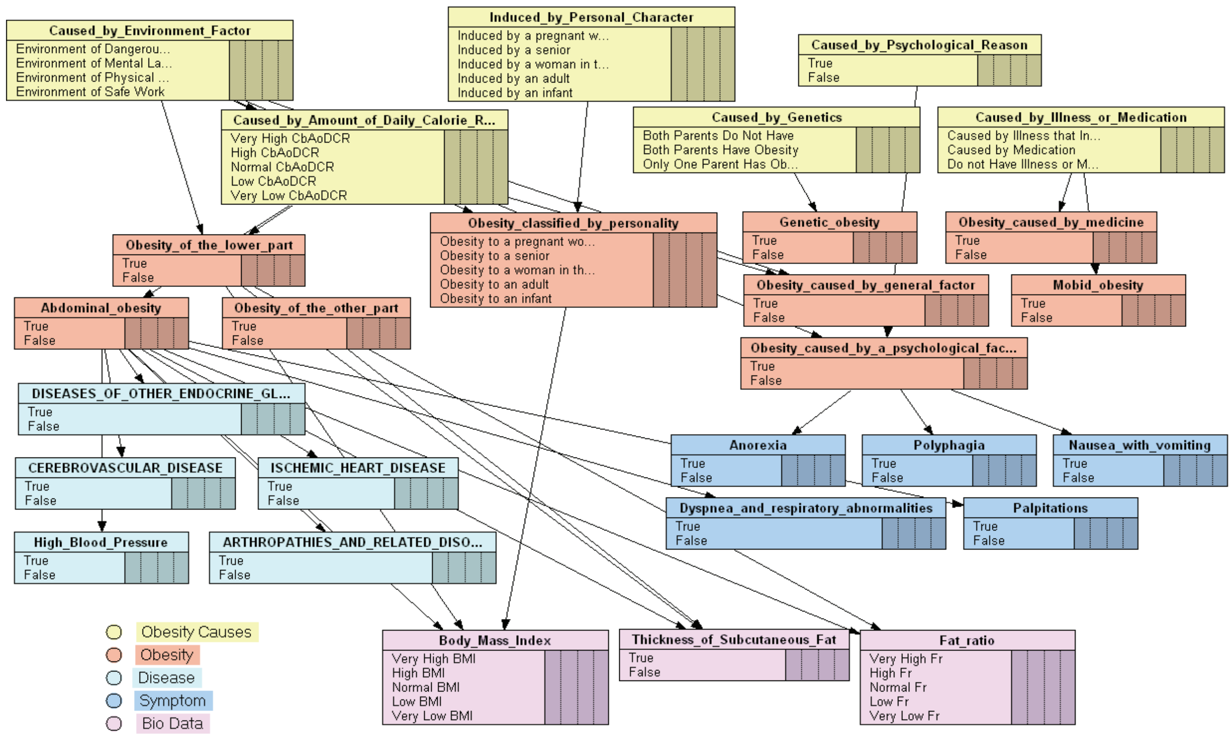

Figure 4. A Bayesian network model for diagnosing obesity