CMAR: Accurate and efficient classification based on multiple class-association rules

Li W., Han J., Pei J.

School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

Abstract: Previous studies propose that associative classification has high classification accuracy and strong flexibility at handling unstructured data. However, it still suffers from the huge set of mined rules and sometimes biased classification or overfitting since the classification is based on only single high-confidence rule. In this study, we propose a new associative classification method, CMAR, i.e., Classification based on Multiple Association Rules. The method extends an efficient frequent pattern mining method, FP-growth, constructs a class distribution-associated FP-tree, and mines large database efficiently. Moreover, it applies a CR-tree structure to store and retrieve mined association rules efficiently, and prunes rules effectively based on confidence, correlation and database coverage. The classification is performed based on a weighted $\chi 2$ analysis using multiple strong association rules. Our extensive experiments on 26 databases from UCI machine learning database repository show that CMAR is consistent, highly effective at classification of various kinds of databases and has better average classification accuracy in comparison with CBA and C4.5. Moreover, our performance study shows that the method is highly efficient and scalable in comparison with other reported associative classification methods. © 2001 IEEE.

Year: 2001
Source title: Proceedings - IEEE International Conference on Data Mining, ICDM
Page : 369-376
Cited by: 275
Link: Scorpus Link
Correspondence Address: Li, W.; School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada; email: wli@cs.sfu.ca
Document Type: Conference Paper
Source: Scopus
Authors with affiliations:

1. Li, W., School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
2. Han, J., School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
3. Pei, J., School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

References:

1. Agrawal, R., Srikant, R., Fast algorithms for mining association rules VLDB'94, Chile, Sept. 1994
2. Clark, P., Niblett, T., The CN2 induction algorithm (1989) Machine Learning, 3, pp. 261-283
3. Dong, G., Zhang, X., Wong, L., Li, J., Caep: Classification by aggregating emerging patterns (1999) LNCS, 1721. , DS'99 Japan, Dec
4. Duda, R., Hart, P., (1973) Pattern Classification and Scene Analysis, , John Wiley & Sons
5. Han, J., Pei, J., Yin, Y., Mining frequent patterns without candidate generation SIGMOD'OO, Dallas, TX, May 2000

6. Lent, B., Swami, A., Widom, J., Clustering association rules ICDE'97, England, April 1997
7. Li, W., (2001) Classification Based on Multiple Association Rules, , M.Sc. Thesis, Simon Fraser University, April
8. Lim, T.-S., Loh, W.-Y., Shih, Y.-S., A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms (2000) Machine Learning, 39
9. Liu, B., Hsu, W., Ma, Y., Integrating classification and association rule mining KDD'98, New York, NY, Aug. 1998
10. Quinlan, J.R., (1993) C4.5: Programs Forkfachine Learning, , Morgan Kaufmann
11. Wang, K., Zhou, S., He, Y., Growing decision tree on support-less association rules KDD'00, Boston, MA, Aug. 2000