AUTOMATED ADAPTIVE MODELING IN BIG DATA RECOMMENDER SYSTEMS: THE CASE OF MOBILE AD PLACEMENT

Donald Kridel, Department of Economics, University of St. Louis-Missouri, St. Louis, MO, 63121, 314-909-6651, kridel@umsl.edu
Daniel Dolk, Graduate School of Operational and Information Sciences, Dept of Information Science, Naval Postgraduate School, Monterey CA 93943, 831-229-0754, drdolk@nps.edu

ABSTRACT

Recommender systems (ReCo's) have become a familiar artifact in cyberspace as a vehicle for increasing revenues while deepening customer loyalty and satisfaction. Typical Reco's are used as cross-sell instruments to encourage existing customers to buy products and services related to previous purchases. However, this form of customer targeting is not the only one which ReCo's can perform. We show how to extend the conventional functionality of ReCo's to a very dynamic form of customer targeting as manifested in Mobile advertisement placement. This type of ReCo must be able to coordinate and access multiple very large databases, perform automated model generation to build large-scale logistic regressions, and implement adaptive modeling in the form of model balancing to reflect current user behavior during marketing campaigns, all in near real-time or extreme real-time. ReCo requirements for predictive real-time analytics necessitate the implementation of current "big data" software and hardware. We discuss the anatomy of such a recommender system designed for placement of Mobile advertising.

INTRODUCTION

Customer targeting in the most general sense is the process employed to determine who to "contact" in the context of a marketing campaign. In much the same way an archer must aim before shooting, a firm must decide which audience and customers to contact before initiating a campaign. Firms encounter the targeting problem in a variety of marketing settings:

- Acquisition: the firm wants to convert prospects into new customers;
- Up- or cross-sell: the firm wants existing customers to buy "bigger" (more or different) products and services;
- Response analysis: the firm wants to contact prospects that are more likely to respond positively to a given offer;
- Retention: the firm wants to entice existing customers to "continue" as a customer or prevent the customer from "churning", or canceling service;
- Conquesting: the reverse of retention, enticing customers to switch brands.

In each case above, a firm must find ways to identify promising contacts at reasonable costs in order for these initiatives to be profit enhancing. Customer targeting (hereafter referred to as simply "targeting") has been shown to be effective in increasing response rates for various marketing campaigns (see e.g. [1] [2]). The essence of the targeting problem is to find prospects that are expected to be the most responsive to a given offer (or contact). Additionally, targeting can also be used to determine "how" to contact (DM, phone, e-mail, mobile advertising, etc.) and "what" to offer (which product, what bundled services, which specific offer, etc.). We have described this process in detail in [3].

In [4] [5] [6] [7] we have described the use of model-based targeting to address the business targeting problem. In each of these cases, automated models were developed, and subsequently deployed (i.e., used to "score" potential customers) to produce a list of targets. In this paper, we extend this work to the construction of a *recommender system (ReCo)* to address the many complex facets of customer targeting in the Mobile advertising (display ads) domain. As we show below, mobile-based ReCo goes well beyond the conventional notion of cross-sell recommender systems as embodied, for example, in Amazon™ and Netflix™ ("people who bought this book/movie also bought the following books/movies") to a much more dynamic, real-time, and "big data" landscape.

