Advertising Effect Assessment

Yuping Yang Jiawei Han Mukesh Singhal

July 19, 1999

Abstract

Assessing advertising effect is difficult because there is no clear and easily definable link between the advertisement and the sale of items. Also, because there are usually multiple advertising channels such as TV, radio, etc, running multiple advertisements at overlapping time periods, it is not clear which advertisement has affected the sale of which item.

This paper takes a rule based approach to assess advertising effect. Strong trends in form of association rules between advertisements and sales are used as evidences of advertising effects. First, frequent item-sets in sales data are searched. Second, strong confidences of association from frequent item-sets in sales data to advertisement-sets are searched and used as basis for assessing the effect of advertisements to the sales. Third, strong confidences of implication from advertisement-sets obtained in the second step to their associated frequent item-sets are searched and used as evidences of advertising effect.

Based on the result of assessing past advertising effect, a precise mathematical model is proposed as the basis for optimizing advertising strategy in the future. The model can be numerically solved using linear programming method to obtain the best advertising strategy.

Key words: data mining, association rule, advertisement, marketing.

1 Introduction

Analyzing advertisement effect is a practical problem with great economical consequences. Often, one advertisement includes several items to be advertised and one item is advertised through many advertisements. To determine which advertisement affecting the sale of which item is usually the guess work of corporate marketing people. Suppose a retailing company collects its advertising data and sales data into two separate relational tables in a relational database. Each advertisement is recorded as a tuple in the advertising table T_{AD}. A tuple of an advertisement is keyed by an identifier $AD\#$ and contains a set of items I_{AD} to be advertised. Each transaction of sales is recorded as a tuple in the sales transaction table T_{SALES}. A tuple of a transaction is keyed by an identifier $S\#$ and contains a set I_S of items sold in the transaction. Each item is represented by an item identifier $I\#$. The following are two of the questions to be answered by advertising effect assessment:

1. which item ($I\#$) should be advertised along with which other item ($I\#$) ?
2. which advertisement ($AD\#$) has the best (or least) effect on the sale of which item ($I\#$) ?

The solution to the first question is a typical association rule mining operation and the rules to be mined are classical boolean association rules [2] in the sales transaction table T_S alone. The solution to the second question requires knowledge about data association because advertising data and sales data which are usually kept in two different relational tables. Our paper focus on the second question.

The main difficult in answering the second question is that there is no clear and definite link between advertisement data and sale data. The sales data can be regarded as "objective" and the advertisement data can be regarded as "subjective". The way advertisements are run (and so the advertisement data) is subjectively determined by the sales organization and the sales data can be regarded as the result of the advertising effort. So, it makes no sense to mine advertising data alone to find any strong trend and mining must start in sales data. Sales data can be mined independently using traditional data mining algorithms such as Apriori algorithm [3].

Once some strong trends such as frequent item-sets in sales data are discovered, then analyzing those advertisements who advertise those items in the discovered frequent item-sets makes sense. The set of advertisements that might have strong effects on the sales of items in the frequent item-sets are determined by data associations. The data associations can be expressed by association rules between advertisement data and sales data with sufficient confidences. The direction of these association rules is, of course, from sales data to the advertisement data because the interested advertisement sets (ad-sets) are unknown at the beginning of this step.

The frequent item-sets and ad-sets in association rules obtained in the above step can be further refined by mining association rules from the ad-set to the frequent item-set of the rule. This step is necessary because the association from ad-sets to the sales data is the goal of this advertising effect analysis.

The above three steps constitute a complete advertising effect assessment procedure proposed in this paper. The three-step mining procedure must be carried out in exactly the order stated above because (1) only frequent item-set can be mined independently; (2) mining ad-set only makes sense from discovered frequent item-sets; (3) The final result must be in the form of association rules from ad-set to item-sets.

This paper also defines two new types of association rules, I-A rules and A-I rules, to be used in the assessment procedure.

Section 2 discusses several difference mechanisms that advertisement may influence the sale. Section 3 through 6 presents the mining steps in the assessment procedure. Section 7 through 9 discusses the suitable cases, validity, and computational cost of the assessment procedure. Section 10 discusses the profitability of running advertisement. Section 11 proposes an optimization procedure based on the result of the assessment procedure. Finally, Section 12 concludes the paper.

2 Advertising Mechanisms and Periods

The way of an advertisement affects the sales of items can be divided into OFFER and INFLUENCE. Here, an OFFER is defined as a mechanism to affect customer behavior by a special arrangement, such as a discount, a favorable payment method, etc. During the effective period of advertisement, it is actually beneficial (for example, cheaper or easier to pay) for customers to buy the advertised item than that in the non-effective period. An INFLUENCE is a mechanism to affect customer behavior by a favorable wording or image in the advertisement so that customers feel good about buying the advertised item or being reminded the usefulness of the advertised item.

If an advertisement works by OFFER mechanism, then the effective period of the advertisement, i.e., the period in which the advertising affect the sales, should be exactly the period of the OFFER. For example, if every Friday a TV advertisement says that if customers buy item IM during Saturday and Sunday, they can get 30% discount of the marked price of this item, then the effective period of this advertisement is Saturday and Sunday.

