

Share Based Measures for Itemsets

Colin L. Carter, Howard J. Hamilton and Nick Cercone

Department of Computer Science
University of Regina
Regina, SK, Canada, S4S 0A2
{carter, hamilton, nick}@cs.uregina.ca

Abstract. We introduce the measures *share*, *coincidence* and *dominance* as alternatives to the standard itemset methodology measure of support. An itemset is a group of items bought together in a transaction. The support of an itemset is the ratio of transactions containing the itemset to the total number of transactions. The share of an itemset is the ratio of the count of items purchased together to the total count of items in all transactions. The coincidence of an itemset is the ratio of the count of items in that itemset to the total of those same items in the database. The dominance of an item in an itemset specifies the extent to which that item dominates the total of all items in the itemset. Share based measures have the advantage over support of reflecting accurately how many units are being moved by a business. The share measure can be extended to quantify the financial impact of an itemset on the business.

1 Introduction

Large amounts of scan code data collected by many businesses represent a potential wealth of information given adequate methods of transforming the data into meaningful information. One class of such data is stored in transaction databases from which all items obtained in a single transaction can be retrieved as a unit. The transactions can then be examined to determine what items customers typically buy together. This gives insight into marketing these products more effectively.

Recent research has focused on determining which groups of items, called *itemsets*, are bought together. From any itemset an *association rule* may be derived which, given the purchase of a subset of the items in the itemset, predicts the probability of the purchase of the remaining items [1,2,4,7]. Several algorithms have been proposed for finding generalized itemsets from items that are classified by taxonomic hierarchies [3,8]. The interestingness of the discovered rules and some methods for pruning uninteresting rules have been addressed in [5,8].

Data managers are primarily interested in itemsets which are bought in sufficient numbers to form a substantial portion of their business. This portion is somewhat reflected in the itemset *support*, which is the ratio of the number of transactions that contain an itemset to the total number of transactions. An itemset is *frequent* if its support falls above a user specified *minimum support*. The *confidence* of an association rule $A \rightarrow B$, where A and B are itemsets and $A \cap B = \emptyset$, is the ratio of the support of the itemset $A \cup B$ to the support of the itemset A . The confidence quantifies the probability that when A is bought, B will also be bought.

To illustrate these concepts, an example database is shown in Table 1. The

Table 1. A transaction database with 5 transactions

Transaction ID	Items
T ₁	A, C, D
T ₂	B, E
T ₃	A, B, C, E
T ₄	B, E
T ₅	B, D, F

Table 3. Itemset supports

Itemset	Number of Transactions	Support $sup(X)$
A, C	2	40%
A, B	1	20%
B, D	1	20%
C, D	1	20%
A, B, C	1	20%
A, B, E	1	20%
A, C, D	1	20%
B, D, F	1	20%
A, B, C, E	1	20%

Table 2. Single item support

Item	Number of Transactions	Support $sup(X)$
A	2	40%
B	4	80%
C	2	40%
D	2	40%
E	3	60%
F	1	20%

Table 4. Association rules and confidences

Association Rule	Confidence $conf(X \rightarrow Y)$
A \rightarrow C	100%
A \rightarrow B	50%
B \rightarrow D	25%
A, B \rightarrow C	100%
A, C \rightarrow B	50%
B, E \rightarrow A	33%

supports of single items are shown in Table 2, and the supports of some itemsets in Table 3. There are 5 transactions in the database, so the support is the number of transactions in which the itemsets occur divided by 5. The confidence values of some association rules are shown in Table 4. The confidence of 100% for the rule A \rightarrow C means that every transaction which contains A also contains C.

The current definition of support is limited in informative feedback since it tells the number of transactions containing an itemset but not the number of items. Some items may be bought in multiples and therefore actually be more frequent than the support measure indicates. The support measure also does not allow for accurate financial calculations or comparisons. Masand and Piatetsky-Shapiro [6] note that for target marketing, measures should take into account the *frequency* of an item contributing to a predictive rule and the *value* of the items in the prediction. The support measure allows for neither of these, so measures based on specific numbers of items, such as percentage of gross sales, costs or net profit, cannot be calculated.

