Copyright © 2009 Institute of Electrical and electronics Engineers, Inc.

All Rights reserved.

Personal use of this material, including one hard copy reproduction, is permitted.

Permission to reprint, republish and/or distribute this material in whole or in part for any other purposes must be obtained from the IEEE.

For information on obtaining permission, send an e-mail message to stds-igr@ieee.org.

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

Individual documents posted on this site may carry slightly different copyright restrictions.

For specific document information, check the copyright notice at the beginning of each document.

Finding a Unique Association Rule Mining Algorithm Based on Data Characteristics

Mohammed M. Mazid, A.B.M. Shawkat Ali, Kevin S. Tickle

School of Computing Sciences, Faculty of Business and Informatics

Central Queensland University, Rockhampton, QLD-4702, Australia

Email : m.mazid@cqu.edu.au; s.ali@cqu.edu.au; k.tickle@cqu.edu.au

Abstract - This research compares the performance of three popular Association Rule Mining algorithms, namely Apriori, Predictive Apriori and Tertius based on data characteristics. The accuracy measure is used as the performance measure for ranking the algorithms. A wide variety of Association Rule Mining algorithms can create a time consuming problem for choosing the most suitable one for performing the rule mining task. A meta-learning technique is implemented for a unique selection from a set of association rule mining algorithms. On the basis of experimental results of 15 UCI data sets, this research discovers statistical information based rules to choose a more effective algorithm.

I. Introduction

Association Rule Mining (ARM) is one of the most important and substantial techniques in machine learning areas. It is particularly important for extracting knowledge from large databases by discovering frequent itemsets and associating item relationships between or among items of a data file. ARM is a powerful exploratory technique with a wide range of applications such as marketing policies, medical diagnosis, financial forecast, credit fraud detection and many other research areas. Machine learning researchers have already proposed a number of ARM algorithms, including Apriori [3], Predictive Apriori [19], Tertius [9], CLOSET [17], MAFIA [8], ELACT [25], CHARM[23] and many others. From among these rapidly increasing arrays of ARM algorithms, users easily become confused in choosing the right algorithm for specific kinds of data. The aim of this research is to determine characteristics of datasets using meta-learning processes and provide users a method to choose the right algorithm based on data characteristics. Of course, it is important to keep in mind Wolpert and Macready's well-known No Free Lunch (NFL) theorem:

"If algorithm A outperforms algorithm B on some cost function, then loosely speaking there must exist exactly as many other functions where B outperforms A" [4].

Credit for development of Association Rule Mining is mostly attributed to Agrawal [3]. Association rule is an expression of $X \rightarrow Y$ (read as 'if X then Y'), where X and Y are itemsets in a database D. The expression can be illustrated as 'if a customer buys item X then likely to buy item Y' or 'if a patient is infected by disease X then likely

to be infected by disease Y', and so on. The itemset of the left hand side of the arrow is called antecedent and itemset of the right hand side of the arrow is called consequent. Each expression is called a rule. A rule can contain from two to an unrestricted number of items with or without conjunctions using of AND or OR operands. An item of a rule is selected from the frequent itemset of the data file. Frequent itemsets are the items that occur more frequently. Basically, ARM follows two major steps to produce rules from a data file: first, find all the frequent itemsets; second, generate strong association rules from the frequent itemsets. The best rules are picked the basis of different types of interestingness measurement rules. Some of these will be discussed later on in this paper.

The Apriori [3] algorithm has become the standard approach and most widely used tool for Association Rule Mining [25]. Many ARM algorithms are being used at present but Apriori has become one of the most prevailing algorithms. However a major problem of Apriori is size of datasets and subsequent frequent itemsets while generating rules. Hegland [11] reviewed the Apriori algorithm and discussed variants for distributed data, inclusion of constraints and taxonomies. The author also talked about some potential tools which deal with long itemsets and considerably reduce the number of (uninteresting) itemsets returned. "A frequent pattern of length l implies the presence of $2^l - 2$ additional frequent patterns as well, each of which is explicitly examined by such algorithms" [24]. For a simple example, a datafile with 4 items could produce 14 itemsets. Details of possible itemsets are shown in Figure 1.

