Divide and Prosper: Comparing Models of Customer Behavior From Populations to Individuals

Tianyi Jiang
Stern School of Business
New York University
tjiang@stern.nyu.edu

Alexander Tuzhilin
Stern School of Business
New York University
atuzhili@stern.nyu.edu

Abstract

This paper compares customer segmentation, 1-to-1, and aggregate marketing approaches across a broad range of experimental settings, including multiple segmentation levels, marketing datasets, dependent variables, and different types of classifiers, segmentation techniques, and predictive measures. Our experimental results show that, overall, 1-to-1 modeling significantly outperforms the aggregate approach among high-volume customers and is never worse than aggregate approach among low-volume customers. Moreover, the best segmentation techniques tend to outperform 1-to-1 modeling among low-volume customers.

1. Introduction

Customer segmentation, such as customer grouping by the level of family income or education, is considered as one of the standard techniques used by marketers for a long time [11]. Its popularity comes from the fact that segmented models usually outperform aggregated models of customer behavior [3]. More recently, there has been much interest in the marketing and data mining communities in building individual models of customer behavior within the context of 1-to-1 marketing [15] and personalization [1]. Although there have been many claims made about the benefits of 1-to-1 marketing [15], there has been little scientific evidence provided to this regard and no systematic studies comparing individual, aggregate and segmented models of customer behavior have been reported in the literature.

In this paper, we address this issue and provide a systematic study in which we compare performance of individual, aggregate and segmented models of customer behavior across a broad spectrum of experimental settings. We found that in general, there exists a tradeoff between the sparsity of data for individual customer models and customer heterogeneity in aggregate models: individual models may suffer from sparse data, while aggregate models suffer from high levels of customer heterogeneity. We study this tradeoff across different experimental settings and show that the individual level models significantly outperform aggregate and segment level models for high-volume customers and are never worse than aggregate models for low-volume customers across these experimental settings. Moreover, the best segmentation techniques perform significantly better than the aggregate and individual level models for low-volume customers. We also present other comparison results for aggregate, segmentation and individual approaches.

2. Problem Formulation

To build predictive models on customer behaviors, we used panelist datasets that track a set of customers' transaction histories over time. Let C be the customer base consisting of N customers, each customer C_i is defined by the set of m demographic attributes $A = \{A_1, A_2, \ldots, A_m\}$, k_i transactions $Trans(C_i) = \{TR_{i1}, TR_{i2}, \ldots, TR_{ik}\}$ performed by customer C_i (such as purchasing transactions), and h summary statistics $S_i = \{S_{i1}, S_{i2}, \ldots, S_{ih}\}$ (such as average dollar amount of purchase), computed from the transactional data $Trans(C_i)$.

Given this data, we learn predictive models of customer behavior of the form

$$Y = \hat{f}(X_1, X_2, \ldots, X_p) \qquad (1)$$

where $X_1, X_2, \ldots, X_p$ are some of the demographic attributes from A and some of the transactional attributes from T, and Y is one of the class labels which we try to predict. Function $\hat{f}$ is a predicative model learned via different types of machine learning classifiers.

Various models of customer behavior can be built at different levels of analysis as customers are grouped into various *segments* of different levels of granularity based on some of their demographic and behavioral characteristics. We consider the following levels of customer analysis:

- *Aggregate level* – the unit of analysis is the *whole* customer base, and only *one* predictive model of type (1) is built for the whole customer base.
- *Segmentation level* – "similar" customers are grouped into progressively finer *segments*, and the model(s) of customer behavior are built at each segment level based on the transactions and the demographic data of *that* particular grouping of customers. In this case, we still use the model of type (1) but learn it from the data pertaining *only* to the selected segment of customers. Moreover, we do this for *each* customer segment. In our study, the degree of customer similarity is determined by clustering summary statistics S_i, across customers.

- *Individual (or 1-to-1) level* – the unit of analysis is an individual customer, the model of customer behavior is built based *only* on the purchase transactions of *that* customer and his or her demographic data.

As we progress from the aggregate to the segmented and then to the individual models of customer behavior, we would create increasingly more “homogenous” customer groups for which predictive models are theorized to be more accurate. However, while we consider more refined customer segments, less data is contained in each such segment, and thus function $\hat{f}$ in (1) is estimated using fewer data points, potentially resulting in less accurate estimates.