Another complication in advertising analysis is that for items whose sales are influenced by the advertisement, the influenced item may or may not be in the advertised item-set. For example, an advertisement may say that the price of milk is 50% less during weekend, but because of this advertisement, the sales of eggs, sausages and many other food items are increased during the weekend. If the affected item is in the advertisement, the influence is called direct influence. Otherwise, it is called indirect influence.

The advertising effect assessment procedure developed in this paper assesses both OFFER and INFLUENCE types of advertisements. For INFLUENCE type of advertisements, both direct and influences are assessed. In the assessment procedure, the difference between OFFER and INFLUENCE is only a difference in figuring out the Expected Effective Period (EEP) of the advertisement. There is no difference in computation in the assessment procedure for direct and indirect influences.

3 Advertisement Data and Sales Data

This paper focuses on effect assessment of periodical advertisements. A periodical advertisement runs periodically. For example, a TV advertisement may run every Friday, Saturday, and Sunday. The advertisement starts at 10:00 pm and repeats 6 times. Each run of the advertisement lasts 45 seconds and the time interval between two consecutive runs is 10 minutes. Information about an advertisement also include the items to be advertised. A sales organization can have many such periodical advertisements. These data are recorded in Table 1 (AD table).

For OFFER type of advertisements, such as advertisement of discount, it is easy to determine the EEP of the advertisement: it is the period when OFFER, such as discount, is effective. For INFLUENCE type of advertisement, because it is almost impossible to numerically compute long-term psychological influence of advertisement, we assume that the advertisements only have short influences. In general, we assume the EEP is the period immediately after the advertisement was run and lasts for a few days. More detailed discussion about EEP will be given in Section 8. In Table 1, the EEP of AD107 is every 1-5, 10-15, 20-25 day of the month; the EEP of AD241 is every Monday, Wednesday, and Friday.

The actual sales transactions are in Table 2 (SALES table). Information about each transaction includes the items to be sold in this transaction and time of the sale.

For assessing advertising effect, ideally, we would like to find an association rule A: X $\Rightarrow$ Y, where X is an advertisement and Y is a item to be sold. The support of

$AD\#$	$AD_{scheule}$	EEP	I_{AD}
$AD103$	AD_{s103}	w6 & w7	I_4, I_{117}
$AD107$	AD_{s107}	m1-m5, m10-15, m20-25	I_4, I_{15}, I_{26}
$AD241$	AD_{s241}	w1, w3, w5	I_{11}, I_{15}

Table 1: (AD table) Time Schedule and Items of Advertisements

$T\#$	T_{time}	I_S
$T1103$	1999.03.13	I_4, I_{15}, I_{100}
$T1104$	1999.03.15	I_{164}, I_{215}, I_{304} I_{362}
$T1105$	1999.03.16	I_{64}, I_{285}, I_{334}
$T1106$	1999.03.17	I_3, I_{75}, I_{117}
$T1106$	1999.03.18	I_17, I_{144}, I_{201}
$T1107$	1999.03.19	I_17, I_{91}

Table 2: (SALES table) Sales Transactions

the rule is the number of times that X and Y appear "together" and the confidence is the ratio of dividing the support by the total number of times the advertisement runs. The first problem encountered in formulating such an association rule is to define what it means that an advertisement and a item are "together" ?

Under our short-term effect assumption, if a transaction and an advertisement is "together", it can only mean that the time of the sale of the transaction is in the EEP of the advertisement. So, the Ad-table and the SALES table can be joined using EEP and T_{time} to obtain a Ad-SALES table as shown in Table 3. Conceptually, a transaction and an advertisement are considered "together" if they appears on the same tuple in the AD-SALES table.

4 Initial Step: Mining "Frequent" Item-Sets

As mentioned in the last section, the mining can start independently within the SALES table. We are only interested in those items or item-sets with strong sales result. This strong sales requirement is directly translated into the language of data mining as the sufficient support for the item or item-set. So, traditional data mining algorithm such as Apriori algorithm [3] can be used to obtain frequent item-sets.

$T\#$	T_{time}	I_S	set of AD#
$T1103$	1999.03.13	I_4, I_{15}, I_{100}	AD103, AD208
$T1104$	1999.03.15	I_{164}, I_{215}, I_{304} I_{362}	AD107, AD582
$T1105$	1999.03.16	I_{64}, I_{285}, I_{334}	AD449, AD35
$T1106$	1999.03.17	I_3, I_{75}, I_{117}	AD241, AD307, AD259
$T1106$	1999.03.18	I_17, I_{144}, I_{201}	AD14
$T1107$	1999.03.19	I_17, I_{91}	AD613

Table 3: AD-SALES table in Transactional Format

Because the task in hand is to assess advertising effect and the ultimate goal is to gain more profit from this assessment, the support needs not be the count of the appearance of the item-set in transactions. Because the ultimate goal of the advertising effect assessment is to earn more profit from the sales, a monetary figure can be used as the support of the rule. If the support is measured by profit, a item-set with larger profit per unit sale needs less transactions to reach a pre-set threshold of support than a item-set with smaller unit profit.

