Item Selection By "Hub-Authority" Profit Ranking

Ke Wang
Simon Fraser University
wangk@cs.sfu.ca

Ming-Yen Thomas Su
Simon Fraser University
tmsu@cs.sfu.ca

ABSTRACT

A fundamental problem in business and other applications is ranking items with respect to some notion of profit based on historical transactions. The difficulty is that the profit of one item not only comes from its own sales, but also from its influence on the sales of other items, i.e., the "cross-selling effect". In this paper, we draw an analogy between this influence and the mutual reinforcement of hub/authority web pages. Based on this analogy, we present a novel approach to the item ranking problem.

We apply this ranking approach to solve two selection problems. In *size-constrained selection*, the maximum number of items that can be selected is fixed. In *cost-constrained selection*, there is no maximum number of items to be selected, but there is some cost associated with the selection of each item. In both cases, the question is what items should be selected to maximize the profit. Empirically, we show that this method finds profitable items in the presence of cross-selling effect.

1. INTRODUCTION

In real life applications, it is often required to rank items with respect to some notion of "profit": a retailer ranks sales items by net profit; a webmaster ranks topics for inclusion in a web page; a soccer coach ranks players by desirability for a tournament, etc. We assume that a collection of historical transactions about items and profit is available to the ranking problem. A transaction contains several items and a non-negative profit for each item. For example, a transaction in the retailing business contains the items checked out at the cashier by a customer, and such transactions convey the historical information on items sold together and profit generated by each item in each transaction.

The difficulty of the ranking problem is that items are not independent of each other in profit generation: the profit of one item not only comes from its own sales, but also from its influence on the sales of other items. Some item itself does not generate a large profit, but it plays some role for other items to generate a good profit. In this paper, it is this kind of "overall profitability" of each item that is used to rank items. For discussion purposes, we consider the retailing business where the terms "item" and "profit" come from, and we refer the mutual influence between items as the "cross-selling effect". However, the framework presented is applicable to other application areas where proper notions of "transaction", "item" and "profit" are defined. We present a solution to this ranking problem.

Related to the ranking problem are two selection problems. In *size-constrained selection*, there is a maximum number of items that can be selected. In *cost-constrained selection*, there is some selection cost associated with the selection of each item. For example, the cost could be the charge for space and an item whose profit does not cover the selection cost generates a negative net profit. In both selection problems, a collection of transactions is given, just like in the ranking problem. The question is what items should be selected to maximize the net profit. We present solutions to these selection problems.

The association rule problem [2, 3] has a similar intention of capturing association between items. However, it is not clear how association rules can be used to solve the ranking problems and selection problems considered here. Recently, Brijs et al. [5, 4] made a move in this direction. They proposed a model celled PROFSET for the size-constrained selection. PROFSET factors in the cross-selling effect by identifying "purchase intentions" in each transaction, in the form of disjoint "maximal frequent itemsets". A *frequent itemset*, introduced in [2, 3], occurs in some specified minimum number of transactions. A *maximal frequent itemset* is a frequent itemset that has no frequent superset. PROFSET then selects "purchase intentions", instead of individual items, using a 0-1 programming system constrained by the specified size.

The approach of PROFSET has several drawbacks. First, PROFSET does not consider the strength of the relationship between items, i.e., the notion of confidence, since it only handles frequent itemsets, i.e. the notion of support. Second, maximal frequent itemsets often do not reflect "purchase intentions" because they do not occur as frequently as their subsets. It is also questionable to assume that the "purchase intentions" in a transaction are disjoint. Third, PROFSET provides no relative ranking of selected items, which is very useful in many situations. Fourth, PROF-

*Research was supported in part by research grants from the Natural Science and Engineering Research Council of Canada

SIGKDD '02 Edmonton, Alberta, Canada

SET keeps track of frequent itemsets, and the size of the 0-1 programming system grows rapidly as the number of such itemsets grows. Finally, if the size constraint changes, a new 0-1 programming system must be solved. This requirement makes it hard to address the cost-constrained selection where there is no size constraint.

The rest of the paper is organized as follows. Section 2 outlines our approach and the issues to be addressed. Section 3 provides an overview of a web page ranking algorithm, HITS. In Section 4, we present the detail of our ranking method. In Section 5, we apply the ranking method to solve the two selection problems. Section 6 presents an empirical study. Finally, we conclude the paper.

