WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

The Problem on Determining "How to Incorporate Knowledge into Information Mining Algorithms"

Nithin.U[1], Likith.U[1], Zafar Ali Khan N[2]

[1]UG Student,Department of Computer Science and Engineering, Presidency University, Itgalpur, Rajanakunte, Bengaluru,
[2]Associate professor, Program head-CSE, Department of Computer Science and Engineering, Presidency University, Itgalpur, Rajanakunte, Bengaluru, Karnataka, 560064, India,
DOI: https://doi.org/10.55248/gengpi.2022.3.6.25

ABSTRACT:

There is a massive amount of data on the Internet right now, but essentially little classification information. The main issue is figuring out how to incorporate knowledge into information mining algorithms. The authors examine information retrieval approaches and discuss their advantages and disadvantages. Although most Web documents are text oriented, there are plenty of them that contain multimedia elements, which are not easily accessible through common search methods. Web information is dynamic, semi-structured, and interwound with hyperlinks. Several advanced methods for Web information mining are analysed: 1) syntax analysis, 2) metadata-based searching using RDF, 3) knowledge annotation by use of conceptual graphs (CGs), 4) KPS: Keyword, Pattern, Sample search techniques, and 5) techniques of obtaining descriptions by fuzzification and back-propagation. The problem of choosing proper keywords is

Introduction:

Users require more sophisticated tools to identify information that is important to them as more documents with multimedia features become available on the Internet. As a result, a variety of language technologies are being used in a variety of information management applications (ref. 3): multilingual search engines, Machine Translation (MT), video access systems, content-based language technologies for information systems, document summarization, robust text processing for document content abstraction, E-commerce, document/database relevancy visualisation, Web browser instrumentation, information extraction, human-computer dialogue systems, and so on. The main goal of this paper is to look at the issues of embedding knowledge into information mining algorithms and to develop appropriate information retrieval approaches. A vast number of multimedia documents and lexical databases have been developed in a single mark-up language.

Syntax analysisLanguage structure investigation Full-text search is most likely the most popular technique, performing string-coordinating (for example utilizing customary articulations) and design matching quests (for example labels, connect names and connection ways) in reports. Web indexes like Altavista, Infoseek, Excite and so forth, for example Web robots build file of watchwords found in records attempting to catch in that way the substance of reports. Questions can be refined by utilizing rationale administrators (AND, OR, NOT). The general benefit is to be quick because of computerized ordering, where no human mediation is required. Drawback is that answers are in many cases unimportant, fragmented or the quantity of results might be exceptionally enormous and, hence, not usable. One method for further developing data recovery is to utilize information portrayal language (KRL) to file Web archives. One of them is RDF - Resource Description Framework, worked over XML which is more machine-comprehensible than intelligible configuration. One more method for facilitating portrayal is to utilize set of natural and combinable orders identical to first-arrange rationale, like Conceptual Graphs (CGs) for ordering any Web data.

Fuzzification

It's undeniably true that data recovery of the ideal data from the Web can be a tedious cycle. The principal reason is the unfortunate grouping of the Web data. Obviously, there are a lot of web search tools which use extraordinary robots in look for new Web pages, and when a page is found it is placed in the 'right' grouping classification relying upon the order technique the web search tool is utilizing. Then again, metadata order can be added to Web objects. This implies that the undertaking of order is to some extent moved to the people who make and keep up with Web components. As a high-level answer for Web arrangement M. Marchiori (ref. 6) proposes the fuzzification strategy. He says that current Web metadata sets really do have credits relegated to objects, yet they either have them or don't have them. All things considered, he contends that ascribes ought to be fuzzified, for example each characteristic ought to be related with a "fluffy proportion of its significance for the Web object, to be specific a number going from 0 to 1". This intends that assuming a characteristic is allotted esteem 0 it isn't appropriate to the reporter Web object. On the off chance that the worth is 0.4 importance of the property to the Web object is 40%. Since grouping without help from anyone else is an estimate, better of more terrible, fuzzification technique permits adaptability inside a predefined

9 arrangement framework giving more nitty gritty positioning and permitting the essential arrangement of ideas to be generally little.

Back-propagation

Investigates a haphazardly picked district of Web objects demonstrated the way that the use of back-proliferation technique can essentially further develop viability of the order. They additionally showed that the minimum amount of Web metadata characterization convenience is accomplished when something like 16% of the Web use metadata order, conversely, with half without fuse of the back-spread technique. Moreover, to accomplish greatness level metadata need 53% of the Web to be arranged, interestingly, with 80% without depicted technique. In particular, the strategy for back-spread, which surmise the fuzzification technique, follows up on top of any arrangement, and requires no type of semantic examination. In this way, it is totally language free which is vital when the quantity of non-English Web pages is continually expanding.

Conclusion:

Because of the lack of fully organised Web pages, which are dynamic, semi-structured, and incorporate multimedia information, current information retrieval techniques are unable to provide precise results. The current tendency is to allow users to utilise metadata languages to annotate documents. Because the issue of accurate indexing and subjective classification is so crucial, it is advised that a generally recognised classification system be used. Manual classification of all Web documents is impossible. Annotating texts with metadata languages such as XML or CGs is becoming more popular as an intermediate step. Conceptual Graphs have a basic, intuitive notation that may be used to annotate any web material (including images) and leaves certain concepts undefined. Users can then represent and query the documents at their desired level. The Webb ontology is proposed to aid in the creation of a Web-accessible knowledge store system. Despite this.