Researchers have proposed two ways to solve this problem. The first one is to mine only maximal frequent itemsets [2, 6, 10, 13] which are typically fewer in order of magnitude than all frequent patterns. The second is to mine only the frequent closed sets [5, 16, 21, 26]. "Closed sets are lossless in the sense that they can be used to uniquely determine the set of all frequent itemsets and their exact frequency" [24]. However, this study found that performance of Apriori is still the best among the three algorithms considered.

Fig. 1 Possible itemsets of four items

In this study, we have examined 15 problems from the UCI Repository [7]. The details of the data sets description is provided in Table 1. We evaluated rules produced by three Association Rule Mining algorithms using accuracy measures. A Java based machine learning tool Weka3.4 [22] was used to perform the experiment. For all the three algorithms, the first 10 best rules were chosen. The machine configuration was Intel Core2 Duo CPU 2.33GHz and 4GB RAM.

The rest of the paper is organized as follows: Section 2 describes briefly the three association algorithms. Section 3 describes the different measurements of interestingness we have used to perform the experiment. The experimental process, data preparation and result are presented in Section 4. Finally we draw conclusions from our research in Section 5.

II. Algorithms

This section provides a brief description of all the algorithms we considered in our experimental design. All of the algorithms belong to the category of unsupervised learning methods, but we can further categorise them into Association Rule Mining algorithms as described in the following sections.

A. Apriori

The Apriori algorithm was first proposed by Agrawal [1]. It uses prior knowledge of frequent tools for association rule mining. The basic idea of the Apriori algorithm is to generate frequent itemsets of a given dataset and then scan the dataset to check if their counts are really large. The process is iterative and candidates of any pass are generated by joining frequent itemsets of the proceeding pass. Apriori is a confidence-based Association Rule Mining algorithm. The confidence is simply accuracy to evaluate rules, produced by this algorithm. The rules are ranked according to the confidence value. If two or more rules share the same confidence then they are initially ordered using their support and secondly the time of discovery.

B. Predictive Apriori

In the case of Apriori, every so often we can find rules with higher confidence but low support on respective items of generated rules. Sometimes, rules are produced with large support but low confidence. Scheffer [19] introduced this algorithm with the concept of “larger support has to trade against a higher confidence”. Predictive Apriori is also a confidence-based ARM algorithm. But rules ranked by this algorithm are sorted according to “expected predicted accuracy”. The interestingness measure of Predictive Apriori suits the requirements of a classification task [15]. It tries to maximise expected accuracy of an association rule rather than confidence in Apriori.

C. Tertius

The Tertius system implements a top-down rule discovery system employing the confirmation measure [9]. Tertius uses a first-order logic representation that deals with structured, multi-relational knowledge. It allows users to choose the most convenient option among several possible options. Generated rules from Tertius are descriptive rather than predictive. Descriptive approaches include clustering, association rule learning, learning of attribute dependencies, subgroup discovery, and multiple predicate learning [9]. Tertius is able to deal with extensional knowledge, either with explicit negation or under the Closed-World Assumption. This algorithm also considers induction of integrity constraints in databases, learning mixed theories of predicate definitions and integrity constraints.

III. Description of Algorithm Performance Measures

In the past few years researchers have contributed to finding the interestingness measurement of a rule. There are numbers of measures for this purpose such as Support, Confidence, Predictive Accuracy, Conviction, Lift, Coverage, etc. In our research, we have used the following measurement for the performance of rules.

A. Support

Support for ARM is introduced by Agrawal [1]. It measures the frequency of association, i.e. how many times the specific item has been occurred in a dataset. An itemset with greater support is called frequent or large itemset. In terms of probability theory we can express support as:

$$\text{Support} = P(A \cap B) =$$

$$\frac{\text{number of transactions containing both A and B}}{\text{Total number of transactions}}$$

where A and B are itemsets in a database D.