2. OUR APPROACH

Our ranking method is motivated by the mutually reinforcing relationship of hubs and authorities adopted in the web pages ranking algorithm HITS [8]. The well known search engine Google (http://www.google.com) has exploited this relationship for ranking query results on the web.

The main idea of HITS is to regard the presence of a hyperlink from page i to page j as conferring authority (on a given topic) to page j. HITS ranks web pages by exploiting the mutually reinforcing relationship of hubs and authorities exhibited by such hyperlinks: a good *hub* is a page that points to many good authorities; a good *authority* is a page that is pointed to by many good hubs. The fixed point, i.e., the converged values, of these goodnesses gives the hub weight and authority weight of web pages.

We draw an analogy between the cross-selling effect and the mutual reinforcement of hub weight and authority weight. An item corresponds to a web page, and the influence of sales between two items corresponds to a hyperlink between two pages. An item gets "authority weight" if it is a "necessary buy" for other items, in that a customer buying other items tends to buy this item. This is analogous to a web page getting authority weight if it is a "necessary destination" of other pages, i.e., pointed by other pages. An item gets "hub weight" if it is an "introductory buy" for other items, in that a customer buying this item tends to buy other items. This is analogous to a web page getting hub weight if it is a "directory page", i.e., points to other pages.

However, this analogy does not quite solve our ranking problem.

- First, the weight of web pages depends merely on the structure of hyperlinks, but the cross-selling effect heavily depends on the strength of "links" between items. This strength measures the likelihood that buying one item influences buying another item, thus, playing an important role in determining the overall profitability of each item.

- Second, there is no obvious place for incorporating the recorded profit of an item, also called the *individual profit* of an item. In fact, the convergence of hub/authority weights for web pages does not depend on the initialization of such weights (provided that all are positive), but the hub/authority weight of items obviously depends on the individual profit of items. Thus, if we treat the individual profit as the initial weight, the converged weights will always be the same regardless of the individual profit.

- Third, it is not clear that the convergence of hub/authority weights will be retained after incorporating the link strength and individual profit.

In this paper, we first address these issues and present a ranking method in the presence of cross-selling effect. We then apply this ranking method to solve the two selection problems. Finally, we present a method for estimating the profitability of a selection of items before applying it to real life applications.

3. AN OVERVIEW OF HITS

We devote this section to review the HITS algorithm in [8]. HITS (Hyperlink-Induced Topic Search) is an algorithm for ranking the relevance of web pages on a given topic. Given a topic, HITS first identifies a (possibly very large) *base set* of pages that are potentially related to the topic. HITS then computes the hub/authority weights of each page in the base set based on the *mutually reinforcing relationship* of hubs and authorities: a good *hub* is a page that points to many good authorities; a good *authority* is a page that is pointed to by many good hubs. This relationship is exploited via an iterative algorithm that updates numerical weights for each page. The details are given below.

Each page i in the base set is associated with a non-negative *authority weight* $a(i)$ and a non-negative *hub weight* $h(i)$. The pages i with larger $a(i)$ and $h(i)$ weight are "better" authorities and hubs. Before each iteration, the weights of each type are normalized so that their squares sum to 1: $\Sigma_p a(i)^2 = 1$ and $\Sigma_p h(i)^2 = 1$. In one iteration, HITS updates $a(i)$ to be the sum of the hub weight of all pages j in the base set that have a hyperlink to i:

$$a(i) = \Sigma_{j \to i} h(j) \tag{1}$$

and updates $h(i)$ to be the sum of the weights of all pages j that i points to:

$$h(i) = \Sigma_{i \to j} a(j) \tag{2}$$

This iterative step continues until $a(i)$ and $h(i)$ converge to stable values. The convergence follows from a property of the matrix formulation below.