In this paper we present alternative measures that provide more informative feedback. These measures incorporate into current itemset algorithms with no decrease in efficiency. We assume a count is associated with each item in a transaction. The *share* of an itemset is the ratio of the total count of items in the itemset when they appear together to the total count of items in the database. The *coincidence* of an itemset is the ratio of the count of the items when they appear together to the total count of the same items in the whole database. The *dominance* of an item in an itemset is a measure that quantifies the total of one item relative to other items in the itemset. Using commercial data, we show that the share measure may give a different view of the relative importance of an itemset than that implied by support.

Table 5. A transaction database with counts

Trans. ID	Item	Item Count
T ₁	A	1
	C	2
	D	2
T ₂	B	1
	E	3
T ₃	A	2
	B	1
	C	2
T ₄	E	1
	B	1
T ₅	E	1
	B	1
T ₅	D	1
	F	1
	F	1

Table 6. Transaction counts and global counts for each item

Item	Transactions Counts $c(I_p, T_q)$					Item Global Counts: $c(I_p)$
	T ₁	T ₂	T ₃	T ₄	T ₅	
A	1	0	2	0	0	3
B	0	1	1	1	1	4
C	2	0	2	0	0	4
D	2	0	0	0	1	3
E	0	3	1	1	0	5
F	0	0	0	0	1	1

2 New Measures

2.1 Share

An intuitive measure which takes into account multiple items is the ratio of the total items in an itemset to the total of all items in the database. We call this measure the *share* of an itemset.

Let $I = \{I_1, I_2, \dots, I_m\}$ be a set of m items and $T = \{T_1, T_2, \dots, T_n\}$ be a set (database) of n transactions, where $T_q \subseteq I$ for each $T_q \in T$. Let the *transaction count* of item $I_p \in I$ in transaction T_q be $c(I_p, T_q)$, which is the number of item I_p purchased in transaction T_q . Each item I_p has an associated set of transactions $T_{I_p} = \{T_q \in T \mid I_p \in T_q\}$ which is the set of all transactions containing item I_p . Let the *global count* of item I_p in the database T be $C(I_p) = \sum_{T_q \in T_{I_p}} c(I_p, T_q)$, which is the sum of the counts of item I_p

in every transaction of the database in which I_p appears. Let the *total item count* of all items in I in the transaction set T be $C = \sum_{p=1}^m C(I_p)$ which is the sum of the global

counts of the individual items in I , or in other words, the total number of items in the database.

We illustrate these measures using the database in Table 5. This is the same database as in Table 1 except that each item has a transaction count. Table 6 shows the transactions counts $c(I_p, T_q)$ and global counts $C(I_p)$ for each item. The total item count for the database is 20.

A *k-itemset* is a set $X = \{x_1, x_2, \dots, x_k\}$, $X \subseteq I$, $2 \leq k \leq m$, of k distinct items that are purchased together in one or more transactions. Each itemset X has an associated set of transactions $T_X = \{T_q \in T \mid T_q \supseteq X\}$ which is the set of transactions that contain X . Let the *local count of an item x_i in the itemset X* be $c(x_i, X) = \sum_{T_q \in T_X} c(x_i, T_q)$, which is

the sum of the transaction counts of the item x_i in all transactions which contain the itemset X . An item will have a separate local count for each itemset in which it appears. Thus for some item I_q , its local count $c(I_q, X)$ in the itemset X will be different from the local count $c(I_q, Z)$ of the same item in the itemset Z , $Z \neq X$.

Let the *local count* of a k -itemset X be $c(X) = \sum_{i=1}^k c(x_i, X)$, which is the sum of the local counts of each item in the itemset. Let the *global count* of the k -itemset X be $C(X) = \sum_{i=1}^k C(x_i)$ which is the sum of the global counts of each item in the itemset.

