Application of PMML to Operationalize Thousands of Predictive Models for Retail

Shafi Rahman
Titanium Building
135 Airport Road,
Bangalore 560017
ShafiRahman@fico.com

Amit Kiran Sowani
Titanium Building
135 Airport Road,
Bangalore 560017
AmitSowani@fico.com

Josh Prismon
181 Metro Drive
Suite 700
San Jose, CA 95110
JoshuaPrismon@fico.com

Andy Flint
200 Smith Ranch Road
San Rafael, CA 94903
AndrewFlint@fico.com

Shalini Raghavan
2665 Long Lake Road
Roseville, MN 55113
ShaliniRaghavan@fico.com

ABSTRACT

Retailers are awash in data that is generated as a result of customer activity. This data comes from a multitude of sources that include but are not limited to point-of-sale systems. Many of the existing analytic methodologies in retail do not adequately capture customer behaviors, often resulting in a lack of relevance for the customer. Analytic Offer Manager [AOM] is an application that allows retailers to achieve a high level of personalization. This level of personalization is made possible by the ability to score thousands of models. Customer behaviors based upon frequency and affinity of transactions are mathematically modeled using transactional and non-transactional data. Operationalizing such a large array of models requires an efficient means to describe and execute models with minimal intervention and perfect accuracy. Our solution trains thousands of scorecard models on Big Data architecture and operationalizes them using PMML for rapid, precise descriptions of each model.

Categories and Subject Descriptors

H.2.8 [Database Management]: Data Mining; G.3 [Probability and Statistics]: Statistical Computing, Statistical Software; I.5.1 [Models]: Statistical Models, Neural Nets.

General Terms

Algorithms, Scorecards, RFM (Recency, Frequency, Monetary), Predictive Model Markup Language (PMML), Documentation, Standardization, Languages.

Keywords

Scorecards, PMML, Predictive Model Markup Language, Predictive Analytics, Data Mining, Decision Management, Time – to – Event, Analytic Offer Manager [AOM].

PMML ’13, Aug 10, 2013, Chicago, IL, USA

1. INTRODUCTION

Most organizations today acknowledge that leveraging customer data is key to their strategies to support customer centricity. Operationalizing this thinking has proved to be challenging for a number of different reasons. Several CRM systems in use today rely heavily on query-based rules to paint a better picture of the customer. Most retailers today continue to use online analytical processing (OLAP) as a means of analyzing and understanding customer behavior. Using OLAP analysis organizations can create rules and do descriptive analyses for specific customer behaviors that have occurred in a historical time-frame. Some CRM systems use the recency, frequency and monetary (RFM) value of transactions of specific purchases to predict future customer behavior. Customer behavior is complex, varied and continually changing. Both of these methods are limiting in that, they do not have a way of encompassing the dynamic behaviors from customers. Many organizations have resorted to using database triggers as an after-the-fact reaction to target specific customer behavior. These methods are proving both time consuming and ineffective and perhaps, the biggest loss, is the inability to understand and become more relevant to their customer base.

Given the dynamic nature of customer behavior, it is critical that organizations have a way to drive action and insight from continually changing datasets. AOM is an application that allows retailers to achieve true personalization through the use of analytics at scale. The application can consume a variety of data-sources, where each has the potential to depict a different aspect of the customer – demographics, transactions, interactions. It supports a number of different kinds of PMML-based Predictive Scorecard Models. AOM allows for predictions that encompass a variety of dimensions of evolving customer behavior such as purchase behavior that accounts for concurrent and non-concurrent purchases, discount and price sensitivities, lifestyle and life-stage events as well as the timing between purchases. It provides organizations the ability to run massively parallel analytics across these varied datasets in both batch and real-time.

2. SCORECARD MODELING APPROACH

Imperfections in real-world data present a challenge to their effective use in predictive models. Dirty data with outliers can lead to a lack of usefulness of the predictive outcome generated by a model. Modeling approaches that can make the most of both the dynamic nature as well as imperfections of real-world data are a must. Often times this aspect of predictive modeling can become cumbersome and can result in delayed model development cycles. In certain instances such as RFM, these aspects are ignored leading to misleading or poor predictive results.

