ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 4, Issue 3 , March 2017

A Literature Survey on Ontology based semantic aware Bayesian Network model

Ramya R, K Pramilarani, Sheba Pari N

P.G. Student, Department of CSE, New Horizon College Of Engineering, Bangalore, India
Senior Assistant Professor, Department of CSE, New Horizon College Of Engineering, Bangalore, India
Assistant Professor, Department of CSE, New Horizon College Of Engineering, Bangalore, India

ABSTRACT: Semantic Data Mining refers to the data mining tasks that systematically incorporate domain knowledge, especially formal semantics, into the process. In the past, many research efforts have attested the benefits of incorporating domain knowledge in data mining. At the same time, the proliferation of knowledge engineering has enriched the family of domain knowledge, especially formal semantics and Semantic Web ontologies. Ontology is an explicit specification of conceptualization and a formal way to define the semantics of knowledge and data. The formal structure of ontology makes it a nature way to encode domain knowledge for the data mining use. In this survey paper, we introduce general concepts of semantic data mining. We investigate why ontology has the potential to help semantic data mining and how formal semantics in ontologies can be incorporated into the data mining process. We provide detail discussions for the advances and state of art of ontology-based approaches and an introduction of approaches that are based on other form of knowledge representations.

KEYWORDS: Ontology, Bayesian network, Data mining

I. INTRODUCTION

Data mining, otherwise called knowledge discovery from database (KDD), is the methodology of nontrivial extraction of verifiable, beforehand obscure, and possibly valuable data from information. With the recent advances in data mining strategies lead to numerous momentous upsets in information investigation and big data. Data mining likewise joins methods from insights, computerized reasoning, machine learning, database framework, and numerous different controls to examine substantial information sets. Semantic Data Mining alludes to data mining errands that deliberately fuse space learning, particularly formal semantics, into the methodology. Past semantic data mining exploration has witness to the positive impact of space learning on data mining. Amid the seeking and pattern generating procedure, area learning can function as an arrangement of former information of requirements to help decrease hunt space and aide the inquiry way

A large volume of web contents is available in unstructured and semi-structured format. This includes the contents available on many online ad portals such as craigslist, eBay, gum tree, etc. Despite providing a reliable and affordable service to a large customer/fan base, these portals contain user-generated posts (data) written in unstructured and ungrammatical format. Thus, standard query languages cannot be used to retrieve relevant information posted on these portals. As a result, users have to rely on key-word-based searches which do not necessarily retrieve the most relevant and accurate information.

Previous semantic data mining research has attested the positive influence of domain knowledge on data mining. For ex-ample, the pre-processing can benefit from domain knowledge that can help filter out the redundant or inconsistent data. During the searching and pattern generating process, domain knowledge can work as a set of prior knowledge of constraints to help reduce search space and guide the search path. Furthermore, the discovered patterns can be cleaned out or made more visible by encoding them in the formal structure of knowledge engineering.

To make use of domain knowledge in the data mining process, the first step must account for representing and building the knowledge by models that the computer can further access and process. The proliferation of knowledge engineering (KE) has remarkably enriched the family of domain knowledge with techniques that build and use domain knowledge

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 4, Issue 3 , March 2017

in a formal way. Ontology is one of successful knowledge engineering advances, which is the explicit specification of a conceptualization. Normally, an ontology is developed to specify a particular domain (e.g., genetics). Such an ontology, often known as a domain ontology, formally specifies the concepts and relationships in that domain. The encoded formal semantics in ontologies is primarily used for effectively sharing and reusing of knowledge and data. Prominent examples of domain ontologies include the Gene Ontology, Unified Medical Language System, and more than 300 ontologies in the National Center for Biomedical Ontology

Research in the area of the Semantic Web has led to quite mature standards for modeling and codifying domain knowledge. Today, Semantic Web ontologies become a key technology for intelligent knowledge processing, providing a framework for sharing conceptual models about a domain. The Web Ontology Language which has emerged as the de facto standard for defining Semantic Web ontologies, is widely used for this purpose. The Semantic Web technologies that formally represent domain knowledge including structured collection of prior information, inference rules, knowledge enriched datasets etc., could thus develop frameworks for systematic incorporation of domain knowledge in an intelligent data mining environment.

