Abstract

This paper shows a new method based on association rule mining and ontology for the classification of web pages. This work is pruning of association rules, generated by mining process. The main complexity arises due to the fact that there are various number of text documents that are considered for generating the association rules using the A-priori algorithm. But these rules that were generated are not based on the semantic knowledge. In order to obtain the most accurate rules we gone for the construction of the ontology, based on the domain knowledge. With this domain knowledge we design various operators which are helpful in reducing the rules generated. Thus the various rules that we get are semantically correct with regards to the domain selected. We use the high confidence value based classifier for classifying the given text document to that particular domain. Association rules are mined from this matrix using A-priori algorithm. Based on the high confidence value, a new text document is classified into one of the predefined classes. In general, from association rule mining, a huge amount of association rules are mined. All the association rules generated may not be useful for the classification purpose. So, In order to reduce the irrelevant association rules, we need semantic knowledge. For this purpose, propose new domain specific ontology to overcome this drawback of association rule mining method.
References

- Abe, H., Tsumoto, S., "Detection of trends of technical phrases in text mining", IEEE International Conference on Granular Computing, GRC '09, Pages: 7 – 12, 2009
- Pak Chung Wong, Whitney, P., Thomas, J., "Visualizing association rules for text mining", Proceedings on Information Visualization (Info Vis &apos;99), Pages: 120 - 123, 152, 1999
- Ping Chen, Rakesh Verma, Janet C. Meininger, Herman Pressler Dr. Houston, "Semantic Analysis of Association Rules", Association for the Advancement of Artificial Intelligence
- FABRIZIO SEBASTIANI, "Machine Learning in Automated Text Categorization", Consiglio Nazionale delle Ricerche, Italy
- Alaa Al Deen, Mustafa Nofal, Sulieeman Bani-Ahmad, "Classification Based On Association-Rule Mining Techniques: A General Survey And Empirical Comparative Evaluation".
- Inhauma Neves Ferraz, Ana Cristina Bicharra Garcia, "Ontology In Association Rules Pre-Processing And Post-Processing", Pages-87-91, IADIS European Conference Data Mining, 2008

Index Terms

Computer Science
Artificial Intelligence
Keywords
Data Mining  Association Rule  Web Data  Web Mining  Classification  Ontology