Suppose that all pages in the base set are numbered by $1, 2, \ldots, n$. The *adjacency matrix* B of the base set is an $n \times n$ matrix with entries of either 0 or 1. The matrix entry $B[i, j]$ is set to 1 if page i has a hyperlink to page j; otherwise, it is set to 0. The hub weights and authority weights are represented by vectors $h =< h_1, \ldots, h_n >$ and $a =< a_1, \ldots, a_n >$. The above update rule can now be written as follows:

$$h = B \cdot a$$
$$a = B^T \cdot h.$$

Unfolding these equations once, corresponding to the first iteration, we obtain:

$$h = BB^T h = (BB^T)h$$
$$a = B^T Ba = (B^T B)a.$$

After k iterations, we have

$$h = B \cdot a = BB^T h = (BB^T)^2 h = \ldots = (BB^T)^k h$$
$$a = B^T \cdot h = B^T Ba = (B^T B)^2 a = \ldots = (B^T B)^k a$$

Linear algebra indicates that the sequence of hub/authority weights converges to the *principal eigenvectors* of BB^T and B^TB, respectively [8]. Furthermore, the converged weights are independent of the choice of initial weights [8]. For a more comprehensive background of linear algebra, we refer the reader to [7].

There are two ways to compute the hub/authority weights: either perform the iterative update of hub/authority weights, as in Equations (1) and (2), until the weights become stable, or apply the standard eigen analysis packages such as [9] to BB^T and B^TB to find their principal eigenvectors. Web pages are ranked by the authority weight. [8] pointed out that the iterative update converges rapidly, usually after 20 iterations. One advantage of the iterative update is that the user can control the number of iterations, therefore, controlling the tradeoff between resources and optimality.

4. HUB-AUTHORITY PROFIT RANKING

We adopt the hub-authority idea of HITS to compute the overall profitability of items. Unlike web pages where hyperlinks are given, we need to model links between items and the strength of such links. The presence of individual profit of items presents another new factor. It is a challenge to incorporate these new requirements in a meaningful way and still maintain the convergence of hub-authority weights.

4.1 The item base

Infrequent occurrences often represent random behaviors and should not be counted on to maximize profit. We specify a minimum support to exclude infrequent items. Precisely, the *support* of an item i, denoted $supp(i)$, refers to the percentage of transactions that contain the item. Similarly, the *support* of a pair i and j, denoted $supp(i, j)$, can be defined. An item or a pair is *frequent* if its support is not less than some minimum support $minsupp$. The set of frequent items forms the nodes in our graph. We assume that the number of frequent items is much larger than the number of items to be selected.

4.2 Create the links

Next, we determine the potential cross-selling links between frequent items. Consider frequent items i and j, not necessarily distinct. A link $i \to j$ represents the cross-selling effect from i to j. The meaning and strength of this link are given by the *confidence* of $i \to j$, denoted $conf(i \to j)$, defined as the percentage of the transactions that contain i also contain j, that is, $supp(i, j)/supp(i)$. The following understandings about $conf(i \to j)$ are relevant to the subsequent discussion:

- (Strength) $conf(i \to j)$ represents the degree that the presence of i implies the presence of j, i.e., the degree that j is necessary for i.
- (Authority) If an item j is necessary for many other items i with a large $conf(i \to j)$, item j is a good authority.
- (Hub) If an item i implies many other items j with a large $conf(i \to j)$, item i is a good hub.

For example, on one extreme, if $conf(i \to j) = 1$, j always occurs whenever i does, in which case j is fully necessary for i. If we do not select j, we will lose not only the profit of j, but also the profit of i. We can recognize this necessity by crediting the authority weight of j by the profit of i. On the other extreme, if $conf(i \to j) = 0$, j never occurs when i does, thus selecting j or not does not affect i. In this case, i has no contribution to the authority weight of j.

We create links as follows. For each pair of frequent items i and j, not necessarily distinct, we create a link $i \to j$ if $supp(i, j) \geq minsupp$ and $conf(i, j) \geq minconf$. $minsupp$ ensures that i and j occur together frequently. $minconf$ ensures that there is enough cross-selling effect from i to j. Note that every frequent item i has a link to itself, i.e., $i \to i$, because $conf(i, i) = 100\%$.

4.3 Model the individual profit

The cross-selling effect of a link $i \to j$ depends not only on the confidence $conf(i \to j)$, but also on the individual profit of i. The term "individual profit of i", denoted $prof(i)$, refers to the recorded profit of i in all transaction. For instance, if j is fully necessary for i, i.e., $conf(i \to j) = 100\%$, and if the individual profit of i is \$500, not selecting j will result in the loss of \$500 of i. If the individual profit of i is \$5000 instead, the loss will be 10 times of that, and for that reason, the authority weight of j should be 10 times as high.

