

A Survey of Ontologies on Disease Classification

Sowkarthikaa .K¹, Sumathi V. P.²

¹M.E – Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore-641049, India

²Assistant Professor, Department Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore-641049, India

Abstract: *In medical ontology, it is often difficult to establish such definitions for diseases classification. This paper addresses the problem of classifying the disease based on medical ontology. Our goal is to provide a method for improving the classification of medical ontology which will allow us to generate a new representation based on concepts. There are many techniques used in medical ontology they are case profile ontology, decision support tool, particle swarm optimization model, etc. To evaluate the effectiveness of the intelligent system, three benchmark medical data sets, viz., Breast Cancer Wisconsin, Pima Indians Diabetes, and Liver Disorders from the UCI Repository of Machine Learning, are used for evaluation. A number of useful performance metrics in medical applications which include accuracy, sensitivity, specificity are computed. The results are analyzed and compared with those from other methods published in the literature. The experimental outcomes positively demonstrate that the hybrid intelligent system is effective in undertaking medical data classification tasks.*

Keywords: Ontology, Breast cancer Wisconsin, Pima Indian Diabetes.

1. Introduction

Ontology defines classes of entities and their interrelations. They are used to organize data according to a theory of the domain. Ontology provides class definitions (i.e., the necessary and sufficient conditions for defining class membership). The semantic web is a major research initiative of the world website consortium (w3c) to create a metadata-rich web of resources that can describe themselves not only by how they should be displayed in html or syntactically by xml, But also by the meaning of the meta data. The semantic web is an extension of the current web in which information is given well defined meaning, better enabling computer and people to work in cooperation. According to the definition, ontology is characterized by the following features:

- **It is formal explicit-** The content of ontology is described using formal mechanisms and mathematical logics.
- **It specifies a conceptualization-** The ontology is used to model some part of the universe and it is responsible for a specific view of the appropriate domain, reflected in certain simplification, abstraction, omissions are other modeling decisions.
- **It is shared-** Ontology reflects a common understanding of the modeled domain, as a consequence of a consensus within a community of an ontology user.

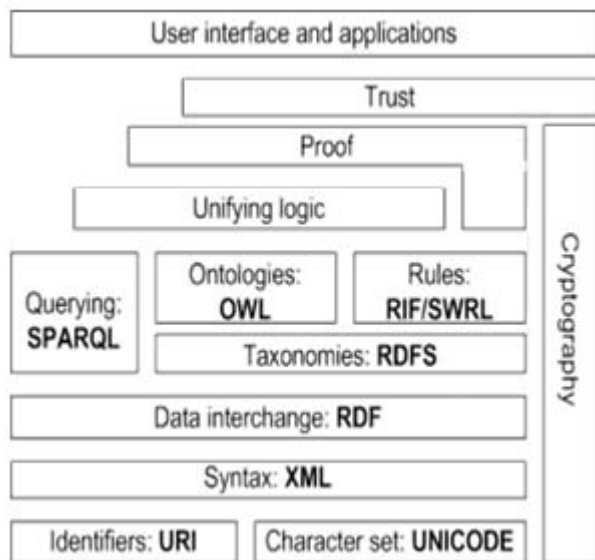
The ongoing rapid growth of the amount of available data and their use in a wide range of domains to solve different tasks requires more sophisticated techniques of intelligent and knowledge management. These efforts have resulted in a significant number of data integration and management systems and tools enriched with Vocabularies, Thesauris, Terminologies and Ontologies. Together with semantic web, the buzzword ontology has been come extremely popular since last few decades. Unfortunately, there are no universally accepted definitions for ontology.

2. Semantic Web Architecture

The first layer, URI and Unicode, follows the important features of the existing WWW. Unicode is a standard of

encoding international character sets and it allows that all human languages can be used on the web using one standardized form. Uniform Resource Identifier is a string of a standardized form that allows to uniquely identify resources A subset of URI is Uniform Resource Locator, which contains access mechanism and a location of a document. Another subset of URI is URN that allows to identify a resource without implying its location and means of dereferencing it. The usage of URI is important for a distributed internet system as it provides understandable identification of all resources. An international variant to URI is Internationalized Resource Identifier that allows usage of Unicode characters in identifier and for which a mapping to URI is defined. In the rest of this text, whenever URI is used, IRI can be used as well as a more general concept.

