Prediction of Advertiser Churn for Google AdWords

Sangho Yoon [*]   Jim Koehler[†]   Adam Ghobarah[‡]

Abstract

 Google AdWords has thousands of advertisers participating in auctions to show their advertisements. Google's business model has two goals: first, provide relevant information to users and second, provide advertising opportunities to advertisers to achieve their business needs. To better serve these two parties, it is important to find relevant information for users and at the same time assist advertisers in advertising more efficiently and effectively. In this paper, we try to tackle this problem of better connecting users and advertisers from a customer relationship management point of view. More specifically, we try to retain more advertisers in AdWords by identifying and helping advertisers that are not successful in using Google AdWords. In this work, we first propose a new definition of advertiser churn for AdWords advertisers; second we present a method to carefully select a *homogeneous* group of advertisers to use in understanding and predicting advertiser churn; and third we build a model to predict advertiser churn using machine learning algorithms.

Key Words: Classification, Machine Learning, Data Mining, Customer Relationship Management, Online Advertising

1. Introduction

In Google AdWords, advertisers participate in auctions to show their advertisements to *users* who come to Google and search for information. Based on many factors (for example, bidding amount and various quality scores), only the winners will show their advertisements. This enables Google to provide more relevant information to users. Google's business model in AdWords is quite different from conventional ones in that Google's customers (or advertisers) participate in auctions to show their advertisements rather than directly selling a product to customers. Since advertisers are Google's customers, the terms "advertiser" and "customer" are used interchangeably in this paper unless otherwise mentioned.

Customer retention is important for Google to maintain its business model. It also benefits users by providing more relevant and high-quality information. This paper focuses on improving customer retention by addressing the customer churn problem in the unique environment of Google AdWords. To this end, we propose a new definition of customer churn for AdWords and a method to select a homogeneous group of customers for better understanding and prediction of customer churn in AdWords. Finally, we build a model to predict customer churn using statistical machine learning algorithms.

Customer churn has been studied in various fields (see references in Section 2). In general, churn can be categorized by contractual relationship (contractual vs. non-contractual) [11] and churn type (voluntary vs. non-voluntary) [10]. In contractual settings, customer churn can be predicted by simply looking at the contract termination date and its non-renewal. However, in non-contractual settings, predicting customer churn is not obvious. Non-voluntary churners are easy to identify

---

[*]shyoon@google.com, Google, Inc., 1600 Amphitheatre Pkwy, Mountain View, CA 94043

[†]jkoehler@google.com, Google, Inc., 2590 Pearl St. Suite 110, Boulder, CO 80302

[‡]ghobarah@google.com, Google, Inc., 1600 Amphitheatre Pkwy, Mountain View, CA 94043

because they are forced to churn by the company who has been accepting their business. Examples of non-voluntary churners are customers who are delinquent in payment or violate AdWords policy. Voluntary churn is more difficult to predict due to its nature: it is the customer's own decision to churn. In this work, we are interested in non-contractual churn because customers in AdWords do not have any contract with Google regarding ending date of using AdWords. In addition, we only consider voluntary churn simply because non-voluntary churn can be easily identified. Therefore, customers are churned if they terminate using AdWords or never win any advertising auction. It is important to note that customers can show seasonal behavior (for example, show advertisements only in the summer season). This must be carefully considered in identifying and predicting customer churn. Otherwise non-churned customers with seasonal behavior can be incorrectly labeled as churned during their seasonal *inactive* period.

We try to solve the problem of predicting customer churn in a supervised fashion. More specifically, customer churn is formulated as a binary classification problem where the ground truth (or dependent variable) is generated based on our definition of churn, which will be explained later.

In the following section, we review the related work in customer churn. In Section 3, we provide the details of our definition of customer churn for AdWords, and the methodology to select candidate samples and features for training and prediction. In Section 4, we explain the details about training classifiers and cross-validated classification results are also presented. Finally, we conclude and summarize our work in Section 5.

