Using Market Basket Analysis to Estimate Potential Revenue Increases for a Small University Bookstore

Bogdan Hoanca
afbh@uaa.alaska.edu
Computer Information Systems

Kenrick Mock
kenrick@uaa.alaska.edu
Computer Science

University of Alaska Anchorage
Anchorage AK 99508, USA

Abstract

Market basket analysis (MBA) is a widely used technique for identifying affinities among items that customers purchase together. MBA metrics are support, confidence, and lift. We show that support and confidence may include misleading information about the nature of the affinity, and that lift is the most useful metric. Starting with the MBA, we use the product affinities to predict ways to increase revenues, and we estimate the magnitude of the possible increases as a function of customer price sensitivity and affinity saturation level. We also point out limitations of the MBA and suggest ways to overcome them. For the case of a small university bookstore, we identify pairings of items that have revenue-increasing potential. Depending on the customers' price sensitivity and affinity saturation level, revenues could be increased by as much as $10,000 or as little as $100 for a $ 410,000 starting level. In particular, we identify pairings where customer price sensitivity might be overcome (one-time situations, for example graduation-related purchases). This case study is the first to provide an actual magnitude of the estimate of potential revenue increases.

Keywords: market basket analysis, scanner panel data, forecasting revenue increases

1. INTRODUCTION

Market basket analysis (MBA) is a well known technique for uncovering affinity relationships among items purchased together. Also known as scanner panel data, the technique identifies associations between items or between categories of items that customers tend to purchase together (complements) or between items customers rarely purchase together (substitutes).

Despite the research on how to best extract and optimize the item associations, nothing has been written about the overall ability of the MBA to increase revenues (or profits), or on the magnitude of such increases for existing businesses. This paper is the first to predict potential revenue increases based on an MBA analysis for a small university bookstore. In the remainder of the paper we refer to increasing the revenues (by focusing on the sales price), but the same approach can be used for optimizing profits (which is done by focusing on the profit margin).

The paper first introduces the basics of MBA and the limitations of MBA metrics in providing information about possible increases in revenues (Section 2). In Section 3, we discuss how the MBA metrics can be used to forecast possible increases in revenues. In Section 4, we then present data from the case study on a small university bookstore, and we discuss the capabilities and the limitations of the augmented MBA technique. We present conclusions in Section 5.

2. MARKET BASKET ANALYSIS

Although the whole reason for conducting a market basket analysis is to increase revenues, a literature review did not uncover any publications on how or to what extent the analysis could lead to potential or actual increases. A possible explanation is the rather sensitive nature of the MBA results, which are likely to be used in-house and closely guarded from competitors, rather than made available for publication.

MBA was first proposed and used in a supermarket in Sweden (Julander, 1992), although most authors cite a later paper on techniques to identify association rules in a database (Agrawal, Imielinski, & Swami, 1993). The technique is typically used for evaluating product affinities for retail stores, but it can also be used to evaluate affinities for any other types of choices: purchases of services, menu choices at a restaurant (Ting, Pan, & Chou, 2010), students' choices of elective classes etc.

More recently, MBA has been combined with additional consumer information including in-store behavior (Schmitt, 2010), visual effects arising from merchandise positioning in the store (including adjacency relations) (Chen, Chen, & Tung, 2006), and choice experiments via mailed surveys (Swait & Andrews, 2003).

Despite the relative simplicity of the MBA, some researchers find it of limited use. The main goal of the MBA is to exploit product affinities by inducing the consumer to purchase additional unplanned products based on an already committed purchase. Some researchers find that unplanned purchasing is less common than generally accepted (Bell, Corsten, & Knox, 2009) and that it depends more on the customers' planning habits and efforts at gathering information, and less on marketing efforts. The direction of the affinity itself is sometimes counterintuitive, as the MBA often fails to distinguish between complements (products customers tend to purchase together) and substitutes (products pairs in which customers tend to substitute one product for another, hence not purchase the two together) (Vindevogel, Poel, & Wets, 2005). Customers often purchase multiple flavors of ice cream or multiple competing brands of soft drinks to address the needs of various family members. This makes functional product substitutes into actual complements. Another problem is the enormous number of transactions in typical larger size businesses, which may make it difficult to carry out MBA unless a well designed subsample of the transactions set is used instead of the entire database (Chandra & Bhaskar, 2011).