Table 7 lists several 2-itemsets from the example database (column 1), the transactions which contain each itemset (column 2), the local counts of each item (columns 3 and 4) and the local count of each itemset (column 5).

Definition 1. The share of an item x_i in the itemset X : $share(x_i, X) = c(x_i, X)/C$

The share of the item x_i in the itemset X is the ratio of the local count of the item x_i in X to the total item count. Table 7 shows the shares of each item for some 2-itemsets. For the sample database, the total item count in the database is $C = 20$.

Definition 2. The share of an itemset X : $share(X) = c(X)/C$

The share of the itemset X is the ratio of the local count of the itemset X to the total item count, or in other words, the ratio of the count of items in the itemset to the total count of items in the database. Table 7 shows the itemset shares for several 2-

Table 7. Some 2-itemsets, their local counts and shares

Itemset X	Containing Transactions T_X	1 st Item Local Count $c(x_1, X)$	2 nd Item Local Count $c(x_2, X)$	Itemset Local Count $c(X)$	1 st Item Share $share(x_1, X)$	2 nd Item Share $share(x_2, X)$	Itemset Share $share(X)$
A, C	$\{T_1, T_3\}$	3	4	7	15%	20%	35%
A, B	$\{T_3\}$	2	1	3	10%	5%	15%
B, D	$\{T_4\}$	1	1	2	5%	5%	10%
C, D	$\{T_1\}$	2	2	4	10%	10%	20%

Table 8. Support and shares for various itemsets in the database

Itemset X	Number of Transactions $ T_X $	Support $sup(X) =$ $ T_X /5$	Itemset Local Count $c(X)$	Share $share(X) =$ $c(X)/20$
A, C	2	40%	7	35%
A, B	1	20%	3	15%
B, D	1	20%	2	10%
C, D	1	20%	4	20%
A, B, C	1	20%	5	25%
A, B, E	1	20%	4	20%
A, C, D	1	20%	5	25%
B, D, F	1	20%	3	15%
A, B, C, E	1	20%	6	30%

itemsets (column 8). Table 8 shows both the support and share values for some itemsets in the database. All itemsets but one have a support of 20%. This would seem to imply that there is no differentiation between these product groups in terms of numbers sold. The shares of the itemsets, however, range from 10% to 30%, showing that some groups outnumber others 3 to 1. Decision makers would not know which of the itemsets with equal support should be focused on, but a decision based on share would be clearer.

An association rule is an implication of the form $X \rightarrow Y$ derived from an itemset Z , where X and Y are itemsets, $X \cup Y = Z$, and $X \cap Y = \emptyset$. The support of the association rule $X \rightarrow Y$ in the transaction set T is $sup(Z)$, the support of the itemset Z . The share of the association rule $X \rightarrow Y$ in the transaction set T is $share(Z)$, the share of the itemset Z .

2.2 Coincidence

Definition 3. The coincidence of an itemset X : $coinc(X) = c(X)/C(X)$

The *coincidence* of an itemset X is the ratio of the local count of the itemset to the global count. Intuitively, a coincidence of 60% for an itemset states that 60% of all the items in that itemset are bought together. Table 9 shows some itemsets from the sample database, their local and global counts, and their coincidences.

While the share measures an itemset's importance relative to all items sold, the coincidence measures the importance of an itemset in terms of all items of that itemset sold. Two itemsets may have the same share, but a higher coincidence in one indicates a stronger relationship between its items. For example, the itemsets $\{A, C, D\}$ and $\{A, B, C\}$ (Table 9) both have a share of 25%, but the coincidence of $\{A, C, D\}$ is 50% and of $\{A, B, C\}$ is 45%. Therefore, relatively more of the first set appear together than the second set, making the first set a more likely target for marketing. On the other hand, itemset $\{A, C\}$ has a share of 35% (see Table 8) and a coincidence of 100%. If market researcher were looking to boost the coincidence of products sold together, there is no sense in promoting the pair $\{A, C\}$ since all of them are already sold together.