RECO AS CUSTOMER TARGETING SERVICE: AUTOMATED ADAPTIVE MODELING

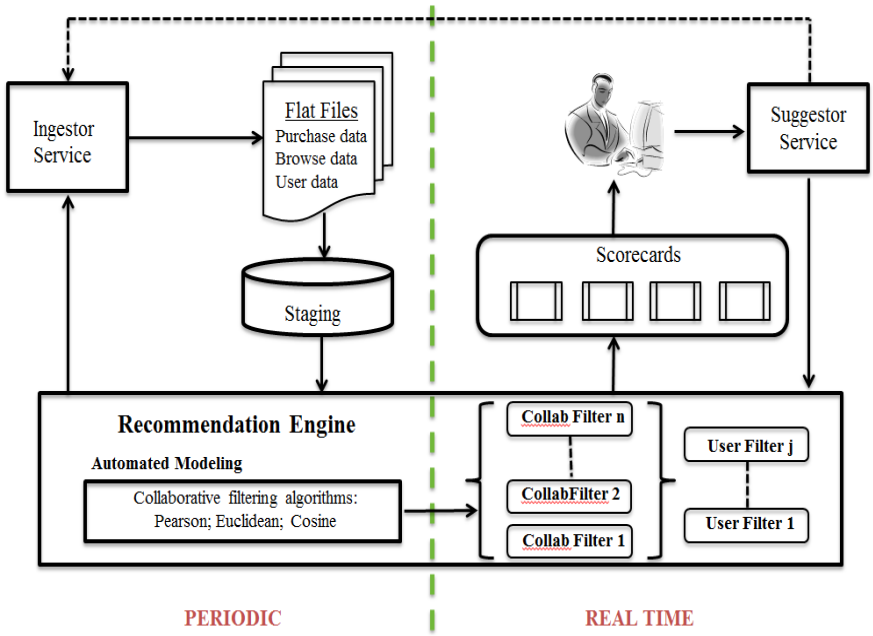
Mobile advertising faces the same customer targeting challenges mentioned above but in a much more demanding computational environment. Mobile advertising relies heavily upon what is called *programmatic marketing* or *computational advertising* [11] where the majority of campaign management processes are conducted via computing intermediaries such as ad networks, ad exchanges, supply side platforms, and demand-side platforms wherein advertisements are delivered on an individual-by-individual basis, often using auction mechanisms to balance demand and supply. In programmatic marketing situations, the problem of mobile device recognition becomes a paramount factor to consider as does the additional constraint that the ReCo must operate in real-time, and in real-time bidding (RTB) situations, extremely constrained real-time. In general ReCo must determine whether to serve an ad to an incoming request from a mobile device and its associated user, and if so, recommend which ad from the portfolio of current campaigns should be served. Further, this requires a response time on the order of 100ms.

Our approach to mobile media targeting is decidedly more dynamic than the collaborative filtering (CF)-based, cross-sell recommendation engine (Carrier ReCo) we have described in [8]. To provide a basis for comparison, Figure 1 shows a high level schematic for the Carrier ReCo consisting of two major temporal components, periodic and real-time, corresponding to the automated modeling and ReCo user interaction functions respectively, and comprising the basic model feedback loop of the overall system. Since user data is highly volatile as the result of users browsing and/or purchasing relevant items, it is necessary to rerun the CF models periodically to synchronize with these data updates (a process we call *model balancing*). Failure to do so may result in pushing recommendations for items which users have already purchased. Two major services are shown in the diagram. The Ingestor service essentially monitors the data volatility in the source flat files, periodically updates the model database discussed above (Staging phase), and reruns the collaborative filtering algorithms (Automated Modeling phase) resulting in a new set of filters. The Suggestor service uses these filters in concert with any user-specified constraint filters (e.g., "don't consider any products from Vendor X") to present current Scorecard recommendations to users. Any subsequent purchases or browsing done by the user is relayed in real time to the Ingestor service.

ReCo also provides more dynamic customer targeting services to clients as well [9] [10] [11]. With the Carrier ReCo example, we assume access to individual user data and their subsequent purchase and browsing behaviors, and subsequently make automated recommendations, identifying products and services we predict they may want to buy. Dynamic targeting considers the inverse of this situation and asks, "for specific products or services (e.g., mobile ads), which customers are most likely to respond and potentially purchase them?" The problem we address in this context is to provide a market service recommending who to target for mobile media advertisements [9]. This is a considerably more difficult

challenge since we are looking beyond just individual purchasing behavior to include additional consumer attributes that try to capture proxies for individual utility functions.

FIGURE 1. AUTOMATED RECOMMENDATION ARCHITECTURE FOR CF-BASED RECO



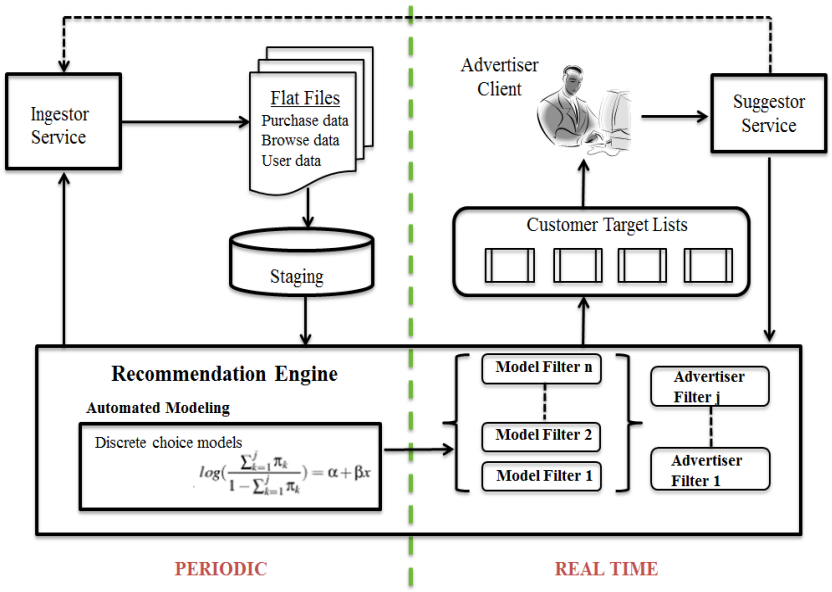
Mobile targeting requires that we adjust the targeting set dynamically as an advertising campaign unfolds allowing us to monitor who is responding to the ads in real time, and making changes to the targeting model "on the fly" if and when required. Although we employ the same basic ReCo architecture as shown in Figure 1, this application requires a more sophisticated recommendation engine including a transition from traditional dimensional data management to advanced "big data" techniques (Figure 2). Specifically, we implement discrete choice econometric models using both logistic and multinomial logistic regressions as the automated modeling component for initially identifying customers, coupled with a near real time, data-driven *model feedback* and *model balancing* loop which monitors the ongoing progress of a campaign and periodically updates the initial model to reflect actual "in-line" consumption.