References :

1. Bindra, S. Pokuri, K. Uppala and A. Teredesai, "Distributed Big Advertiser Data Mining," 2012 IEEE 12th International Conference on Data Mining Workshops, 2012, pp. 914-914, doi: 10.1109/ICDMW.2012.73.
2. A. K. Tiwari, G. Ramakrishna, L. K. Sharma and S. K. Kashyap, "Neural Network and Genetic Algorithm based Hybrid Data Mining Algorithm (Hybrid Data Mining Algorithm)," 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2019, pp. 95-99, doi: 10.1109/ICCCIS48478.2019.8974485.
3. Z. Dong, "Research of Big Data Information Mining and Analysis : Technology Based on Hadoop Technology," 2022 International Conference on Big Data, Information and Computer Network (BDICN), 2022, pp. 173-176, doi: 10.1109/BDICN55575.2022.00041.
4. A. J. Chamatkar and P. K. Butey, "Implementation of Different Data Mining Algorithms with Neural Network," 2015 International Conference on Computing Communication Control and Automation, 2015, pp. 374-378, doi: 10.1109/ICCUBEA.2015.78.
5. D. Khan, "CAKE – Classifying, Associating and Knowledge DiscovEry - An Approach for Distributed Data Mining (DDM) Using PArallel Data Mining Agents (PADMAs)," 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008, pp. 596-601, doi: 10.1109/WIIAT.2008.236.
6. J. S. Challa, P. Goyal, S. Nikhil, A. Mangla, S. S. Balasubramaniam and N. Goyal, "DD-Rtree: A dynamic distributed data structure for efficient data distribution among cluster nodes for spatial data mining algorithms," 2016 IEEE International Conference on Big Data (Big Data), 2016, pp. 27-36, doi: 10.1109/BigData.2016.7840586.
7. L. Li, "Research and Application of Data Mining Classification Algorithm Based on Multi-classifier Fusion," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2022, pp. 651-654, doi: 10.1109/ICSCDS53736.2022.9761013.
8. K. K. Pandey and D. Shukla, "Mining on Relationships in Big Data era using Improve Apriori Algorithm with MapReduce Approach," 2018 International Conference on Advanced Computation and Telecommunication (ICACAT), 2018, pp. 1-5, doi: 10.1109/ICACAT.2018.8933674.
9. M. A. Qadeer, N. Akhtar and F. Khan, "Comparison of Tools for Data Mining and Retrieval in High Volume Data Stream," 2009 Second International Workshop on Knowledge Discovery and Data Mining, 2009, pp. 252-255, doi: 10.1109/WKDD.2009.176.
10. P. V. Subba Reddy, "Fuzzy MapReduce Data Mining algorithms," 2018 International Conference on Fuzzy Theory and Its Applications (iFUZZY), 2018, pp. 304-310, doi: 10.1109/iFUZZY.2018.8751692.
11. P. K. Kotturu and A. Kumar, "Data Mining Visualization with the Impact of Nature Inspired Algorithms in Big Data," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), 2020, pp. 664-668, doi: 10.1109/ICOEI48184.2020.9142979.
12. M. Zaffar, M. A. Hashmani and K. S. Savita, "Performance analysis of feature selection algorithm for educational data mining," 2017 IEEE Conference on Big Data and Analytics (ICBDA), 2017, pp. 7-12, doi: 10.1109/ICBDAA.2017.8284099.
13. S. Sharma, A. K. Sharma and D. Soni, "Enhancing DBSCAN algorithm for data mining," 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), 2017, pp. 1634-1638, doi: 10.1109/ICECDS.2017.8389724.
14. Anoopkumar M and A. M. J. M. Z. Rahman, "A Review on Data Mining techniques and factors used in Educational Data Mining to predict student amelioration," 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), 2016, pp. 122-133, doi: 10.1109/SAPIENCE.2016.7684113.
15. K. R. Thakre and R. Shende, "Implementation on an approach for mining of datasets using APRIORI hybrid algorithm," 2017 International Conference on Trends in Electronics and Informatics (ICEI), 2017, pp. 939-943, doi: 10.1109/ICOEI.2017.8300845.

16. R. M. Pai and V. S. Ananthanarayana, "A novel data structure for efficient representation of large data sets in data mining," 2006 International Conference on Advanced Computing and Communications, 2006, pp. 547-552, doi: 10.1109/ADCOM.2006.4289952.
17. A. Sheshasaayee and D. Sridevi, "Fuzzy C-means algorithm with gravitational search algorithm in spatial data mining," 2016 International Conference on Inventive Computation Technologies (ICICT), 2016, pp. 1-5, doi: 10.1109/INVENTIVE.2016.7823259.
18. Y. V. Bodyanskiy, A. O. Deineko, Y. V. Kutsenko and O. O. Zayika, "Data streams fast EM-fuzzy clustering based on Kohonen's self-learning," 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP), 2016, pp. 309-313, doi: 10.1109/DSMP.2016.7583565.