Multiple uses of frequent sets and condensed representations
Extended abstract

Heikki Mannila and Hannu Toivonen
University of Helsinki
Department of Computer Science
P.O. Box 26, FIN-00014 University of Helsinki, Finland
Heikki.Mannila@cs.Helsinki.FI, Hannu.Toivonen@cs.Helsinki.FI

Abstract

In interactive data mining it is advantageous to have condensed representations of data that can be used to efficiently answer different queries. In this paper we show how frequent sets can be used as a condensed representation for answering various types of queries.

Given a table r with 0/1 values and a threshold σ, a frequent set of r is a set X of columns of r such that at least a fraction σ of the rows of r have a 1 in all the columns of X. Finding frequent sets is a first step in finding association rules, and there exists several efficient algorithms for finding the frequent sets. We show that frequent sets have wider applications than just finding association rules.

We show that using the inclusion-exclusion principle one can obtain approximate confidences of arbitrary boolean rules. We derive bounds for the errors in the confidences, and show that information collected during the computation of frequent sets can also be used to provide individual error bounds for each clause. Experiments show that this method enables one to obtain different forms of rules from data extremely fast. Furthermore, we define a general notion of condensed representations, and show that frequent sets, samples and the data cube can be viewed as instantations of this concept.

Introduction

Knowledge discovery in databases is an interactive process: one has to be able to look at the data from several different angles. For large data sets, it is not efficient to go back and read the data every time the user wants to see something new from it. A condenced representation of the data for answering most of the needs of the user would be very useful.

One of the most researched areas in data mining is the problem of finding association rules from binary data (Agrawal, Imielinski, & Swami 1993; Houtsma & Swami 1993; Klemettinen *et al.* 1994; Han & Fu 1995; Holsheimer *et al.* 1995; Park, Chen, & Yu 1995; Savasere, Omiecinski, & Navathe 1995; Srikant & Agrawal 1995; Agrawal *et al.* 1996; Srikant & Agrawal 1996; Toivonen 1996). Assuming we have a relation (table) with 0/1-valued attributes R, an association rule is an expression $X \Rightarrow Y$, where $X, Y \subseteq R$. The meaning of such a rule is that whenever a row has a 1 in all the columns of X, it tends to have a 1 also in all the columns of Y.[1]

The algorithms for finding association rules all proceed by first finding frequent sets, i.e., sets X of attributes such that there are sufficiently many rows containing a 1 in each column of X. From the number of such rows it is easy to compute the confidences of association rules.

In this paper we consider additional uses for frequent sets. We show that the collection of frequent sets can actually serve as a condensed representation of the input data for rules with disjunction and negation, a much larger class of rules than simple association rules. These simple observations lead to a considerable widening in the use of frequent sets. Experiments show that this method enables one to obtain generalized rules from data extremely fast.

We analyze formally the use of frequent sets in answering queries concerning complex rules. Our basic tool is the inclusion-exclusion principle. The frequent sets provide some of the terms of this sum. Using the concept of the border of the collection of frequent sets, we are able to derive bounds on the error caused by omitting the rest of the terms.

Finally, we consider the general theory of condensed representations of data. We define what it means for one class of structures to serve as a representation for another with respect to a class of queries. We dis-

[1] Note that the input data need not be explicitly in 0/1 form: one can just as well consider association rules based on derived attributes. For example, if the relation contains the attribute Age, one can look for rules containing derived binary attributes such as "Age ≤ 40".

cuss how the use of frequent sets, samples and the data cube (Gray *et al.* 1996) can be viewed as instances of this general concept. Our concept can be viewed as a generalization of the notion of ε-approximations from computational geometry (Haussler & Welzl 1987; Mulmuley 1993).

General rules and frequent sets

In this section we describe a simple and basically well-known rule formalism of general boolean rules, show that knowledge of the frequencies of sets is sufficient to determine rule confidences, and give some examples.

Given a set R of 0/1-valued attributes R, a *boolean formula* over R is a formula φ built from the atomic formulae $A = 0$ and $A = 1$, where $A \in R$, using the connectives $\wedge$, $\vee$, and $\neg$, and parenthesis. Given a 0/1 relation r with attributes R, the *frequency* $s(\varphi, r)$ of φ in r is

$$\frac{|\{t \in r \mid \varphi \text{ is true of } t\}|}{|r|},$$

i.e., the fraction of rows where φ holds.

