Item-Basket Revenue Maximization

Abhijeet Mohapatra, Abhishek Gupta

ABSTRACT

Setting prices for items is difficult because one needs to look at prices other substitutable as well as complimentary items. In a co-purchasing network, co-occurence of items can be exploited to tweak item prices for revenue maximization. In our report, we investigate the effects of co-occurrence of items on their prices. In general this problem seems to be NP-Hard. We propose heuristics based on properties of the co-purchasing graph and compare them on the basis of their overall revenue gain and convergence to good solutions.

1. INTRODUCTION

In a co-purchasing network, there is a huge opportunity to exploit tweaks in item prices to maximize overall revenue of a service. To get a flavor of such an opportunity consider the following example. Given a pair of items A and B that are bought together frequently. Assume that while purchasing these two items together, buyer's main intention is to purchase item A and since item B was conveniently located near A, the buyer ended up picking B as well. In such a scenario, decreasing the price of A by say 5%, while increasing the price of B by say 10% could possibly generate more revenue.

In a general co-purchase network, the effect of price changes of items cascades to its *neighbors*. Neighbors of an item A in a co-purchasing graph are the items that are co-purchased with A. Consider the following case where items A_1 and A_2 co-occur, A_2 and A_3 co-occur and so on. In this scenario, a pair-wise price changes of A_1 and A_2 needed to maximize revenue influences pairwise changes of A_2 and A_3 and so on. Therefore, changing the price of a an item influences the revenue due to other items.

We assume that we're given a collection of shopping carts with items and their relative prices. We transform this collection into a co-purchasing graph. However, to be able to analyze this problem, we need a model that captures the frequency of items changes with respect to change in the price of the items. This model with its underlying assumptions is described in section 2. Refinements to this model are proposed in section 3. Solution strategies based on local search are discussed in section 4. Section 5 describes the experimental setup and results and section 6 delves into future work.

2. PRELIMINARIES

A co-purchasing graph is an undirected graph $G(V, E)$, where V is the set of items and $u - v \in E$ *iff* items u and v are co-purchased together. Associated with the co-purchaing graph are the following three functions: $price : V \to \mathbb{R}$, $f : V \to \mathbb{R}$ and $g : E \to \mathbb{R}$. The function $f(u)$ represents the number of times item u has been purchased alone and function $g(u, v)$ represents the number of times u and v are co-purchased.

Revenue in a co-purchasing network $G(V, E)$ described above is given by $\sum_{u \in V} price(u) \times (f(u) + \sum_{u,v \in E} g(u, v))$. Now when the price of an item u changes to $price(u) \times (1 + \Delta(u))$, $f(u)$ changes to $f(u) \times (1 + \Delta(f(u)))$ and $g(u, v)$ changes to $g(u, v) \times (1 + \Delta(g(u, v)))$.

2.1 Assumptions

The first assumption that we make is to restrict price changes to within a small threshold. In a real-life setting, factors as cost of manufacturing, marketing, distribution etc. are inherent in setting a price for an item. Hence, it is realistic to assume that the once an item is priced, significant increase or decrease of the price is not permitted. This assumption leads to a simplification when modeling change in frequency of items based on change in prices. This threshold is given by a constant χ such that $\forall u \in V, |\Delta(u)| \leq \chi$.

After restricting ourselves to small price changes, we assume a linear change in frequency with respect to change in price. When the price of an item u changes to $price(u) \times (1 + \Delta(u))$, the relative change $\Delta(f(u))$ in the isolated frequency $f(u)$ is $-\Delta(u)$. When the price of an item u changes to $price(u) \times (1 + \Delta(u))$ and that of v changes to $price(v) \times (1 + \Delta(v))$, the relative change $\Delta(g(u, v))$ in the number of times u and v are co-purchased $g(u, v)$ is $-(weight_u \times \Delta(u) + weight_v \times \Delta(v))$ for some constants $weight_u$ and $weight_v$ respectively.