The Scorecard modeling approach offers a number of features that allow for an improved predictive result. Scorecards allow for the capture of complex often non-linear relationships in the underlying data. They allow for a high level of robustness by allowing for the engineering of the model to account for selection bias as well as customization of ranges that are meaningful to specific business problems. This is critical, since it allows for the injection of domain knowledge into building the model. Furthermore by allowing for the automation of repetitive tasks such as variable binning, and score scaling the scorecard modeling approach can become a key ingredient in how scalable modeling approaches are constructed.

For the retailer it is most important to capture the time frame when the customer's propensity to purchase a product is at the peak. For example, a customer's propensity to buy baby care products lasts for a short duration, when there is a small baby in the family. Most any offers on baby care products would be relevant only in that time duration. The time-related nature of customer behavior is an important aspect to understand and action upon. Time to Event (TTE) ScoreCards allow for the prediction of not only what a customer will buy; but when they will buy. These models are a combination of FICO's scorecard technology and survival analysis.

Figure 1 provides an illustration of how TTE scorecards are used to predict propensities for a specified future time period.

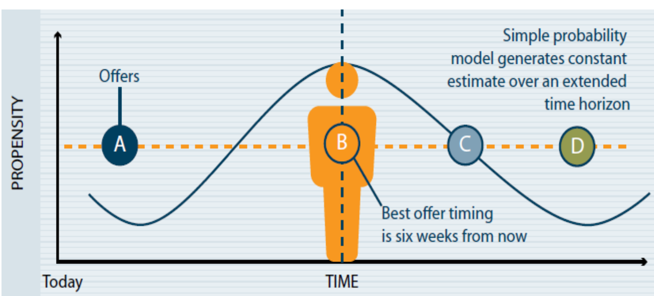

Figure 1. TTE Scoring Methodology

TTE models are truly a big-data methodology, in that they can use a variety of data sources transactional and non-transactional to enable a better prediction. Customer purchase patterns are extremely complex. Rule-based systems that capture domain knowledge are typically successful at leveraging obvious relationships in behavior patterns. While these rule-based methods are successful at capturing obvious and known behaviors they are not able to ascertain non-obvious behavior. TTE models are able to discover and predict such seemingly non-intuitive relationships like the purchase of landscape lighting leading to the purchase of fireplaces, purchase of spray paints leading to the purchase of water proofers and many more. Such relationships can be effectively captured to establish a trigger target relationship within the model thereby augmenting the predictive power of the model. Some customer attributes have a complex interaction with other attributes. Cross interactions of different attributes are captured though segmentation of the cross variables. Also they produce the greatest response and redemption rates for time bound offers with a defined validity period.

Fig 2 illustrates a simple scorecard model, similar to those that might be used to rank order credit accounts by repayment risk.

Variable	Bin	Score Weight	
Years of experience with credit	Missing, unknown	53	
	0	53	
	1 – 4	64	64
	5 – 9	75	
	≥ 10	81	
	Baseline score: 81		
Months since last late payment	Missing, unknown	78	
	No missed payments	85	85
	0 – 1	50	
	2 – 5	63	
	6 – 11	71	
	12 – 23	76	
	≥ 24	79	
	Baseline score: 85		
Bank account references	Checking & savings	56	
	One account only	52	52
	None	41	
	Missing, unknown	43	
	Baseline score: 56		

