10742-2016

Build Recommender system with SAS to improve cross-selling etc for online business

Miao Nie , ABN AMRO bank; Shanshan Cong , SAS Institute

Abstract

Nowadays, the recommender system is a popular tool for online retailer business to predict the customers' next-product-to-buy (NPTB). Based on the information collected by the retailer and statistical techniques, an efficient recommender system can suggest customers with meaningful NPTB. A useful suggestion can reduce customer's searching time for a wanted product and improve buying experience, which would increase the chance of cross-selling for online retailers, helping them build customer loyalty. Within a recommender system, how to combine advanced statistical techniques with information available (customer profiles, product attributes, good selling products, etc.) is the key part to bridge the retailer's database with the useful suggestion of NPTB for customers. This paper illustrates how to create a recommender system with SAS RECOMMEND procedure for online business. Using the recommender system, we can produce the prediction, compare the performance with different predictive models (i.e. decision tree, multinomial discrete-choice models etc), and make business-oriented recommendations from the analysis.

Introduction

A recommender system refers to software tools and techniques that provide suggestions for an item to be adopted by a user. Item is the general term used to denote what the system recommends to users, which can include products, services and so on. Recommender system development initiated from a rather simple observation: individuals often rely on recommendations provided by others in making routine, daily decisions. For instance, individuals may rely on the comments of film critics when selecting for a movie to watch, or ask their friends for opinions with regard to which hotel to book. Recommender system mimics this behavior and produces useful and effective suggestions for the users who lack sufficient experience or capacity to evaluate the overwhelming number of alternatives.

A recommender system can play various roles depending on whom is using it. Generally speaking, a service provider could benefit from it by increasing the number of items sold, selling more diverse items, increasing the users' satisfaction and fidelity, and better understanding what the users want. On the other hand, a user wants to use the recommender system in order to find some or all good items, recommend a sequence or bundle, just browse, find credible recommender, express himself, or even influence others.

In this article, we will demonstrate how recommender system can be deployed to improve cross-selling, a term describing the action or practice of selling an additional product (category) or service to the existing customers. Cross-selling is a term which refers to the "sales of a number of related products which can occur concurrently with the sale of a major product"(Nash and Sterna-Karwat, 1996). Deighton et al. (1994) and Nash (1993) describe cross-selling as "encouraging a company's customers who have already bought product A to buy its product B". Typical examples of cross-selling include a life Insurance company suggesting its customer sign up for car or health insurance, a wholesale mobile retailer suggesting a customer choose a network or carrier after one purchases a mobile, a laptop seller offering a customer a mouse, pen-drive, and/or accessories.

Companies use cross-selling as a tool of customer relationship management (CRM), and It not only helps increase company revenue, but also reduce/prevent the churn. In order to succeed in cross-selling, it is important to create customer intimacy through knowledge about the customer's preferences and needs. However, customers may react in a negative way when they are aware of they are observed. The perceived risk may explain why sometimes there is lack of information about socio-demographics in the customers' database.

In comparison to using an intuitive model it is effective to use a Next-Product-To-Buy (NPTB) model to optimize the cross-selling. An NPTB model will use information about the customer to predict what product (category) they are most likely to purchase from next. There are some steps in developing an NPTB model: 1) compile data which consists of requiring customer-specific data of a certain time period t, being the independent or predictor variables, and an observation of a purchase at time t+1, being the dependent variable. 2) select a statistical model (i.e. regression, decision tree, neural network etc predictive models. 3) estimate and evaluate the model: estimate the coefficients of the model needed to make the predictions. 4) score and target customers: customers are scored with probabilities generated by the best performing model and targeted based on the scores (Knott, Hayes and Neslin, 2002).

Methodology

In this section, we will first elaborate on the technique of the recommender system. Secondly we will discuss the discrete-choice model (especially the use of the multinomial logit model) and the decision tree model. Then we compare the performance of the above predictive models built for an online business case and make business-oriented recommendations.

Recommender system

Our cross-selling model is built upon an individual customer's previous purchase behavior. It is more comparable to explore the user-based collaborative filtering to build recommender systems. In order to implement its core function, identifying the useful items for the user, a recommender system must predict that an item is worth recommending. In order to do this, the system must be able to predict the utility of some of them, or at least compare the utility of some items, and then decide what items to recommend based on this comparison. One major feature used to classify the recommendation systems is especially in regard to the recommendation algorithm, i.e., how the prediction of the utility of a recommendation, is made. Burke, R (2007) distinguishes between six different types of recommendation approaches.