2.2 Problem Statement

Our goal is to compute the following:

$$\arg \max_{|\Delta(u)| \leq \chi, |\Delta(v)| \leq \chi} \sum_{u \in V} p'(u) \times (f'(u) + \sum_{u,v \in E} g'(u, v)))...(1)$$

$$\text{where} \begin{cases} p'(u) & = & price(u) \times (1 + \Delta(u)) \\ f'(u) & = & f(u) \times (1 - \Delta(u)) \\ g'(u,v) & = & g(u,v) \times (1 - (w_u \Delta(u) + w_v \Delta(v)) \end{cases}$$

Our maximization problem turns out to be an instance of range estimation in quadratic polynomials. In general, range estimation in quadratic polynomial is NP-Hard as discussed in [2] and so it seems that our problem would be NP-Hard as well. Instead of looking for approximation algorithms, we look at the properties of the graph and employ local search procedures to iteratively improve the overall revenue given our model for price changes.

Note the problem is not interesting under any arbitrary assignment of weights w_u and w_v. For e.g. when all items are equally influential, $\forall u \in V, w_u = 1$, any change in price will lead to a reduction in overall revenue. This is obvious from equation 1.

3. REFINEMENTS

In the previous section, we saw that if the weights of both the items were equal then the problem degenerates into a case when the optimal price change is 0 for every item. Here comes the notion of *influence*. Let us define influence, by referring to the example previously used in section 1: Lets say A and B are co-purchased together. On decreasing the price of A by say 5% and increasing the price of B by say 10% we find that we generate more revenue. In fact, $g(A, B)$ increases due to this new price assignment. This entails that change in price of A dominates the value of $g(A, B)$ rather than change in price of B. In this example, A is more *influential* than B.

However, there were problems, discussed in the previous section, when we try to arbitrarily set the *influences* i.e. w_u for each item u. In fact we solve for constraints on w_u and w_v, we observe that if $\forall u, v,\ w_u + w_v < 1$, the total revenue mostly increases with decrease in price of both items, and vice-versa. Therefore, to reduce bias or to distribute it evenly across these two cases, we choose influences $\forall u, w_u$ randomly between 0 and 2.

The only missing piece in the model now is the computation of 'f'' and 'g'' values. We solve this using a procedure that takes as input a set of shopping carts $\{S_i\}, 1 \leq i \leq k$ and constructs a co-purchasing graph $G(V, E)$. Each cart S_i is of the form $\{item_j\}, 1 \leq j \leq l_i$ where l_i is the number of items in shopping cart S_i. We make the following simplifying assumptions:

1. For sake of simplicity, the item frequencies in the cart are assumed to be 1.
2. If two carts S_i and S_j contain an item u then we assume that the same $price(u)$ was paid for both carts.

Translate$(S[][])$
{
for $i \leftarrow 1$ to $|S|$
if $|S[i]| = 1$ then $f(S[i][j]) \leftarrow f(S[i][j]) + 1$
else
$K = |S| + \binom{|S|}{2}$
for $j \leftarrow 1$ to $|S[i]|$
$f(S[i][j]) \leftarrow f(S[i][j]) + 1/K$
for $k \leftarrow j + 1$ to $|S_i|$
$g(S[i][j], S[i][k]) \leftarrow g(S[i][j], S[i][k]) + 1/K$
}

The functions $f(u)$ and $g(u, v)$ now represent frequency scores instead of actual frequency. This change is necessary to ensure that total revenue in $G(V, E)$ is equal to the total revenue across the shopping carts $\{S_i\}$.

4. LOCAL SEARCH HEURISTICS

Since the problem is hard in general, we seek to employ local search heuristics based on properties of the co-purchasing graph. In this study, we investigate the following four strategies:

1. **Strategy1:** Let u be the item whose price change yields maximum revenue gain. Fix the price of u and iterate over the rest of the items.
2. **Strategy2:** Let $u - v$, be an edge in the graph G such that locally changing the price of u and v yields maximum revenue gain. Break ties arbitrarily. Fix the price of u and v and iterate over the graph induced by rest of the items.
3. **Strategy3:** Same as *strategy 2* except we break ties based on the number of triangles in which the edge participates.
4. **Strategy4:** Order the items statically based on the maximum increase in revenue that can achieved by varying respective item prices. Pick the items in this order and locally fix their prices.

