Cross-sell: A Fast Promotion-Tunable Customer-item Recommendation Method Based on Conditionally Independent Probabilities

Brendan Kitts
Vignette Corporation
19 Newcrossing Road
Reading, MA. 01867. USA
Ph: +1 781 942-3600
bkitts@vignette.com

David Freed
Vignette Corporation
19 Newcrossing Road
Reading, MA. 01867. USA
Ph: +1 781 942-3600
dfreed@vignette.com

Martin Vrieze
Vignette Corporation
19 Newcrossing Road
Reading, MA. 01867. USA
Ph: +1 781 942-3600
mvrieze@vignette.com

ABSTRACT

We develop a method for recommending products to customers with applications to both on-line and surface mail promotional offers. Our method differs from previous work in collaborative filtering [8] and imputation [18], in that we assume probabilities are conditionally independent. This assumption, which is also made in Naïve Bayes [5], enables us to pre-compute probabilities and store them in main memory, enabling very fast performance on millions of customers. The algorithm supports a variety of tunable parameters so that the method can address different promotional objectives. We tested the algorithm at an on-line hardware retailer, with 17,400 customers divided randomly into control and experimental groups. In the experimental group, clickthrough increased by +40% ($p<0.01$), revenue by +38% ($p<0.07$), and units sold by +61% ($p<0.01$). By changing the algorithm's parameter settings we found that these results could be improved even further. This work demonstrates the considerable potential of automated data mining for dramatically increasing the profitability of on and off-line retail promotions.

Keywords

Imputation, cross-sell, collaborative filtering, recommendation

1. INTRODUCTION

Most customer recommendation algorithms can be understood as performing some kind of imputation [13]. Some of the customer's interests are known because they have entered "star ratings" or have bought a product, but most are not. The problem of deciding what product to recommend next involves finding out what the customer's attitudes would be toward the missing values, by analyzing the statistical patterns of the population. For example, say that Joe purchased science&nature and mystery. We can look for all other customers who bought the same two items, and possibly other purchases. Joe's probability of interest might now be calculated by taking the average of the interest of the donor customers in the new category, or in other words, what Joe's "soul mates" thought on average about the other category. This particular method of filling in missing values is known in the statistics literature as conditional mean imputation [18].

Formally, let a customer profile x consist of a binary vector $x = [0,1] \in \Re^N$ where a x_{si}=1 means that customer s purchased/clicked product/web-page i, and a 0 means that the customer did not, and N is the number of variables in the profile[1]. We are trying to predict the customer's interest in variable j. Let $x_{sj}=MV$, which stands for "missing value". Conditional mean imputation is defined as:

$$\text{Given that } x_{sj} = MV$$

$$D = \{d\} : (x_{si} = 1 \rightarrow x_{di} = x_{si}) \forall i$$

$$x_{sj}^{+} = \frac{1}{\# D} \sum_{d \in D} x_{dj}$$

Other typical imputation algorithms include regression imputation [17,19], the EM algorithm, and hot-deck imputation [7,14,4]. Regression imputation selects donor cases in exactly the same way, and then calculates a least squares estimate:

$$x_{sj}^{+} = [x_{si} 1] \cdot w$$

$$\text{where } w = [x_{di} 1_{\# D}]^{-1} \cdot x_{dj}$$

[1] This is not the only choice for profile; for instance, we could have used a profile of revenues, percentages of spending, or page hits. We will use binary profiles in this article because this is what we have used in our experiments reported later.

Collaborative filtering systems [8,9,16] implement a nearest neighbor variant of the above strategy. The donor set is restricted to the k closest matching customer profiles to a candidate.

$$D = lowest(k)\ |x_{s1..N}, x_{d1..N}|$$

$$x_{sj}^{\ +} = \frac{1}{k}\sum_{d \in D} x_{dj}$$

Various alterations to this procedure have been proposed including weighting users, products or star ratings, and adding significance tests for measuring the reliability of recommendations [16].

2. PROBLEMS WITH COLLABORATIVE FILTERING

In all the above methods one needs to calculate a match between the candidate and every other customer in the population, before blending the donor data to arrive at a score. In practice one needs to perform this computation quickly. One option is to calculate it when a customer visits a site. The time complexity of this operation is O(CN) where C are the number of customers and N is the size of the profile.

The alternative is to pre-compute probabilities and store in memory or on disk for faster lookup using an index or hash table. Unfortunately, there are usually too many match patterns to store for this to be feasible. We need to store results for each

$$\Pr(x_{ci}=X_i \mid x_{c1}=X_1, x_{c2}=X_2, x_{c3}=X_3, \ldots, x_{cN}=X_N) \text{ where } X_i \in [0,1]$$

There are N possible items in the term before the bar - variable in the profile that we are estimating a probability for. The term after the bar, or pattern of conditional purchases in the customer's profile - contains a string of N variables which can take the value 0 or 1. This means that there are 2^N different condition patterns. The total number of combinations is $N2^N$, which grows as an exponential function of N.