This application differs from the standard CF-based ReCo in several important ways:

1. First, the analytical models are more sophisticated. Discrete choice models are employed to predict whether a particular individual is likely to click through a mobile ad appearing on his/her mobile device.
2. To make these predictions, models generally require considerably more detailed user demographic attributes which significantly increases the data collection effort. Once the relevant data are collected, it is then necessary to determine from the population of user attributes which ones are statistically meaningful independent variables.
3. Further complicating the process, the set of significant independent variables may change as the advertising campaign progresses and we can examine who is actually clicking through the ad and who is not. One of the consequences of this is that the feedback balancing loop is much more dynamic and therefore much tighter in this application than with the CF-based ReCo. Typically the

targeting model is recalibrated every 4-5 hours as opposed to the 24 hour refresh interval used in ReCo.

FIGURE 2. AUTOMATING CUSTOMER TARGETING USING ECONOMETRIC MODELS



MOBILE RECO AND REAL-TIME PREDICTIVE ANALYTICS: MOBILE MEDIA AND "BIG DATA"

The Mobile ReCo system has extenuating data and modeling requirements. In the data domain, One of the defining characteristics of the Mobile ReCo we are describing is the immense size of the underlying databases coupled with the very high degree of volatility these databases undergo. In the most general mobile advertising setting, there may be millions of customers arriving in ad hoc fashion on hundreds of publisher websites across hundreds of advertising campaigns running in parallel. Due to the potentially billions of signals per day generated within the mobile advertising ecosystem, traditional dimensional data management models are ill-equipped to store and manage the large volume of data directed to the predictive models described above. Rather it is necessary to resort to highly parallel and distributed computing techniques. In addition to logging the customer activity occurring during a campaign, there will typically be a large portfolio of supporting databases that help define user profiles subsequently used in developing the propensity models described below. These include, but by no means are limited to, demographics data (e.g., Acxiom™ household data of ~350M US households), mobile device data (database of ~300M+ mobile devices in the US), historical campaign data, proprietary advertiser-specific data ("1st party" data) on their customers and associated engagement profiles, and industry-specific databases (e.g., Polk™ automotive data sources). All these databases need to be coordinated and integrated to identify relevant user attributes which can be used in the propensity models which generate customer scorecards in real-time.

Modeling in this extreme environment requires a radical departure from conventional analytic modeling and data mining approaches. For example, there will typically be on the order of a hundred active ad

campaigns running simultaneously, each with its own associated propensity model. The propensity models themselves change many times during a campaign as the model adjusts to the data streaming in about who responds to an ad and who does not. Each propensity model may undergo roughly a hundred different instantiations over the lifetime of a single campaign (~30 days). Thus, there may be in the neighborhood of 10,000 active models in play during this overall time period. The model management environment of Mobile ReCo then is a highly real-time predictive analytics-driven setting which in turn requires the automatic model generation and adaptive model balancing e have described earlier.

The data and model dimensions of Mobile ReCo clearly require a "big data" solution with respect to hardware and software implementation. Specifically, this application requires a high degree of parallelized computing using a variety of big data solutions. We summarize our current suite of tools in Table 1. The data dimension of the system uses Spark streaming in combination with Kafka which enables large amounts of data to be processed in shorter amounts of time than with conventional Hadoop-based Map Reduce jobs.

