

Associative Classification

By

Dr. T. Sree Sharmila,
Associate Professor, Dept. of IT,
SSN College of Engineering



Associative Classification

- Associative classification
 - Association rules are generated and analyzed for use in classification
 - Search for strong associations between frequent patterns (conjunctions of attribute-value pairs) and class labels
 - Classification: Based on evaluating a set of rules in the form of

$$P_1 \wedge p_2 \dots \wedge p_l \rightarrow "A_{\text{class}} = C" (\text{conf}, \text{sup})$$

- Why effective?
 - It explores highly confident associations among multiple attributes and may overcome some constraints introduced by decision-tree induction, which considers only one attribute at a time
 - In many studies, associative classification has been found to be more accurate than some traditional classification methods, such as C4.5



Typical Associative Classification Methods

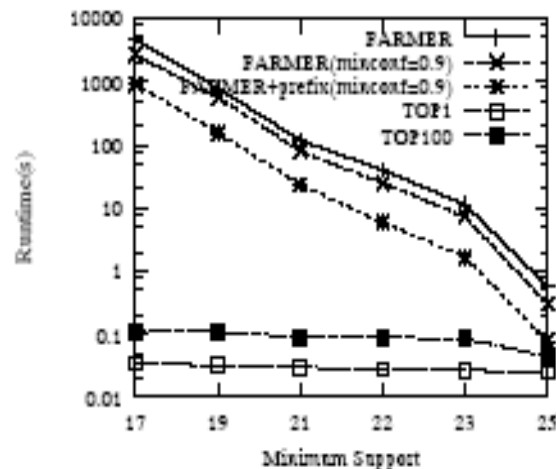
- CBA (Classification By Association: Liu, Hsu & Ma, KDD'98)
 - Mine association possible rules in the form of
 - Cond-set (a set of attribute-value pairs) → class label
 - Build classifier: Organize rules according to decreasing precedence based on confidence and then support
- CMAR (Classification based on Multiple Association Rules: Li, Han, Pei, ICDM'01)
 - Classification: Statistical analysis on multiple rules
- CPAR (Classification based on Predictive Association Rules: Yin & Han, SDM'03)
 - Generation of predictive rules (FOIL-like analysis)
 - High efficiency, accuracy similar to CMAR
- RCBT (Mining top- k covering rule groups for gene expression data, Cong et al. SIGMOD'05)
 - Explore high-dimensional classification, using top- k rule groups
 - Achieve high classification accuracy and high run-time efficiency



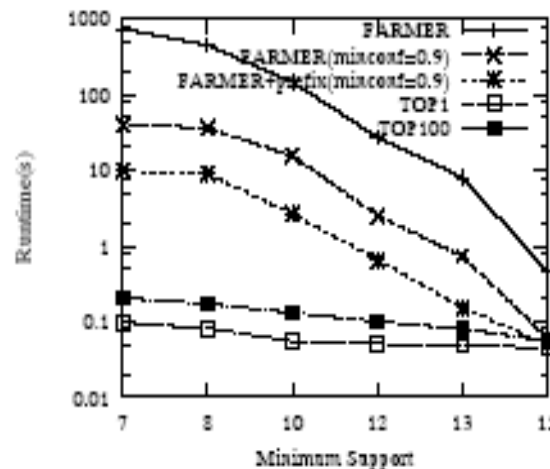
Associative Classification May Achieve High Accuracy and Efficiency (Cong et al. SIGMOD05)

Dataset	RCBT	CBA	IRG Classifier	C4.5 family			SVM
				single tree	bagging	boosting	
AML/ALL (ALL)	91.18%	91.18%	64.71%	91.18%	91.18%	91.18%	97.06%
Lung Cancer(LC)	97.99%	81.88%	89.93%	81.88%	96.64%	81.88%	96.64%
Ovarian Cancer(OC)	97.67%	93.02%	-	97.67%	97.67%	97.67%	97.67%
Prostate Cancer(PC)	97.06%	82.35%	88.24%	26.47%	26.47%	26.47%	79.41%
Average Accuracy	95.98%	87.11%	80.96%	74.3%	77.99%	74.3%	92.70%

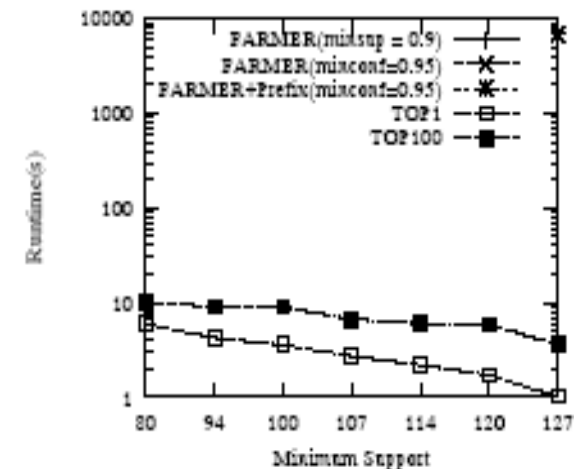
Table 2: Classification Results



(a) ALL-AML leukemia



(b) Lung Cancer



(c) Ovarian Cancer

Summary

- Studied different types of associative classification methods