Extensible Markup Language layer with XML namespace and XML schema definitions makes sure that there is a common syntax used in the semantic web. XML is a general purpose markup language for documents containing structured information. A XML document contains elements that can be nested and that may have attributes and content. XML namespaces allow to specify different markup vocabularies in one XML document XML schema serves for expressing schema of a particular set of XML documents. All data in the semantic web use RDF as the primary representation language. To allow standardized description of taxonomies and other ontological constructs, a RDF Schema (RDFS) was created together with its formal semantics within RDF.



Semantic web architecture

A core data representation format for semantic web is Resource Description Framework (RDF). RDF is a framework for representing information about resources in a graph form. It was primarily intended for representing metadata about WWW resources, such as the title, author, and modification date of a Web page, but it can be used for storing any other data. It is based on triples *subject-predicate-object* that form graph of data. RDFS can be used to describe taxonomies of classes and properties and use them to create lightweight ontologies.

More ontologies can be created with Web Ontology Language_OWL. The OWL is a language derived from description logics, and offers more constructs over RDFS. It is syntactically embedded into RDF, OWL comes in three species - OWL Lite for taxonomies and simple constrains, OWL DL for full description logic support, and OWL Full for maximum expressiveness and syntactic freedom of RDF. Since OWL is based on description logic, it is not surprising that a formal semantics is defined for this language.

RDFS and OWL have semantics defined and this semantics can be used for reasoning within ontologies and knowledge bases described using these languages. To provide rules beyond the constructs available from these languages, rule languages are being standardized for the semantic web as well. Two standards are emerging - RIF and SWRL. For querying RDF data as well as RDFS and OWL ontologies with knowledge bases, a Simple Protocol and RDF Query Language (SPARQL) is available. SPARQL is SQL-like language, but uses RDF triples and resources for both matching part of the query and for returning results of the query. Since both RDFS and OWL are built on RDF, SPARQL can be used for querying ontologies and knowledge bases directly as well.

3. Literature Review

In this paper, David Riano has introduced an ontology for the care of chronically ill patients and implement two personalization process and a decision support tool. the first personalization process adapts the contents of the ontology of

the particularities observed in the health care record of a particular patients. Personalised ontology will automatically contains only the clinical information that is relevant for health care professionals to manage the patients. The second personalization process uses the personalize ontology of a patients to automatically transform intervention plans describing general health care treatments into individual intervention plans. The case profile ontology is an ontology for the management of chronically ill patients at home that integrates the general medical knowledge that is specific for the kind of patients. The decision support tool to help physicians to change treatments when it was required.

Alireza Rahimi validated an automated ontology-based semantic query of collected data from Electronic Health Records to diagnose and analyze the patients with Type 2 Diabetes Mellitus. The accuracy and a direct relationship to data quality metrics has established and demonstrated. The ontology based approach reduced the collision of the incompleteness of the data set.

In medical area, researchers have tried to use different methods to improve the accuracy of data classification. Pei-Chann Chang in his research has developed by integrating a case based reasoning approach and particle swarm optimization model. the researchers are interested in finding potential weights of disease factors. Here GA is applied to find the weights of each factor in order to derive the best weights of each attribute.

In this study Cynthiya Brandt present novel features engineering that encoded in the Unified Medical Language System to improve machine learning based clinical text classification. Critical steps in clinical text classification include identification of features relevant to the classification task. They have developed novel information theoretical techniques that utilize the taxonomical structure of the Unified Medical System to improve the feature ranking and developed semantic similarities of a projects clinical text into a features space to improve classification. This method has improved the performance of other machine learning based clinical text classification system.