The MBA is nonetheless widely used, in part because of its conceptual simplicity. Given a set of N transactions involving two or more items, the MBA starts by identifying those transactions that involve pairs of items. For example, given items A and B, the number of transactions involving item A is N_A, the number of transactions involving item B is N_B, and the number of transactions involving both A and B is N_{AB}. These numbers can be obtained via simple queries in a database of transactions.

The pairing may be for synchronous purchases, where the two items are bought during the same transaction or for asynchronous purchases, where a customer who has purchased item A in the past might be targeted for purchasing item B. For synchronous applications, sales staff might be directed to suggest additional relevant items (item B), given a set of items the customer has already selected in this transaction (item A). For asynchronous applications, the application is to suggest additional purchases over time, for example by sending coupons to get the customer back to the store, sending reminders for consumables given a typical usage or maintenance schedule, or sending offers for relevant accessories for an item previously purchased. Although the synchronous and the asynchronous MBA approaches are similar, there are also significant differences. One of the main differences is in the ability to track purchases over time to a particular customer, which allows for the asynchronous MBA. This customer tracking is typically done via a loyalty program (frequent buyer card, membership store etc).

The MBA is based on three metrics: support, confidence and lift. All three metrics are derived from the transactions record for the business.

Support

The first metric defined for MBA is *support*, which is the probability of an association (probability of the two items being purchased together). Given the number of times items A and B occur together in the same transaction and the total number of transactions as above, the support is $p_{AB} = {}^{N_{AB}}/_{N}$.

Although it would seem that a high support is relevant, this metric has several limitations. The MBA approach is based on analyzing past data to forecast the future sales. Unless the support values are increasing over time, the only forecast that can be made is that future sales will be similar to past ones. High support values do not indicate a potential for increasing revenues.

Moreover, some items may end up being purchased together in a large number of transactions, but not because there is an affinity. Although bread and milk are not necessarily consumed together, customers purchase both items so often that the pairing is likely to have a high support for most supermarkets. The value of affinity relationships is when there is an inherent value in purchasing the items together, because they achieve some synergistic goal and because the pairing is more likely than the chance level.

Finally, the chance of items being purchased together is governed by randomness. For small values of support, it is possible that the appearance of an affinity is the result of a small number of coincidences. As an illustration, the author recalls purchasing a highly technical textbook on Amazon a decade or so ago and noticing a suggestion on the website that customers who purchased that book also purchased the latest Harry Potter volume. While the affinity was deemed statistically significant by Amazon's software at the time, it was clearly just the result of coincidence. To reduce the likelihood of such spurious associations, a minimum support value is typically used in calculations. A support value corresponding to an $N_{AB} \geq 10$ is often used.

Confidence

The next metric typically defined in the MBA is the conditional probability of an item to be purchased, given that another one has already been purchased. For example, a customer who has already placed item A in her shopping cart will have a different probability of purchasing item B than if she had not decided to purchase A. Mathematically, confidence is given by $C_{AB} = {}^{p_{AB}}/_{p_A}$.

Confidence can be used to assess the probability of purchasing item B in the same visit (this is the synchronous approach we mentioned earlier: if item A is in the shopping cart, what is the probability that the customer will also purchase item B at the same time?) Alternatively, confidence could be used to assess the chance that the customer who already owns item A (having purchased it in a previous transaction) may be willing to purchase item B at this time (asynchronous approach).

Again, a high confidence may seem beneficial, but confidence by itself does not indicate that additional revenues are possible. On the other hand, a high confidence may signal to the sales staff that a customer who has already purchased item A is an "easy target" to be sold item B. This is somewhat misleading, as the example of milk and bread shows: the confidence is high that a customer who already purchased bread will also purchased milk, but this is because most customers will purchase milk anyway. As such, not only is there no clear way to increase revenues, but the effort of the sales staff to make another sale is unnecessary, as the customer is highly likely to purchase item B anyway.

Lift

A third metric in the MBA approach is the lift, $L_{AB} = {}^{C_{AB}}/_{p_B}$. Because the lift is a ratio of probabilities, it mitigates some of the problems with the earlier example of milk and bread.

There are two ways to look at lift. First, lift is the ratio of the conditional probability of purchasing B to the simple probability of purchasing B. As such, lift is a measure of how much more likely the customer is to purchase B now that she intends to purchase A, as compared to a customer who is not purchasing A.