3. CONDITIONALLY INDEPENDENT RECOMMENDATIONS

The approach in this paper differs from previous work on collaborative filtering in the following respect. We do not calculate interest probabilities conditional upon meeting all of the criteria of a customer's profile, as is required in conditional mean imputation and collaborative filtering. Instead, we operate under the assumption of conditional independence of the past behavior of the customer in question. Formally:

Definition: Conditional independence

$$\Pr(b|a) = \Pr(b|a,c,d,e,\ldots,n)\ \forall a,b,c,d,e,\ldots,n$$

This is somewhat unrealistic. If a customer has bought five scifi books, we would expect their probability of being interested in a new scifi book to be higher than another customer with one scifi book with ten gardening. Never the less, we will adopt the constraint. This assumption is also made in some other prediction algorithms such as Naïve Bayes [5].

The disadvantage of this constraint is that accuracy can be lower because we are ignoring interactions. The advantage is that storage requirements are tiny, and as a result algorithm speed can be greatly increased. The recommendation of interest will be some function of the customer's profile and single-condition events, $\Pr(b|a)$. This means that the storage complexity for those probabilities is only $N*N = O(N^2)$ which is polynomial in N. Storage can be decreased further since only half of a co-occurrence counter matrix needs to be stored, and low-frequency pairs can be ignored below a certain threshold [1]. With low memory requirements, lookup can be achieved using fast hash tables.

4. THE CROSS-SELL RECOMMENDATION ALGORITHM

This section will describe how individual conditional probabilities are combined to create a customer recommendation. Let a *driver* be an item the customer has purchased before, $x_{ci}=1 \rightarrow driver(c,i)$. Let R be the number of recommendations the customer needs to be provided with. Our recommendation algorithm simply considers each driver, and then reads off the top R cross-sell items with the highest promotion objective score described below, subject to various parameter settings also described below.

4.1 Promotion objective score

Retail businesses rarely have a single promotional goal. After a web site is first opened discount offers might be presented with the aim of generating traffic/clickthrough. Later, maximizing profit might become important. For new users with little data it might be best to offer products with the highest response probabilities across the population, such as Whitney Houston or Britney Spears CDs. But for veteran loyal customers, understanding their exact needs might be crucial. For these reasons, the recommendation score of an item is customizable to the promotional objective.

We have developed four criteria to measure the value of an item recommendation: 1. probability of customer responding to item, 2. lift or degree of mutual attraction between item and customer, 3. expected profit from item, 4. incremental profit from item.

4.1.1 Response Probability

Response probability is the probability of an item b being bought, given a customer's purchase of item a. Interestingly, using this method for scoring item desirability, the most probable item a customer will buy after a hammer, might *not* be nails! It could be a magazine. If there exist items in the store which have very high baseline rates of being purchased, these can be recommended frequently, and seemingly without regard to the drivers in the customer's purchase history. Figure 1 illustrates how the most frequently purchased items dominate the recommendation value. Lift described below "fixes" the problem of high probability items dominating recommendations. The formula for response probability is:

RecommendationValue(b) = Pr(b|a)

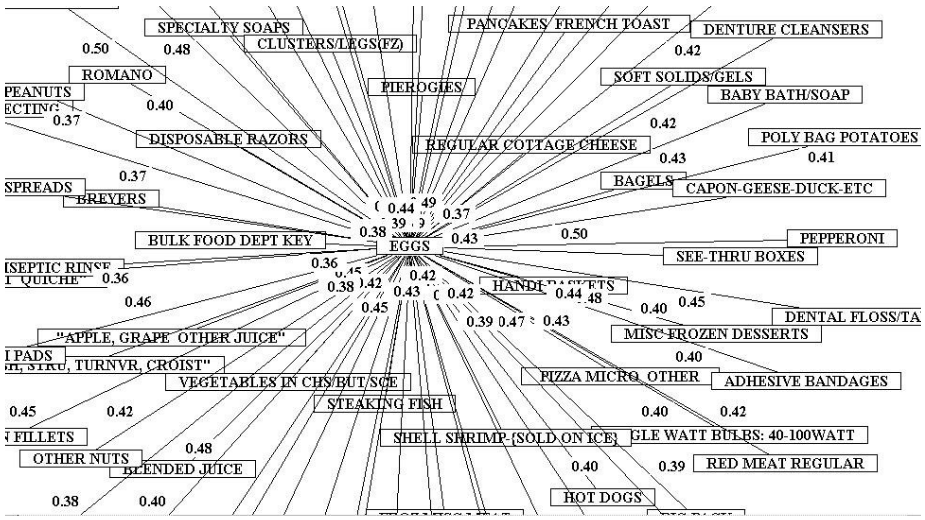


Figure 1: Graph of the highest conditional probabilities in a grocery store. If we take any item in the store and list the conditional probabilities from largest to smallest, the top three "cross-sell" products are nearly always eggs, bread and milk. This is because eggs, milk and bread have a high probability of appearing in any basket. In a recommendation system, utilizing cross-sell probabilities would result in these very high baseline probability items being recommended again and again, regardless of the drivers appearing in the customer's purchase profile.