To take the role of individual profit into account, an immediate approach is treating it as the initial authority weight in HITS. Unfortunately, this is equivalent to ignoring the individual profit because the final weights computed by HITS, which are completely determined by hyperlinks, are independent of the choice of initial weights [8]. In a sense, HITS finds the densest communities of hubs and authorities, thereby, ignoring any initial weights.

We solve this problem by "incorporating" the individual profit into links. In particular, if there is a link $i \to j$, we credit the authority weight of j by $prof(i) \times conf(i, j)$. Intuitively, $prof(i) \times conf(i, j)$ represents the part of $prof(i)$ lost by not selecting j, which should be viewed as the credit for the authority weight of j, just as a hyperlink to a page j is viewed as the credit for the authority weight of j. With this consideration, we modify the update Equation (1) as

$$a(i) = \Sigma_{j \to i} prof(j) \times conf(j, i) \times h(j) \qquad (3)$$

and modify the update Equation (2) as

$$h(i) = \Sigma_{i \to j} prof(i) \times conf(i, j) \times a(j) \qquad (4)$$

While Equation (3) has the above "crediting authority interpretation", Equation (4) has the following "rewarding hub interpretation": $prof(i) \times conf(i, j)$ rewards i as a good hub to j if i has a high individual profit and a strong link $i \to j$, which makes sense. Alternatively, $prof(j) \times conf(j, i)$ and $prof(i) \times conf(i, j)$ can be viewed as the "quality" of links $j \to i$ and $i \to j$, and a better quality gives more endorsement of the hub/authority reinforcement.

The corresponding matrix formulation is straightforward. We set the entry $B[i, j]$ in the matrix B as follows. If there is a link $i \to j$,

$$B[i, j] = prof(i) \times conf(i, j) \qquad (5)$$

If there is no link $i \to j$,

$$B[i, j] = 0 \qquad (6)$$

Note that $B[i, i] = prof(i)$ for all items i. B is called the *cross-selling matrix*. Since BB^T and B^TB are symmetric,

from [7], the sequence of hub/authority weights converges to the principal eigenvectors of BB^T and B^TB.

4.4 An example

We illustrate the algorithm using an example

EXAMPLE 4.1. Given the following information of frequent items X, Y, and Z,

$prof(X) = \$5$,
$prof(Y) = \$1$,
$prof(Z) = \$0.1$,
$conf(X \to Y) = 0.2$,
$conf(Y \to X) = 0.06$,
$conf(X \to Z) = 0.8$,
$conf(Z \to X) = 0.2$,
$conf(Y \to Z) = 0.5$,
$conf(Z \to Y) = 0.375$.

The cross-selling matrix B is as follows:

	X	Y	Z
X	5.0000	1.0000	4.0000
Y	0.0600	1.0000	0.5000
Z	0.0200	0.0375	0.1000

where $B[i,i] = prof(i)$ and $B[i,j] = prof(i) \times conf(i,j)$. For instance, $B[X,Z] = prof(X) \times conf(X \to Z) = 5.0000 \times 0.8000 = 4.0000$. Here, we assume that $minconf = 0$. B^T is

	X	Y	Z
X	5.0000	0.0600	0.0200
Y	1.0000	1.0000	0.0375
Z	4.0000	0.5000	0.1000

BB^T is

	X	Y	Z
X	42.0000	3.3000	0.5375
Y	3.3000	1.2536	0.0887
Z	0.5375	0.0887	0.0118

B^TB is

	X	Y	Z
X	25.0040	5.0608	20.0320
Y	5.0608	2.0014	4.5038
Z	20.0320	4.5038	16.2600

The hub weight is given by the principal eigenvector of BB^T and the authority weight is given by the principal eigenvector of B^TB. Figures 1 and 2 show the eigenvectors of BB^T and B^TB, in columns, computed by the standard analysis software package [9]. The eigenvalues for X, Y, Z columns are 42.2725, 0.9902, 0.0027, respectively. So the X columns (in bold face) are the principal eigenvectors because they correspond to the largest eigenvalue. The principal eigenvector $(0.767264, 0.165708, 0.619554)$ of B^TB gives the authority weights to items X, Y, Z, thus ranking X, Z, Y in that order.

Interestingly, this ranking is different from the ranking X, Y, Z according to the individual profit. The reason is that Z is 80% necessary for X and 50% necessary for Y because $conf(X \to Z) = 0.8$ and $conf(Y \to Z) = 0.5$. This cross-selling effect increases the overall profitability of Z. Note that we can ignore the sign in the principal eigenvectors because if ω is the principal eigenvector, so is $-\omega$ [7].