In this paper Siaw-Teng-Liaw has examined the accuracy of a Type 2 Diabetes Mellitus phenotyping algorithm and its ontology infrastructure in finding cases of Type 2 Diabetes Mellitus in individual Electronic health Records's and a multi- Electronic health record data repository. Our clinical informatics program enables the reuse of the knowledge already represented in SNOMED CT-AU to perform semantic retrievals for different applications and clinical domains. The goal to create knowledge driven models to integrate disparate datasets and knowledge bases to support the integrated care of patients with multiple chronic diseases and on multiple medications has been partially achieved. The ontologically driven approach can improve the accuracy of Electronic health Records-based disease registers and potentially lead to improved and coordinated chronic disease management, patient safety, and quality outcomes.

The researcher is more interested in finding potential disease factors. In this researchers Chin-Yuan-Fan took a different approach by applying a case-based fuzzy decision tree to

diagnose the illness symptoms. A stepwise regression (SRA) method is applied to select the most important factors from the set of inputs. A weighted clustering method is adopted to divide the case base into a smaller case. In each case, a more homogeneous data are grouped into together. Therefore, these data react to the illness symptom detection. And at last GA is applied to find the fuzzy terms of each factor in order to derive the best fuzzy decision tree from each case.

In this paper, a hybrid intelligent system that consists of the Fuzzy Min–Max neural network, the Classification and Regression Tree, and the Random Forest model has proposed, and a decision support tool for medical data classification is examined. It is able to learn from data samples explains its predicate output and achieve high classification performance. To evaluate the effectiveness of the hybrid intelligent system, three benchmark medical data sets, viz., Breast Cancer Wisconsin, Pima Indians Diabetes, and Liver Disorders from the UCI Repository of Machine Learning, are used for evaluation. A number of useful performance metrics in medical applications which include accuracy, sensitivity, specificity, and the area under the Receiver Operating Characteristic curve are computed. The results are analyzed and compared with those from other methods. The experimental results demonstrate that the hybrid intelligent system is effective in undertaking medical data classification tasks. The hybrid intelligent system is able to produce good results and its knowledge base with a decision tree. As a result, domain users are able to understand the prediction given by the hybrid intelligent system; hence accepting its role as a useful medical decision support tool.

4. Conclusion

In this survey shows, many issues that ontology researchers in semantic integration to hold similarities to the issues that database and information-integration researchers have been addressing. Some of the approaches are also similar although the ontology community relies more on the higher expressive power of ontology languages and on reasoning techniques. With ontologies, using a common upper ontology or reference ontology to make easier the integration problem is also a common approach. The emerging Semantic Web can prove to be an excellent test bed for scalability of various approaches and a common ground for experimenting with hybrid approaches. This premises of the Semantic Web is that the use of machine-interpretable ontologies defined in formal languages agreed to reasoning will provide the next generation of services. On the other hand, the scale of the Semantic Web will certainly require well-tested approaches from the database community.

References

- [1] David Riano , Francis Real , Joan Albert Lopez-Vallverdu , Fabio Campana , Sara Ercolani , Patrizia Mecocci , Roberta Annicchiarico , Carlo Caltagirone ,” An ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients”, (2012), pp 429–446.
- [2] Alireza Rahimi , Siaw-Teng Liaw, Jane Taggart, Pradeep Ray, Hairong Yu “Validating an ontology-based

- algorithm to identify patients with Type 2 Diabetes Mellitus in Electronic Health Records”, 2014, pp 768–778.
- [3] Vijay N. Garla, Cynthia Brandt, “Ontology- guided feature engineering for clinical text classification”, (2012), PP 992–998.
- [4] Manjeevan Seera , Chee Peng Lim,” A hybrid intelligent system for medical data classification”, (2014), PP 2239–2249.
- [5] Chin-Yuan Fana, Pei-Chann Chang, Jyun-Jie Lin, J.C. Hsieh,”A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification”, (2011), PP 632–644.
- [6] Siaw-Teng Liaw, Jane Taggart, Hairong Yu , Simon de Lusignan , Craig Kuziemsy , Andrew Hayen, ”Integrating electronic health record information to support integrated care: Practical application of ontologies to improve the accuracy of diabetes disease registers”, (2014), PP 364–372.
- [7] Pei-Chann Chang, Jyun-Jie Lin, Chen-Hao Liu, “An attribute weight assignment and particle swarm optimization algorithm for medical database classifications”, (2 0 1 2), PP 382–392.