- Content-based: The system learns to recommend items that are similar to the ones that the user liked in the past. The similarity of items is calculated based on the features associated with the compared items.
- Collaborative filtering: The simplest and original implementation of this approach recommends to the active user the items that other users with similar tastes liked in the past. The similarity in taste of two users is calculated based on the similarity in the rating history of the users.
- Demographic: This type of system recommends items based on the demographic profile of the user. The assumption is that different recommendations should be generated for different demographic niches.
- Knowledge-based: Knowledge-based systems recommend items based on specific domain knowledge about how certain item features meet users' needs and preferences, and ultimately how the item is useful for the user.
- Community-based: This type of system recommends items based on the preferences of the users' friends. Evidence suggests that people tend to rely more on recommendations from their friends than on recommendations from similar but anonymous individuals.

- Hybrid recommender systems: These RSs are based on the combination of the above mentioned techniques. A hybrid system combining techniques A and B tries to use the advantages of A to fix the disadvantages of B.

The collaborative filtering is the main technique used by PROC RECOMMEND. In general, user-based recommender system recommends the top N products by computing the highest purchase probability for a particular group of customers. Firstly we identify the active customers who are the most similar to the group of customers in the database. Principal Components Analysis is used to reduce the predictors due to the high number of variables and their high multicollinearity. Then hierarchical clustering is applied to cluster the customers of the training sample. Afterwards k-mean clusters are created to obtain the many similar users. Each product category is rated based on their frequency inside a group/cluster. Then the highest weight item is selected for each customer and finally the union of the products purchased by each cluster/group can be computed. In order for it to be comparable to our other predictive cross-sell models we chose to recommend product categories instead of products and predicted only one category as opposed to multiple categories. As a final step, we select certain product categories from the union of each cluster. We recommend the item with the highest rating that has not been purchased yet to the active customers.

Within this framework, the method used to determine the k most similar users and the scheme used to determine the importance of the different items plays the most critical role in the overall performance of the algorithm. Commonly, the similarity between the users is computed by treating them as vectors in the item-space and by measuring similarity via the correlation coefficient functions, whereas the importance of each item is determined by how frequently it was purchased by the k most similar users (Han and Karypis, 2005).

Discrete-choice model

As people have finite resources, they tend to have a priority structure of products. Products for more basic objectives are generally acquired before products for more advanced objectives. Acquisition pattern analysis provides insight into the order in which customers acquire such products. In fact, acquisition patterns deducted from cross-sectional data can predict future product acquisitions by individuals, though not the timing of acquisition. Also, cross-sectional data do not provide insight into divergent orders of acquisition. Divergence means that different segments of customers in a population follow different orders for acquiring products (Paas, Vermunt and Bijmolt, 2007). These drawbacks can be overcome by the use of certain types of models; this will be explained in the following paragraphs.

The discrete-choice process states that each individual will choose the option that gives him/her the greatest utility. The most popular discrete choice models are the multinomial probit model and the generalized extreme value (GEV) models. The probit class can model more flexible covariance structures for disturbances, but is not intuitive to use. The GEV class, containing the Multinomial Logit (MNL) and nested logit models, requires homoscedastic disturbances (Zeng, 2000).

Multinomial logit is a popular discrete choice model which is easy to implement. The many alternatives of choices can easily be taken care of in a computationally convenient way. In addition, the likelihood function is a concave curve, which facilitates the computations (Hausman and Mcfadden, 1984). Multinomial logit is a frequently used method to predict the next-product-to-buy (NPTB). It is especially useful when only one product can be purchased at a time, proceeding from the fact that the customer will make a choice for the next purchase out of all categories offered by the company. The multinomial logit model takes into account that not every consumer can choose from all options, while making a choice for NPTB. This implies that every consumer has a subset of the total alternatives of choices. The MNL is expressed by

$$\mathrm{P}_i(\mathrm{j}) = \frac{\mathrm{e}^{x'_{ij}\beta}}{\sum_{\mathrm{k} \in \mathrm{C}} \mathrm{e}^{x'_{ik}\beta}} \quad (1)$$

Where k is a subset of the total of choices C and where x is the matrix of the independent variables and β is the set of parameters estimated by the model. It is very clear that when $k = 2$ this equation is reduced to a binary logit (Akiva and Lerman, 1991).