	X	Y	Z
X	**−0.996695**	0.080723	0.009106
Y	**−0.080213**	−0.995678	0.046808
Z	**−0.012845**	−0.045923	−0.998862

Figure 1: Eigenvectors of BB^T

	X	Y	Z
X	**0.767264**	0.344655	0.540850
Y	**0.165708**	−0.921226	0.351971
Z	**0.619554**	−0.180432	−0.763936

Figure 2: Eigenvectors of B^TB

5. SELECTION PROBLEMS

We apply the new ranking method to solve the two selection problems. Also, we propose a method for estimating the profitability of selected items.

5.1 Item selection

For the size-constrained selection, the ranking of items suggests selecting the top s items in the ranked list, where s is the size constraint. So we focus on the cost-constrained selection.

For the cost-constrained selection, the selection of items depends on the selection cost. We consider a restricted case where the ranked list can be used to produce a solution. In particular, we assume that the selection cost for all items is uniform (i.e., the same for all items). This uniform cost ensures that the ranking after considering the selection cost remains the same as the ranking before considering it. Thus, a solution to the cost-constrained selection can be obtained by cutting off the ranked list at some point.

The question is: where should the ranked list be cut off. If it is cut off too soon, items that bring positive net profit will be cut off. If it is cut off too late, items whose profit does not cover the selection cost will be selected. In both cases, the net profit is not maximized. Obviously, the cutoff point should maximize the net profit of selected items after considering the selection cost.

Suppose that we can estimate the total profit of a set of selected items without considering the selection cost. Then we have a way to find the optimal cutoff point: we cut off the ranked list at the point that maximizes the difference $(estimated\ profit - \#selected\ items \times selection\ cost)$, where "selection cost" is the (uniform) cost charged for each selected item.

5.2 Estimate profitability

Estimating the profitability of selected items is an important problem by itself. This profitability can be used to derive a solution to the cost-constrained selection, as discussed above, to evaluate the solution produced by different methods, as in Section 6, or simply to tell the user how "good" his/her selection is. What makes this problem interesting, and difficult as well, is the fact that, if some items necessary for a selected item i were not selected, the profit of i will not be fully generated. The key is determining the part of profit that will be generated. Without generating

new transactions with only the selected items available, this information can only be estimated. We present an estimation model based on the confidence and support observed in the given transactions.

Consider a database of transactions T and a set of selected items S. We estimate the profit generated by S as follows. For each transaction, t, in the database, let

$$t' = t \cap S,$$
$$d = t - t'.$$

t' represents the items selected in the transaction t and d represents the items not selected in the transaction t. If d is empty, all items in t are selected, and the profit of t remains unchanged. If t' is empty, no item of t is selected, and t generates no profit. The other case is that d and t' are non-empty. In this case, we estimate the profit of the selected items in t' as follows.

First, for each transaction t, we find the confidence $conf(\neg d \rightarrow \neg a)$ for each item a in t'. This confidence estimates the probability that the customer will not buy item a given that the items in d are not selected. $\neg d \rightarrow \neg a$ are called *loss rules*. Suppose that the original profit of item a in t is $prof(a,t)$. We estimate the loss of the profit on item a in transaction t by $conf(\neg d \rightarrow \neg a) \times prof(a,t)$, due to the absence of the items in d. Thus, the *estimated profit* of item a in t' is $prof(a,t)(1 - conf(\neg d \rightarrow \neg a))$. The *estimated profit* of t is the sum of the estimated profit for all items in t'. The *estimated profit* of S is the sum of the estimated profit for all transactions in the database. This is summarized below.

DEFINITION 5.1. The *estimated profit* of a selection S is

$$\Sigma_t \Sigma_{a \in t'} prof(a,t)(1 - conf(\neg d \rightarrow \neg a)).$$

6. EMPIRICAL STUDY

To evaluate our method, we implemented HAP, PROFSET, and the naive approach in C++. We used mathematic software packages [9, 6] for the eigen analysis of HAP and for the 0-1 programming of PROFSET. The same result was also obtained using iterative computation.