B. Confidence

Confidence measures the strength of the association [5]. It determines how frequently item B occurs in the transaction that contains A. Confidence expresses the conditional probability of an item. The definition of confidence is

$$\text{Confidence} = P(A | B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{\text{number of transactions containing both A and B}}{\text{number of transactions containing B}}$$

C. Predictive Accuracy

Predictive accuracy is another way to measure interestingness of a generated rule. Basically this accuracy is used for the Predictive Apriori rule measurement. According to Scheffer [19], the definition of predictive accuracy is:

Let D be a data file with r number of records. If $[x \rightarrow y]$ is an Association Rule which is generated by a static process P then the predictive accuracy of $[x \rightarrow y]$ is

$$c([x \rightarrow y]) = Pr[r \text{ satisfies } y | r \text{ satisfies } x]$$

where distribution of r is governed by the static process P and the Predictive Accuracy is the conditional probability of $x \rightarrow r$ and $y \rightarrow r$.

IV. Experiments

Our research compares three algorithms of ARM which are Apriori, Predictive Apriori and Tertius. We organised our research into three main steps: first we compare the

algorithms across a number of different measures of accuracy providing a comprehensive empirical evaluation of the performance of three association algorithms on 15 datasets. We then characterise the datasets using the central tendency measures and some other basic statistical measures. Finally, the empirical results are combined with the dataset characteristic measures to generate rules describing which algorithm is best suited to which type of problems.

A. Data Preparation

After selecting data files without any missing values, we have applied three algorithms of Association Rule Mining that are Apriori, Predictive Apriori and Tertius using WEKA 3.4. For Association Rule Mining, WEKA uses 'unsupervised discretized' value of attributes. We have used 3-bin discretization and the other parameter settings of Discretize option of the software have been kept as default. As well, parameter settings for each algorithm were the default and we always chose the first 10 best rules.

B. Meta Learning Process

Meta learning is one of the sections of Machine learning technique. In this technique automatic learning algorithms are applied on pre processed data to conduct machine learning experiments. In our experiment we converted nominal data to numeric by using Java programming and analysis those data using Matlab [14]. Then a powerful data mining tool See5 [18] had been used to get final rules.

Table1 : Basic Properties of 15 data files

Data file Name	No of Instance	No of Attribute			
		Total	Nominal	Numeric	No. of Class
page-blocks	5473	11	1	10	1
vehicle	846	19	1	18	1
Market Basket	1000	12	3	10	2
liver disorders	345	7	1	6	1
ionosphere	351	35	1	34	1
Zoo	101	18	17	1	1
ecoli	336	8	1	7	1
cmc	1473	10	2	8	X*
breast cancer	286	10	10	0	1
prostoperative patient data	90	9	9	0	1
bridges version1	107	13	10	3	1
Iris	150	5	1	4	1
tae	151	6	3	3	1
haberman	306	4	2	2	1
car	1728	7	7	0	1

* Denotes none

C. Experimental Design

In the first step of our experiment, we have compared average confidence of the 10 best rules of 15 data files. On the basis of confidence, Apriori always showed supremacy compared to the other two algorithms (Predictive Apriori and Tertius). The details of the comparative study are provided in Figure 2- 4. Rules generated from Tertius are worst in terms of confidence comparison. Then we compared between Apriori and Predictive Apriori in terms of average confidence and predictive accuracy of the initially selected 15 data files (The details of the data sets description are provided in Table 1). We have converted Nominal data to numeric values to perform the statistical analysis. We have considered simple, statistical and information theoretic measures to identify the dataset characteristics. The details of statistical characteristics comparisons are provided in Table 2. Some of the statistical formulation is available in MATLAB Statistics Toolbox [14]. By using See5 [18] data mining tools with the most co-related attributes, we generate rules to identify which

algorithm is suitable for which type of problem. Finally, we examined the rules by the statistical measures. The details of the rule descriptions are provided in Table 5.

D. Performance Analysis

In the experiment, we have compared Apriori, Predictive Apriori and Tertius with ‘confidence’ in the first step. Apriori always shows the best performance in this experiment. The details of the ‘confidence’ based performance of 3 algorithms are shown in Figure 2, 3, and 4.

As in the first step of this experiment, Apriori and Predictive Apriori showed better performance. Tertius had the worst among those in terms of ‘confidence’ based comparisons. So we compared Apriori and Predictive Apriori with ‘confidence’ and ‘predictive accuracy’ respectively. Then we found out which algorithm is better between them. The details of the data sets comparisons are provided in Table2.