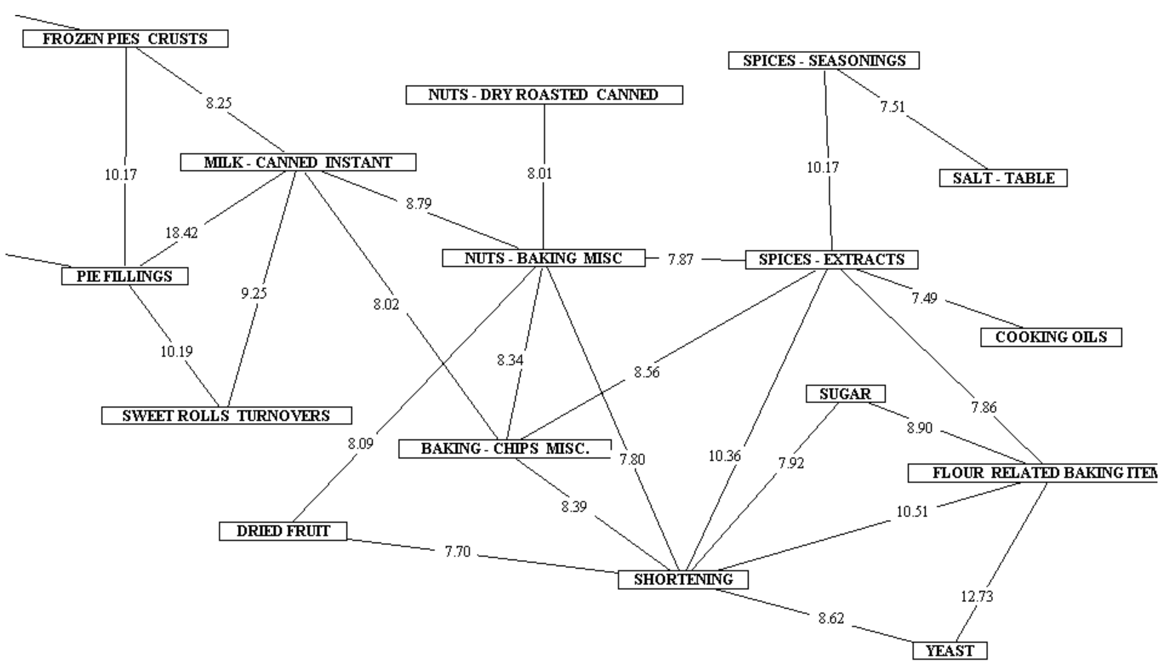


Figure 2: Graph of top lift affinities for the same grocery store. Lift is very effective in revealing which products have strong two-way purchase relationships.

4.1.2 *Lift or Mutual Affinity*

The idea of lift is to promote products which have high mutual attractions to each other. For instance, an "air conditioning unit" and "air conditioning unit accessory" might be very rarely bought with other items, but might be bought together frequently. Although lift does not necessarily maximize sales probability or profit, previous work has indicated that profit can be generated by cross-selling products with high lift scores. In a past experiment we optimized shelf-layout by moving high lift items together. This resulted in a +40% increase in profit for items that were moved together [11]. The formula for lift is

$$\text{RecommendationValue}(b) = \Pr(b|a)/\Pr(b) = \Pr(a,b)/[\Pr(a)*\Pr(b)]$$

Lift is a symmetric measure, so Lift(a,b)=Lift(b,a). A number greater than one is interpreted as the number of times higher than random that two items occur together. A fractional number can be inverted and interpreted as the number of times lower than random that two items occur. Interestingly, lift is related to the Mutual Information Criterion (MIC) from information theory [3]. MIC is equal to log of lift. We favor the untransformed lift score because it is easier to interpret for the user.

4.1.3 Expected Profit

If we assume mutual independence between products, then the expected profit after buying a product a is equal to the probability of buying b given a, $\Pr(b|a)$ multiplied by the profit Π of b. As a result this is the formula:

$$\text{RecommendationValue}(b) = \Pr(b|a) * \Pi(b)$$

4.1.4 Incremental Profit

The idea behind incremental profit is to maximize the profit minus the profit you would expect to receive due to the natural course of a customer's purchasing. For example, say a customer comes into a store and buys a hammer (product a). You have two choices: nails, or screwdrivers. Analysis of customer purchase patterns may indicate that nails are almost certainly going to be bought in the future, since these have a 20% chance of being bought by any customer. Therefore, instead of promoting something that we know the customer will be buying anyway, we go for the purchase that has a higher incremental profit – the screwdriver. Incremental profit maximizes the profit of the item, minus the baseline profit associated with the item. Thus incremental profit is similar to lift, except it subtracts the base probability, rather than dividing by it.

$$\text{RecommendationValue}(b) = [\Pr(b|a)-\Pr(b)]\ \Pi(b)$$

4.2 Driver Diversity

This parameter directs the algorithm to recommend at least one item from each driver, or will pool all of the recommendations together, and will select those with the highest promotional objective scores[2].

4.3 Driver Recency

Driver recency forces the recommendation algorithm to consider more recent purchases preferentially over purchases in the past. This was implemented by considering drivers in time order from most recent purchase to oldest purchase, until the required number of recommendations was filled.

4.4 Other Parameters

The level of analysis, level of recommendation, number of duplicate recommendations tolerated, and recommendation of products already in customer history are also customizable. For example, in almost every retailer an item hierarchy is available. If no recommendations can be made confidently at the item level, it is possible to switch to examining affinities at the sub-category or category level, and reading off recommendations at that level. This strategy for walking the hierarchy was implemented, but not used in the experimental test that follows since the retailer was only interested in targeting specific items.

