Book Title	Encyclopedia of Machine Learning	
Book CopyRight - Year	2010	
Title	Apriori Algorithm	
Author	Particle	
	Given Name	**Hannu**
	Family Name	**Toivonen**
	Suffix	
	Email	hannu.toivonen@cs.helsinki.fi
Affiliation	Division	Department of Computer Science
	Organization	University of Helsinki
	Street	P.O. Box 68 (Gustaf Hallstromin katu 2b)
	Postcode	FI-00014
	City	Helsinki
	Country	Finland

A

Apriori Algorithm

Hannu Toivonen
Department of Computer Science,
University of Helsinki,
Helsinki, Finland

Definition

Apriori algorithm (Agrawal, Mannila, Srikant, Toivonen, & Verkamo, 1996) is a ►data mining method which outputs all ►frequent itemsets and ►association rules from given data.

Input: set $\mathcal{I}$ of items, multiset $\mathcal{D}$ of subsets of $\mathcal{I}$, frequency threshold *min_fr*, and confidence threshold *min_conf*.

Output: all frequent itemsets and all valid association rules in $\mathcal{D}$.

Method:

1: level := 1; frequent_sets := $\emptyset$;
2: candidate_sets := $\{\{i\} \mid i \in \mathcal{I}\}$;
3: while candidate_sets $\neq \emptyset$
3.1: scan data $\mathcal{D}$ to compute frequencies of all sets in candidate_sets;
3.2: frequent_sets := frequent_sets $\cup$ $\{C \in$ candidate_sets $\mid$ frequency$(C) \geq$ *min_fr*$\}$;
3.3 level := level + 1;
3.4: candidate_sets := $\{A \subset \mathcal{I} \mid |A| =$ level and $B \in$ frequent_sets for all $B \subset A, |B| =$ level $- 1\}$;
4: output frequent_sets;
5: for each $F \in$ frequent_sets
5.1: for each $E \subset F, E \neq \emptyset, E \neq F$
5.1.1: if frequency(F)/frequency$(E) \geq$ *min_conf* then output association rule $E \rightarrow (F \setminus E)$

The algorithm finds frequent itemsets (lines 1-4) by a breadth-first, general-to-specific search. It generates and tests candidate itemsets in batches, to reduce the overhead of database access. The search starts with the most general itemset patterns, the singletons, as candidate patterns (line 2). The algorithm then iteratively computes the frequencies of candidates (line 3.1) and saves those that are frequent (line 3.2). The crux of the algorithm is in the candidate generation (line 3.4): on the next level, those itemsets are pruned that have an infrequent subset. Obviously, such itemsets cannot be frequent. This allows Apriori to find all frequent itemset without spending too much time on infrequent itemsets. See ►frequent pattern and ►constraint-based mining for more details and extensions.

Finally, the algorithm tests all frequent association rules and outputs those that are also confident (lines 5-5.1.1).

Cross References

►Association Rule
►Basket Analysis
►Breadth-First Search
►Constraint-Based Mining
►Frequent Itemset
►Frequent Pattern

Recommended Reading

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. I. (1996). Fast discovery of association rules. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy (Eds.), *Advances in knowledge discovery and data mining* (pp. 307–s328). Menlo Park: AAAI Press.

C. Sammut, G. Webb (eds.), *Encyclopedia of Machine Learning*, DOI 10.1007/978-0-387-30768-8, © Springer-Verlag Berlin Heidelberg, 2010