2. Related Work

Customer churn has been gaining significant attention in various market segments including subscription based businesses (newspapers or pay-tv) [3][4], insurance, finance and banking [16][17][24][13][25], internet service providers, telecommunications [21][27][19][20][18][1][23][12][2], business-to-business (B2B), and retail [11]. In the aforementioned articles, data mining technology and statistical data processing and evaluation methods are applied to identify and predict customer churn. In general, there are five stages in analyzing and building a predictive model for customer churn:

1. Select samples for analysis

2. Define churn and select features (potential explanatory variables)

3. Process data: transform features and impute missing values

4. Build predictive models

5. Evaluate trained models

This five stage modeling for customer churn management was first introduced in [5].

The first step of sample selection is crucial for correct understanding and prediction of customer churn. For example, in our work we are only considering voluntary churners. A common issue encountered in sample selection is that the population size of churned customers is often significantly less than that of non-churned customers. The authors in [25] report that a better prediction of customer churn is

achieved after class distribution is balanced with a re-sampling method. In [13], two different sampling schemes (proportional sampling and balanced sampling [14]) are compared in predicting customer churn. However, predictions can be biased if a model is trained on a balanced sample [18]. To overcome this bias problem, appropriate bias correction methods need to be considered [15][18]. *In the second step*, customer churn is defined and necessary features for predicting churn are collected.

In the third step, features can be processed via linear or non-linear transformation to generate more discriminating or relevant features to predict churn. In [25], linear combinations of features obtained from linear discriminant analysis (LDA) are used to maximize separation between churn and non-churn classes. Principal component analysis is performed to select relevant features in [18]. The author in [8] mentions that data preprocessing to derive new relevant features for prediction is a key to success in his application. In addition to processing features, missing values can be handled either by imputation or removing features with missing values. *In the fourth and fifth steps*, predictive models are trained and evaluated to choose the best one. There have been two different approaches to predicting customer churn. One concentrates on modeling repeat-buying using a stochastic defection process [6][7][11] and the other formulates customer churn as a binary classification problem where ground truth is either churn or non-churn (see [11][2] for example). We follow the latter approach as we detail in the next section.

3. Selection of Samples and Definition of Churn

In this section, we first explain how we sample and process the data. Then, we present our definition of churn for AdWords.

3.1 Data

We are interested in predicting customer churn for voluntary churners in a non-contractual setting. Moreover, we focus on customers who already have sustained a specified level of spending in AdWords, which reflects their sustained success in winning auctions. Unlike new customers, for which we have very little information, existing and established customers have more historical data from which we can better understand and predict customer churn. Therefore we use the following eligibility criteria to select a *homogeneous* group of customers:

1. Have similar tenure at a specified reference time point

2. Have similar spend range across a specified time period

3. Located in similar geographical regions

4. Manage their own accounts directly rather than through third party agencies

In this work, we decide to measure advertisers' tenure on 2008/01/01 and their spend levels are estimated based on their spend activity in the period of 2008/01/01 $\sim$ 2008/3/31. We also consider advertisers located in North America. The four filtering conditions described above are applied to select advertisers.

We consider two different types of features: static and time varying. Static features include customer static information and are based on the information that customers provided Google when they opened their accounts in AdWords. Time varying features are based on the customers' activity and we collect time varying

features within the time period of 2008/01/01 ∼ 2008/03/31. More over, to capture changes in customer behavior over time, we aggregate the time varying features monthly.

Note that the same period is used to filter customers and to collect monthly features. Therefore time varying features are collected when both churned and non-churned advertisers have similar spend levels. In other words, future churned and non-churned advertisers are active in this period. This enables our predictive models to capture inherently different characteristics between churned and non-churned advertisers. Due to the confidentiality of the data, we can not provide further details about features used in our work.

3.2   Definition of Churn

Customers can churn for various reasons, and churn can happen at any time. Churn prediction is made more difficult by the fact that customers can show seasonal behavior. Therefore, careful consideration should be given in defining customer churn to avoid incorrect identification of churned customers. To this end, we take into account the duration of the latest lapse period and changes in customer activity during the same season year over year.

Definition Consider customers who have shown advertisements in a season s in year y. We define customer churn for these customers one year after the last date of s in y. A customer is *churned* if none of their advertising campaigns have shown any advertisement for more than 181 consecutive days on the last date of s in year $y + 1$.