The big drawback of this method is the independence from irrelevant alternatives (IIA). This implies that the ratio of the probabilities of choosing any two alternatives is independent of the attributes of any other alternative in the choice set (Hausman and Mcfadden, 1984). In other words, the unobservable components of utilities should be mutually independent and homoscedastic (Munizaga, Heydecker and Ortùzar, 2000).

Decision tree

Decision tree is a predictive model which represents the mapping between the object value and the object attributes. A decision tree consists of two main elements, namely a node and a branch. Each node represents an object; this node point corresponds to the attribute value which indicates a path from the root of the leaf node to the end of the leaf node. Every branch path represents a certain attribute value. Therefore, decision tree is a powerful tool for profiling customers with respect to a particular target variable which is usually categorical or outcome. Moreover, Berry and Linoff (2004) also stated out the decision tree is not only used to calculate the probability that a given record belongs to each of the categories, but also as a classification tool. This last purpose means that a record can be classified by assigning it to the most likely classification.

Classification and regression trees are the usual practice of describing the partitioning process which could be graphically represented as decision tree. In the classification tree the dependent variable should take a finite number of unordered values, with the prediction error measured in as a misclassification cost. In the regression tree, the dependent variables take continuous or ordered discrete values. The prediction error is typically measured by the squared difference between the observed and the predicted values (Loh, 2010).

However, sometimes irrelevant attributes are ignored. To avoid the over-fitting issue, the technique of pruning is applied. By pruning the decision trees, more compact trees with better classification accuracy can be produced. Moreover, pruning decision trees could handle the noisy data, because they will ignore attributes rendered meaningless by noise (B. Chandra, Ravi Kothari & Pallath Paul, 2010). Another reason for pruning a tree is to make them smaller, so that they are more intuitive and efficient to use.

An online retailer business application

Our model is based on a dataset of 1044532 transactions, which is generated by 103916 unique customers. We use full year of 2014 as the dependent period, starting from the 1st of Jan 2014 until the 31 of Dec 2014. We sampled our customers and based our end of the dependent period on the date of the dependent purchase. In our research we computed RFM (recency, frequency and monetary values) variables for our analysis. i.e. those provides the in-depth insights of length of the relationship, how often the customer purchase the product (categories), and (total) amount of purchase on the product (categories).

The following table gives an overview of Abbreviations of the dependent variables.

Product group	Abbreviations
TV	TV001
Home cinema/theatre system	TV002

Video /DVD player	VID01
Camcorder (video camera)	VID02
Camcorder accessories	VID03
Video accessories	VID04
Camcorder tapes	VID05
Recorded DVDs	VID06
Speakers	HIFI1
Car stereo/auto audio systems	HIFI2
Radio	HIFI3
Recorded music CDs	HIFI4
Audio accessories	HIFI5

Then we re-group the above purchase of individual products into product categories:
'HIFI'- ('HIFI1','HIFI2','HIFI3','HIFI4','HIFI5);
'TV'- ('TV001','TV002');
'VIDEO'- ('VID01','VID02','VID03','VID04','VID05','VID06').

We have built predictive modellings like recommender system, multinomial logit model and decision tree model. To investigate the quality of prediction in terms of sensitivity and specificity, PCC (percentage correctly classified)/ accuracy is examined. Based on the PCC values, we can say that the multinomial logit model is of better performance, namely 53.15%, than the recommender system and the decision tree, respectively 48.51% and 33.33%. However, we have to be careful with comparing both models, since we had to limit the recommender systems to a certain extent. We evaluate the NPTB model in terms of its ability to correctly classify cases in all classes. Given the objective and the class balance, we conclude that the multinomial logit model has a better performance.

Model	PCC
Multinomial logit model	53.15%
Decision tree	33.33%
Recommender system	48.51%

Multinomial logit model

observed _category	Predicted category		
	1	2	3
1	20%	4%	39.07%
2	1%	1%	7.72%
3	24%	3%	52.64%

The majority of the observations were predicted as Video, namely 52.64%. Another 20% were predicted as HIFI. The best predicted category is Video with a correct classification of 99%.