The architecture for predictive modeling uses two forms of processing:

- *Feature selection* employs a typical Map Reduce framework for pruning the training dataset of all possible attributes (independent variables in the logistic regression) into the set of most important variables.
- *Propensity modeling* accepts the output of Feature Selection and iterates until all model conditions have been satisfied and generates a scorecard that is used by the real-time scoring engine. Kafka and Spark within this context of propensity modeling provide a "near real-time" approach for performing adaptive modeling. It fuels a process whereby the propensity model's predictors are updated with new coefficients and elasticities.

The final component of the architecture supports real-time scorecard generation which assesses the user's likelihood to engage with ad content. The architecture used to perform real-time scoring utilizes a Reactor pattern and is implemented in a compiled language to avoid performance problems associated with garbage collection [12].

TABLE 1. CURRENT "BIG DATA" SW SUITE FOR MOBILE RECO

System Function	Tools	Description
Data Collection Cascading technology High performance data aggregation ETL	Spark; Kafka	Hadoop-like "in-memory" data streaming; Open source high throughput distributed "publish subscribe" messaging framework
Predictive Analytics Modeling Feature selection Adaptive modeling and model balancing	Hadoop MapReduce; Kafka/Spark; Horizontally scalable algorithm	Proprietary nonlinear optimization algorithm for solving logistic models on parallel machines concurrently
Real-time Customer Scoring	Reactor patterns	Pattern-oriented software for network and concurrent objects [12]

The field of "big data" software tools is changing quickly as might be expected in such a fast-growing field. More recent deployments are experimenting with newer tools such s Scala and Scalding which perform ETL operations, data synthesis and feature selection on the very large numbers of data signals stored in log files on a Hadoop Distributed File System (HDFS). This approach preserves the integrity of the dimensional data model while minimizing the amount of rework necessary in other software components and services. We are also actively investigating the use of GPU's (graphical processing units) which show significant promise for dramatically parallelizing all dimensions of the Mobile ReCo.

CONCLUSIONS AND FUTURE RESEARCH

We've indicated how automated adaptive modeling in combination with big data tools can be deployed to address the daunting challenges of real-time predictive analytics for mobile advertising applications. Much interesting work remains to be done, especially in the area of devising near-optimum model balancing reformulation strategies which is the current performance bottleneck in the system. The creation of a simulation test bed for emulating campaigns and exploring alternative strategies would be a valuable asset in campaign planning and execution.

REFERENCES

[1] Rossi, P., McCulloch, R., Allenby, G.. The value of purchase history data in target marketing. *Marketing Science, Vol. 15, No. 4,* 1996, pp. 321-340.

[2] Shaffer, G. and Zhang, Z. Competitive coupon targeting. *Marketing Science, 14, 4,* 1995, 395-416.

[3] Kridel, D. and Dolk, D. Automated self-service modeling: Predictive Analytics As a Service, Information Systems and e-Business Management, March 2013, Volume 11, Issue 1, pp 119-140.

[4] Kridel, D. and Dolk, D. An on-line marketing consultant for small and medium businesses. *Proceedings of the 32nd WDSI Conference*, Lihue, HI, April 2003.

[5] Kridel, D. and Dolk, D. Using intelligent profiling to generate smart lists: An empirical test. *Proceedings of the 33rd WDSI Conference*, Manzanilla, Mexico, April 2004.

[6] Kridel, D. and Dolk, D. A self-service automated targeting portal: an example of model as a service. *Proceedings of the 38th WDSI Conference*, Lihue, HI, April 2009.

[7] Edmiston, E. and Kridel, D. "Automated Modeling and Sales Targeting: Case Study for AT&T Advertising and Publishing," Proceedings of DMA07 (Direct Marketing Association Conference and Exhibit), Chicago, IL, October 2007.

[8] Kridel, D., Dolk, D., Castillo, D. Recommender systems as a mobile marketing service, *JSSM, Vol 6, No. 5A*, December 2013, pp. 32-48.

[9] Kridel, D., Dolk, D., Castillo, D. Adaptive Modeling for Real Time Analytics: The Case of 'Big Data' in Mobile Advertising, *48th Annual Hawai'i International Conference on System Sciences,* Lihue, Hawaii, January 2015.

[10] Yuan, S.T. and Tsao, Y.W. A recommendation mechanism for contextualized mobile advertising. *Expert Systems with Applications, Vol. 24, No. 4,* 2003, pp. 399–414.

[11] Broder, A.Z. Computational advertising and recommender systems. *Proceedings of the 2008 ACM Conference on Recommender Systems,* ACM, New York, NY, 2008, 1-2.

[12] Schmidt, D., Stal, M., Rohnert, H., Buschmann, F. *Pattern-Oriented Software Architecture Volume 2: Patterns for Concurrent and Networked Objects*. September 2000, Wiley.