Figure 2. Illustration of Scorecard Model

Figure 3 shows the corresponding PMML representation for the first set of computations, as applied to variable "Years of experience with credit", which we translated to input variable "experienceInYears" in the PMML code.

```
<Characteristics>
   <Characteristic name="experienceInYearsScore"
       baselineScore="81">
      <Attribute partialScore="53" reasonCode="UK1">
         <SimplePredicate field="experienceInYears"
             operator="isMissing"/>
      </Attribute>
      <Attribute partialScore="53" reasonCode="EX1">
         <SimplePredicate field="experienceInYears"
             operator="lessThan" value="1"/>
      </Attribute>
      <Attribute partialScore="64" reasonCode="EX1">
         <CompoundPredicate booleanOperator="and">
            <SimplePredicate field="experienceInYears"
                operator="greaterOrEqual" value="1"/>
            <SimplePredicate field="experienceInYears"
                operator="lessThan" value="5"/>
         </CompoundPredicate>
      </Attribute>
      <Attribute partialScore="75" reasonCode="EX1">
         <CompoundPredicate booleanOperator="and">
            <SimplePredicate field="experienceInYears"
                operator="greaterOrEqual" value="5"/>
            <SimplePredicate field="experienceInYears"
                operator="lessThan" value="10"/>
         </CompoundPredicate>
      </Attribute>
      <Attribute partialScore="81" reasonCode="UND">
         <SimplePredicate field="experienceInYears"
             operator="greaterOrEqual" value="10"/>
      </Attribute>
   </Characteristic>
   ...
</Characteristics>
```

Figure 3. PMML code used to represent the computations involving a scorecard characteristic.

PMML provides a facile method of representation of scorecard models. They are highly portable making it easy to deploy and support in any system As such it allows for the deployment of thousands of models. The portability as well as the ease of implementation allow for the periodic deployment of models in the production environment.

3. PMML MODEL TRAINING

Retail data can be described by the three V's of Big Data - volume, variety and velocity. A single retailer can have millions of customers, with billions of transactions across thousands of products. Mining meaningful and actionable patterns from such sparse data can present a challenge. Furthermore, it is very important to ensure that the prediction of these patterns do not degrade over time. In order to overcome these challenges it is critical to train models over this very vast space of potential patterns and extract the most meaningful predictions.

The modeling platform is a highly automated, scalable, distributed modeling platform. All steps including customer data integration, exploratory analyses, sampling and feature extraction are fully automated.

Figure 4 illustrates the full-suite of automation that is applied to model training.

A distributed platform allows the processing of large datasets and uses stratified sampling to condense the relevant information by generating rolled up features. Scorecard models are an ideal candidate for automation as opposed to several other model types. This because the models can be easily parameterized and the numbers of parameters is smaller than certain other model types.

It is not unusual for automated algorithms to generate over-fitted or poorly extrapolated models. This is avoided by automated mechanisms to automatically identify and fix these issues. One simple technique uses a test dataset and computes a fitness-metric like RMSE on both training and test datasets, after each iteration of the fitting process. An increase in the RMSE on the test dataset accompanied by a corresponding decrease on the training dataset, is indicative of an over-fitted model. It is possible to validate each predictor on a test dataset and compute its marginal contribution. The marginal contribution of a variable represents the change in the model’s statistical performance, with and without the said variable. If this metric is significantly lower on test dataset then the predictor is deemed not to validate well and is automatically discarded before next iteration of model fitting.

In the case of several model-fitting algorithms, multi-collinearity can lead to misleading predictions. TTE ScoreCard models mitigate this risk of picking two or more highly correlated variables as predictors. Since the model fitting algorithm uses marginal contribution as the metric for inclusion of a variable, the selection of one variable by the algorithm renders the marginal contribution of the other correlated variable to zero.

The automated framework requires minimal intervention by the modeler and creates scorecard models for the multitude of products in parallel. The entire process can be controlled by setting up a few parameters which are defined by the business problem to be solved.