Conclusions and Recommendations

In this paper we aim to build predictive models to increase cross-selling for online retailer business. Cross-selling can help companies increase the revenues gained from their existing customers, as an existing customer delivers a higher profit than a prospect (Prinzie and Van den Poel, 2004). Cross-selling can be also used as a CRM-tool in order to reduce/prevent customer churn. An NPTB model will use customers' information to predict which product (category) they are most likely to purchase in the future (Knott, Hayes and Neslin, 2002).

An overview of the methodology is presented regarding the creation of a NPTB model. One method frequently used for predicting the next-product-to-buy is the recommender system. 'Feature-Based Recommendation System' and 'Collaborative Filtering' are the two main techniques applied in the recommender system. Feature-Based Recommendation System makes recommendations based on products which have similar features (Han and Karypis, 2005). Collaborative filtering recommends items which are purchased by similar users (Yu, Xu and Kriegel, 2001). Another popular discrete-choice model is the multinomial logit model because it is easy to implement. This model takes into account that not every customer can choose from all options, while making a choice for the next-product-to-buy (Zeng, 2000). Then we introduced the decision tree model, which is a predictive model that represents the mapping between the object value and the object attributes. A decision tree consists of two elements, namely a branch and a node. This technique is not only used to predict probability that a given record belongs to each of the categories, but is also used as a classification tool (Berry and Linoff 2004).

Then we presented a case study of online retailer as an example. The first step was to determine the time window. We defined the dependent period as starting from 1 Jan 2014 until 31 Dec 2014. With regards to the independent variables, we computed the following groups of variables based on previous research in this domain: broadness of purchase, depth of purchase and recency.

Practical insight of the implementation of the models has been discussed. Based on PCC, our best model is multinomial logit model. We can conclude that 99% of Video is the predicted product category.

Based on these results we would like to make a few recommendations to this online retailer:

First, we recommend them to invest in improving customers' data. Since the socio- demographical information available on their database is limited, it will be beneficial to collect such information to facilitate predictive modeling, which will eventually lead to better performance of the recommender system, i.e. understand what has impact on customers' purchasing behavior. Secondly, we suggest to collect the information such as Promotion Sensitivity to observe how customers react on the promotion.

References

http://support.sas.com/documentation/cdl/en/inmsref/67306/HTML/default/viewer.htm#p1ikis8zzd3iszn1t1hsxun24qe3.htm

Sample *53925:* Building Explicit Recommendations: http://support.sas.com/kb/53/925.html

Akiva, M.B, Lerman, S. and Akiva, B. (1991), "Discrete Choice Analysis: theory and application to travel demand", *Cambridge MA: Mit Press.*

Akçura, M. T and Srinivasan K. (2005), "Research Note: Customer Intimacy and Cross-Selling Strategy", *Management Science*, Vol. 51, No. 6, pp. 1007-1012.

Baesens, B., Verstraeten G., Van den Poel, D., Egmont-Petersen, M., Van Kenhove, P. and Vanthienen, J. (2004), "Bayesian Network Classifiers for Identifying the Slope of the Customer-Lifecycle of Long-Life Customers", *European Journal of Operational Research*, Vol. 156, pp. 508-523.

Ching, W. and Ng M.K. (2005), “Markov Chains models, algorithms and applications”, Springer.

Chintagunta, P. K. and Haldar, S. (1998), “Investigating purchase timing behavior in two related product categories“, Journal of Marketing Research, Vol. 35, No. 1, pp. 43-53.

Chi-Wing, R. and Wai-Chee Fu, A.(2005), “Data Mining for Inventory Item Selection with Cross-Selling Considerations”, Data Mining and Knowledge Discovery, Vol. 11, pp. 81-112.

Cohen, M.D. (2004), “Exploiting response models—optimizing cross-sell and up-sell opportunities in banking”, *Information Systems*, Vol. 29, pp. 327-341.

Degeratu, A.M., Rangaswamy, A. and Wu, J. (2000), “Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes”, *International Journal of Research in Marketing*, Vol. 17, pp. 55-78.

Demiriz, A., Kula, U. and Akbilek, N. (2009), “A Framework For Balanced Service And Cross-Selling By Using Queuing Science”, *Journal of Intelligent Manufacturing*, Vol. 20, pp. 249-257.

Durbin, R., Eddy, S., Krogh, A. and Mitchison, G. (1998), “Biological Sequence Analysis. Probabilistic Models of Proteins and Nucleic Acids”, *Cambridge University Press*.