We introduce some shorthand notations for having *all* or *at least one* attribute of a set X equal to 1: For $X = \{A_1, \ldots, A_k\}$ we denote $a(X, r) = s(A_1 = 1 \wedge \cdots \wedge A_k = 1)$ and $o(X, r) = s(A_1 = 1 \vee \cdots \vee A_k = 1)$; we simply use notations $a(X)$ and $o(X)$ if the relation is clear from the context.

A *boolean rule* over R is an expression of the form $\varphi \Rightarrow \psi$, where φ and ψ are boolean formulae. The *confidence* of the rule is $s(\varphi \wedge \psi)/s(\varphi)$.

The following proposition is of course well-known.

Proposition 1 For all boolean formulae φ over R there exists sets $W_1, \ldots, W_m \subseteq R$ and coefficients $e_1, \ldots, e_m \in \{-1, +1\}$ such that for all relations r

$$s(\varphi, r) = \sum_{i=1}^{m} e_i a(W_i, r).$$

□

That is, knowing the special terms $a(X, r)$ is sufficient to determine the number of rows satisfying any boolean formula, and hence also the confidence of any boolean rule.

Example 2 We give some examples of boolean rules and how their confidences can be expressed in terms of the frequencies of conjunctions.

An *association rule* has the form $X \Rightarrow Y$, where $X, Y \subseteq R$, and the confidence is defined to be $a(X \cup Y)/a(X)$.

A rule with *disjunctive right-hand side* has the form $X \Rightarrow Y \vee Z$ and expresses that if a row $t \in r$ has a 1 in each column of X, then it has a 1 in each column of Y or in each column of Z. The confidence of the rule is

$$\frac{a(X \cup Y) + a(X \cup Z) - a(X \cup Y \cup Z)}{a(X)}.$$

Similarly, we can write a rule with a *disjunctive left-hand side*: $X \vee Y \Rightarrow Z$. The rule states that if a row $t \in r$ has a 1 in each column of X or in each column of Y, then it has a 1 in each column of Z. The confidence of the rule is

$$\frac{a(X \cup Z) + a(Y \cup Z) - a(X \cup Y \cup Z)}{a(X) + a(Y) - a(X \cup Y)}.$$

Rules can also have *negation in the left-hand side*, as in $X \wedge \neg Y \Rightarrow Z$. This means that if a row $t \in r$ has a 1 in each column of X and does not have a 1 in each column of Y, then it has a 1 in each column of Z. The confidence of such a rule is

$$\frac{a(X \cup Z) - a(X \cup Y \cup Z)}{a(X) - a(X \cup Y)}.$$

Negation in the right-hand side is just as easy. The confidence of a rule of the form $X \Rightarrow Y \wedge \neg Z$ is

$$\frac{a(X \cup Y) - a(X \cup Y \cup Z)}{a(X)}.$$

□

Example 3 One of our example data sets r is an enrollment database of courses in computer science. The data set consists of registration information of 4734 students. There is a row per student, containing the courses the student has registered for. On average, a row has 4 courses. The total number of courses is 127. We first discovered all sets X of courses with $a(X, r) \geq 0.05$. The number of frequent sets is 489. We then computed association rules, rules with disjunctions (two disjuncts) on the left or on the right-hand side, and rules with negations, and then used some simple selection tools to locate interesting rules from these sets.

While already association rules produce lots of interesting and sometimes even surprising information about the data set, looking for more complex rules makes it possible to locate more subtle phenomena. For example, we can notice that the confidences of many rules of the form

$X \Rightarrow$ *C programming* $\vee$ *C and Unix*

have significantly higher confidences than the individual rules

$X \Rightarrow$ *C programming*
$X \Rightarrow$ *C and Unix.*

Inspecting the history of the curriculum, this becomes clear: the courses are alternatives to each other. □

Given a relation r with attributes R and a threshold $\sigma \in [0,1]$, the *σ-frequent sets* of r are the subsets X of R such that $a(X,r) \geq \sigma$; this collection is denoted by $Fr(r,\sigma)$ or just $Fr(r)$, if σ is clear from the context. The algorithms for finding association rules compute the collection $Fr(r,\sigma)$, and the quantity $a(X,r)$ for all frequent sets. The time required is roughly proportional to the product of $|Fr(r,\sigma)|$ and $|r|$. These methods are applicable whenever $|Fr(r,\sigma)|$ is not too large.