In this survey paper, we study the advances and state of art of semantic data mining. We specifically focus on the ontology-based approaches. The ontology-based approaches for semantic data mining attempt to make use of formal ontologies in the data mining process. This is generally achieved by using the formal definition of concepts and relationship.

II. TYPES OF ONTOLOGIES

A variety of different types of ontologies are considered in the literature. These types may be characterized according to their granularity, formality, generality and computational capability. In terms of granularity, an ontology can be defined as either coarse-grained or fine-grained. Coarse-grained ontologies facilitate the conceptualization of a domain at the macro-level, and are typically represented in a language of minimal expressivity. Fine-grained ontologies, on the other hand, allow the conceptualization of a domain at the micro-level, and tend to be represented in a language of significant expressivity. In terms of formality, ontologies may be classified as being highly informal, semi informal, semi-formal or rigorously formal. At one end of the formality spectrum, highly informal ontologies are expressed in natural language. At the other end of the spectrum, rigorously formal ontologies are defined in a language with a formal semantics and with desirable computational properties such as soundness and completeness. In term of generality, ontologies may be classified as being top-level ontologies, mid-level ontologies, task ontologies, domain ontologies and application ontologies.

• Top-level ontologies (also called upper ontologies or foundational ontologies) are high level, domain-independent ontologies.
• Mid-level ontologies (also called utility ontologies) serve as a bridge between top-level ontologies and domain ontologies; theyserve a purpose analogous to that of software libraries in the object-oriented programming paradigm.
• Domain ontologies specify concepts and inter-concept relations particular to a domain of interest.
• Task ontologies are ontologies developed for specific tasks.
• Application ontologies are ontologies used in specific applications. They typically utilize both domain and task ontologies.

III. PERFORMANCE EVALUATION AND APPLICATIONS

As a formal specification of domain concepts and relation-ships, ontology can assist in the data mining process in various perspectives. It is reasonable to expect a performance gain in ontology-based approaches compared with the data mining approaches without using ontologies or other form of domain knowledge. Many semantic data mining research efforts have attested such improvements. With well-designed algorithms, previous research either reports performance improvement or accomplishment of data mining tasks that could not be achieved without using ontologies. In this section, we give a brief summarization of the performance improvement in ontology-based approaches and their applications.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 4, Issue 3 , March 2017

A. PERFORMANCE GAIN IN PRECISION, RECALL, AND CONSISTENCY OF DATA MINING RESULTS

Many previous ontology-based efforts have reported performance gain in the data mining results. Ontology-based approaches have been reported to have better precision and recall than the traditional approaches in various tasks such as text clustering information extraction link prediction and recommendation systems.

Research in recommendation system suggests that ontology-based recommendation systems have better prediction precision than traditional recommendation methods .With the enriched semantics and reduced search space, execution speed gain has been reported in the gene clustering task from microarray experiments with ontology-based clustering .In the web usage mining and next page prediction task, semantics aware sequential pattern mining algorithms was proved to perform 4 times faster than regular and non-semantics-aware algorithms .Ontology-based approaches improve the consistency of data mining results as well.

B. SEMANTICS RICH DATA MINING RESULTS

Ontology can also assist in enriching data mining results with formal semantics. Semantics rich data mining results are expected from ontology-based approaches compared with approaches without using ontologies. For example, OBIE is able to extract the information with similar or close semantics that does not directly appear in the data source.
Without knowing semantics of the attributes or itemsets, association rule mining usually generate too many rules or even inconsistent rules. Ontology-based association rule mining bridges the semantic gap of the domain knowledge and the association rule mining algorithm. It results in better collection and representation of association rules by pruning the results or reducing the search space. The top ranked rules also result in high support measure for the targeting domain.