Ehrenberg, A.S.C.(1959), “The Pattern of Consumer Purchases”, *Appl. Statistics*, 8, pp. 26-41.

Gupta, S. (1988), “Impact of Sales Promotions on When, What and How Much to Buy”, *Journal of Marketing Research*, 25, pp. 342-355.

Han, E.-H. and Karypis (2005),“Feature-Based Recommendation System”, *Association for Computing Machinery*, pp. 446-452.

Harrison, T. and Ansell J. (2002), “Customer retention in the insurance industry: using survival analysis to predict cross-selling opportunities”, Journal of Financial Services Marketing, Vol. 6. 3, pp. 229-239.

Hausman, J. and Mcfadden, D. (1984), “Specification test for the multinomial logit model”, *Econometrica*, vol 52, pp. 1219-1240.

Hill, W., Stead, L., Rosenstrein, M. and Furnas, G. (1995), “Recommending and evaluating choices in a virtual community of use”, *Association for Computing Machinery*, pp. 194-201.

Jain, D. and Vilcassim, N. (1991), “Investigating household purchase timing decisions: a conditional hazard function approach”, *Marketing Science*, 10(1), pp. 1-22.

Jarrar, Y. F. and Neely, A.(2002), “Cross-selling in the financial sector: Customer profitability is key”, *Journal of Targeting, Measurement and Analysis for Marketing,* Vol. 10, 3, pp. 282–296.

Kamakura, W. A., Kossar, B. S., Wedel, M. (2004), “Identifying innovators for the cross-selling of new Products”, *Management Science*, Vol. 50, No. 8, pp. 1120-1133.

Kamakura, W. A., Wedel, M., de Rosa, F. and Mazzo, J.A. (2003), “Cross-selling through database marketing a mixed data factor analyzer for data augmentation and prediction”, International Journal of Research in Marketing, Vol. 20, pp. 45-65.

Kitts, B., Freed, D. and Vrieze M.(2000), “Cross-sell: a fast promotion-tunable customer-item recommendation method based on conditionally independent probabilities”, *International Conference on Knowledge Discovery and Data Mining*, pp. 437-446.

Knott, A., Hayes, A. and Neslin S. A. (2002), “Next-Product-To-Buy Models For Cross Selling Appilications”, *Journal of interactive marketing*, Vol. 16, pp. 59-75.

Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R. and Riedl, J. (1997), “GroupLens: applying collaborative filtering to Usenet news”, pp.77-87.

Kumar, V., Morris, G. and Pancras J. (2008), “Cross-buying in retailing: Drivers and consequences”, Journal of Retailing, Vol. 84, pp. 15-27.

Larivière, B. and Van den Poel, D. (2004), “Investigating The Role Of Product Features In Preventing Customer Churn, By Using Survivial Analysis And Choice Modeling: The Case Of Financial Services”, *Expert Systems with application*, Vol. 27, pp. 277-285.

Li, S. , Sun, B. and Wilcox R. T. (2005), “Cross-selling Sequentially Ordered Products: An Application to Consumer Banking Services”, *Journal of Marketing Research*, Vol. XLU, pp. 233-239.

Liu, C.H and Cai S. (2007), “Customer Cross-Selling Model Based on Customer Maturity and Product Grade”, International Management Review, Vol. 3, No. 4, pp. 50-62.

Loh, W.Y. (2010), “Classification and regression trees”, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*.

Manchanda, P., Ansari, A. and Gupta S. (1999), “The "Shopping Basket": A Model for Multicategory Purchase Incidence Decisions, *Marketing Science*, Vol. 18, No. 2, pp. 95-114.

McFall J. (1969), “Priority Patterns and Consumer Behavior“, *The Journal of Marketing*, Vol. 33, No. 4, Part 1, pp. 50-55.

Morrison D. G. (1969), “On the interpretation of discriminant analysis”, *Journal of Marketing Research*, Vol. 6, No. 2, pp. 156-163.

Nash, D. and Sterna-Karwat, A. (2001), “An Application Of DEA To Measure Branch Cross Selling Efficiency”, *Computers and Operations Research*, Vol. 23, pp. 385-392.

Netessine, S., Savin, S. and Xiao W. (2004), “Revenue Management through Dynamic Cross-Selling in E-commerce Retailing”, Available at SSRN: http://ssrn.com/abstract=628505.