Figure 2 : Performance of Apriori based on average confidence

1	ionosphere
2	bridges version1
3	Zoo
4	Ecoli
5	postoperative patient data
6	liver disorders
7	vehicle
8	Iris
9	Market Basket
10	tae
11	haberman
12	breast cancer
13	cmc
14	car
15	page-blocks

Figure 3 : Performance of Predictive Apriori based on average confidence

1	page-blocks
2	vehicle
3	Market Basket
4	liver disorders
5	ionosphere
6	Zoo
7	ecoli
8	cmc
9	breast cancer
10	postoperative patient data
11	bridges version1
12	Iris
13	tae
14	haberman
15	car

Figure 4 : Performance of Tertius based on average confidence

1	Zoo
2	Iris
3	ionosphere
4	ecoli
5	Market Basket
6	cmc
7	bridges version1
8	breast cancer
9	liver disorders
10	vehicle
11	postoperative patient data
12	tae
13	car
14	page-blocks
15	haberman

Table 2 : Comparisons of algorithms

Data file Name	Apriori	Predictive Apriori	Better Performance Among two
	Ave. Confidence of 10	Ave. accuracy of 10 rules	
bridges version	0.946	0.861495	Apriori
Car	1	0.994993	Apriori
Cmc	0.947	0.994989	predApriori
Ecoli	0.984	0.994923	pred Apriori
Haberman	0.96	0.969948	predApriori
ionosphere	0.988	0.994017	predApriori
Iris	1	0.9929	Apriori
liver disorders	0.970833333	0.968394167	Apriori
Market Basket	0.959	0.971457396	predApriori
page-blocks	1	0.9949	Apriori
prostoperative patient data	0.94	0.983791366	predApriori
Tae	0.950	0.991509	predApriori
Vehicle	0.982	0.994866	predApriori
Zoo	0.959	0.993823	predApriori
breast cancer	0.944	0.994545	predApriori

V. Algorithm selection

A. Dataset Characterization Measurement

Each dataset can be described by a number of simple, statistical and information theoretical measures [20]. We average some statistical measures of all the attributes and take these as a global measure of the dataset characteristics (The details of the data sets statistical characteristics are provided in Table 4). The Table 3 lists some of the statistical measures used in our experiment to find out data characteristics.

Table 3: Statistical measures for characterisation of each dataset

Measure	Notation
Arithmetic Mean	Mean
Median	Median
Mode	Mode
Geometric mean	GM
InterQuartile Range	IQR
Range	Range
Variance	Var
Standard deviation	Std
Z-score	Z_score

A brief description of the above measures are as follows:

Arithmetic mean

The arithmetic mean or the simple mean of a list of sequence $\bar{X}$ is:

$$\frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + \dots + x_n)$$

Mode

The Mode is the most frequent incident number in a sequence. For example, the mode of 16, 15, 18, 9, 18, 11, 18 and 33 is 18.

Median

The Median is the middle number in a list of sequence; that is, half the numbers have values that are greater than the median, and half the numbers have values that are less than the median. For example, the median of 16, 15, 18, 9, 18, 11, 18 and 33 is 17.

Geometric mean (GM)

The geometric mean of a sequence $\{X_{j=1}^n\}$ is:

$$\left[\prod_{i=1}^n X_i \right]^{1/n} = \sqrt[n]{X_1, X_2, \dots, X_n}$$

InterQuartile Range (IQR)

The InterQuartile Range (IQR) is the difference between the 75th percentile and the 25th percentile.

Variance (Var)

The variance is used to characterize the dispersion among the measures in a given population. It calculates the mean of the scores, and then measures the amount that each score deviates from the mean and then squares that deviation for a given population. Numerically, the variance equals the average of the squared deviations from the mean.

Standard Deviation (Std)

The std measures the spread of a set of data as a proportion of its mean

$$\text{std} = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2}$$

where $\bar{x}$ is Arithmetic mean.