If $L_{AB} > 1$, items A and B have an affinity that may lead to additional sales. Returning to the earlier example, assume that the lift for two items is $L_{AB} = 10$. Armed with this information, a sales person who sees the customer with item A in her shopping cart knows that the probability to sell her item B is now ten times higher than the probability of selling item B to a customer who is not intent on purchasing item A. A high lift value might also induce a store manager to place items together (or far apart, to induce customers to walk across the store), to advertise them together, or to bundle them.

Clearly, if $L_{AB} < 1$, the effort to sell item B to the customer is higher than the effort to sell it to a customer who has not purchased A. This might indicate that items are substitutes: a customer who intends to purchase A has no need or interest in purchasing B. Alternatively, the items could be in different segments (a luxury item and a discount version). When there is such a choice, the store manager may choose to discontinue the item with lower revenues.

The other way to look at lift is as a ratio of three probabilities. Substituting the formula for the confidence in the formula for lift leads to $L_{AB} = {}^{p_{AB}}\!/_{(p_A \cdot p_B)}$. From probability theory, if items A and B are independent (customers purchase one independently of whether they also purchase the other), the probability of the two being purchased together is $p_{AB_{independent}} = p_A \cdot p_B$, so $L_{AB} = {}^{p_{AB}}\!/_{p_{AB_{independent}}}$. If $L_{AB} > 1$, items are more likely to be purchased together than just by chance ($p_{AB_{independent}}$), again indicating the affinity, whereas if $L_{AB} < 1$, the items are even less likely to be purchased together than just by chance. In other words, when comparing customer's choices with the choices or a random purchasing agent (an agent that picks up items at random), pairs of items with if $L_{AB} < 1$ will be less likely to appear in the rational customer's shopping cart than in that of the random agent.

The discussion above suggests that finding pairs of items with high lift could lead to increased revenues by selling more items. Still, lift by itself is just an indication of an affinity, but does not show how much additional revenues can be expected. Lift can also be misleading, identifying as complements products that are actually substitutes, but are purchased together for customer dependent reasons (Vindevogel, Poel, & Wets, 2005). Finally, lift is dependent on the number of transactions, which leads to high values for infrequently purchased products. Substituting the number of transactions in the probabilities in the formula for lift leads to $L_{AB} = \frac{N \cdot N_{AB}}{N_A \cdot N_B}$. For a given number of transactions involving items A and B separately and together, the lift increases with the total number of transactions in the set. As such, the more transactions there are in a supermarket database, the higher the value of lift calculated. Standardizing (or normalizing) lift circumvents this problem, as described next.

Normalized lift

A modified metric that addresses the problem above is the normalized lift proposed in (McNicholas, Murphy, & O'Regan, 2008). The number of transactions involving both items A and B must be smaller than the number of transactions involving either item alone: $N_{AB} \leq \min(N_A, N_B)$. Also, the number of transactions involving both items has a lower bound (it needs to be a positive or zero number). When the sum of the numbers of transactions for items A and B exceeds the total number of transactions, some transactions will by necessity involve both items. Hence the lower bound for the number of transactions involving both items is $N_{AB_{min}} \geq max(0, N_A + N_B - N)$.

These values are used to calculate $L_{AB_{min}} = \frac{N \cdot max(0, N_A + N_B - N)}{N_A \cdot N_B}$ and $L_{AB_{max}} = \frac{N \cdot \min(N_A, N_B)}{N_A \cdot N_B}$. With a bit of algebra, it is easy to show that $L_{AB_{min}} \leq 1 \leq L_{AB_{max}}$. With these values, the normalized lift is $\Lambda_{AB} = \frac{L_{AB} - L_{AB_{min}}}{L_{AB_{max}} - L_{AB_{min}}}$. Substituting the numbers of transactions leads to the final formula for the normalized lift $\Lambda_{AB} = \frac{N_{AB} - max(0, N_A + N_B - N)}{\min(N_A, N_B) - max(0, N_A + N_B - N)}$.

This normalized lift can have values in the interval [0, 1]. For small values of *N*, the normalized lift does depend on the total number of transactions, but for large values, once $N_A + N_B \leq N$, the normalized lift is independent of the total number of transactions, $\Lambda_{AB} = \frac{N_{AB}}{\min(N_A, N_B)}$.