We show next how the rule confidences can be bounded using frequent sets. The idea is very simple: to get a lower bound on the confidence, just replace every term $a(W)$ such that $W \notin Fr(r)$ in the above formulae either by 0 or σ, depending on whether $a(W)$ occurs positively or negatively within the formula and whether it occurs in numerator or denominator.

For example, to find the confidence of the disjunctive rule $X \Rightarrow Y \vee Z$, we need to know $a(X)$ and $a(X \cup Y) + a(X \cup Z) - a(X \cup Y \cup Z)$. If all these terms are known, we can directly compute the frequency and confidence of the rule. What if some terms are not known? If either $a(X \cup Y)$ or $a(X \cup Z)$ is less than the frequency threshold, then the sum of the unknown terms in the numerator is less than the threshold, and its effect on the confidence can not be very strong. An interesting case is when $a(X \cup Y \cup Z)$ is not known. Then we can simply assume the worst, which is that $a(X \cup Y \cup Z)$ is just below the threshold, and we get a lower bound of $(a(X \cup Y) + a(X \cup Z) - \sigma)/a(X)$ for the confidence of the rule.

Approximate inclusion-exclusion

In the previous section we gave simple examples of how the confidences of general propositional rules can be estimated on the basis of the information about the frequencies of the frequent sets. Such estimations are, however, fairly useless, unless one is able to give bounds on the errors in the estimation. In this section we show how the information collected during the computation of the frequent sets can be used to derive tight bounds on the quality of the approximate confidences; the bounds depend on the boolean formula, and are hence especially accurate.

For simplicity, consider the following approximation problem. Given $W \subseteq R$, we want to know how many rows of r have a 1 in at least one of the columns of W, i.e., we want to compute $o(W)$. This is of course trivial to do by reading the data once, but this can be very costly. What could be done using only the information collected during the computation of the frequent sets? The inclusion-exclusion formula gives us the expression

$$o(W) = \sum_{\emptyset \neq X \subseteq W} (-1)^{|X|+1} a(X).$$

However, we do not know all terms in the above sum. What we know is the truncated sum

$$\bar{o}(W) = \sum_{\emptyset \neq X \subseteq W, X \in Fr(r)} (-1)^{|X|+1} a(X).$$

The absolute error $e(W)$ caused by using the truncated sum is

$$e(W) = |o(W) - \bar{o}(W)|.$$

We want to bound $e(W)$ somehow.

A similar type of problem has been considered in (Linial & Nisan 1990; Kahn, Linial, & Samorodnitsky 1995), where it is shown that knowing the terms $(-1)^{|X|+1}a(X)$ of the inclusion-exclusion formula for all X with $|X| \leq \sqrt{|W|}$ is sufficient for getting good approximations to the whole sum. Interestingly, the approximation is not formed by simply computing the truncated sum; rather, one uses a linear combination of the terms of the truncated sum. Their results are not, however, applicable to our problem, as the frequent sets do not necessarily provide the required subset of the terms of the inclusion-exclusion formula. See also (Grable 1993; 1994) for related results.

Given r over R, and σ, the *negative border* $Bd^-(r,\sigma)$ of $Fr(r,\sigma)$ consists of sets $W \subseteq R$ not in $Fr(r,\sigma)$ but such that all proper subsets of W are in $Fr(r,\sigma)$. In other words, the negative border consists of the minimal sets that are not frequent. The negative border also defines which sets are frequent and which are not. Given a set X, if X is a superset of some set in the negative border, then X is *not* frequent, otherwise it is frequent.

Example 4 Let $R = \{A,B,C,D\}$. Let r be a relation with attributes R and assume $Fr(r,\sigma) = \{\{A,B\},\{B,C\},\{A,C\},\{A\},\{B\},\{C\},\emptyset\}$. Then $Bd^-(r,\sigma) = \{\{A,B,C\},\{D\}\}$. □

Algorithms that compute the collection of frequent sets also compute the frequencies of sets in the negative border, as the negative border consists of exactly those sets which, on the basis of other information, could be frequent, and whose frequency should therefore be checked. In certain restricted models of computation one can even show that the sets in the border have to be examined (Mannila & Toivonen 1996).