2.3 Dominance

In an itemset, the item local counts may not contribute equally to the itemset count.

Table 9. Coincidence measures for itemsets in Table 3

Itemset X	Local Count $c(X)$	Global Count $C(X)$	Coincidence $c(X) / C(X)$
A, C	7	7	100%
A, B	3	7	43%
B, D	2	7	29%
C, D	4	7	57%
A, B, C	5	11	45%
A, B, E	4	12	33%
A, C, D	5	10	50%
B, D, F	3	8	38%
A, B, C, E	6	16	38%

For example, for the itemset {B, D} in Table 7, the local counts for B and D are both 1. However, the local counts of A and C in {A, C} are different, being 3 and 4 respectively. We measure the proportion of the item's local count to the itemset count with a *dominance* measure.

The dominance could be represented by the simple percentage of the local count of the item relative to the local count of the itemset. For example, for the itemset {A, B} in Table 9, the item A with a local count of 2 represents 66% of the itemset's local count of 3. However, this measure would have to be compared to the size of the itemset to be appreciated. Instead, we normalize the dominance of an item by multiplying the percentage of the local count of an item by the number of items in the itemset.

Definition 4. The dominance of an item x_i in the k -itemset X :

$$dom(x_i, X) = c(x_i, X) / c(X) * k$$

The dominance of item x_i in the k -itemset X normalizes the item proportions according to the assumption of uniform distribution of items in the itemset. In a k -itemset X with local count $c(X)$, items with a count of $c(X)/k$ have a dominance of 1.0. Items dominating the count of an itemset of size k have a dominance approaching k . For example, the items in the itemset {B, D, F} each have local counts of 1, and therefore each item has a dominance of 1.0. However, in the itemset {A, C}, which has an itemset count of 7, the item A has a local count of 3, and C has a local count of 4 (Table 7). The dominance of A would be $3/7 * 2 = 0.86$, and C would be $4/7 * 2 = 1.14$. Items with a dominance less than 1.0 have a lower than average count, and those with greater than 1.0 have a greater than average count.

2.4 Share-based Confidence

In standard itemset methodology, the confidence of an association rule $X \rightarrow Y$ quantifies the strength of prediction the antecedent X has in reference to the consequent Y . The higher the confidence, the more likely the items in Y will be purchased when the items in X are purchased. This same intuition can be carried into itemsets where counts of the items are available.

Definition 5. The share-based confidence of the association rule $X \rightarrow Y$ derived from

$$\text{an itemset } Z = \{z_1, z_2, \dots, z_k\} = X \cup Y: \text{conf}_{share}(X \rightarrow Y) = \frac{\sum_{z_i \in X} c(z_i, Z)}{c(X)}.$$

The share-based confidence of the rule $X \rightarrow Y$ is the ratio of the sum of the local counts of the items comprising X in Z to the local count of X . Note that the local counts $c(z_i, Z)$ of the items common to both X and Z are counts of the items z_i in the context of the itemset Z , while the local count $c(X)$ of the itemset X is the sum of the counts of the same items in the context of the itemset X . For example, from Table 8, consider the itemsets $X = \{A, C\}$ with local item counts of 3 and 4 respectively and a local itemset count of 7, and $Z = \{A, C, D\}$ with local item counts 1, 2 and 2 respectively. The sum of the local counts of items A and C in Z is 3, and the local count of the itemset $X = \{A, C\}$ is 7. The share-based confidence of the association rule $\{A, C\} \rightarrow \{D\}$ is therefore $3/7 = 43\%$. In contrast, the support-based confidence

is 50% because $\{A, C, D\}$ has a support of 1, and $\{A, C\}$ has a support of 2, which gives a confidence of $1/2$. Consider also the itemset $Z = \{A, B, C\}$ with item local counts of 2, 1 and 2. Again the itemset $\{A, C\}$ has a local count of 7, while the local counts of items A and C in the itemset Z have a sum of 4. The rule $\{A, C\} \rightarrow \{B\}$ has a share-based confidence of $4/7 = 57\%$ and a support-based confidence of 50%. The share-based confidence values can therefore be more or less than those based on support.