The normalized lift also suffers from a major limitation: it does not directly show whether an affinity exists or not. The unnormalized lift is greater than one when the two items have an affinity, which corresponds to a normalized lift $\Lambda_{AB} \geq \Lambda_{AB_{threshold}} = \frac{1 - L_{AB_{min}}}{L_{AB_{max}} - L_{AB_{min}}}$.

In conclusion, the best use of the MBA is to first use the support to determine whether an association has statistical significance (the number of times two items are associated is

©2011 EDSIG (Education Special Interest Group of the AITP)

large enough not to be due to chance alone). The lift can be used to determine if the affinity is positive (complementary pairing) or negative (substitution). Finally the normalized lift gives a measure of the strength of the affinity, independent of the number of transactions in the data set.

This knowledge about the affinities between pairs of items in the data set does not suggest how much more revenues could be obtained. In the next section, we use the MBA metrics to estimate the additional revenues.

3. USING MARKET BASKET ANALYSIS TO INCREASE REVENUES

Given a set of transactions, an opportunity for increasing revenues arises when there are affinities among the items in the data set. At the most basic level, an affinity will allow sales of one item to drive up sales of another. If selling items A and B, with $N_A \gg N_B$, an affinity between the items would allow selling more of item B, by piggybacking on the sales of item A.

While we refer to pairings below, the way to use the results of an MBA can be multifaceted. The information about product affinities from the MBA must be used in pricing strategies, in product placement and in marketing campaigns, if it is to result in any increase in revenues.

As we already mentioned in the section about lift, two items with a positive affinity can be packaged together, could be placed in close proximity or could be priced as a bundle (buy one item and get the other one for a discount). Similarly, items with a negative affinity (substitutes) can either be the object of market segmentation or could increase revenues by discontinuing one of the items, if done properly (Vindevogel, Poel, & Wets, 2005). For simplicity, the analysis here assumes fixed pricing, and does not discuss any pricing strategies. MBA can also be used to adjust prices for individual items or for item bundles to account for affinities.

Impact of lift and affinity saturation level

It is intuitive that the higher the lift, the more likely that additional sales of item B could be made to a customer who has committed to purchasing item A. On the other hand, the potential for additional sales might not be realized if the affinity is saturated for sales of item B. For example, assume that women tend to purchase item B after purchasing item A, but men do not associate items the same way. If most women already purchase the items together, there is little if anything to gain from the association, unless men could also be induced to associate the two items. The saturation of an association could be based on any number of demographic or behavioral traits of the customer, including previous purchases, income etc. The affinity saturation level is typically unknown, although it can be estimated either empirically (for example via customer surveys) or via thought experiments (e.g., modeling customer behavior).

Each possible pairing of two SKUs has a different nature, hence also a different affinity saturation. To complicate matters even further, the affinity saturation can change over time, for example as more customers purchase the items together as a result of a successful marketing campaign.

In the model developed here, the affinity saturation level is a variable ρ that multiplies the potential revenue increase. For a saturated affinity, $\rho = 0$, while for an affinity with maximum potential to increase revenues, $\rho = 1$. For simplicity, we assume affinity saturation level for all the pairings in our data set to be $\rho = 1$. This means that any actual revenue increase will be lower than what we estimate below in the case study.

Impact of relative pricing and customer price sensitivity

Another factor in the ability to increase revenues is the relative pricing of the two items as it relates to the customer price sensitivity. Intuitively, customers who already purchased a high priced item might be somewhat willing to purchase a lower priced related product. When the situation is reversed, customers are usually less likely to purchase a higher priced product that has an affinity with the lower priced product they already intend to purchase. We use a linear model for the likelihood that customers would be willing to purchase more expensive items. The probability of making a purchase of item B if the customer is already purchasing A is $p_{B_{price}} = max(0, 1 - \frac{P_B}{k \cdot P_A})$, where *k* is a measure of the price sensitivity of the customer (while the lower case *p* earlier stands for probability, while upper case *P* here refers to price). The customer is not willing to purchase item B if the price of the item is more than *k* times the price of item A

$(P_B > k \cdot P_A)$. This is a statement about the relative strength of the affinity as compared to the customer's price sensitivity. If the customer was already predisposed to purchase item B, he or she would do so regardless of the pricing, not as a result of the affinity between items A and B.