Z Score

Z-score is derived from subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation. It helps to measure, how far the observation is from the mean in units of standard deviation. The value of Z-score is greater than 3 indicates that the data distribution has outliers [4].

$$\text{Zscore} = \frac{x - \bar{x}}{\sigma}$$

Table 4: Using statistical formulation and comparison between confidence and ‘predictive accuracy’ we have generated the following table

	Mean	Median	Mode	GM	IQR	Range	Var	Std	Zscore	Class
Breast-Cancer	584.19	607.2	607.2	576.85	96.3	252.7	8416.6	64.934	23.462	2
bridge_version	212.48	233.75	225.92	198.64	55.938	193.58	6034.9	54.408	6.7602	1
car	312.88	250.92	222.67	282.95	207.25	268	16230	116.15	3.91	1
cmc	5.4593	5.4444	4.7778	4.7094	2.5833	7.1111	8.559	1.7011	17.247	2
ecoli	187.87	192	192.43	186.47	3.6429	149	326.72	17.149	3.263	2
haberman	39.779	38.667	36.667	38.032	9	38.667	59.655	7.0808	11.44	2
ionosphere	269.18	337.68	51.176	209.21	200.76	344.26	18575	132.11	4.4927	2
iris	149.19	149	148	149.16	4.25	11	7.8595	2.7971	40.272	1
liver-disorders	112.88	112	111.17	112.36	4.2083	71	114.82	9.8548	18.25	2
market-basket	175.81	177.67	178.5	173.69	18.125	138.75	661.93	24.643	15.954	2
page-blocksW	160.81	158.9	151.8	157.86	32.6	142.8	750	24.799	16.018	1
pos-patient-data	497.88	435.31	421.13	486.96	142.88	230.25	15001	92	5.8092	2
tae	71.115	71	71.4	69.56	20	34.8	233.42	11.303	3.2965	2
vehicle	123.06	122	121.72	122.37	11.5	32.889	123.46	8.2291	10.275	2
zoo	472.73	480.71	463.35	469.99	57.368	96.059	3139.7	40.134	10.75	2

B. Rule Generation

As a part of the meta-learning process, we have used data from Table 4 and generated rules using See5 data mining tools. See 5 basically classified

the data characteristics on the basis of Mode and Geometric Mean. The details of the characteristics rules are provided in Table5.

Table 5 : Characteristics rules generated by See5.

```
See5 [Release 2.03]
Options:
  Rule-based classifiers
  Pruning confidence level 99%
Read 15 cases (9 attributes)
Rules:
Rule 1: (6/2, lift 2.3)
  Mode > 121.72 & Geomean <= 282.95 -> class 1 [0.625]
Rule 2: (6, lift 1.2)
  Mode <= 121.72 -> class 2 [0.875]
Rule 3: (3, lift 1.1)
  Geomean > 282.95 -> class 2 [0.800]
Default class: 2
Evaluation on training data (15 cases):
  Rules
  -----
  No      Errors
  3      2 (13.3%) <<
  (a)    (b)    <-classified as
  -----
  4      9      (a): class 1
  2      9      (b): class 2
Time: 0.0 secs
```

Rule for Apriori

If

Mode > 121.72 and Geomean <= 282.95 of a data file then choose Apriori.

Rules for Predictive Apriori

if

Mode <= 121.72 or Geomean > 282.95 then choose Predictive Apriori

VI. Conclusion

Comparative study of three algorithms showed that Apriori had performed best on confidence based ranking and Predictive Apriori had performed better on accuracy based ranking. However the main aim of this research is to assist in the selection of an appropriate Association Rule Mining algorithm without the need for trial-and-error testing of the vast array of available algorithms. Based on statistical formulation measurement and meta-learning process, we can recommend an algorithm by

analysing Mode and GM of a data file. Although we have used nine statistical measurements, the rules emphasise only two of them. We aim to continue this research by considering more association problems, extracting better rules using other participating statistical measurements.