6.1 Datasets

The first dataset is acquired from a large drug store in Vancouver over a period of 3 months. This dataset carries 26,128 different items and 193,995 sales transactions. A transaction contains 2.86 items on average. The relatively short transaction length implies that the cross-selling effect is limited in this dataset. Each item in a transaction is associated with a profit, computed by $(price \times quantity - cost)$ of the item. The "cost" refers to the cost of the item itself, which is different from the selection cost in the cost-constrained selection problem. Every item has a positive profit, with the total profit of $1,006,970.

We also generated a synthetic database of stronger cross-selling effect. This dataset is generated using the IBM synthetic data generator [1] with the following parameters: 1,000 items, 10,000 transactions, 10 items per transaction on average, and 4 items per frequent itemset on average. We generated the profit of items for a single quantity as follows: 80% of items have a medium profit ranging from $1 to $5, 10% of items have a high profit ranging from $5 to $10, 10% of items have a low profit ranging from $0.1 to $1. This is a simplified version of the normal distribution. The exact profit of each item is determined by randomly picking a dollar amount from the respective profit range. We consider only single quantity sales for each item. The total profit in this dataset is $317,579.

	Drug store		Synthetic	
Tran.#	193,995		10,000	
Item #	26,128		1,000	
total profit	$1,006,970		$317,579	
minsupp	0.1%	0.05%	0.5%	0.1%
Freq. items	332	999	602	879
Freq. pairs	39	115	15	11322

Table 1: The summary of the two datasets

6.2 Profitability results

We measure the profitability of each method by the estimated profit of selected items in Definition 5.1

The drug store dataset. The results on the drug store dataset are shown in Figure 3 for minimum support of 0.1%, and in Figure 4 for minimum support of 0.05%. The profit generated and items selected are in the percentage of total profit and total number of items in the dataset. As the number of selected items increases, HAP enlarges its lead over the other two methods. However, the lead is limited because with the average of 2.86 items per transaction, there is no strong cross-selling effect between items. The cross-selling effect increases to some extent at the lower minimum support of minimum support of 0.05%. At the same selection size, the smaller minimum support generates a larger profit, due to more cross-selling effect considered. The y/x (for profit/item) ratio indicates the "profit effectiveness" of selected items. When very few items are selected, this ratio is high because only the most profitable few items will be selected.

The synthetic dataset. The results for the synthetic dataset are shown in Figure 5 for the minimum support of 0.5%, and in Figure 6 for the minimum support of 0.1%. At the minimum support of 0.1%, all three methods perform similarly because there are only 15 frequent pairs (and few frequent itemsets). If this happens, there is not much association among items, and HAP and PROFSET degenerate to the naive approach, i.e., selecting items based on their individual profit. However, at the minimum support of 0.1% (Figure 6), with 11,322 frequent pairs or links, the cross-selling effect becomes strong and HAP generates a considerably higher profit than the other methods after 1/3 of the items are selected. This increase comes from a better selection based on the overall profitability of each item in the presence of cross-selling effect.

Surprisingly, PROFSET performed worse than the naive approach on this dataset. One reason is that, as pointed out in the Introduction, maximal frequent itemsets used by PROFSET do not necessarily represent strong associations. This is particularly so when the minimum support is low, in which case PROFSET tends to select long itemsets that have a small support. In the extreme case of the lowest minimum support, for example, PROFSET will select the whole transaction to represent a "purchase intention". A consequence of this purchase intention model is that, as the