5. EXPERIMENTAL TEST

We tested our algorithm at an on-line and catalogue hardware retailer based in California. This retailer had accumulated 11 years of data on customer transactions, with approximately 60 million rows. The company ran an opt-in direct email list, and distributed email messages to around 65,000 customers each week. Revenues accrued from the direct email campaign average around $2.18 per customer emailing, with clickthrough probability equal to 1.78%.

We took 14,770 customers and divided them randomly into control and experimental groups, with 6,999 and 7,771 respectively. The experimental group customers received automated recommendations, while the control group customers received the weekly scheduled promotion, put together by this company's marketing department.

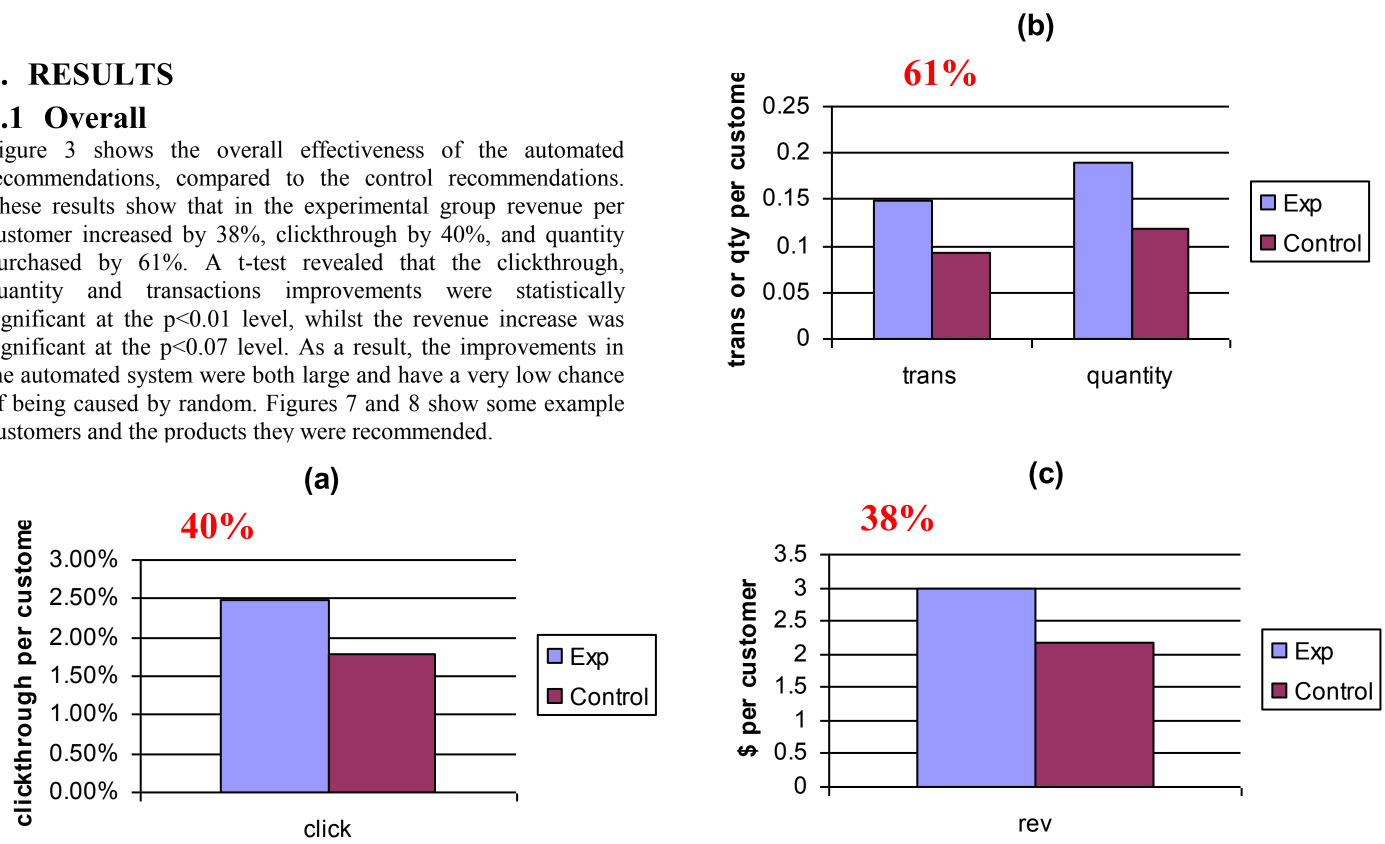
To test the various parameter settings for this algorithm, we allocated recommendations to the experimental group based on a variety of parameter settings. We did not allow a customer to be recommended with a product that they already owned, and did not allow duplicates to be recommended.

[2] Forcing the algorithm to make a recommendation based on each of the customer's historical purchases can be beneficial, because the largest RecommendationValue scores might come from just one product in the customer's profile (eg. one which has a high baseline probability). Thus all recommendations would be based on a single purchase, when that customer's profile might contain much more information, for instance, 10 purchases of scifi books.