Another difference in mining operation between our proposed method and previous boolean association rule mining is that for each potential frequent item-set S_I to be tracked during scanning of SALES table, the transactions which contain S_I should be remembered.

Assuming that the SALES table has n transactions and there are no more than P potential frequent item-sets to be tracked during the scanning, then, a bit-matrix BIT with n rows and P columns can be set up. Each transaction T takes a row i and each potential item-set S_I takes a column j in BIT. If S_I appears in T, the (i, j) bit in BIT is set to "1". Otherwise it is set to "0".

Under the circumstance that the appearance of item-sets in transactions in the SALES table must be recorded, the bit-matrix is the smallest data structure to accomplish that.

The scope of searching can be limited by business decisions. For example, if it is determined that we are only interested in the effect of advertisements to some individual items, then only those item-set of size 1 containing these items are selected as potential frequent item-sets.

Optionally, after frequent item-sets are mined, the AD-SALES table can be trimmed. For each item in a transaction in SALES, if this item is not in any frequent item-set, then both this item (in SALES) and its associated advertisement (in AD) can be deleted from AD-SALES table. This trimming requires one scanning of AD-SALES table and may substantially cut the size of the AD-SALES. On the other hand, if all or most of the items in SALES are in some frequent item-sets, this trimming has no or little effect on reducing the size of the AD-SALES table. Consequently, this trimming may be skipped.

The juxtaposition of the AD-SALES table and the bit-matrix BIT is a new table. For convenience of reference, let this new table be called AD-SALES-BIT table.

The mining of frequent item-sets in SALES table greatly reduces the data volume to be processed in the subsequent steps of advertising effect assessment.

5 Rule: Item-Set $\Rightarrow$ Advertisement-Set

This section discuss what type of association rule is needed for assessing advertising effect. Suppose boolean association rule [2] is used. Assume an association rule A: X $\Rightarrow$ Y is discovered with support 10% and confidence 56%, where X = {AD382, AD413} and Y = greeting cards. This rule may be interpreted as: 10% of the total sale transactions fall within the EEP of advertisements AD382 and AD413, and whenever these two advertisements run, there is 56% probability that a customer will buy greeting cards.

The trouble with boolean association rule is: greeting cards may be popular items among buyers and many more customers buy greeting cards with or without the advertisement.

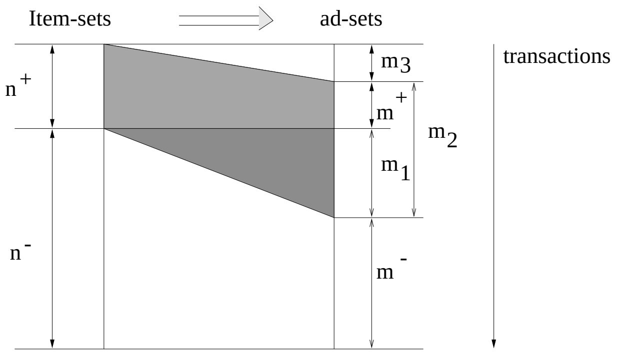

Figure 1: An I-A Rule

Only a negative association can remove such a possibility: if during the time period AD382 and AD413 are not run (more precisely, the time period outside the EEP of these two advertisements), transactions usually do not contain greeting cards, then it is much more reliable to say that advertisements AD382 and AD413 did have an effect on the sale of greeting cards.

An association rule useful for assessing advertising effect is defined as follows. It is an association A: {item-set} $\Rightarrow$ {AD-set}. For each tuple in AD-SALE table, {item-set} takes one of two possible values: appear (+) or not appear (-); {AD-set} also takes one of two possible values: being run (+) or not being run (-). Let n^+ be the number of tuples contain {item-set} and n^- be the number of tuples do not contain {item-set}. Also, let m^+ be the number of tuples that contain {item-set} as well as {AD-set}; let m^- be the number of tuples that contain neither {item-set} nor {AD-set}. The support of the rule is defined as m^+ and the confidence of the rule is defined by two ratios: m^+/n^+ and m^-/n^-, which are called positive and negative confidences, respectively. The quantitative relationship among parameters of an I-A rule is shown in Fig.1. m_1 is the number of transactions that do not contain S_I but contain SA. m_2 is the number of transactions that contain SA. m_3 is the number of transactions that contain S_I but not SA.

I-A rules can be mined as follows. For each frequent item-set S_I, the ad-set S_A that advertise at least some of the items in the item-set is a fixed and usually small set. After frequent item-sets are discovered and their actual appearance in transactions are recorded in a bit-matrix BIT as described in Section 4, the AD-SALES-BIT table is scanned. The number of transactions in which any advertisement or any subset of advertisements SA in S_A (which is the set of advertisement that contain items in S_I) appears together with the appearance of S_I is accumulated (by checking BIT matrix in the AD-SALES-BIT table) as m^+ for possible I-A rule $S_I \Rightarrow SA$. If the positive confidence m^+/n^+ exceeds a pre-determined threshold pc_0 for positive confidence for an advertisement or a subset of advertisement SA of S_A, then $S_I \Rightarrow SA$ is a candidate of I-A rule. After this scanning, those potential I-A rules with insufficient positive confidences are excluded from further consideration.