Impact of competing bundling

A final consideration must include the effect of multiple pairings of items on the customer's choice. Research has shown that too many choices will distract the customer and might lead either to no decision or to a suboptimal decision (Lehrer, 2010). As such, the number of pairings should be kept relatively small or localized, so each individual decision should be made with a manageable number of choices.

Overall impact on revenues

To summarize, pairing two items makes sense when the number of times item A is sold is greater than the number of times item B is sold: $N_A > N_B$. Second, the two items must have a positive affinity (they must be complements) and the affinity must be below saturation level. Finally, the relative pricing of the items must be within the customer's price sensitivity range.

With the considerations above, the additional revenues from pairing item B with as many of the item A as possible will be given by $R_{AB} = (N_A - N_B) \cdot P_B \cdot \rho \cdot max(0, 1 - \frac{P_B}{k \cdot P_A}) \cdot \Lambda_{AB}$. The terms in the formula include in order: the additional number of item B sold $(N_A - N_B)$, the sales price for item B, P_B, the affinity saturation level, ρ, and the price sensitivity probability ($p_{B_{price}}$ from above). Most generally, the affinity saturation level, ρ, and the customer price sensitivity, k, will be different for each product pairing. In the case study below, we assume these parameters to be the same for all product pairings in the data set, for simplicity.

The total impact on revenues is the sum of additional revenues for all the possible pairings.

4. CASE STUDY AND DISCUSSION

We now apply the MBA approach to a set of cash register transactions from the UAA Bookstore. Transaction dates are from February and March 2010. There are 13,916 transactions involving 28,462 line items and 4,332 individual SKUs (stock-keeping units), including textbooks, apparel, snacks and office products. The total sales for the two month period amount to $ 410,000. To protect confidential information, the number of units sold has been multiplied by a randomly generated scaling factor that affects only the overall amounts, but not the MBA affinities or the relative increase in revenues.

For the data set, 17.8% of the SKUs bring in 90% of the revenues and 8% of the SKUs bring in 80% of the revenues. This distribution shows many more active SKUs than the typical 80/20 distribution where 20% of the SKUs bring in 80% of the revenues. A possible explanation is that a university bookstore carries many specialty books that sell in very small volumes, but might be carried because they are written by faculty or by local authors. Also, some of the SKUs tend to be very detailed, differentiating between various colors of a sunburst T-shirt, which artificially increases the number of SKUs. Interestingly enough, transactions for the month of February involve much fewer textbook purchases and show much closer agreement with the 80/20 rule (17% of the SKUs are responsible for 80% of the revenues).

The MBA can be applied at the SKU level or at the category level (i.e., looking for associations for 2% Horizon Organic milk – SKU level – or for milk – category level). The data set available did not include categories, the number of SKUs was too large, and many SKUs were difficult to classify. As such, the MBA we carried out was only at the SKU level. The same analysis at the category level would give additional insights.

There were 17,231 different pairs of items that were sold together in at least one transaction, but only 88 pairs were sold together in at least 10 transactions, 250 were sold together in at least 5 transactions and 586 were sold together in at least 3 transactions. As mentioned earlier, the higher the number of times items are sold together, the more likely that the affinity is statistically significant. We consider all three sets of values for comparison purposes, realizing that probably only data for the item pairs sold in at least 10 transactions is to be trusted.

The data set included the end of the academic year, and most of the strong affinities involve graduation paraphernalia. Graduation is a unique event in the life of a student (and of the parents of that student), so customers might be more inclined to make purchases they would otherwise not be willing to make (recall the earlier discussion about upselling a more expensive item when the customer is already willing to purchase a less expensive one).

Support

When considering the support, three pairings stand out (Table A.1 in the Appendix) with support close to 0.03 (all other pairings have less than 0.01 support):

Tassels and caps
Tassels and gowns
Caps and gowns

These pairings are rather expected, and as explained before, they do not give much information about increasing revenues.

Confidence

The list of the highest confidence pairings includes several with confidence higher than 0.8, but many of the pairings are rather obvious again (masters hood and tassel) as shown in Table A.2 in the Appendix. Many pairings occur a small number of times, and many prices are such that the customer is unlikely to purchase the more expensive item: the first line in Table 2 is for an expensive textbook ($200) as the upsell item for $0.15 test sheets. Clearly, the affinity is coincidental and not significant.