Thus the general research question is to *determine which level of analysis would provide better prediction of customer behavior*, as defined by some measure of predictive performance of models of type (1). The answer to this question depends on the tradeoff between the sparsity of data for individual customer models and customer heterogeneity in aggregate models. In this paper, we study this tradeoff experimentally by comparing predictive models of type (1) across the three levels of analysis (i.e. individual vs. aggregate, individual vs. segmentation, and segmentation vs. aggregate) and six dimensions of different types of data sets; types of customers (high vs. low-volume); types of predictive models; dependent variables; performance measures; and segmentations techniques:

Types of datasets. We used two real world marketing datasets, ComScore panelist dataset from Media Merix on Internet buying behavior and Beverage panelist dataset from a major market research firm on household beverage purchase behavior. These datasets differed greatly in terms of the type of purchase transactions (Internet vs. physical purchases), variety of products, number of individual families covered, and demographics.

Types of customers. We partitioned our datasets into high and low volume customers to study the effect of data sparsity (details of data partitioning is described in [9])

Types of predictive models. We build predictive models using three different types of classifiers via Weka 3.4 system [18]: C4.5 decision tree [16], Naïve Bayes [10], and rule based classifier RIPPER [5] for building predictive models. These are chosen because they represent popular and fast classifiers.

Dependent variables. We built various models to make predictions on transactional variables, TR_{ij}, to compare discussed approaches across different experimental settings. The data we used to train any one model are customer C_i's independent variables $X_1, X_2, \ldots, X_p$, except TR_{ij}.

Performance measures. We use the following performance measures: percentage of correctly classified instances (CCI), root mean squared error (RME), and relative absolute error (RAE)[18].

Given models α and β, α is considered “better” than β when $\left(CCI_\alpha > CCI_\beta\right) \wedge \left(RME_\alpha < RME_\beta\right) \wedge \left(RAE_\alpha < RAE_\beta\right)$

Segmentation techniques. We use clustering algorithms to generate customer groupings across five levels of segment/sub-segment hierarchy. For Random clustering, customers are grouped together arbitrarily. The clustering techniques used are: *Random*, *SimpleKMeans[18]*, *FarthestFirst[8]*, and *Expectation Maximization.* Detailed description is presented in [9].

3. Related Work

The problem of building individual and segmented models of customer behavior is related to the work on (a) user modeling and customer profiling in data mining, (b) customer segmentation in marketing, and (c) building local vs. global models in statistics. We examine the relationship of our work to these areas in this section.

There has been much work done in data mining on modeling customer behavior and building customer profiles. Customer profiles can be built in terms of simple factual information represented as a vector or as a set of attributes. Customer profiles can also be defined by sets of rules defining behavior of the customer [2], sets of sequences such as sequences of Web browsing activities [7, 14, 17], and signatures, used to capture the evolving behavior learned from data streams of transactions [6].

There has also been some work done on modeling personalized customer behavior by building appropriate probabilistic models of customers [4, 12]. However, all these approaches focus on the task of building good profiles and models of customers and do not study the performance of individual vs. segmented and vs. aggregate models of customer behavior.

Comparison of segmentation vs. aggregate models of customer behavior has also been done by marketing researchers who demonstrated that segmented models of customer behavior exhibit better performance characteristics than aggregate models [3]. However, this work has not been extended to the 1-to-1 case and no comparison has been made between aggregate and individual, and between individual and segmented models.

Our work is also related to the work on clustering that partitions the customer base and their transactional histories into homogeneous clusters for the purpose of building better models of customer behavior using these clusters [19]. However, we go beyond simple partitioning to compare performance of aggregated vs. segmented and vs. individual models of customer behavior.

Finally, our work is related to the problem of building local vs. global models in data mining and statistics [7]. Rather than building one global aggregated model of customer behavior, it is often better to build several local models that would produce better performance results. Furthermore, this method can be carried to the extreme when a local model is built for *each* customer, resulting in 1-to-1 customer modeling. We pursue this approach and compare the performance of aggregate, segmented and individual models of customer transactions.

4. Comparing Individual vs. Aggregate Levels of Customer Modeling

In this section, we compare individual vs. aggregate levels of customer modeling. More specifically, we compare predictive accuracy of function (1) estimated from data *Trans*(C_i) for all the *individual* customer models and compare its performance with the performance of function (1) estimated from the transactional data for the *whole* customer base. In particular, we explore the aforementioned tradeoff between the heterogeneity of customer base and the sparsity of data.

To determine whether individual modeling performs statistically better than aggregate level modeling, we use a variant of the non-parametric Mann-Whitney rank test [13] to test whether the accuracy score of the one aggregate model is statistically different from a random variable with a distribution generated from individual accuracy results of the individual level models.