C. PERFORMING DATA MINING TASK THAT IS UNACHIEVABLE WITH TRADITIONAL DATA MINING METHODS

Certain data mining tasks that are not achievable by traditional data mining methods can be accomplished by ontology-based approaches. For example, traditional classification task usually requires at least reasonable amount of labeled data as prior knowledge. Using ontology as the specification of prior knowledge, classification task without enough labelled data is proved to have a comparable performance compared with traditional classification methods. Using the ontology as a conceptual consistency constraint, the model with unlabeled data can be tuned into the one that have the best consistency with the prior knowledge (i.e., ontology). Classification task without labeled or annotated data is also reported in the ontology-based text classification task.

IV. DIFFERENTAPPROACHES IN SEMANTIC DATA MINING

A.AUTOMATING KNOWLEDGE DISCOVERY WORKFLOWCOMPOSITION THROUGH ONTOLOGY-BASED PLANNING

The methodology consists of two main ingredients. The first on is a formal conceptualization of knowledge types and algorithms implemented in the KD ontology following up on state-of-the-art developments of a unified data mining theory, experiments and to enable reasoning on the results to facilitate reuse of workflows and results. Manual annotation of algorithms, which require extending the core ontology with new knowledge classes Can be time consuming and requires expertise in semantic modeling. However, more often algorithms working with already defined knowledge classes are added.

B.ONTOLOGY BASED TEXT-MINING

Clustering Based on Similarities Using Text Mining After the classification is done according to the keywords Text mining technique is used to cluster the papers in each domain. The five steps are performed to cluster the project

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 4, Issue 3 , March 2017

papers. Which are collections of 1. Text document, 2. Text document processing, 3. Encoding of text document, 4.vector dimension reduction and 5. Vector clustering. Self-organized mapping (SOM) algorithm is used cluster the new proposals.
A methodology ontology is constructed to categorize the concept terms in different domain and to from relationship among them. It facilitates text-mining and optimization technique to cluster papers based on their similarities and then to balance them according to the size of domain.

C. ONTOLOGY BASED LINK PREDICTION

In this methodology the information are initially expounded by controlled vocabulary terms from ontologies. The annotation connections between the information and predicates in ontology font an annotation chart. They proposed heuristics for edge weighting that depend by implication on the semantics of substance and property sorts in the ontology and on qualities of the occurrence information. The showcase p-diagram era calculation was proposed to concentrate a little association sub graph from the data chart.

D. ONTOLOGY BASED RECOMMENDATION SYSTEM

Recommender frameworks or proposal frameworks are the frameworks that devote to anticipate the inclination or evaluations that a client would provide for athing. Suggestion frameworks have ended up greatly famous lately and been connected in an assortment of utilizations including films, music, news, books, examination articles, seek questions, and social labels. In a decent proposal framework, heterogeneous data from different sources is typically needed. Ontology can incorporate the utilization of heterogeneous data and aide the proposal inclination.

V. CONCLUSION

The advances in learning building and data mining advance semantic data mining, which conveys rich semantics to all phases of data mining methodology. Numerous examination endeavours have validated the playing point of joining space learning into data mining. Formal semantics encoded in the ontology is all around organized which is simple for the machine to peruse and process accordingly make it a nature approach to utilize ontologies in semantic data mining. Utilizing ontologies, semantic data mining has points of interest to extension semantic holes between the information, applications, data mining calculations, and datamining results, give the data mining calculation with former learning which either controls the mining process or decreases the inquiry space, and to give a formal approach to speaking to the data mining stream, from information preprocessing to mining results.

To handle and control the enormous information have brought serious examination up in the data mining group. With the improvement of learning designing, particularly Semantic Web methods, mining expansive sum, semantics rich, and heterogeneous information rises as an imperative examination theme in the group. Numerous specialists have brought up, work along semantic data mining is still in its initial stage.