Lift

Lift values range from 0.16 for test sheets and snacks to 242 for envelope seals and announcement cards. Of all the item pairings that involve at least 3 transactions, 567 have lift greater than one, indicating some affinity. Also, 248 are pairings involving at least 5 transactions, and 88 involve at least 10 transactions.

Normalized lift

Normalized lift values range from 0.01 to 1, with 38 pairings having normalized lift greater than 0.9. Recall that normalized lift by itself is not an indication of product affinity. Instead, 567 of these normalized lift values are larger than the threshold (indicating that there is some affinity between the items).

Overall impact on revenues

Table A.3 in the Appendix includes the top revenue generator pairings assuming a price sensitivity coefficient $k = 3$ (customers willing to purchase items priced up to three times higher than the price of the item they are buying). The data shows all the pairings that have a sufficiently large support for statistical significance, regardless of the confidence level. Indeed, some pairings might lead to high revenue increases despite their lower confidence.

Because most of the pairings in Table A3 in the Appendix are related to graduation paraphernalia, we expect customers to be willing to pay more for certain items that are appropriate for the occasion. For example, customers who rent a bachelor gown might be willing to pay for the almost twice as expensive announcement packs. In general, the customer's willingness to pay for more expensive items will depend on the nature of the transaction, but might involve a price sensitivity coefficient as low as zero.

Table 1 below includes possible additional revenues for a series of price sensitivity coefficients ranging from 0.1 to 3. Because not all affinity pairings might be meaningful (given the sometimes small number of transactions involving a certain pairing), we list the possible additional revenues for various numbers of minimum transactions per pairing.

Table 1. Expected additional revenues for various values of the price sensitivity factor (*k*) and of the minimum number of occurrences of a pairing for a statistically relevant affinity

k	Minimum number of occurrences of the MBA pairing		
	3	5	10
0.1	$ 287.97	$ 266.09	$ 58.06
0.3	$ 2,974.39	$ 1,925.20	$ 738.32
1	$ 18,809.62	$ 13,889.18	$ 5,267.52
3	$ 52,410.14	$ 40,039.05	$ 18,327.20

Based on these results, for the pairings we identified in the context of graduation season, a reasonable expected value for the additional revenues would be in the $5,000-$10,000 range, if customers are really less price sensitive. On the other hand, with price sensitive customers, the revenues gain could be lower than $100.

©2011 EDSIG (Education Special Interest Group of the AITP)
www.aitp-edsig.org

An extension of the analysis presented could involve combinations of more than two items. MBA can be used to develop association rules between larger groups of SKUs, when such combinations of several SKUs in one transaction are frequent enough to be statistically significant. Because of the small data set and the relatively low numbers of combinations of two SKUs, this paper did not consider combinations of more than two SKUs.

The authors are currently working with the bookstore on designing a marketing campaign based on the MBA results, to evaluate the saturation level for the affinities, as well as the hypothesis of lower customer price sensitivity for special-occasions purchases. Because most of the revenue increases in the model result from marketing graduation paraphernalia, a test of the model will need to wait for the next graduation season. We will report on the results in a future paper.

5. CONCLUSIONS

Market basket analysis (MBA) is a widely used technique for identifying affinities among items customers purchase together. In this paper we outline the limitations of the MBA metrics in identifying sources of revenue growth, and we suggest ways to overcome the limitations. For the case of a small university bookstore, we quantify the MBA metrics for the most promising pairings of items and we estimate the potential revenue increase.

The data set covered two months in 2010, and included almost 14,000 transactions involving 28,000 line items and close to 4,300 individual SKUs (stock-keeping units), including textbooks, apparel, snacks and office products. Because of the specialized nature of the business, 17.8% of the SKUs bring in 90% of the revenues when considering the textbook selling season, and a more typical 80% of the revenue when considering months in the middle of the semester.

Although close to 18,000 different pairings could be identified, only 88 of them occurred sufficient times to be considered statistically significant. All of these 88 pairings showed an affinity (lift >1). Many pairings suggested selling a more expensive item to a customer who has purchased a less expensive one. Although this would not be very feasible in general, our estimate is that such a pairing might be acceptable to graduating students as a one-time “splurge” for graduation related paraphernalia.

Depending on the customers’ price sensitivity and on the saturation level of the affinities uncovered, revenues can increase by as much as $10,000 or as little as $100, which is 0.025% to 2.5% of the overall revenues.

Acknowledgement The authors would like to thank Alessandra Vanover and Anna Gage at the UAA Bookstore for making available transactional data.