The advantage of a count based confidence is increased precision. Confidence values based on support for the preceding two examples were both 50% and allowed no differentiation between the two. Confidence values based on share were 43% and 57%, which would allow a data manager to more precisely rank the value of the two association rules.

2.5 Frequent Itemsets

In standard itemset methodology, an itemset is only considered interesting if its support falls above a user specified minimum support value. The minimum support specifies the minimum percentage of transactions in which the complete itemset must be contained to be considered *frequent* [1]. For share based measures, we redefine the concept of frequency as follows:

Definition 6. An itemset X is *frequent* if: $\forall x \in X, share(x, X) \geq minshare$

That is, an itemset is frequent if the share of every item in the itemset falls above a user defined minimum share value. Intuitively, we are interested in items that are individually bought frequently and also bought with other frequent items. We do not define frequency in terms of the share of the itemset as a whole since this could lead to items with very small individual support being included in a frequent itemset where the total of the remaining items is still relatively high.

3 Extensions of Share Based Measures

Share based measures can be extended to include financial data to give a more informative feedback about the relative importance of various items. Many transactions include not only an item count field, but item cost and profit fields. An itemset task which takes into account the amount of money generated will rate more profitable items higher than less profitable ones and give data managers a more realistic view of what products drive the business.

To empirically test this intuition, we ran an itemset task on commercial data from one of our corporate sponsors. We implemented the Apriori algorithm [1,2] with both support and two share based measures reflecting the number of items sold and the gross income generated from these. The database represented 427,620 customer accounts and 2200 items. A minimum support and share of 0.25% was used. An itemset was considered frequent if it was above the minimum support, if the share based on the count of each component item was above 0.25% of the total number of items in the database, or if each item's summed gross value was above 0.25% of the total income in the database.

There were 110 frequent items, 258 frequent 2-itemsets, 188 3-itemsets, 29 4-itemsets, and no larger itemsets. Table 10 shows the rankings of three sets of single

Table 10. Single items ranked by support, share of count and share of amount

Top 10 Items Ranked by Support			Top 10 Items Ranked by Share of Count			Top 10 Items Ranked by Share of Income		
Support Ranking	Count Ranking	Income Ranking	Support Ranking	Count Ranking	Income Ranking	Support Ranking	Count Ranking	Income Ranking
1	1	1	1	1	1	1	1	1
2	3	10	102	2	9	10	10	2
3	4	12	2	3	10	4	8	3
4	8	3	3	4	12	31	26	4
5	6	21	82	5	20	100	102	5
6	13	18	5	6	21	59	44	6
7	14	8	26	7	63	35	17	7
8	11	85	4	8	3	7	14	8
9	15	15	108	9	37	102	2	9
10	10	2	10	10	2	2	3	10

items ranked by support, share of count and share of income. Each set of three columns represents 10 items ranked by one of these criteria. We assigned each item three different ordinals representing its ranking by support, share of count and share of income respectively. We then sorted the itemsets by support and extracted the rankings of the top 10 of these. This is shown in the first three columns of the table. For example, the item which ranked second by support was ranked 3rd by share of count and 10th by share of income. While this item was highly ranked by both the number of customer accounts in which it was purchased, nine other items actually made more income than it did. Similarly, the item that ranked 8th by support and 11th by share of count, only ranked 85th out of 110 in terms of income generated. The support and share of count, therefore, do not necessarily give an accurate picture of what really is of interest to a business.