6. RESULTS

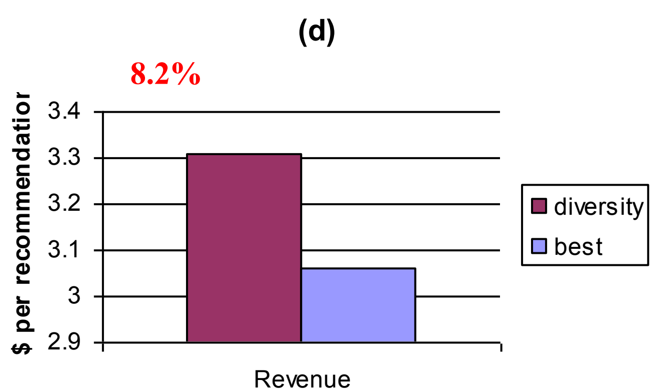
6.1 Overall

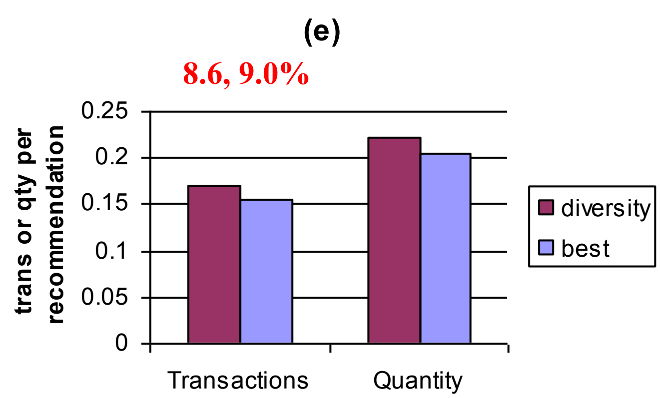
Figure 3 shows the overall effectiveness of the automated recommendations, compared to the control recommendations. These results show that in the experimental group revenue per customer increased by 38%, clickthrough by 40%, and quantity purchased by 61%. A t-test revealed that the clickthrough, quantity and transactions improvements were statistically significant at the $p<0.01$ level, whilst the revenue increase was significant at the $p<0.07$ level. As a result, the improvements in the automated system were both large and have a very low chance of being caused by random. Figures 7 and 8 show some example customers and the products they were recommended.



		click			trans			quantity			rev		
Group	Count	Mean	StdDev	p	Mean	StdDev	p	Mean	StdDev	p	Mean	StdDev	p
Exp	7771	0.02484	0.15564	<0.01	0.14786	1.20643	<0.01	0.19071	1.7400	<0.01	3.0009	28.863	0.0676
Control	6999	0.01772	0.13193		0.09244	0.97856		0.11830	1.2575		2.1776	25.537	

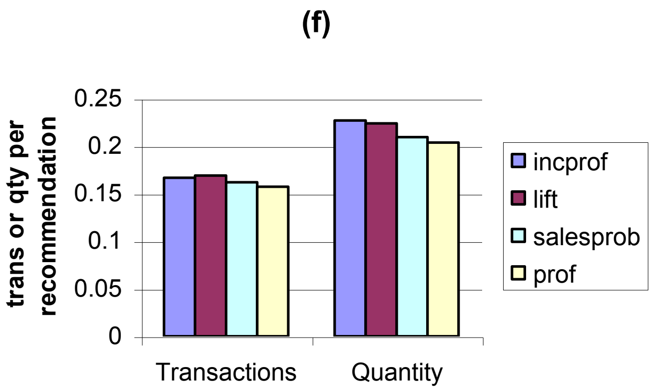
Figure 3: Main results from experiment aggregated over all tested parameter settings (charts a, b and c)

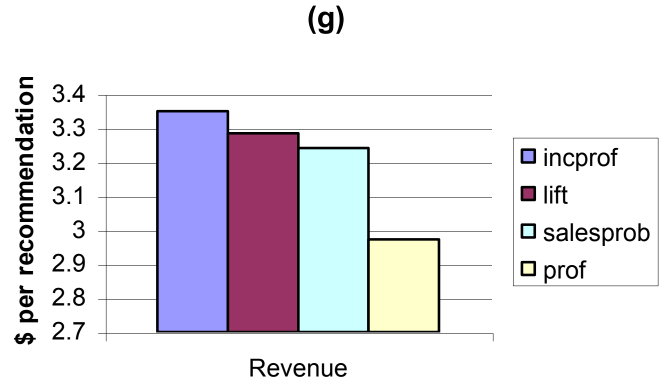


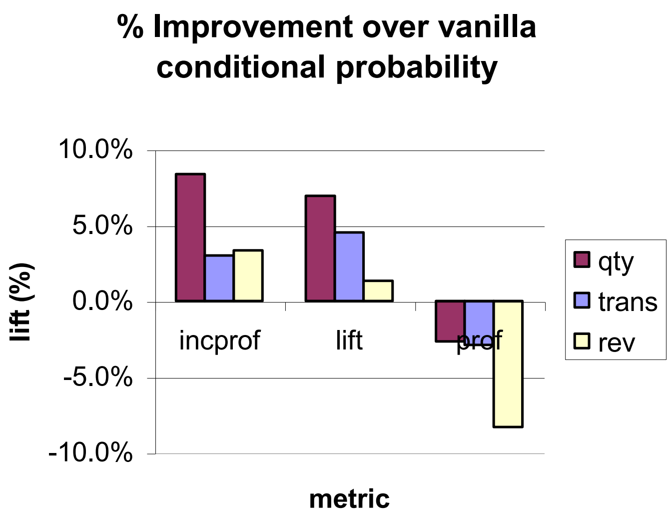


		trans			quantity			rev		
Group	Count	Mean	StdDev	P	Mean	StdDev	p	Mean	StdDev	p
best	61639	0.15594	1.22594	p 0.0512	0.20453	1.83855	0.0755	3.05911	28.2182	0.1152
diversity	76801	0.16928	1.29586		0.22289	1.96555		3.30926	30.2522	