Then, the AD-SALES-BIT table is scanned again. Each candidate rule $S_I \Rightarrow SA$ is examined during the scanning and the number of transactions that neither S_I nor SA appears is accumulated by checking the BIT matrix in the SALES-BIT table. After scanning, this number is the m^- of the candidate I-A rule. If the negative confidence m^-/n^- exceeds a pre-determined threshold nc_0 for negative confidence, then, we accept $S_I \Rightarrow SA$ as an I-A rule.

The above mining order cannot be reversed, nor can any of the steps be skipped. Without first mining frequent item-sets, the number of item-sets to be considered is simply too large to be worked on.

Also, for each I-A rule, its m^+ must be computed before its m^-. Although searching for I-A rule based on frequent item-sets reduces the search scope, the number of potential I-A rules, in usual cases, are still too large for actual computation. To elaborate this point, we should in mind that each item in each frequent item-set may have several corresponding advertisements and the set SA of all the advertisements corresponding to the item-set S_I usually have a very large power set PSA, i.e., the set of all subsets of SA. For example, if $\mid SA \mid = 4$, then $\mid PSA \mid$ is $C_4^4 + C_4^3 + C_4^2 + C_4^1 = 1 + 4 + 6 + 4 = 15$; when $\mid SA \mid = 6$, $\mid PSA \mid = 63$; when $\mid SA \mid = 10$, $\mid PSA \mid = 1023$.

So, it is computationally infeasible to pick each and every subset SSA from SA, and tracking $S_I \Rightarrow SSA$ during the scanning of SALES-BIT table. But to count m^- requires exactly just that. For example, if S_I has five items, each item is advertised in two advertisements and none of these advertisements advertise more than one item. Then, $\mid SA \mid = 10$ and $\mid PSA \mid = 1023$. This means 1023 sets of advertisements have to be checked for each transaction that does not contain S_I.

When scanning to accumulate m^+, only those advertisements actually appear in the row where S_I appears are checked. This greatly reduces the amount of computation. Again, using the above example, for a transaction that contains S_I, the number of advertisements advertise items in S_I at the time of sale is usually less than 10 (usually, for economical reason, not all advertisements are kept running at all times). If there are only four advertisements in SA are effective (see Section 3) at the time of sale, then, only 15 (instead of 1023) sets of advertisements need to be checked. This saving of computation coupling with the fact that the number of transactions containing S_I is usually much less than the number of transactions which do not contain S_I, results in order of magnitude reduction in computation.

To conduct next step of analysis, the transactions that both S_I and SA of an I-A rule $S_I \Rightarrow SA$ appear should also be remembered. This can be done by setting up another bit-matrix B to record the appearance of ad-sets. B has n rows and A columns. n is the number of rows in AD-SALES-BIT table and A is the number of ad-sets (frequent ad-sets) in I-A rules. Each row of B corresponds to a transaction in AD-SALES-BIT table and each column of B corresponds to a frequent ad-set. Similar to BIT (the bit-matrix for frequent item-set), B can also be juxtaposed with the AD-SALES-BIT table as shown in Fig.2. Each discovered I-A rule links a frequent item-set S_I to a frequent ad-set SA. A column of n (n is the number of transactions in SALES table) bits can be used to indicate the appearance of both S_I and SA: "1" indicates

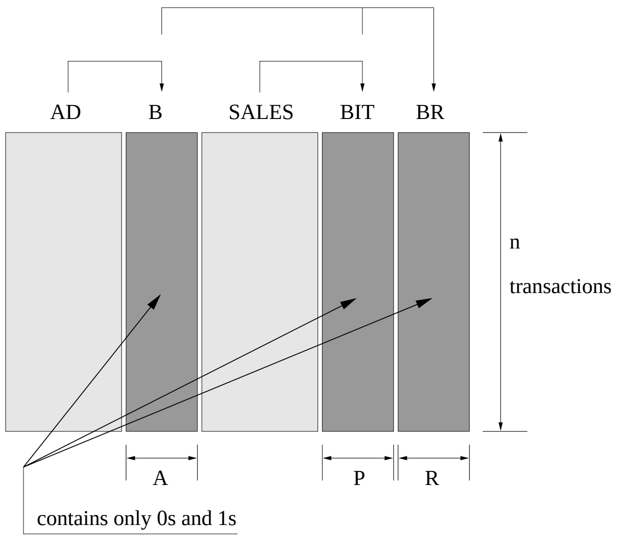

Figure 2: An ABSBB table

that both S_I and SA appears in a transaction and "0" otherwise. If there are R I-A rules be discovered, then a bit-matrix BR with n rows and R columns can be made to associate the transactions to I-A rules. Let the resulting table be called an ABSBB (AD-B-SALES-BIT-BR) table. An ABSB table has five sub-tables: a table AD of ad-sets, a bit-matrix B for frequent ad-sets, a table SALES of transactions, a bit-matrix BIT for frequent item-sets, and a bit-matrix for I-A rules. The ABSBN table contains the same number of rows as the AD-SALES-BIT table.