The second and third sets of three columns are similar except the second set shows the top 10 items ranked by share of count, and the third set shows the top 10 items ranked by share of income. Note that 3 of the top 10 items ranked by share of count ranked 82nd, 100th and 102nd by support. These items are typically bought in multiples, but the share measure has no way of determining this and gives them a much lower ranking than is justified by the number of items sold. Also, the item ranked 7th by share of count is ranked 63rd by share of income. This item, represented a large percentage of the items sold, but was a relatively inexpensive item since it was ranked well down by share of income. From the third set of three columns, the item that was ranked 5th by share of income was ranked 100th by support and 102nd by share of count. This is a costly item that although it is not bought as frequently as some other items, makes the company a relatively large amount of money in comparison to some other more frequently purchased items. Again, the support measure has no way of determining this.

Table 11 shows three sets of itemsets where each itemset is again assigned 3 ranking ordinals similar to Table 10, but each set is only sorted by share of income. For 2-itemsets and 3-itemsets, we again see that the support measure ranks many itemsets much lower than either share of income or share of count. Since the ranking of 4-itemsets is only out of 29, not much variation is evident.

Table 11. Itemsets with 2, 3 and 4 items ranked by share of amount

Top 10 2-itemsets by Income (out of 258)			Top 10 3-itemsets by Income (out of 188)			Top 10 4-itemsets by Income (out of 29)		
Income Ranking	Support Ranking	Count Ranking	Income Ranking	Support Ranking	Count Ranking	Income Ranking	Support Ranking	Count Ranking
1	1	1	1	2	2	1	1	1
2	2	2	2	1	1	2	2	2
3	5	5	3	164	21	3	3	3
4	75	34	4	165	28	4	4	4
5	3	3	5	4	4	5	7	8
6	6	6	6	157	20	6	8	9
7	4	4	7	172	12	7	5	5
8	138	15	8	5	5	8	6	6
9	54	18	9	3	3	9	13	13
10	193	57	10	131	35	10	14	15

Table 12. Top 5 2-itemsets ranked by share of amount, showing coincidence

Itemset Ranking	Share of Income	Coincidence
1	6.58%	30.93%
2	4.85%	87.35%
3	3.97%	20.78%
4	3.82%	35.05%
5	4.47%	19.34%

Table 12 shows the top 5 2-itemsets as ranked by share of income and their coincidences. The most frequent itemset generated about 6.6% of the income of the company, and of all instances of the two items sold by the company, 31% of these were sold together. The two items that comprise this itemset, therefore, may have room for increased coincidence through a marketing campaign. However, the second most frequent itemset has a coincidence of about 87%. It would be unprofitable to promote the sale of one of these items to customers who have already bought the other, since they are already very often sold together. Without the coincidence measure, data managers would not have enough information to make an informed choice about this issue. The standard itemset measure of support is not able to provide information of this nature.

Table 13 shows the top 10 2-itemsets ranked by share of income, the amount of income each of the items in the itemset represents, the share of the itemset overall and the dominance of each item.

Finally, in Table 14, we present the number of frequent itemsets that were detected using only one minimum support or minimum share at a time. In the itemset task run on commercial data described in this paper, an item or itemset was considered frequent if the number of transactions in which the item or itemset appeared exceeded 0.25% of the total number of transactions (minimum support), if the count of items in the itemset exceeded 0.25% of the total number of items (minimum share of count), or if the income generated by the item or itemset exceeded 0.25% of the total income represented in the database (minimum share of income). Any one of these criteria was

Table 13. Top 10 2-itemsets ranked by share of income and dominance measures

Itemset ID	Item 1 Income	Item 2 Income	Itemset Share of Income	Item 1 Dominance	Item 2 Dominance
1	\$1,097,719	\$256,343	6.58%	1.62	0.38
2	\$250,432	\$746,943	4.85%	0.50	1.50
3	\$597,209	\$323,156	4.47%	1.30	0.70
4	\$337,775	\$496,104	4.05%	0.81	1.19
5	\$668,445	\$149,147	3.97%	1.64	0.36
6	\$568,486	\$230,070	3.88%	1.42	0.58
7	\$619,434	\$165,268	3.82%	1.58	0.42
8	\$336,872	\$271,158	2.96%	1.11	0.89
9	\$225,148	\$347,055	2.78%	0.79	1.21
10	\$347,415	\$91,904	2.14%	1.58	0.42