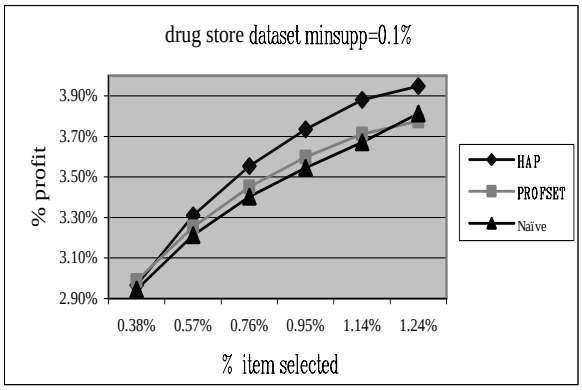

Figure 3: The drug store dataset, minimum support=0.1%

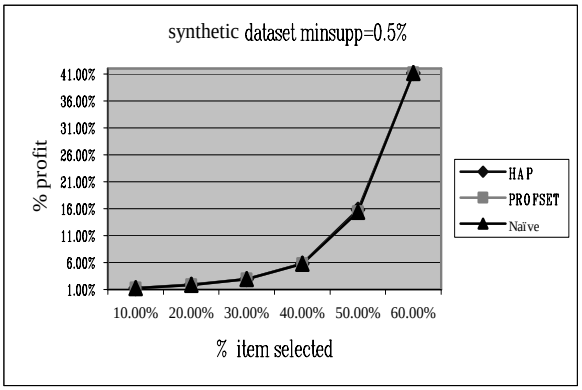

Figure 5: The synthetic dataset, minimum support=0.5%

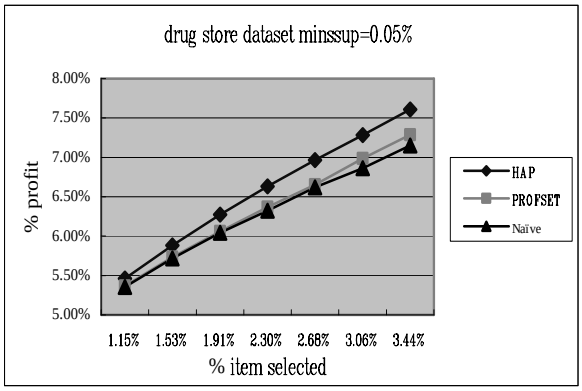

Figure 4: The drug store dataset, minimum support=0.05%

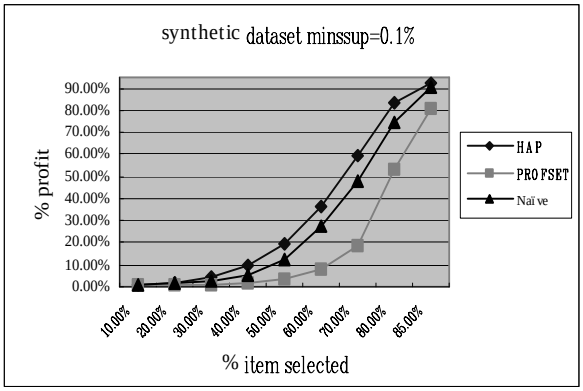

Figure 6: The synthetic dataset, minimum support=0.1%

minimum support decreases, PROFSET often selects many items that are not strongly associated. On the other hand, such items do not necessarily have high profit themselves. As a result, PROFSET loses to the naive approach.

7. CONCLUSION

Our work benefits from HITS's mutual reinforcing relationship between hubs and authorities considered for web page. However, several unique requirements in the context of item ranking present new issues to deal with. In particular, the profit of items and the strength of cross-selling effect are new factors in our problems, but there is no obvious place to model them in the web page ranking framework. Another challenge is that any modification to the web page ranking model must preserve the convergence of the iterative computation. The work presented can be considered as a generalization of the hub-authority web page ranking in that nodes/links have non-uniform weights to start with. We have shown that decision making problems often require such a generalized modeling. The two item selection problems considered are such examples.

8. REFERENCES

[1] R. Agrawal. Ibm synthetic data generator. In *http://www.almaden.ibm.com/cs/quest/Syndata.html #assocSynData*. IBM.

[2] R. Agrawal, T. Imilienski, and A. Swami. Mining association rules between sets of items in large datasets. In *SIGMOD*, pages 207–216, 1993.

[3] R. Agrawal and R. Srikant. Fast algorithm for mining association rules. In *VLDB*, pages 487–499, September 1994.

[4] T. Brijs, B. Goethals, G. Swinnen, K. Vanhoof, and G. Wets. A data mining framework for optimal product selection in retail superma rket data: the generalized profset model. In *ACM SIGKDD*, August 2000.

[5] T. Brijs, G. Swinnen, K.Vanhoof, and G. Wets. Using association rules for product assortment decisions: a case tudy. In *KDD*, pages 254–260, August 1999.

[6] CPLEX. Ilog cplex. In *http://www.ilog.com/products/cplex/*.

[7] G. Golub and C. F. V. Loan. *Matrix computations*. Johns Hopkins University Press, 1989.

[8] J. M. Kleinberg. Authoritative sources in a hyperlink environment. In *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms*, pages 668–677. ACM, 1998.

[9] MINILAB. Minilab. In *http://www.minilab.com/*.