Ngobo, P.V. (2004), “Drivers of customers’ cross-buying intensions’, *European Journal of Marketing*, Vol. 38, No. 9, pp.1129-1157.

Paas, L. and Kuijlen T. (2001), “Acquisition pattern analyses for recognizing cross-sell opportunities in the financial services sector”, Journal of Targeting, *Measurement and Analysis for Marketing*, Vol. 9, 3,pp. 230-240.

Paas, L. and Molenaar, I.W. (2005), “Analysis of acquisition patterns: A theoretical and empirical evaluation of alternative methods”, *International Journal or Research in Marketing*, Vol. 22, pp. 87-100.

Paas, L.J., Vermunt , J.K. and Bijmolt, T.H.A. (2007), "Discrete time, discrete state latent Markov modeling for assessing and predicting household acquisitions of financial products", *Journal of Royal Statistic Society*, 170, Part 4, pp. 955-974.

Prinzie, A. and Van den Poel, D. (2009), "Dynamic Bayesian Networks for Acquisition Pattern Analysis: A Financial-Services Cross-Sell Application", *Lecture notes in artificial intelligence*, pp. 123-133.

Prinzie, A. and Van den Poel, D. (2008), "Random Forests for Multiclass classification: Random Multinomial Logit", *Expert Systems with Applications*, Vol. 34, pp. 1721-1732.

Prinzie, A. and Van den Poel, D. (2007), "Predicting home-appliance acquisition sequences: Markov/Markov for Discrimination and survival analysis for modeling sequential information in NPTB models", *Decision Support Systems*, Vol. 44, pp. 28-45.

Prinzie, A. and Van den Poel, D. (2007), "Random Multiclass Classification: Generalizing Random Forests To Random MNL And Random NB", *Lecture Notes in Computer Science,* pp. 349-358.

Prinzie, A. and Van den Poel, D. (2006), "Investigating Purchasing-Sequence Patterns For Financial Services Using Markov, MTD And MTDg Models", *European Journal of Operational Research*, Vol.1 170, pp. 710-734.

Reichheld, F. and Sasser, W. (1990), "Zero Defections: Quality Comes to Services", *Harvard Business School Review*, pp. 105-111.

Resnick, P., Lacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. (1994), "GroupLens: an open architecture for collaborative filtering of netnews", *Association for Computing Machinery*, pp. 175-186.

Rossi, P. E., McCulloch, R.E. and Allenby, G. M. (1996), "The Value of Purchase History Data in Target Marketing“, *Marketing Science*, Vol. 15, No. 4, pp. 321-340.

Russel ,G. J., Ratneshwar, S. and Shocker, A. D. (1999), "Multiple-Category Decision-Making: Review and Synthesis", *Marketing Letters*, 10:3, pp. 319-332.

Schafer, J., Konstan, J., and Riedl, J. (1999), " Recommender systems in e-commerce", *Association for Computing Machinery*, pp. 158-166.

Shaw, M. J., Subramaniam, C., Woo Tan, G. and Welge, M. E. (2001), “Knowledge management and data mining for marketing”, *Decision Support Systems*, Vol. 31, pp. 127-137.

Shardanand, U. and Maes, P. (1995), “Social information filtering: algorithms for automating “word of mouth”, *Association for Computing Machinery*, pp. 210-217.

Tirenni, G., Labbi, A., Berrospi, C., Elisseeff, A., Bhose, T., Pauro, K. and Poyhonen, S. (2007), “Customer Equity and Lifetime Management (CELM) – Finnair Case Study”, *Marketing Science*, Vol. 26, No. 4; pp. 553-565.

West, D. , Dellana, S. and Qian, J. (2005), “Neural network ensemble strategies for financial decision applications”, *Computers & Operations Research*, Vol. 32, pp. 2543-2559.

Yu, Xu and Kriegel (2001), “Selecting Relevant Instances for Efficient and Accurate Collaborative Filtering”, *Association for Computing Machinery*, pp. 239-246.

Acknowledgements

The authors would like to thank the anonymous online retailer for providing the data, and our colleagues and our friends for their patience and support.

Contact Information

Your comments and questions are valued and encouraged.
Contact the authors:

Miao Nie
ABN AMRO bank
Emily.nie@nl.abnamro.com

Shanshan Cong
SAS Institute
Shanshan.cong@sas.com