6. REFERENCES

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining Association Rules between Sets of Items in Large Databases. *ACM SIGMOD International Conference on Management of Data (SIGNOD ’93)* (pp. 207-216). Washington DC: ACM.

Bell, D. R., Corsten, D., & Knox, G. (2009, January). Unplanned Category Purchase Incidence: Who Does It, How Often, and Why. *Knowledge@Wharton* .

Chandra, B., & Bhaskar, S. (2011). A new approach for generating efficient sample from market basket data. *Expert Systems with Applications , 38*, 1321–1325.

Chen, Y.-L., Chen, J.-M., & Tung, C.-W. (2006). A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales. *Decision Support Systems , 42* (2006), 1503-1520.

Julander, C.-R. (1992). Basket analysis: A new way of analysing scanner data. *International Journal of Retail & Distribution Management , 20* (7), 10-18.

Lehrer, J. (2010). *How we decide.* Boston: Mariner Books : Houghton Mifflin Harcourt.

McNicholas, P., Murphy, T., & O’Regan, M. (2008). Standardising the lift of an association rule. *Computational Statistics and Data Analysis , 52*, 4712–4721.

Schmitt, J. (2010). Drawing Association Rules between Purchases and In-Store Behavior: An Extension of the Market Basket Analysis. *Advances in Consumer Research , 37*, 899-901.

Swait, J., & Andrews, R. L. (2003). Enriching Scanner Panel Models with Choice Experiments. *Marketing Science , 22* (4), 442-460.

Ting, P.-H., Pan, S., & Chou, S.-S. (2010). Finding Ideal Menu Items Assortments: An Empirical Application of Market Basket Analysis. *Cornell Hospitality Quarterly , 51* (4), 492-501.

Vindevogel, B., Poel, D. V., & Wets, G. (2005). Why promotion strategies based on market basket analysis do not work. *Expert Systems with Applications , 28*, 583-590.

Appendix. Data tables

Table A1. Pairings with highest support values

SKU1 Description	SKU1 price	SKU2 Description	SKU2 price	SKU1 Trans	SKU2 Trans	SKU 1&2	Support
TASSEL GREEN/BGOLD	$ 5.00	GRADUATION CAPS	$ 7.00	522	471	406	0.029175
GRADUATION CAPS	$ 7.00	BACHELOR GOWN	$ 28.00	471	353	346	0.024863
TASSEL GREEN/BGOLD	$ 5.00	BACHELOR GOWN	$ 28.00	522	353	341	0.024504

Table A2. Pairings with highest confidence values (all pairings have confidence = 1)

SKU1 Description	SKU1 price	SKU2 Description	SKU2 price	SKU1 Trans	SKU2 Trans	SKU 1&2
BETTELH/INTRO.TO G	$ 202.10	TEST SHEETS	$ 0.15	4	912	4
PACKAGE B ANNOUNCE	$ 107.95	SHIPPING	$ 15.00	13	198	13
THANK YOU NOTES 25	$ 11.75	TASSEL GREEN/BGOLD	$ 5.00	6	522	6
THANK YOU NOTES 25	$ 11.75	GRADUATION CAPS	$ 7.00	6	471	6
MASTERS HOODS	$ 28.00	SHIPPING	$ 15.00	4	198	4
THANK YOU NOTES 25	$ 11.75	BACHELOR GOWN	$ 28.00	6	353	6
CERT. OF APPRECIAT	$ 16.75	SHIPPING	$ 15.00	9	198	9
THANK YOU NOTES 25	$ 11.75	SHIPPING	$ 15.00	6	198	6
ANNOUNCEMENT COVER	$ 11.75	SHIPPING	$ 15.00	7	198	7
MASTERS HOODS	$ 28.00	TASSEL GREEN/BGOLD	$ 5.00	4	44	4
COMMENCEMENT	$ 1.83	SHIPPING	$ 15.00	11	198	11
SCRUBS V-NECK TUNI	$ 13.95	PORTFOLIO 2 PKT A	$ 0.50	3	19	3
SCRUBS V-NECK TUNI	$ 13.95	REPORT COVER ASST	$ 0.50	3	17	3
FOWLER/POLICY STUD	$ 123.30	SPRING/AMERICAN ED	$ 56.25	7	8	7
KIM/LOST NAMES	$ 14.25	KATSU/MUSUI'S STOR	$ 15.85	3	8	3
TECLA/DIONISIO AGU	$ 25.00	TECLA/FERNANDO SOR	$ 23.00	4	6	4
GIANTMICROBES GIGA	$ 24.95	GIANTMICROBES GIGA	$ 24.95	4	6	4
HAYCOX/FRIGID EMBR	$ 21.95	HAYCOX/ALASKA AMER	$ 18.95	3	4	3
JOHNSON/HOW DO I L	$ 17.75	COREY/I NEVER KNEW	$ 135.45	3	4	3
REF CHART ALGEBRA	$ 5.95	REF CHART ALGEBRA	$ 5.95	6	7	6
REF CHART MED TERM	$ 5.95	REF CHART MED TERM	$ 5.95	3	4	3
ADOBE P/ADOBE ACRO	$ 59.99	STEWART/PROFESSION	$ 81.60	3	3	3
HAMILTO/TARASCON P	$ 24.95	ROTHROC/TARASCON P	$ 14.95	3	3	3
MONTGOM/CHICAGO GU	$ 12.75	WILHOIT/BRIEF GUID	$ 59.75	3	3	3
ROTHROC/TARASCON P	$ 14.95	HAMILTO/TARASCON P	$ 24.95	3	3	3
STEWART/PROFESSION	$ 81.60	ADOBE P/ADOBE ACRO	$ 59.99	3	3	3
WILHOIT/BRIEF GUID	$ 59.75	MONTGOM/CHICAGO GU	$ 12.75	3	3	3