The first condition that checks year over year behavior is to avoid incorrect identification of customer churn due to seasonal behavior. And the second condition of lapsing more than 181 consecutive days is to make sure that customers are not temporarily pausing their campaigns but continue to be inactive for at least 6 months. Figure 1 illustrates an example where we select active customers from a season in 2008/1/1 ∼ 2008/3/31 and define customer churn on 2009/3/31.

4. Classification

We randomly select 35,000 customers from the pool of customers who satisfy the eligibility criteria in Section 3.1. This selected subset focuses on a specific segment within AdWords rather than the AdWords customer base as a whole. Thus the numbers presented in this Section are illustrative rather than representative of Google. Following our definition of churn, we label the sample. Note that our churn definition in Section 3.2 tracks customers' activity over a year as opposed to tracking only a few months as in other studies (see [11][26] for example). It is natural to observe a higher churn rate as we track customers for a longer period of time. Each customer is given a class label based on his lapse status as of 2009/3/31. If a customer had not shown advertisement for more than 181 days consecutively as of 2009/3/31 then the customer is churned and labeled 1. Otherwise not churned and labeled 0. Note that the customers sampled in our work have won advertising auctions in the period of 2008/01/01 ∼ 2008/03/31 and had a similar and positive spend range in the same period. Our ultimate goal is to identify inherent differences between churned and non-churned customers using features extracted while both
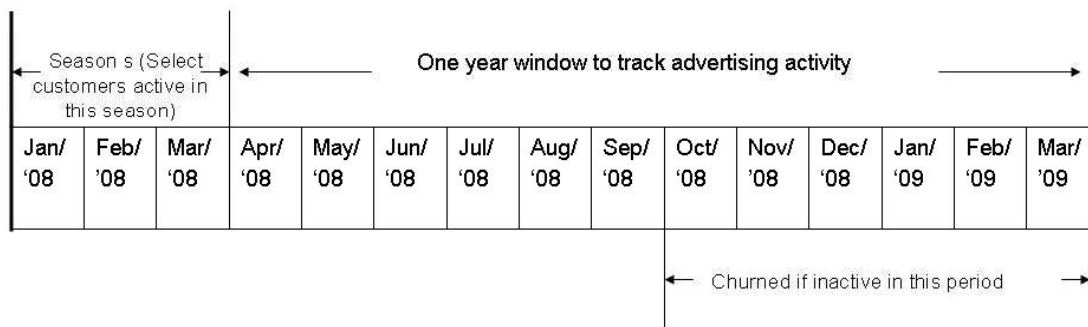
Figure 1: An example of definition of customer churn

groups of non-churned and churned customers were active. Overall we collect 103 features to train and predict customer churn.

Four classification algorithms are considered for building churn prediction models: 1. Boosted trees 2. Random forests 3. Logistic regression with L1 penalty 4. C4.5. All of these algorithms are implemented in R using standard R-packages: 1. gbm package for boosted tree 2. randomForest package for Random forests 3 glmnet package for Logistic regression with L1 penalty 4. RWeka package for C4.5. Note that the gbm package extends the gradient boosting algorithm [9]. See [22] for more details about the gbm package. Two-fold cross-validation is performed to compare their classification performance. We use true positive rate (TPR) and false positive rate (FPR) to measure classification performance, and TPR and FPR are defined as follows

$$\text{TPR} = \frac{\text{churned customers correctly classified}}{\text{total churned customers}}$$
$$\text{FPR} = \frac{\text{non-churned customers incorrectly classified}}{\text{total non-churned customers}}$$

Figure 2 shows the ROC classification performance of the four classification algorithms. As can be seen from Figure 2, boosted tree and Random forests outperform the other algorithms. This result is consistent with previous work in which boosting or Random forests based algorithms achieve the best classification performance in prediction of customer churn (see [18][25][28] for example.)

