Abstract

Web usage mining aims to discover interesting user access patterns from web logs. Web usage mining has become very critical for effective web site management, creating adaptive web sites, business and support services, personalization and so on. In this paper, an efficient approach for frequent pattern mining using web logs for web usage mining is proposed and this approach is called as HFPA. In this approach HFPA, the proposed technique is applied to mine association rules from web logs using normal Apriori algorithm, but with few adaptations for improving the interestingness of the rules produced and for applicability for web usage mining. This technique is applied and its performance is compared with that of classical Apriori-mined rules. The results indicate that the proposed approach HFPA not only generates far fewer rules than Apriori-based algorithms (FPA), the generated rules are also of comparable quality with respect to three objective performance measures, Confidence, Lift and Conviction. Association mining often produces large collections of association rules that are difficult to understand and put into action. In this paper effective pruning techniques are proposed that are characterized by the natural web link structures. Experiments showed that interestingness measures can successfully be used to sort the discovered association rules after the pruning method was applied. Most of the rules that ranked highly according to the interestingness measures proved to be truly valuable to a web site administrator.
References

- Tianyi Wu, Yuguo Chen, and Jiawei Han, "Association Mining in Large Databases: A Re-Examination of Its Measures", In Proceedings of PKDD-2007, 11th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), Warsaw, Poland, September 17-21, 2007, pp 621-628.
- J. Pei, J. Han, B. Mortazavi-Asl, and H. Zhu, "Mining access patterns efficiently from web logs", in PADTK (aka:00: Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Current Issues and New Applications. London, UK: Springer-Verlag, 2000, pp. 396-407
- P. Batista, M. ario, and J. Silva, "Mining web access logs of an on-line newspaper", 2002
Index Terms

Computer Science  Information Sciences

Keywords

Web Usage Mining  Web Logs  Web Personalization  Association Rules  Interestingness Measures