Ontology based semantic data mining is by all accounts one of most encouraging methodologies. The real testis to grow more programmed semantic data mining calculations and frameworks by using the full quality of formal ontology that has very much characterized representation dialect, formal semantics, and thinking instruments for rationale surmising and consistency checking.

REFERENCES

[1] B. Berendt, A. Hotho, D. Mladenic, M. van Someren, M. Spiliopoulou, G. Stumme, A roadmap for web mining: from web to Semantic Web, Web Mining: FromWeb to SemanticWeb (2004) 1–22.
[2] B. Berendt, A. Hotho, G. Stumme, Towards Semantic Web mining, in: Proceedings of the First International Semantic Web Conference on The Semantic Web, Springer-Verlag, 2002, pp. 264–278.
[3] A. Tjoa, A. Andjomshoaa, F. Shayeganfar, R. Wagner, Semantic Web challenges and new requirements, in: Sixteenth International Workshop on Database and Expert Systems Applications, 2005, pp. 1160–1163.
[4] M. Wilson, B. Matthews, The Semantic Web: prospects and challenges, in: 7th International Baltic Conference on Databases and Information Systems, 2006, pp. 26–29.
[5] P. Mika, Social Networks and the Semantic Web, Springer, 2007.
[6] G. Antoniou, F.V. Harmelen, A Semantic Web Primer, MIT Press, 2004.
[7] Q.N. Rajput, S. Haider, N. Touheed, Information extraction from unstructured and ungrammatical data sources for semantic annotation,

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 4, Issue 3 , March 2017

International Journal of Information Technology 5 (3) (2009) 189–197.
[8] Q.N. Rajput, S. Haider, Use of Bayesian network in information extraction from unstructured data sources, International Journal of Information Technology 5 (4) (2009) 207–213.
[9] G. Fiumara, Automated information extraction from Web sources: a survey, in: Proceedings of the Workshop between Ontologies and Folksonomies (BOF), Michigan, USA, 2007.
[10] A.H.F. Laender, B.A. Ribeiro-Neto, A.S.D. Silva, J.S. Teixeira, A brief survey of web data extraction tools, Sigmod Record 31 (2002) 84–93.
[11] L. Reeve, H. Han, Survey of semantic annotation platforms, in: Proceedings of the 2005 ACM Symposium on Applied Computing, Santa Fe, New Mexico, pp. 1634–1638.
[12] J. Tang, J. Li, H. Lu, B. Liang, X. Huang, K. Wang, iASA: learning to annotate the Semantic Web, Journal on Data Semantics IV (2005) 110–145.
[13] M. Motta, E. Motta, J. Domingue, M. Lanzoni, F. Ciravegna, MnM: ontology driven semi-automatic and automatic support for semantic markup, in: Gomez-Perez (Ed.), The 13th International Conference on Knowledge Engineering and Management (EKAW), Springer-Verlag, 2002, pp. 379–391.
[14] B. Popov, A. Kiryakov, D. Ognyanoff, D. Manov, A. Kirilov, KIM-a semantic platform for information extraction and retrieval, Natural Language Engineering 10 (2004) 375–392.
[15] D.W. Embley, D.M. Campbell, Y.S. Jiang, S.W. Liddle, D.W. Lonsdale, Y. Ng, R.D. Smith, Conceptual-model-based data extraction from multiple-record Web pages, Data Knowledge Engineering 31 (1999) 227–251.
[16] Y. Ding, D. Embley, S. Liddle, Automatic creation and simplified querying of Semantic Web content: an approach based on information-extraction ontologies, in: The Semantic Web – ASWC 2006, 2006, pp. 400–414.
[17] D.W. Embley, C. Tao, S.W. Liddle, Automating the extraction of data from HTML tables with unknown structure, Data and Knowledge Engineering. 54 (2005) 3–28.
[18] M. Michelson, C.A. Knoblock, Creating relational data from unstructured and ungrammatical data sources, Journal of Artificial Intelligence Research 31 (2008) 543–590.