Figure 4: Effect of using driver diversity (charts d and e)







		trans			quantity			rev		
Group	Count	Mean	StdDev	p	Mean	StdDev	p	Mean	StdDev	p
incprof	25500	0.166784	1.28724	0.3505	0.22698	2.08233	0.1516	3.35024	32.0955	0.107
lift	34258	0.169245	1.28464	0.2023	0.223948	1.98448	0.1516	3.28502	29.6433	0.1354
prof	38313	0.157257	1.24604	NA	0.203847	1.7922	NA	2.97227	26.7753	NA
salesprob	40369	0.161931	1.25259	0.5999	0.209443	1.83902	0.6658	3.24182	29.6677	0.1817

Figure 5: Performance of different objective scores (charts f and g). The table shows that maximizing incprof and lift resulted in the best performance on most metrics, where-as maximizing “prof” resulted in the lowest performance on all metrics. The significance test is the probability that a group (eg. lift) is significantly different from the lowest group (prof). The bottom figure shows that incprof and lift generated 6-8% more revenue than base response probability.

6.2 Parameter Selection

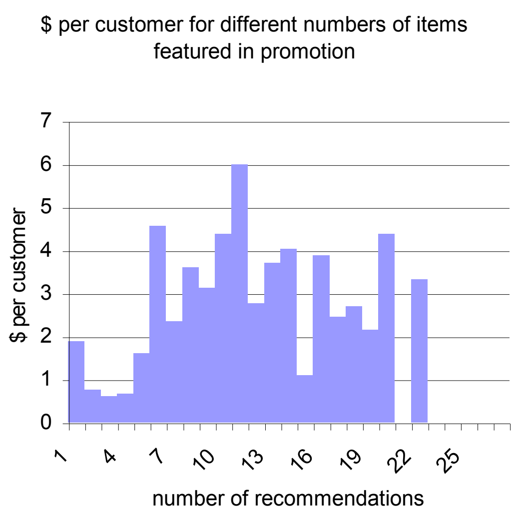
Incremental profit and lift both outperformed conditional probability and profit maximization in all behavioral measurements including revenue, transactions, and quantity purchased (figure 5).

The fact that incremental profit and lift out-performed the other methods is interesting. Lift is the conditional probability divided by the baseline probability. Now consider that incremental profit is the conditional probability minus the baseline probability. These two measures are similar in that both are discounting the baseline probability in some way.

[2] also found that discounting base rating frequencies increased accuracy in predicting interest in test data. Their "inverse user weighting" scheme increased accuracy in all 24 experiments they ran on test data. Further experiments will be needed to identify (a) if this principle holds true in general, (b) the best way to account for base probabilities ([2] divided by a log inverse probability score, where-as we have proposed dividing / subtracting the base rate), and (c) under what conditions base interests should be favored over lifted interests (the base probabilities might be effective on new users with little data, and lift affinities for veteran customers; however, this experiment needs to be performed).

Driver diversity increased revenue, transactions and quantity-purchased by 8%, 9% and 8.6% respectively, per recommendation. The increase in transactions was significant at the $p<0.06$ level. (figure 4)

A histogram of revenue versus number of recommendations is shown in figure 6. Although the distribution is noisy, it appears that an optimal number of recommendations is around 11 per email, which results in \$5.99 revenue per customer. The company currently uses 15 recommendations per email message.



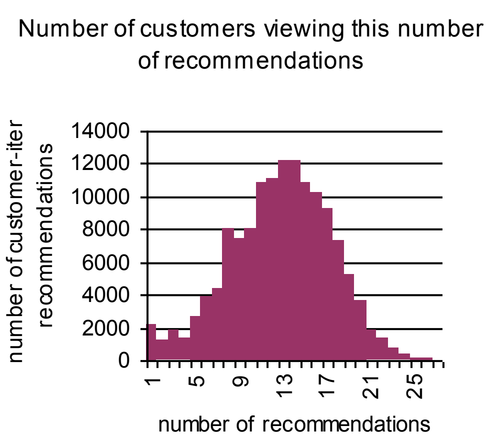


Figure 6: Revenue resulting from different numbers of recommendations

Table 4.1. Complete Purchase history for Customer A

SKU	Date	Qty	Price	Description
2776	6/18/99	1	19.99	Lathe Bits AR-6 10 PK
2901	6/18/99	1	9.99	O-Ring Assortment 382 PC
33684	6/18/99	1	329.99	Lathe-7" X 10" Mini
36954	6/18/99	1	9.99	Retaining rings-225PC

Table 4.2. Recommendations for Customer A

Driver	Recommendation
Lathe bits AR-6 10 PK	Tool set-indexable Carbide
Lathe-7" X 10" Mini	Lathe Toolkit-Quick change
O-Ring assortment 382 PC	Lock nut storehouse-150PC
Retaining rings-225PC	Spring asst-200 PC

Table 4.3. Customer A purchases three days after offer sent

SKU	Qty	Price	Description
3629	1	8.99	7 PC. Forstner bit set
35140	1	72.99	Quick change lathe toolkit
39424	1	18.99	40 PC. Tungsten alloy SAE tap & die set
39931	1	14.99	5 PC. Indexable carbide Tool set
Total	4	115.96	All

Figure 7: On June 18, 1999 Customer A bought a \$329.99 Mini Lathe, along with some replacement cutting bits, a toolkit of O-rings and Retaining rings. In response the system recommended an additional set of carbide lathe cutting bits, a Lathe quick-change toolkit, and toolkits with locknuts and springs. After receiving these offers through email, the customer bought four products including the lathe parts.