The arrows at the top of Fig.2 represent sources from which some parts of ABSBB table are generated: B represents frequent ad-sets and is generated by scanning data in AD; BIT represents frequent item-sets and is generated by scanning data in SALES; BR represents I-A rules and is generated by scanning data in both B and BIT.

Additional data structures are needed to interpret the contents of B, BIT, and BR. For example, the seventh column in BR may represent an I-A rule from an item-set S_I and an ad-set SA, which are represented by the 11-th column in BIT and the third column in B, respectively. The fact that an I-A rule from the 11-th column in BIT to the third column in B is recorded in the seventh column in BR should be stored somewhere in a separate data structure. We do not elaborate on the details of these data structures because the need for their existence is obvious and their structures and contents can be easily figured out from B, BIT, and BR. Theses additional data structures are many orders of magnitude smaller than B, BIT, or BRs in large databases.

6 Rule: Advertisement-Set ⇒ Item-Set

The goal of assessing advertising effect is to determine which advertisement have successfully influenced which item, that is, we are actually interested in rules such as $SA \Rightarrow S_I$ (A-I rule) instead of $S_I \Rightarrow SA$ (I-A rule).

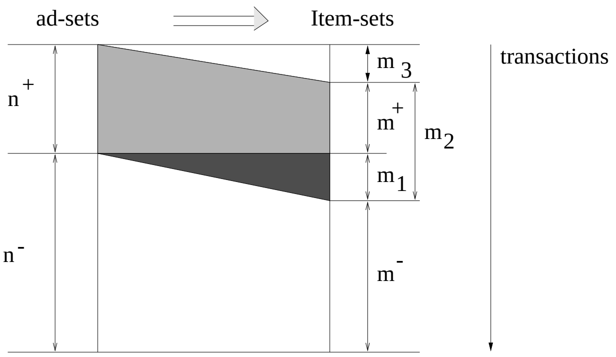

Figure 3: An A-I Rule

In principle, the A-I rule to be defined should be able to tell, quantitatively, to what extent an ad-set SA affect the sales of an item-set S_I. When SA appears on a row in the AD part of an ABSBB table, the corresponding transaction in the SALES part of the ABSBB table may or may not contain S_I. As shown in Fig.3, an A-I rule is described by four numbers: m^+, m^-, n^+, and n^-. Similar to I-A rules, m^+ of an A-I rule is the count of transactions contain both SA and S_I and m^- is the count of transactions contain neither SA nor S_I. n^+ and n^- are defined differently between I-A rules and A-I rules. For an A-I rule, n^+ is the number of rows in ABSBB contain SA and n^- is the number of rows in ABSBB that do not contain SA. The positive confidence of an A-I rule is defined as m^+/n^+ and the negative confidence of the A-I rule is defined as m^-/n^-. To accept an A-I rule, both positive and negative confidence must exceed some pre-determined thresholds.

A-I rules cannot be mined before I-A rules are mined from AD-SALES-BIT table for the same reason (explained in Section 5) that m^- cannot be computed before m^+: there are simply too many (orders of magnitudes more) ad-sets to be checked.

After I-A rules are mined, the mining of A-I rules become possible and can be conducted as follows. The ABSBB table obtained after mining I-A rules are scanned. For each frequent ad-set (which is the consequent of some I-A rule $S_I \Rightarrow SA$ and is represented in the B part of the ABSBB) SA, all corresponding (the correspondences are recorded in the BR part of the ABSBB) frequent item-sets (which are recorded in the BIT part of the ABSBB) are checked during the scanning of the ABSBB table.

Each frequent ad-set SA may correspond to several frequent item-sets (such as S_I) as recorded in the BR part of the ABSBB table. For each SA, counts n^+ and n^- are obtained after scanning of ABSBB. For each (SA,S_I) pair, counts m^+ and m^- are also obtained after scanning. If both m^+/n^- and m^-/n^- are above some pre-determined threshold pc_0 and nc_0, respectively, then $SA \Rightarrow S_I$ is an acceptable A-I rule.

The computational part of the advertising effect assessment is completed when all such A-I rules are discovered. The rest of the work is to interpret the A-I rules by business analysts. As an example, an A-I rule $SA \Rightarrow amodelofcar$ has $m^+/n^+ = 70\%$ and $m^-/n^- = 80\%$ may be interpreted as whenever the set of advertisements SA was run, there is a 70% chance that a car dealership would sell some cars of the advertised model and whenever the set of advertisements SA was not run, there is a 80% chance that a car dealership would not sell any car of that model. So, if it is not too costly to run SA, based on this A-I rule, it looks like that the advertisements SA did have a significant impact on the sale of that model of cars and it would be a good idea to keep running this set of advertisements.