Theorem 5 For all $W \subseteq R$ denote $\mathcal{K} = \{Y \in Bd^-(r) \mid Y \subseteq W\}$. If $\mathcal{K} = \emptyset$, then $e(W) = 0$. Otherwise, let $C = \max\{2^{|\mathcal{K}|-2}/|\mathcal{K}|, 1\}$. Then

$$e(W) \leq C \sum_{Y \subseteq W, Y \in Bd^-(r)} a(Y).$$

Relative error $e(W)/\bar{o}(W)$	Number of cases $\|W\|$			
	2	3	4	5
< 1 %	77	57	20	0
$1 - 5$ %	16	29	48	10
$5 - 10$ %	3	6	10	16
$10 - 20$ %	3	7	13	16
≥ 20 %	1	1	9	58
Total	100	100	100	100
Average error (%)	1.3	2.5	6.2	38.3

Table 1: Distribution of relative error for random disjunctions W of different sizes

Example 6 Continuing Example 4, if we want to approximate the frequency of rows such that at least one attribute has value 1 (i.e., if we want to know $o(\{A, B, C, D\})$, the theorem states that the error is at most $a(\{A, B, C\}) + a(\{D\})$. □

Example 7 We made extensive experiments with a a telecommunications alarm database, where we looked at over 3,300,000 windows of an alarm sequence, and located the frequent sets with threshold 0.0001. There are 128 frequent sets and 713 sets in the negative boundary; only 5 of these, however, have more than 2 elements.

To evaluate the effectiveness of Theorem 5, we generated 100 random disjunctions of size 4 where each disjunct is a frequent alarm, and we used the bound in the theorem to obtain estimates on the size of error. The maximum error bound for the frequency of the disjunction was 0.00041. In 53 cases the absolute error was less than 0.0001, and in 28 cases less than 0.00005.

More interestingly, on average the error bound was only 6.2 % of the estimated size of the disjunction. For 20 cases the error bound was less than 1 % and for another 48 cases less than 5 %; the maximum relative error was 35 %. Table 1 shows the average value and a rough distribution of the relative error for random disjunctions of sizes 2, 3, 4, and 5.

The table shows that the bounds are very accurate for disjunctions of size 2 to 4. In our experiments the bounds were not very useful for disjunctions with 6 or more terms. Experiments with very different alarm databases gave results very similar in nature. Obviously also the frequency threshold has a significant influence on the quality of the bounds: the lower the threshold, the better are the estimates. □

Example 8 In the course enrollment database r we have $|Bd^-(r, 0.05)| = 336$. Theorem 5 gives, e.g., the following results from our database.

- For the set W of courses *Data Structures* and *Database Systems I*, all three terms of the sum are known and $\bar{o}(W) = o(W) = 0.25$ and $e(W) = 0$.
- For a set W of three 'C' or Unix-related courses, 4 out of 7 terms are greater than the threshold and 2 are in the negative border. We have $\bar{o}(W) = 0.24$ and $e(W) \leq 0.017$.
- For a set W of three different intermediate courses, *Theory of Computation*, *Artificial Intelligence*, and *Computer Graphics*, we have 3 out of 7 terms of the sum, and another 3 are in the negative border. The estimated frequency is $\bar{o}(W) = 0.21$ and the error bound $e(W) \leq 0.079$.
- For a W consisting of six programming projects we have $\bar{o}(W) = 0.31$ and $e(W) \leq 0.012$. In this case 7 out of 63 terms exceed the threshold, and another 3 are in the negative border. □

Theorem 5 can also be used to obtain bounds on the error that are independent of the exact frequencies of sets in the border.

Corollary 9 For all $W \subseteq R$ let $K = |Bd^-(r, \sigma) \cap \mathcal{P}(W)|$ and $C = \max\{2^{K-2}, K\}$. Then

$$e(W) \leq C\sigma.$$

Bounds on the error of confidence can be found by assuming opposite errors in the frequencies in the numerator and in the denominator.

Concise representations

In this section we outline a general definition of what it means to represent a class of structures by another, with respect to a set of queries. We discuss how the use of samples, the use of frequent sets, and the data cube (Gray *et al.* 1996) can could be viewed as examples of the same general idea. The concept is based on the notion of ε-approximations widely used in computational geometry and in computational learning theory (Haussler & Welzl 1987; Mulmuley 1993).