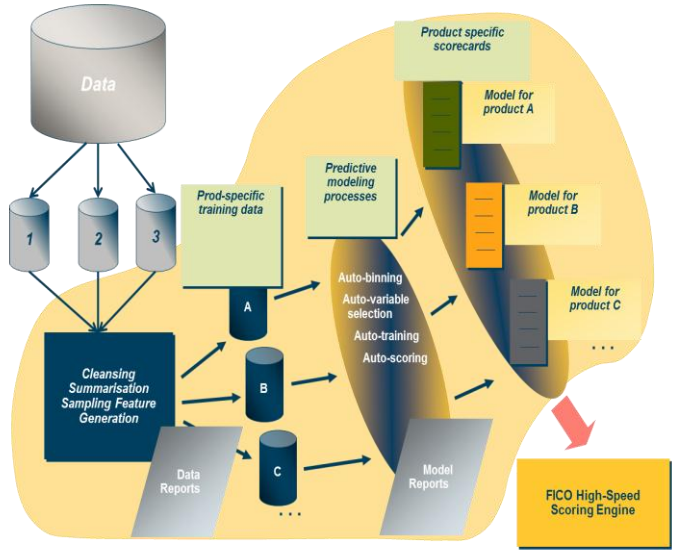

Figure 4. Model Training

3.1 Model Refresh

Automation of the modeling process comes with its own trade-offs in terms of the shelf life of models. It enables the minimization of intervention from the modeler and significantly reduces the runtime for creating the models. With this approach the models are not as optimally tuned as models with manual binning, variable selection and score engineering. The model performs well on the training datasets however with dynamic changes occurring in the data they are bound to degrade over a period of time. This necessitates periodic refresh of the models every four to six months to keep the models robust.

4. SCORECARD EXECUTION

Given the number of models, and the even greater number of predictors for those models, calculating characteristics presents a unique challenge.

For example, several unique events can act as predictors for a consumer’s propensity for a specific event. Events are temporal and transitory in nature. The fact that an event has occurred is important but the frequency and recency of the events are of greater importance. These events also are prone to shifts that may influenced by seasonality, changing behaviors and a change in temporal characteristics.

The universe of possible purchases, store visits, online web clickstreams and other transactional behavior is not easily summarized. AOM’s execution capabilities do not require summarized data; instead characteristics are generated on an as-needed basis. Since characteristics are specified and declared in the PMML directly, the system dynamically derives from the characteristic name how to generate the particular characteristic. For example a characteristic named f_product_10566_LOYALTY_USAGE, indicates that this is a frequency variable that should be calculated via the observation of product 10566. AOM’s characteristic system also supports an arbitrary system that is virtual in nature. Individual events may be classified into a number of hierarchies, including marketing hierarchies, product hierarchies, brand as well as analytically derived hierarchies.

Using techniques derived from complex event processing, the system maintains a list of characteristics that must be generated, and updates these characteristics during each scoring run. To handle the immense amount of data, AOM is typically distributed among many different machines, and workload is portioned out to each node individually. This allows AOM to literally generate hundreds of billions of scores representing thousands of events that can be fed into tens of thousands of models. Each score represents the intersection of a predicted event and an individual consumer. Since the system can generate billions of scores, it becomes possible to build a cohesive view of a customers predicted activities. This big data approach significantly increases the flexibility of the solution and the understanding of the consumer.

5. CONCLUSION

Scorecard models can be a powerful method to help organizations predict relevant actionable outcomes. They are successful at capturing complex and dynamic behavior, while accounting for imperfections in the data. The ability to express these models in PMML makes them very portable. This portability makes it possible to both train and score them in a massively parallel manner. AOM as a decision application is built to leverage the portability of PMML. As a true Big Data application it allows for the scoring of many thousands of models offering organizations a way to engage with their customers in a timely and relevant manner.

REFERENCES

[1] A. Guazzelli, M. Zeller, W. Lin, G. Williams. PMML: An Open Standard for Sharing Models. The R Journal, Volume 1/1, May 2009.

[2] A. Flint & A. Guazzelli (2011). Scorecard Element in PMML 4.1 Provides Rich, Accurate Exchange of Predictive Models for Improved Business Decisions. In Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.

[3] Scorecard for FICO ModelBuilder Whitepaper, http://www.fico.com/en/firesourceslibrary/introduction_scorecard_model_builder_1991wp.pdf