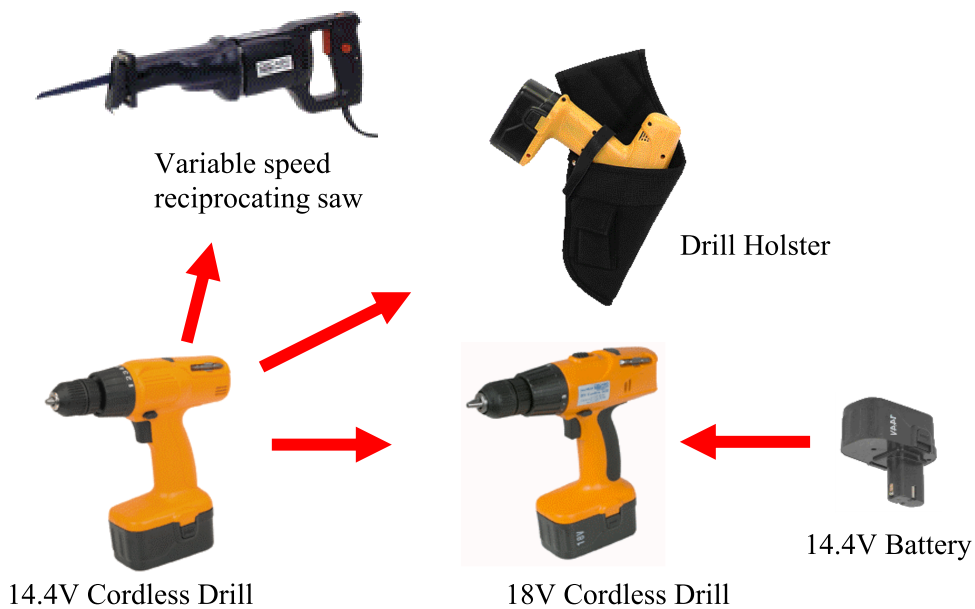


Table 8.1. Historical purchases for Customer B

Customer	Qty	Rev	Responses	First date	Return rev	Days active
B	82	561.74	11	4/22/96	0	1165

Table 8.2. Recommendations for Customer B

Driver	Recommendation	Criterion
Drill-14.4V	Recip saw	Incprof
Drill -14.4V	Recip saw	Lift
Drill -14.4V	Drill Holster	Incprof
Drill -14.4V	Drill Holster	Lift
Battery-14.4V	Drill -18V	Salesprob
Drill -14.4V	Drill -18V	Prof
Battery -14.4V	Drill -18V	Incprof
Battery -14.4V	Drill -18V	Lift

Table 8.3. Customer B purchases three days after offer sent

Qty	Price	Description	Date
1	59.99	Drill-18V	2/19/00

Figure 8: Customer B previously purchased a 14.4V Drill and replacement battery. The system recommended an 18V Drill and the customer purchased it.

6.3 Lifetime factors

Because we had access to a long period of customer history, we were also able to analyze the effect of previous responses to promotions on the likelihood of responding to this promotion. We identified a 25 factors, listed in figure 9. The best predictor for high revenue in the promotion is a high quantity purchased per catalogue received (R=0.38) followed by other lifetime revenue and quantity variables. The response probability of items recommended was correlated with customer revenue (R=0.13).

Variables that indicated low revenue included quantity returned as a percent of total ordered (R=-0.33), and revenue returned as a percent of total (R=-0.29). In other words, customers who returned large numbers of goods were poor responders to future promotions. Perhaps this was due to dissatisfaction, and this might have indicated that an alternative strategy should be used for these customers.

Figure 9. Impact of lifetime factors on promotion performance

Factor	R	Description
qtty per catalogue	0.384	quantity ordered per catalogue received
Log rev per catalogue	0.326	log of revenue per catalogue
rev per catalogue	0.260	revenue generated per catalogue
prof per catalogue	0.232	profit generated through orders per catalogue
lifetime qty	0.224	quantity ordered in lifetime
lifetime rev	0.142	revenue generated in lifetime
lifetime prof	0.131	profit generated in lifetime
mean probability	0.130	average response probability for recommendations this customer was given
avg price	0.121	average price of products purchased by this customer
response rate	0.106	number of orders divided by number of catalogues received
days active	0.098	days since customer made first purchase
Resp	0.089	number of orders
return qtty	0.080	number of items returned
Nocatalogues	0.073	number of catalogues received
mean profit	0.058	average profit for recommendations this customer viewed in the email promotion
return rev	0.053	dollar amount of products returned
Meanrank	0.014	average rank of recommendations this customer viewed
rev per day	0.001	revenue generated / days active
qtty per day	-0.009	quantity generated / days active
prof per day	-0.028	profit generated / days active
NumberOfRecommendations	-0.032	number of recommendations this customer viewed
DistinctRecommendations	-0.060	number of distinct recommendations this customer viewed
response per day	-0.091	orders / days active
profit as % of revenue	-0.096	for each dollar this customer spends, how much of that is profit
rev returned as % of totalrev	-0.293	percentage of customer's spending that returns to the store
qtty returned as % of totalqty	-0.330	percentage of products that the customer returns to the store