Bibliography

- [1] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," Proc. of the 20th Int'l Conference on Very Large Databases, Santiago, Chile, September 1994.
- [2] R. Agrawal, C. Aggarwal, and V.V.V. Prasad, "Depth First Generation of Long Patterns," Proc. Seventh Int'l Conf. Knowledge Discovery and Data Mining, August 2000.
- [3] R. Agrawal, T. Imielinski, and Swami A. "Mining associations between sets of items in large databases," In Proc. of the ACM SIGMOD Int'l Conference on Management of Data. - Washington D.C., pp. 207-216, May 1993
- [4] S. Ali and K. A. Smith, "On learning algorithm for classification", Applied Soft Computing, pp. 119-138, December 2004.
- [5] Y. Bastide, R. N. Taouil, Y. Pasquier, G. Stumme, and L. Lakhal, "Mining Frequent Patterns with Counting Inference," SIGKDD Explorations, vol. 2. December 2000.
- [6] R.J. Bayardo, "Efficiently Mining Long Patterns from Databases", Proc. ACM SIGMOD Conf. Management of Data. - June 1998.
- [7] C. Blake and C.J. Merz, UCI Repository of machine learning databases, University of California, 2007. available at <http://archive.ics.uci.edu/ml/>. viewed on Feb 2008.
- [8] D. Burdick, M. Calimlim and J. Gehrke, "MAFIA: a maximal frequent itemset algorithm for transactional databases," In International conference on Data Engineering. April 2001.
- [9] P.A. Flach and N. Lachiche, "Confirmation-guided discovery of first-order rules with Tertius," Kluwer Academic Publishers. - The Netherlands, Vol. 42. - pp. 61-95, 2001
- [10] K. Gouda and M.J. Zaki, "Efficiently Mining Maximal Frequent Itemsets," in Proc. First IEEE Int'l Conf. Data Mining. November 2001.
- [11] M. Hegland, "Algorithms for Association Rules," Lecture Notes in Computer Science, Vol. 2600. pp. 226 – 234. January 2003.
- [12] T.-S. Lim, Knowledge Discovery Central datasets. - 2002. available at <http://www.kdcentral.com/>.
- [13] D-I. Lin and Z.M. Kedem, "Pincer-Search: A New Algorithm for Discovering the Maximum Frequent Set," in Proc. Sixth Int'l Conf. Extending Database Technology, March 1998.
- [14] Matlab Statistics Toolbox User's Guide [Book]. - USA : The MathWorksInc, version 6.2., 2008
- [15] S. Mutter, M. Hall and E. Frank, "Using Classification to Evaluate the Output of Confidence based Association Rule Mining," Lecture Notes in Artificial Intelligence, Advances in Artificial Intelligence - AI 2004. Berlin, Springer, vol. 3339. - pp. 538-549. 2004.
- [16] N. Pasquier, Y. Bastide, R. Taouil, & L. Lakhal, "Discovering Frequent Closed Itemsets for Association Rules," Proc. Seventh Int'l Conf. Database Theory, January 1999.
- [17] J. Pei, J. Han and R. Mao, "Closet: An Efficient Algorithm for Mining Frequent Closed Itemsets", Proc. SIGMOD Int'l Workshop Data Mining and Knowledge Discovery. May 2000.
- [18] RuleQuest, RuleQuest Research Pty Ltd, available at <http://www.rulequest.com/download.html>. November 2007, viewed on April 2008.
- [19] T. Scheffer, "Finding Association Rules that Trade Support Optimally Against Confidence" in proceedings of the 5th European Conference on Principles and Practice of Knowledge Discovery in Databases(PKDD'01). - Freiburg, Germany : Springer-Verlag, pp. 424-435. September 2001
- [20] K.A. Smith, F. Woo, R. Ibrahim, and V. Ciesielski, "Matching data mining algorithm suitability to data characteristics using a selforganising map", Hybrid Information Systems, Heidelberg : Physica-Verlag, pp. 169-180, 2002.
- [21] J. Wang, J. Han and J. Pei "Closet+: Searching for the Best Strategies for Mining Frequent Closed Itemsets," in Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, August 2003.
- [22] I.H. Witten and Frank E., "Data Mining: Practical Machine Learning Tool and Technique with Java Implementation," San Francisco : Morgan Kaufmann, 2000.
- [23] M. J. Zaki and C.J. Hsiao, "CHARM: An efficient algorithm for closed association rule mining," Computer Science Department ; Rensselaer Polytechnic Institute. - New York, October 1999.
- [24] M. J. Zaki and C-J. Hsiao, "Efficient Algorithms for Mining Closed Itemsets and Their Lattice Structure," IEEE Computer Society, vol. 17, April 2005.
- [25] M. J. Zaki, "Scalable algorithms for association mining," IEEE Transaction on Knowledge and Data Engineering, vol. 12. May 2000.
- [26] M.J. Zaki, "Generating Non-Redundant Association Rules," in Proc. Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining. August 2000.