7 Discussion: Periodical and Long-Running Advertising

In previous sections, each assessment step is presented. For simplicity of discussion, we did not raise the issue of periodical advertising. In real life sales activities, sales volume is affected by many factors, such as seasons, quality of the items, changing work force, etc. Suppose the sales data and advertising data are accumulated during spring and summer seasons, and most advertisings were done in the spring season. For some advertised items, sales volume is high for the spring and low in summer season. It is difficult to tell whether the high sales volume in spring was due to the advertising or due to the good season.

The factors such as seasonal influence in the effect assessment can be largely eliminated when assessing periodical advertisements. Periodical advertising, if arranged properly, can eliminate most of the influence factors such as weather, season, holiday or weekend, basically by averaging them out. For example, if an advertisement was run many times in spring, and also was run many times in summer. Though in general spring sale volume is higher than summer sales volume, the assessment (using steps in Sections 4 through 6) may show that in spring season, whenever the advertisement was run, the sales are much better than the time when the advertisement was not run.

If the same pattern is detected for data in summer season, then, it is pretty reliable to say that this advertisement had a good impact on the sale of the items.

So, our proposed assessment procedure is better suited for assessing the effect of periodical or long running advertising. We propose to break the SALES table into several segments. The assessment is done independently for each segment of data. The overall assessment is taken as the consensus of the assessment results from the several segments. This will achieve a much more reliable assessment then applying the assessment procedure once to the entire SALES table.

8 Discussion: Validity of the Proposed Method

Advertising effect remains largely an art, not a science. What we can provide by mining S-I rules is to reduce the amount of data that requires the attention of business analysts. After applying the proposed assessment procedure, the massive sales data and advertising data are reduced to some A-I rules. It is much faster and easier to analyze these A-I rules than directly examing massive sales data.

The validity of each step of the assessment procedure has been discussed in previous sections each time an assessment step is described. In this section, we take a backward approach to further explain the proposed method.

The task of this advertising effect assessment is to find some rules, to tell business analysts that when some advertisements are run, what is their effect on the sale of some items. Because sales volume is the result of many factors, such as the quality of the items, weather, holiday, competitor influences, etc, the effect of advertisement alone is very difficult to assess.

It is far from enough to just find out (as a traditional association rule would) that whenever a set of advertisements SA was run, the sales volume of items S_I are high. It is the difference that makes more sense to tell the effect of the advertising. Based on past sales and advertising data, the most reliable way to assess the advertising effect is to compare the difference of sale between the period that advertising was run and the period that advertising was not run. Based on this understanding, A-I rule is defined as the goal of this assessment procedure.

However, directly mining A-I rule means for every possible subsets of advertisements, and for every possible subsets of item-sets that advertised by these advertisements, AD-SALES table should be scanned to accumulate the necessary count counts to determine whether this is a good A-I rule. An item-set of k has $2^k - 1$ subsets.

This is an NP-hard problem and essentially insolvable for reasonably large problems. For example, when $k = 30$, $2^k - 1 = 1.07 \times 10^9$.

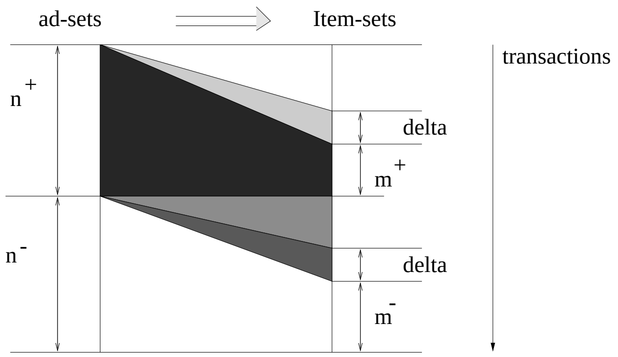

Figure 4: Correlation between Strengths of A-I Rule and I-A Rule

To find a feasible way to compute A-I, we can take advantage the fact that when the association from an ad-set SA to an item-set S_I is strong, then so is the association from S_I to SA. As shown in Fig.4, this correlation of rule strengths can be quantitatively explained as follows. If the number of transactions containing SA as well as the number of transactions containing S_I remain the unchanged and the positive confidence m^+/n^+ of $SA \Rightarrow S_I$ is increased to $(m^+ + \delta)/n^+$, then its negative confidence will increase from m^-/n^- to $(m^- + \delta)/n^-$. As a result, the positive confidence of the corresponding I-A rule $S_I \Rightarrow SA$ will increase from $m^+/(m^+ + n^- - m-)$ to $(m^+ + \delta)/(m^+ + n^- - m-)$ and its negative confidence will increase from $m^-/(n^+ - m^+ + m-)$ to $(m^- + \delta)/(n^+ - m^+ + m-)$. So, strong A-I rules imply strong I-A rules and vice versa.

Because of this correlation, mining I-A rules can be used as a pre-step to mine A-I rules. Mining I-A rules is much easier than mining A-I rules.

The mining of frequent item-sets is a necessary pre-step for mining I-A rules because (1) without the pruning of selected frequent item-sets, we could again run into the incomputability problem because the number of item-sets is exponential to the number of items; (2) we are only interested in numerically significant item-set anyway.