Table 14. Count of frequent items and itemsets for different minimum support or shares exceeded

	Exceeded Any Minimum	Exceeded Minimum Support	Exceeded Minimum Share of Count	Exceeded Minimum Share of Income
Items	110	80	58	60
2-itemsets	258	258	65	22
3-itemsets	188	188	16	3
4-itemsets	29	29	0	0

sufficient to include the item or itemset in the frequent list. The number of frequent items or itemsets detected under these conditions is shown in the *Exceeded Any Minimum* column of Table 14. The remaining three columns show the number of items or itemsets detected using only one of the minimum support, minimum share of count or minimum share of income at a time. For this specific task, the greatest numbers of frequent items and itemsets were detected using minimum support, followed by minimum share of count and minimum share of income respectively. More testing on various data sources is necessary to see if this pattern is common to different data sources. In this case, however, where income generated by an itemset is most relevant, the smaller numbers of itemsets detected by using only the minimum share of income would allow for a more focused concentration on these for a more effective marketing campaign.

4 Conclusion

We have presented a new set of *share-based* measures to augment the standard measure of support. The support of an itemset is the ratio of the number of transactions containing an itemset to the total number of transactions. The share of an itemset is the ratio of the count of items in an itemset to the total count of items in the database. The coincidence of an itemset is the ratio of those items that are purchased together to the total number of the same items purchased in the database. The coincidence measure is useful to determine which of several itemsets with approximately the same share are more interesting. Itemsets with high coincidence

may not be interesting since most of them are bought together anyway and a marketing campaign would not boost sales substantially. Itemsets with a large share and moderate coincidence are good candidates for marketing in that there is room for coincidence to grow. The dominance of an item in an itemset is a measure of how its count dominates the count of the itemset as a whole, either locally in the itemset or relative to global proportions of the same items. An itemset whose proportions deviate substantially from database patterns may be of interest. All share based measures can be easily incorporated into current itemset generation algorithms.

Share based measures are both intuitively reasonable and understandable. They are based on the stable, unchanging baseline of the total number of items sold in the given set of transactions. Since they take into account the number of items purchased by customers, they allow the capability of relating the number of items to the financial impact of the sales. This financial extension is ultimately the most relevant measure when making marketing decisions.

References

- [1] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," in *Proceedings of the 20th VLDB Conference*, Santiago, Chile, 1994, 487-499.
- [2] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen and A.I. Verkamo, "Fast Discovery of Association Rules," in U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, eds., *Advances in Knowledge Discovery and Data Mining*, AAAI Press / MIT Press, Menlo Park, CA, 1996, 307-328.
- [3] J. Han and Y. Fu, "Discovery of Multiple-Level Association Rules from Large Databases," in *Proceedings of the 21st VLDB Conference*, Zurich, Switzerland, 1995.
- [4] M. Houtsma and A. Swami, "Set-Oriented Mining for Association Rules in Relational Databases," in *Proceedings of IEEE International Conference on Data Engineering*, March, 1995, 25-33.
- [5] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen and A. I. Verkamo, "Finding Interesting Rules from Large Sets of Discovered Association Rules," in *Proceedings of the Third International Conference on Information and Knowledge Management*, Gaithersburg, Maryland, Nov., 1994.
- [6] B. Masand and G. Piatetsky-Shapiro, "A Comparison of Approaches for Maximizing Business Payoff of Prediction Models," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, August, 1996, 195-201.
- [7] J. S. Park, M. S. Chen and P. S. Yu, "An Effective Hash Based Algorithm for Mining Association Rules," in *SIGMOD Record*, **24:2**, 1995, 175-186.
- [8] R. Srikant and R. Agrawal, "Mining Generalized Association Rules," in *Proceedings of the 21st VLDB Conference*, Zurich, Switzerland, 1995.