We consider a class of structures $\mathcal{S} = \{s_i \mid i \in I\}$, where the index set I can be finite or infinite. Examples of classes of structures are $\mathcal{S}_{R,01}$, the class of all 0/1 relations over the attributes R, $\mathcal{S}_{R,D}$, the class of all relations over the domain D and attributes R, and $\mathcal{S}_E$, the set of all event sequences with event types from the set E.

A query class $\mathcal{Q}$ for $\mathcal{S}$ is a finite set $\{Q_1, \ldots, Q_p\}$ of queries. The value of a query $Q \in \mathcal{Q}$ on a structure $s \in \mathcal{S}$ is denoted by $Q(s)$; the value is assumed to be a real number in the interval $[0, 1]$. Examples of query classes for $\mathcal{S}_{R,01}$ are the *conjunctive* queries $\mathcal{Q}_\wedge = \{Q_W : r \mapsto$

$a(W,r) \mid W \subseteq R\}$ and the *disjunctive* queries $\mathcal{Q}_\vee = \{Q'_W : r \mapsto o(W,r) \mid W \subseteq R\}$.

An ε-*adequate representation* for $\mathcal{S}$ with respect to $\mathcal{Q}$ is a class $\mathcal{R} = \{r_i \mid i \in I\}$ of structures and a query evaluation function $m : \mathcal{Q} \times \mathcal{R} \to [0,1]$ such that for all $Q \in \mathcal{Q}$ and $s_i \in \mathcal{S}$ we have

$$|Q(s_i) - m(Q, r_i)| \leq \varepsilon.$$

That is, the values of queries from $\mathcal{Q}$ on any structure from $\mathcal{S}$ can be evaluated using the corresponding representation from $\mathcal{R}$ and the query evaluation function m.

Example 10 Consider as the class of structures $\mathcal{S}_{R,01}$ the class of all 0/1 relations over the set of attributes R. Consider the query class $\mathcal{Q}_\wedge = \{Q_W \mid W \subseteq R\}$, where for $s \in \mathcal{S}_{R,01}$ we have $Q_W(s) = a(W,s)$, i.e., the fraction of rows of s such that the row has a 1 in each column of W. Then the collection $\{Fr(s,\varepsilon) \mid s \in \mathcal{S}_{R,01}\}$ of the frequent sets of s, for each $s \in \mathcal{S}_{R,01}$, provides an ε-adequate representation of $\mathcal{S}_{R,01}$ with respect to $\mathcal{Q}_\wedge$.

For the telecommunications alarm database we have a 0.0001-adequate representation that consists only of 128 sets. In the case of our course enrollment database r the size of a 0.05-adequate representation is $|Fr(r, 0.05)| = 489$ sets. Note that the number of frequent sets does not depend on the number of rows in the database. □

The notion of an ε-adequate representation is very closely related to the concept of ε-approximations and ε-nets widely used in computational geometry (Haussler & Welzl 1987; Mulmuley 1993). Roughly, given a set N and a collection $\mathcal{S}$ of subsets of N, an ε-approximation is a subset $M \subseteq N$ such that for each $X \in \mathcal{S}$ we have

$$\left|\frac{|X \cap M|}{|M|} - \frac{|X|}{|N|}\right| \leq \varepsilon.$$

Thus an ε-approximation is a subset that gives a good estimate of the sizes of all subsets in $\mathcal{S}$. The ε-adequate representations differ in two respects: the representation does not have to be of the same form as the original data, and we have additionally required that such representations exist for every structure in a class of structures. Still, the results on ε-approximations help in obtaining results about adequate representations.

The existence result for ε-approximations was proved by Vapnik and Chervonenkis, and the sizes of such approximations are naturally connected with the VC-dimension (Vapnik 1982). One can show that samples provide an ε-adequate representation for finite query classes.

The data cube (Gray *et al.* 1996) is a recently introduced summarizing representation for relations with arbitrary values. Using the concepts above, the data cube is 0-adequate representation for the class of queries containing all aggregate functions. An interesting possibility is to diminish the space and time complexity of the data cube by allowing some error. It seems that this can be accomplished by using a similar strategy as in the computation of frequent sets.

The preceding discussion of adequate representations is quite tentative: the usefulness of the notion has yet to be conclusive demonstrated. It seems to us, however, that this concept could serve as a unifying point of view to look at several different types of approximate ways of representing information.