The biggest uncertainty of the assessment comes from the determination of EEP, the expected effective period of the advertisement. We believe that this is an area where no computers can provide complete help. As described in Section 2, some effects of advertising are purely psychological and it is entirely up to the business analysts to determine how long is the EEP. The bottom lines is, if the assumption of EEP are reasonably good and advertising does have some effect, the result of the assessment will show some trends. In other cases, such as discount is offered in the advertisement, it is easier to determine the EEP: it is the period when discount is offered.

9 Discussion: Computational Cost

As discussed in the last section, part of the reasons the assessment procedure has to go through the steps of finding frequent item-sets, finding I-A rules, and eventually finding A-I rules is to reduce the computational (both CPU computation and disk IOs) cost of the assessment. These carefully designed steps in the assessment procedure reduce the computational cost to an manageable magnitude. At least, it is no longer an NP-hard problem.

With all those cost-reduction steps, this assessment is no doubt still a significant computational task. The joining of AD table and the SALES table is the first step requires extensive computation. This is essentially a one dimensional (over the time axis) spatial join and there are many ready methods to speed it up.

The second major step is to mine frequent item-sets. The classical algorithm for this is the Apriori algorithm [3]. Recently, some improvements have been proposed to speed it up. Among the improvements, a method to mine the m best rules could be particularly useful in reducing the mining work. Also, because the assessment is a practical problem, some business knowledge can be used to reduce the number of item-sets to be tracked. For example, an item-set over five items may simply be useless to business analysts.

The third major step is to mine I-A rules. The selection of thresholds m_0^+ and m_0^- determines the cost in this step. Choosing m_0^+ and m_0^- too low may cause too much extra computation and too high may cause some of the useful A-I rules (will be mined in the next step) to be missed. In this step, it should be emphasized that because of the different prices, profit margins, advertising costs, etc., the thresholds m_0^+ and m_0^- may not be the same for all potential I-A rules.

The fourth major step is to mine A-I rules, which are the result of this assessment. Again, the choice of thresholds m_0^+ and m_0^- determines the cost in this step. As this is the final step and the number of I-A rules should be reasonably small, the choice of thresholds is less hampered by computational consideration and more dominated by business needs.

This assessment procedure also requires large temporary storage space for B, BIT and BR bit-matrices in ABSBB table (see Section 5). Because these are bit-matrices, the additional space needed should not be too large to be accommodated in the system, especially when the number of frequent item-sets is limited.

10 Does the Advertising Worth Its Cost ?

Running advertisements incurrs cost and is expected to bring extra sales and resulting in additional profit. However, if the additional profit is too small compared to the cost of running advertisement, the advertisement is not worth running. So, it is necessary to compute the benefit of running an advertisement. This can be done from discovered A-I rules.

For a discovered A-I rule $A \Rightarrow I$, let m^+/n^+ be called the sales ratio of I in EEP of A and m^-/n^- be called the sales ratio of I in non-EEP of A. Sales ratio represents the percentage of transactions containing a specific item or item-set in a specific period.

The additional revenue of increased sale of I caused by running advertisement A can be computed as R = $(m^+/n^+ - m^-/n^-) \times n^+ \times$ *unit price*. The increased profit (without including the cost of running advertisement) can be computed as P = $(m^+/n^+ - m^-/n^-) \times n^+ \times unit\ profit$. The profit after deducting the cost $COST_{ad}$ of running advertisement is P - $COST_{ad}$.

The way to compute the cost of discount type of advertisement is a little more complicated: the cost of discount type of advertising is the sum of running advertisement (which tells customer there is a discount being offered) and the discount multiples the number of items sold under discounted price. The more items sold under the discounted price, the bigger the cost of offering the discount. The extra profit for discount type of advertisement can be computed as $P_{discount} = m^+ \times p_{discount}$ - $m^-/n^- \times n^+ \times p_{original}$ - $COST_{ad}$.

11 Optimizing Future Advertising Effect

Section 4 through Section 6 proposes a computational procedure to assess the effect of advertisements. To further save time and effort for business analysts and to more objectively and precisely adjust the advertising strategy, we further propose a method to compute optimal advertising strategy based on the discovered A-I rules.

Suppose the set of advertisements in all discovered A-I rules is $\{A_1, A_2, A_3, ..., A_k\}$ and the set of items in all discovered A-I rules is $\{I_1, I_2, I_3, ..., I_t\}$. For the simplicity of discussion, each A_i is assumed to be an individual advertisement. However, it could be a set of advertisements. Also, each I_i is assumed to be an individual item although it could be an item-set.

As evidenced by the discovered A-I rules, the advertisement A_j benefits the sale of items I_i and generates additional profit (calculated by the revenue minus both the cost of the item and the cost of the sale. This calculation does not include the cost of the advertising). To run an advertisement also incurs a cost. Apparently, it is interesting to business analysts how to maximize the benefit of the advertising with limited advertising budget.