5. EXPERIMENTS

We randomly generated 10 co-purchasing graphs with 10,000 items for our experiments. 10 million shopping carts were generated for each case. We executed the four heuristics on each of the co-purchasing graphs and recorded the final revenue, increase in revenue with every price change up to 1,000 price changes. The degree distribution of the randomly generated co-purchasing network was found to mimic a power law similar to Amazon's co-purchasing data set [1]. We set the threshold for price change to be 7%. The influences of the nodes were randomly generated for each experiment.

Surprisingly for every random co-purchasing network, Heuristic 4 performed the best - both in terms of overall revenue and in terms of convergence to a *good* revenue. The heuristics in order of performance are: Heuristic 4, Heuristic 1, Heuristic 2, Heuristic 3 (Heuristic 4 being the best). The results are shown in figure 1.

6. CONCLUSION AND FUTURE WORK

Co-occurence of items in a co-purchasing network induces certain dependencies on *optimal* pricing of items. Intelligently tweaking the prices of the items may substantially increase overall revenue.

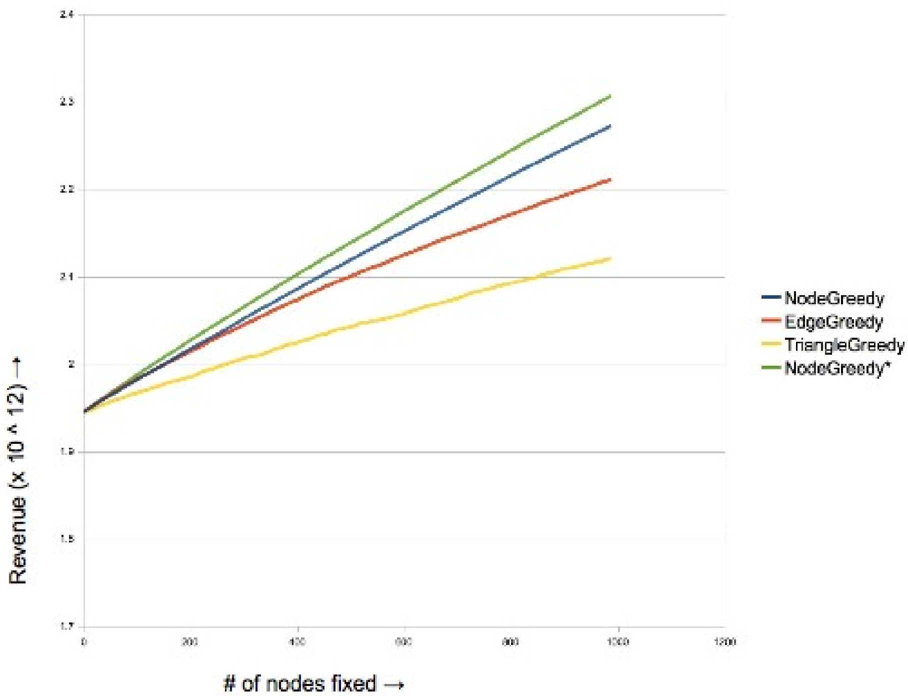

Figure 1: Comparison of performance of Local Search Heuristics

From our experiments over random co-purchasing graphs, we conclude that a greedy static ordering of nodes based on potential revenue maximization leads to higher overall revenue amongst the local search heuristics investigated. If prices of all items are more or less similar, then greedily picking nodes based on their influences and locally fixing their prices would lead to a good overall revenue gain.

We would like to extend the expressivity of the model. For e.g., we would like to be able to capture the following scenario: item A and item B are co-purchased together. However due to price changes of A and B, some co-purchasers of A and B shift to buying A alone, some shift to B and others buy neither. It is important to model each of these transitions to mimic consumer behavior.

We're currently investigating these topics. It would also be interesting to evaluate these heuristics over other real co-purchasing data-sets.

7. REFERENCES

[1] `http://snap.stanford.edu/data/amazon0601.html`.

[2] V. Kreinovich. Range estimation is np-hard for e2 accuracy and feasible for e2-delta. *Reliable Computing*, 8(6), December 2002.