We conducted significance test for all customer type datasets across different dependent variables and classifiers. While performances of classifiers vary, our results clearly show that *for high-volume customers, modeling customer behavior at the individual level will yield significantly better results than the aggregate case*. In fact, modeling low-volume customers at the individual level will not be worse off than the aggregate level approach. The details of our experimental results are reported in [9].

5. Comparing Individual vs. Segmentation vs. Aggregate Levels of Customer Modeling

In this section, we compare individual vs. segmentation and aggregate vs. segmentation levels of customer modeling. More specifically, we compare predictive accuracy of function (1) estimated from the transactional data *Trans*(C_i) for the segmentation level models, and compare its performance with the performance results obtained in Section 4. To do this, we generate progressively finer customer sub-segment levels as explained in Section 2.

5.1 Performance Curves

To compare the clustering algorithms against aggregate and individual level models, we first compute the *best performing segmentation level* for a clustering algorithm as follows:

$$\text{Best Segment Level} = \arg\max\left(\overline{CCI}_l - \overline{RME}_l - \overline{RAE}_l\right),$$

where $l = 1 \ldots 5$ levels, and $\overline{CCI}_l, \overline{RME}_l, \overline{RAE}_l$ are the average CCI, RME, and RAE for all the groups at level l as defined in Section 2. We took the difference between these performance measures in order to compare the performance of various models as explained in Section 2.

As we compare various segmentation models against that of aggregate and individual models, we found that different clustering algorithms can produce significantly different patterns. To gain a better understanding of the various factors that influence the performance of our models across the three levels of analysis, we plot the *performance curves*, which plot a performance measure across different segmentation levels. For example, in Fig. 1, the performance curves plot the $\overline{CCI}$ measure across all levels of segmentation ranging from aggregate to the individual level. Out of the overall 260 performance curves of $\overline{CCI}$ generated in our experiments, three dominating patterns are presented in Fig. 1: (A) for *high-volume* customers and "well-behaved" clustering (clustering that performed significantly better than Random clustering), algorithms, we see a monotonically increasing curve; (B) for *low-volume* customer datasets and "well-behaved" clustering algorithms, we see a convex curve; (C) for low-volume customer datasets and "badly-behaved" clustering algorithms, we see a "concave" pattern. These dominant performance patterns help guide the interpretation of our subsequent findings.

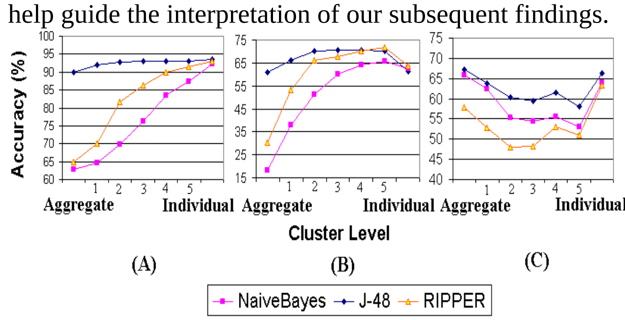

Figure 1. General Performance Curve Shapes

5.2 Segmenting Customers Using Clustering

We compare aggregate model to the best segment level using the same Mann-Whitney rank test we employed in Section 4. Our results show that best Segment Level significantly dominates aggregate level models across all customer types (Fig. 1-A,B). However, there is a significant number of instances where the aggregate level models performed better than best segment level models (Fig. 1-C). Further clustering performance analysis showed that this occurred because of some of the clustering algorithms produced poor performance results (see detailed results in [9]).

Comparisons of individual models to the best segment level models show that individual level models significantly dominate best segment level models for high-volume customers (Fig. 1-A). However, for low-volume customers, the best segment level models performed better in more instances than individual level models (Fig. 1-B). In particular, clustering analysis showed that the best segment level models in the best performing clustering algorithms significantly dominate individual level models.

We found significant differences in performances among the 3 non-random clustering techniques when compared against that of Random clustering method described in Section 2. When we focused our analysis on "well-behaved" clustering algorithms, such as FarthestFirst, we found that, as expected, the best segment level model performs significantly better than aggregate model across all customer types (Fig. 1-A,B). However, while individual level outperforms best segment level for the *high-volume* customers (Fig. 1-A), best segment level clearly dominates for the *low-volume* customers (Fig. 1-B). As mentioned in Section 4, there is general tradeoff between customer heterogeneity and data sparsity when building customer segmentation models. Our results strongly suggest that *aggregation of idiosyncratic customers with insufficient data outperforms individual level models*.