Let x_j denote the cost of running advertisement A_j and y_i denote the additional profit from selling more item I_i when some of the advertisements in $\{A_1, A_2, A_3, ...\}$ were run. Information of y_i can be directly obtained from discovered A-I rules. For example, an A-I rule $A_p \Rightarrow I_r$ had m^+ = 10,000, n^+ = 20,000, m^- = 80,000, n^- = 100,000. This rule says that without advertising, the transaction of sales for item I is 20%, with advertising, the transactions of sales for item I is 50%. The difference between 20% and 50% can be regarded as the benefit of the advertising. If each unit sale of item I brings \$7 profit, then the benefit in terms of profit is computed as (50% - 20%) × 20,000 × 7 = \$42,000.

The way that an advertisement A_j contributes to the sale of I_i is denoted by c_{ij}. $c_{ij} = 0$ if I_i and A_j are not associated by any discovered A-I rules. On the other hand, if I_i and A_j are associated by some A-I rules, then $c_{ij} > 0$ and its value is to be determined by further computations.

With above notations, the way the sales of $\{I_1, I_2, I_3, ...\}$ are affected by $\{A_1, A_2, A_3,$ $...\}$ can be written in the following linear system:

$$y_i = \Sigma_{j=1}^{t} c_{ij} \times x_j, \qquad i = 1, 2, ..., k. \tag{1}$$

Let X=$\{x_j\}$, Y=$\{y_i\}$, and C=$\{c_{ij}\}$, then above system can be written as Y = AX. Solving this system, we obtain C.

Suppose the total budget for advertising is K, This condition can be written as $\Sigma_{j=1}^{t} x_j \leq K$. Now, the task of optimizing the advertising strategy can be attributed to a typical linear programming problem: under the restriction $\Sigma_{j=1}^{t} x_j \leq K$, to find the best allocation. There are many mature computational procedures and packages for solving linear programming problems. Solving this linear programming problem, we obtain the best advertising strategy (which advertisement should be allocated how much from the fund of total size K) under the restriction of that the total fund for advertising should not exceed K.

Now the question is: how reliable is it ? Can we trust the result of this computation ? Assuming that the result from assessing the advertisement effect, i.e., the A-I rules, are sufficiently reliable, the reliability of the optimizing procedure as proposed in this section can be mathematically described as the stability of the system Y = CX. The stability of Y = CX can be determined by diagonalizing C to a diagonalized matrix D. If D is dominated by diagonal elements, then the solution of Y = CX is stable (not easy to change due to small changes in C). We will not elaborate on this for the space limit of the paper. The point we want to deliver about this reliability discussion is that the reliability is numerically computable and have mature mathematical procedure to accomplish the computation.

12 Conclusion

In this paper, a computational procedure for assessing advertising effect as well as a computational procedure for optimizing advertising effect is proposed. The assessment procedure is essentially a multi-step association rule mining procedure. The association rules, I-A rules and A-I rules, used in the assessment procedure are specially defined to suit the task of advertising effect assessment. The mining of I-A rule is an intermediate step towards the mining of A-I rules, which are the final result of this mining procedure.

The design of this assessment procedure is dominated by the considerations of (1) usefulness of the computational result; (2) computational feasibility. Without careful pruning, the computation can easily turn into an NP-hard problem and become essentially uncomputable for reasonably sized problems. So, the mining of useful A-I rules must first be transformed into the mining of I-A rules. The validity of the transformation is provided in Section 8. However, Because it is possible that even the mining of I-A rules can be NP-hard, in the assessment procedure, the mining of I-A rules is based on frequent item-sets.

The mining of frequent item-sets is a potential NP-hard problem itself. However, past experiences show that using fast algorithms such as Apriori [3], the mining can usually be completed within reasonable time. New improvements to Apriori such as the method to mine the k best rules can completely control the complexity of frequent item-sets mining.

The proposed optimizing procedure relies on the result of the assessment from the first proposed procedure. The optimizing procedure consists of a step to solve a system of linear equations and a step based on linear programming. Both are computationally intensive method. However, these computations are based on discovered A-I rules and the amount of computation can be controlled by the number of I-A rules used. The reliability of this optimizing procedure depends entirely on the stability of matrix C and can be computed mathematically.

Both assessing and optimizing advertising effect are extremely important issues for business communities. The automation of these task, such as the proposed procedures in this paper, will have a significant impact in marketing operations in business.

References

[1] G. P.-Shapiro: Discovery, Analysis, and Presentation of Strong Rules. In Knowledge Discovery in Databases, Editors G. P.-Shapiro and W. J. Frawley, pp. 229-248, AAAI Press/MIT Press, 1991.

[2] R. Agrawal, T. Imielinksi, A. Swami: Mining Association Rules between Sets of Items in Large Databases. In Proceedings of (ACM SIGMOD) Int'l Conf. on Management of Data, 1993, 207-216.

[3] R. Agrawal, R. Srikant: Fast Algorithm for Mining Association Rules. In Proceedings of the 20th Int'l Conf. on Very Large Database, 1994.

[4] R. Srikant and R. Agrawal: Mining Quantitative Association Rules in Large Relational Tables. In Proceedings of the (ACM SIGMOD) Int'l Conf. on Management of Data, 1996, pp. 1-12.