Table A3. Pairings with the highest possible additional revenues (k = 3, only pairings with additional revenues in excess of $1000 are shown). The highlight is on the rows where the pricing relationship in the affinity is reversed: customers who have already purchased a lower priced item would need to be induced to purchase another more expensive item.

SKU1 Description	SKU1 price	SKU2 Description	SKU2 price	SKU1 Trans	SKU2 Trans	SKU 1&2	Additional revenues
BACHELOR GOWN	$ 28.00	ANNOUNCEMENTS 25PK	$ 45.75	353	47	26	$ 3,526.47
BACHELOR GOWN	$ 28.00	THANK YOU NOTES 25	$ 11.75	353	6	6	$ 3,506.92
BACHELOR GOWN	$ 28.00	RETURN LABELS 50PK	$ 13.25	353	11	7	$ 2,428.82
GRADUATION CAPS	$ 7.00	THANK YOU NOTES 25	$ 11.75	471	6	6	$ 2,406.65
BACHELOR GOWN	$ 28.00	ANNOUNCEMENT COVER	$ 11.75	353	7	4	$ 1,998.18
GRADUATION CAPS	$ 7.00	ENVELOPE SEALS 25P	$ 6.75	471	22	19	$ 1,776.14
GRADUATION CAPS	$ 7.00	TASSEL GREEN/BGOLD	$ 5.00	471	9	9	$ 1,760.00
TASSEL GREEN/BGOLD	$ 5.00	ENVELOPE SEALS 25P	$ 6.75	522	22	20	$ 1,687.50
BACHELOR GOWN	$ 28.00	ANNOUNCEMENTS 5PK	$ 8.95	353	21	13	$ 1,643.45
GRADUATION CAPS	$ 7.00	RETURN LABELS 50PK	$ 13.25	471	11	7	$ 1,431.40
GRADUATION CAPS	$ 7.00	ANNOUNCEMENTS 5PK	$ 8.95	471	21	13	$ 1,430.63
BACHELOR GOWN	$ 28.00	ENVELOPE SEALS 25P	$ 6.75	353	22	15	$ 1,400.94
GRADUATION CAPS	$ 7.00	ANNOUNCEMENT COVER	$ 11.75	471	7	4	$ 1,372.27
GRADUATION CAPS	$ 7.00	TASSEL GREEN/BGOLD	$ 5.00	471	44	37	$ 1,367.88
TASSEL GREEN/BGOLD	$ 5.00	THANK YOU NOTES 25	$ 11.75	522	6	6	$ 1,313.65
TASSEL GREEN/BGOLD	$ 5.00	ANNOUNCEMENTS 5PK	$ 8.95	522	21	15	$ 1,291.80
POSTAGE	$ 136.34	S/S SCALLOP SEAL	$ 29.95	66	9	6	$ 1,054.76