6. Conclusions

We conducted a comparative study of aggregate, segmentation, and individual level modeling across multiple dimensions of analysis such as different types of datasets, customers, predictive models, dependent variables, performance measures, and segmentation techniques. We identified four factors that significantly influence the prediction outcomes of customer behavior models: customer heterogeneity, data sparsity, quality of segmentation techniques, and levels of segmentation.

Our results show that, given sufficient transactional data, 1-to-1 modeling significantly outperforms other types of models of customer behavior. However, when modeling customers with very little transaction data, segmentation dominates individual modeling for the best segmentation techniques and the best level (granularity) of segmentation. What is surprising is that 1-to-1 modeling is never worse than aggregate level modeling in our experiments, even in the case of sparse data. However, we do not want to over-generalize this claim to arbitrary datasets since this phenomenon needs further study. We also showed that poor segmentation techniques could lead to poor performance results that are comparable to random segmentation.

We performed further analysis of the four influencing factors by plotting *performance curves* across all levels of customer segmentation and observed three dominating patterns presented in Fig. 1. The first monotone pattern occurs for high-volume customers and "well-behaved" clustering algorithms, and shows that we can build models of idiosyncratic customer behavior all the way to the individual level without running into the data sparsity problem. The second convex pattern occurs for low-volume customers and "well-behaved" clustering algorithms, and shows that we will eventually run into the problem of data sparsity while trying to build progressively finer models of customer behavior. The last concave pattern occurs primarily for low-volume customers and "poorly-behaved" clustering algorithms (i.e. clustering algorithms yielding statistically equivalent performance results as that of Random clustering). It occurs because heterogeneous customers are grouped into same segments by poorly-behaved clustering algorithms.

In the future, we would like to study the problem of predicting customer behaviors via different levels of segmentation under a more general class of experimental settings. We also need to gain a better understanding on the nature of the tradeoff between customer heterogeneity and data sparsity at a more theoretical level.

7. References

[1]. *Comm of ACM*, in *Special Issue on Personalization*. 2000.
[2]. Adomavicius, G. and A. Tuzhilin, *Expert-driven validation of rule-based user models in personalization applications.* Data Mining and Knowledge Discovery, 2001. **5**((1/2)): p. 33-58.
[3]. Allenby, G.M. and P.E. Rossi, *Marketing Models of Consumer Heterogeneity*. J. of Econometrics, 1999. **89**.
[4]. Cadez, I.V., P. Smyth, and H. Mannila, *Predictive profiles for transaction data using finite mixture models* 2001 UC Irvine
[5]. Cohen, W.W. *Fast Effective Rule Induction*. *ICML* 1995
[6]. Cortes, C., et al. *Hancock: a language for extracting signatures from data streams*. *ACM SIGKDD* 2000
[7]. Hand, D., H. Mannila, and P. Smyth, *Principles of Data Mining*. 2001: MIT Press. Sections 6.3.2-6.3.3
[8]. Hochbaum, S.D. and B.D. Shmoys, *A Best Possible Heuristic for the K-Center Problem.* Mathematics of Operational Research, 1985. **10**(2): p. 180-184.
[9]. Jiang, T. and A. Tuzhilin, *Divide and Prosper: Comparing Models of Customer Behavior From Populations to Individuals,* Stern *Working Paper, CeDER-0405*. NYU. 2004
[10]. John, G.H. and P. Langley. *Estimating Continuous Distributions in Bayesian Classifiers*. *UAI* 1995
[11]. Kotler, P., *Marketing Management*. 11 ed. 2003: Pren Hall
[12]. Manavoglu, E., D. Pavlov, and C.L. Giles. *Probabilistic User Behavior Models*. *ICDM* 2003
[13]. Mendenhall, W. and R.J. Beaver, *Introduction to probability and statistics*. 9th ed. 1994: Thomson Publishing
[14]. Mobasher, B., et al. *Using Sequential and Non-Sequential Patterns for Predictive Web Usage Mining Tasks*. *ICDM* 2002
[15]. Peppers, D. and M. Rogers, *The One-to-One Future*. 1993, New York, NY: Doubleday
[16]. Quinlan, R., *C4.5: Programs for Machine Learning*. 1993
[17]. Spiliopoulou, M., et al., *A Framework for the Evaluation of Session Reconstruction Heuristics in Web Usage Analysis.* INFORMS Journal of Computing, 2003: p. 15(2).
[18]. Witten, I.H. and E. Frank, *Data Mining: practical machine learning tools and techniques with Java implementations*. 2000
[19]. Yang, Y., B. Padmanabhan. *Segmenting Customer Trans. Using a Pattern-Based Clustering Approach*. *ICDM* 2003