7. RELATED WORK

Other researchers have reported similar results to those in our experiment. [10] reported a lift in clickthrough from 8.3% to 13.2% for market basket analysis (possibly similar to the method in this paper), and 13.96% for nearest neighbor method, in direct email campaigns (59% and 68% respectively). [15] reported a lift in revenue of 60% at a catalogue company in the United Kingdom using a nearest neighbor method. Because of these large improvements, we are confident that our results are typical of results achieved by implementing intelligent customer-item recommendation methods at other on-line retailers.

8. CONCLUSION

On-line retailers face a difficult situation. Customer acquisition costs are high, and competitor stores are a mouse-click away. As on-line retailers struggle to survive in this environment, we believe this will lead to a burgeoning market for data mining techniques that can analyze large volumes of data, develop quality individualized customer services such as recommendation, price optimization, and notification; and increase profitability of customers. We have shown in this paper that implementation of such a system can significantly increase profitability and re-visit propensity by as much as 38% and 40% respectively at a low volume retailer, and without a finely tuned system. This kind of improvement cannot be ignored, and we predict that all web sites will install systems of a type like that in this paper to increase their customer satisfaction, re-visit frequency, and most importantly, the bottom-line profitability of their web business.

9. ACKNOWLEDGMENTS

Thanks to Vignette Corporation for making possible this research.

10. REFERENCES

[1] Agrawal, R. and Srikant, R, Fast algorithms for mining association rules, *Proceedings of the 20th International Conference on Very Large Databases*, Santiago, Chile. (1994).

[2] Breese, J., Heckerman, D., Kadie, C., Empirical Analysis of Predictive Algorithms for Collaborative Filtering, Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, Madison, WI. Morgan Kaufmann, (July, 1998).

[3] Chan, P., A non-invasive learning approach to building web user profiles, Workshop on Web usage analysis and user profiling, *Fifth International Conference on Knowledge Discovery and Data Mining*, San Diego. (August, 1999).

[4] David, M., Little, R., et. al., Alternative methods for CPS income imputation, *Journal of the American Statistical Association*, Vol. 81, No. 393, pp. 29-41. (1986).

[5] Elkan, C., Boosting and Naïve Bayesian Learning, Technical Report No. CS97-557, Department of Computer Science and Engineering, University of California, San Diego. (1997).

[6] Ford, B., An Overview of Hot-Deck Procedures, in Madow, W. et. al. (eds), *Incomplete data in sample surveys*, Vol 2., Academic Press, NY. pp. 185-206. (1980).

[7] Herlocker, J., Konstan, J., Borchers, A., Riedl, J., An Algorithmic Framework for Performing Collaborative Filtering, Department of Computer Science and Engineering, University of Minnesota, http://www.cs.umn.edu/Research/GroupLens (2000).

[8] Hey, J., System and method of predicting subjective reactions, Patent number 4,870,579, US Patent and Trademark Office, http://164.195.100.11 (1989).

[9] Hey, J., System and method for recommending items, Patent number 4,996,642, US Patent and Trademark Office, http://164.195.100.11 (1991).

[10] Kaupp, R., The New Science of eMarketing: Targeted Email Communications to keep customers Coming Back, Digital Impact Corporation, San Mateo, CA, *Proceedings of the Personalization Summit 2000*, Boston. (April, 2000).

[11] Kitts, B., The Ecology of the Retail Store, Working paper, Vignette Corporation. (2000a)

[12] Kitts, B., Diamonds in the Rough, Working paper, Vignette Corporation. (2000b)

[13] Little, R. and Rubin, D., *Statistical Analysis with Missing Data*, John Wiley and Sons, NY. (1987).

[14] Oh, H. and Scheuren, F., Estimating the Variance impact of Missing CPS Income Data, *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 408-415. (1980).

[15] Riedl, J., Soul Mate Searches, Net Perceptions Corporation, *Proceedings of the Personalization Summit 2000*, Boston. (April, 2000).

[16] Robinson, G., Automated collaborative filtering system, US Patent and Trademark Office, http://164.195.100.11 (1998).

[17] Rubin, D., Multiple imputation after 18+ years, *Journal of the American Statistical Association*, Vol. 91, pp. 434, 473-489. (1996).

[18] Schafer, J., *Analysis of Incomplete Multivariate Data*, Chapman and Hall, London. (1997).

[19] Schafer, J. and Olsen, M., Multiple imputation for multivariate missing-data problems: a data analyst's perspective, *Multivariate Behavioural Research*, Vol. 33, pp. 545-571. also available from http://www.stat.psu.edu/~jls (1997).

[20] Schafer, J., Multiple imputation: a primer, *Statistical Methods in Medical Research*, Vol. 8, pp. 3-15. (1999).