Semantic Web Mining

Vidushi Singh
Galgotias College of
Engineering &
Technology
vidushisingh.singh@g
mail.com

Shaurya Pratap Singh
Galgotias College of
Engineering &
Technology
shauryasingh.singh@g
mail.com

Shubham Singh
Galgotias College of
Engineering &
Technology
shubham2102@gmail.c
om

Tej Prakash Gautam
Galgotias College of
Engineering &
Technology

Abstract:- Integrating domain knowledge is one of the most perplexing complications in data mining. The Semantic Web technologies are promising to offer solutions to formally capture and efficiently use the domain knowledge.This paper offers the incorporation of semantic information drawn from a web application's domain knowledge into all segments of the web convention mining process like pre-processing, pattern discovery, and prediction). The goal is to have an intellectual semantics-aware web usage mining framework. This is accomplished by using semantic information in the sequential pattern mining algorithm to crop the search space and partially release the algorithm from sustenancetotalling. Low order Markov model is used in the prediction phase which will help solve vague difficulties. Semantic Web tools and (semi-)automatically extract knowledge hidden in the vast amounts of Web data using Web Mining tools.

Keywords: Semantic web, web mining, unstructured data.

Introduction

The most challenging problem in data mining is integration of domain knowledge. It is important to explore the blend of them when both data and domain knowledge are available. The informal manner of the representation hinders proficient usage of the codified knowledge. The management of unstructured data is recognized as one of the most critical unsolved problems in data management now days. The major reason or this unresolved problem is primarily because of the fact that the methods, systems and related tools that work so successfully on structured information simply fail when they are implemented on unstructured information. Discoveryof new methods and approaches is the need of the hour.

A huge amount of information is shared by the organizations across the world over the Web. This information exchange across the globe has emphasized the need of opening a lot of new avenues to create tools for data management and business. In this paper, we explore the challenges being faced by information system developers during mining of unstructured data in the relation to semantic web and web mining. Opportunities to handle these challenges need to be developed as early as possible.

Semantic Web

The present WWW has aenormousquantity of statistics that is often amorphous andusually only human comprehensible. The Semantic Web objects to discourse this delinquent by providing machine interpretable semantics to deliver superior machine sustenance for the user. The Semantic Web has a layer structure that defines the levels of abstraction pragmatic to the Web. At the nethermost level is the familiar World Wide Web, then continuing to XML, RDF, Ontology, Logic, Proof and Trust. The main tackles that are presently being used in the Semantic Web are ontologies based on OWL (Web Ontology Language) and its subordinate reasoners.

Web Mining

Web Mining is the application of data mining techniques to the content, structure and usage of Web resources.

The three main areas of Web Mining are:
– Content Mining - Analyses the content of Web resources. Mainly based on text mining techniques, but postponements

to multimedia content is commencement to emerge in the research.

– Structure Mining - Scrutinizes the hyperlink edifice between Web pages.

– Usage Mining - Investigates the user's snaps from Web server logs.

Semantic Web Mining

The semantic web is based on the visualization of Tim Berners-Lee , the inventor of the World-Wide-Web (WWW). According to him, "The semantic web is not at all visualized as a separate web but it is an expansion of the existing one, in which information is given well-defined sense and significance, better enabling PCs and people to work in cooperation."

Current Scenario of Data Mining Over Web

A structured approach is being followed by data administrators and managers across the world when we talk about data mining over web. Below are the major fields which are being explored in terms of finding knowledge from un-structured data:

1 .Data mining with a focus towards mining unstructured data

2. A lot of unstructured data is noisy text

3. Spontaneous communication (such as e-mails, discussion forums, SMS, blogs, and collaborative Web portals) contains noisy text and processing noise. Vast proportion of this unstructured data contains informal and semi-formal, internal and external Communications of a given organization. Usually humans can understand such text straightaway. However, with enormous quantity of such data content being available nowadays, both online and inside the enterprise, it becomes critical to mine such text using computers as it becomes very difficult and complex for a human being to mine huge data manually.

We can think of using available data mining generalized models to represent unstructured dataalso but with very less efficiency and proper outcome. There are a few algorithms availabletoextract useful information from unstructured

data or from noisy text but a generalized, rugged approach is still missing.

Conclusion and Future Work

In this paper, we introduce semantic data mining as an area we look upon emergingas the solution to systematic incorporation of domain knowledge in data mining with the help of the Semantic Web technologies. The large amount of information stored in (semi-)structured sources is calling for attention to develop pioneering approaches to solve following challenges: 1) how to impart knowledge encoded in ontologies into the (semi-) structured data, and 2) exploration of more meaningful ways to utilize the knowledge.Two fast developing research areas, Semantic Web and Web Mining, are summarised above. The combined area of Semantic Web Mining offers new techniques to improve both areas. Semantics can improve the results of Web Mining by taking advantage of structures in the Web. Web Mining can improve the Semantic Web by finding new semantic structures to enrich the semantics. The application of each area to the other creates a feedback loop, where the goal of Semantic Web Mining is realised.

References

1. Stumme, G., Hotho, A., Berendt, and B.: Semantic Web Mining: State of the art andfuture directions. Web Semantics: Science, Services and Agents on the World Wide Web 4(2) (2006) 124 – 143 Semantic Grid –The Convergence of Technologies.

2. Berendt, B., Hotho, A., Mladenic, D., van Someren, M., Spiliopoulou, M., Stumme, G.: A Roadmap for Web Mining: From Web to Semantic Web. Web Mining: From-Web to Semantic Web Volume 3209/2004 (2004) 1–22

3. Liu, B.: Web Data Mining - Exploring Hyperlinks, Contents, and Usage Data. Springer Berlin Heidelberg (2007)

4. H. Alani and C. Brewster. Ontology ranking based on the analysis of concept structures.In K-CAP '05: Proceedings of the 3rd international conference on Knowledge capture, pages 51–58, New York, NY, USA, 2005. ACM.

5. L. Ding, T. Finin, A. Joshi, Y. Peng, R. Pan, and P. Reddivari. Search on the Semantic web. IEEE Computer, 10, 2005.

6. L. Ding, R. Pan, T. Finin, A. Joshi, Y. Peng, and P. Kolari. Finding and ranking knowledge on the semantic web. In In Proceedings of the 4th International Semantic Web Conference, pages 156–170, 2005.