Conclusions

We have shown how the collection of frequent sets can be used as a condensed representation for a relation with 0/1 values. The collection makes it possible to approximate the confidences of arbitrary boolean rules. We have given a strong theorem about the sizes of errors caused by the approximation using the concept of the border of the frequent sets, and have given experimental evidence that the bound is typically extremely good. We have also outlined a possible approach to a general theory of condensed representations and showed how frequent sets, sampling, and also the data cube can be viewed as instances of this concept.

There are several open questions. On the theoretical side, the development of the general notions of condensed representation seems useful. From the practical point of view, a more thorough investigation on the actual sizes of the errors in the approximations could be worthwhile.

References

Agrawal, R.; Mannila, H.; Srikant, R.; Toivonen, H.; and Verkamo, A. I. 1996. Fast discovery of association rules. In Fayyad, U. M.; Piatetsky-Shapiro, G.; Smyth, P.; and Uthurusamy, R., eds., *Advances in Knowledge Discovery and Data Mining*. Menlo Park, CA: AAAI Press. 307 – 328.

Agrawal, R.; Imielinski, T.; and Swami, A. 1993. Mining association rules between sets of items in large databases. In *Proceedings of ACM SIGMOD Conference on Management of Data (SIGMOD'93)*, 207 – 216.

Grable, D. A. 1993. Sharpened bonferroni inequalities. *Journal on Combinatorial Theory, Series B* 57(1):131 – 137.

Grable, D. A. 1994. Hypergraphs and sharpened sieve inequalities. *Discrete Mathematics* 132:75 – 82.

Gray, J.; Bosworth, A.; Layman, A.; and Pirahesh, H. 1996. Data Cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-totals. In *12th International Conference on Data Engineering (ICDE'96)*, 152 – 159.

Han, J., and Fu, Y. 1995. Discovery of multiple-level association rules from large databases. In *Proceedings of the 21st International Conference on Very Large Data Bases (VLDB'95)*, 420 – 431.

Haussler, D., and Welzl, E. 1987. Epsilon-nets and simplex range queries. *Discrete Comput. Geom.* 2:127–151.

Holsheimer, M.; Kersten, M.; Mannila, H.; and Toivonen, H. 1995. A perspective on databases and data mining. In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD'95)*, 150 – 155.

Houtsma, M., and Swami, A. 1993. Set-oriented mining of association rules. Research Report RJ 9567, IBM Almaden Research Center, San Jose, California.

Kahn, J.; Linial, N.; and Samorodnitsky, A. 1995. Inclusion-exclusion: exact and approximate. Manuscript.

Klemettinen, M.; Mannila, H.; Ronkainen, P.; Toivonen, H.; and Verkamo, A. I. 1994. Finding interesting rules from large sets of discovered association rules. In *Proceedings of the Third International Conference on Information and Knowledge Management (CIKM'94)*, 401 – 407. Gaithersburg, MD: ACM.

Linial, N., and Nisan, N. 1990. Approximate inclusion-exclusion. *Combinatorica* 10(4):349 – 365.

Mannila, H., and Toivonen, H. 1996. On an algorithm for finding all interesting sentences. In *Cybernetics and Systems, Volume II, The Thirteenth European Meeting on Cybernetics and Systems Research*, 973 – 978.

Mulmuley, K. 1993. *Computational Geometry: An Introduction Through Randomized Algorithms.* New York: Prentice Hall.

Park, J. S.; Chen, M.-S.; and Yu, P. S. 1995. An effective hash-based algorithm for mining association rules. In *Proceedings of ACM SIGMOD Conference on Management of Data (SIGMOD'95)*, 175 – 186.

Savasere, A.; Omiecinski, E.; and Navathe, S. 1995. An efficient algorithm for mining association rules in large databases. In *Proceedings of the 21st International Conference on Very Large Data Bases (VLDB'95)*, 432 – 444.

Srikant, R., and Agrawal, R. 1995. Mining generalized association rules. In *Proceedings of the 21st International Conference on Very Large Data Bases (VLDB'95)*, 407 – 419.

Srikant, R., and Agrawal, R. 1996. Mining quantitative association rules in large relational tables. In *Proceedings of ACM SIGMOD Conference on Management of Data (SIGMOD'96)*.

Toivonen, H. 1996. Sampling large databases for finding association rules. In *Proceedings of the 22nd International Conference on Very Large Data Bases (VLDB'96)*. To appear.

Vapnik, V. 1982. *Estimation of Dependencies Based on Empirical Data.* New York: Springer-Verlag.