5

5. Conclusions

Google AdWords has thousands of advertisers participating in auctions daily to show their advertisements to users. It is important to Google to maintaining a good relationship with customers, and to help them reach their advertising goals. Moreover this objective benefits users by allowing Google to provide more relevant information to users. To this end, we built a model to predict customer churn for advertisers in AdWords. We first proposed a new definition of churn for customers in AdWords and then followed a binary classification framework to predict customer churn. Comparison of four state-of-the-art classification algorithms showed that tree-based ensemble algorithms (boosted tree and Random forests) outperform the other algorithms considered in this paper. This predictive model of churn can help identifying and assisting customers who are at risk of churning.
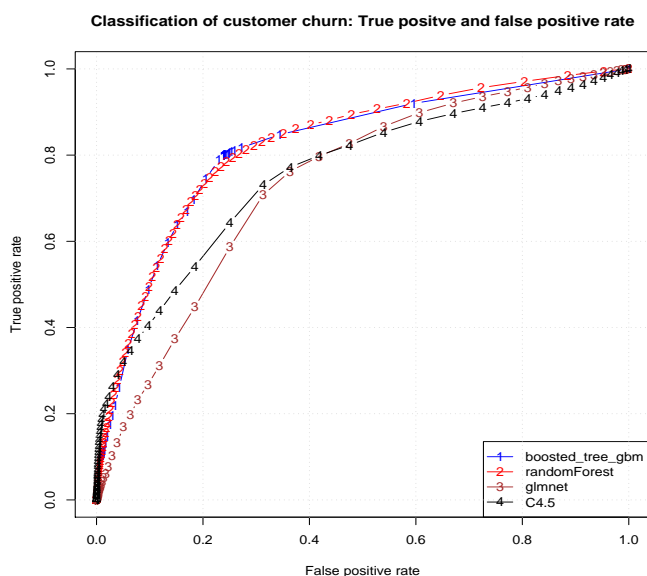

Figure 2: Classification of customer churn: true positive rate and false positive rate

References

[1] Wai-Ho Au, K.C.C. Chan, and Xin Yao (2003), "A novel evolutionary data mining algorithm with applications to churn prediction," *IEEE Transactions on Evolutionary Computation*, vol.7, no.6, pp. 532-545

[2] Indranil Bose , and Xi Chen (2009), "Hybrid Models Using Unsupervised Clustering for Prediction of Customer Churn," *Journal of Organizational Computing and Electronic Commerce*, vol. 19, issue 2, pp. 133-151

[3] J. Burez and D. Van den Poel (2007), "CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services," *Expert Systems with Applications*, 32, 277-288

[4] K. Coussement, and D. Van den Poel (2008), "Churn prediction in subscription services: An application of support vector machines while comparing two parameter selection techniques," *Expert Systems with Applications*, 34, 313-327

[5] Piew Datta, Brij M. Masand, D. R. Mani, and Bin Li (2001), "Automated Cellular Modeling and Prediction on a Large Scale," *Artificial Intelligience Review,* vol. 14, no. 6, pp. 485-502.

[6] A.S.C. Ehrenberg (1959), "The Pattern of Consumer Purchases," *Applied Statistics*, 8, 26-41

[7] A.S.C. Ehrenberg (1972), "Repeat Buying: Theory and Applications," North-Holland Pub. Co., American Else, Amsterdam

[8] Timm Euler (2005), "Churn Prediction in Telecommunications Using MiningMart," in *Proceedings of the Workshop on Data Mining and Business at the 9th European Conference on Principles and Practice in Knowledge Discovery in Databases*,

[9] Jerome H. Friedman (2001), "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics* vol. 29, no. 5, pp. 1189-1232.

[10] John Hadden, Ashutosh Tiwari, Rajkumar Roy, and Dymitr Ruta (2007), "Computer Assisted Customer Churn Management: State-of-The-Art and Future Trends." *Computers & Operartions Research*, v34(10), 2902-2917

[11] Jorg Hopmann and Anke Thede (2005), "Applicability of Customer Churn Forecasts in a Non-Contractual Setting," in *Innovations in Classification, Data Science, and Information Systems,* Baier Daniel and Wernecke Klaus-DIeter, eds. Berlin: Springer-Verlag, 330-37

[12] S. - Y. Hung, D.C. Yen, and H.-Y. Wang (2006), "Applying data mining to telecom churn management," *Expert Systems with Applications*, vol. 31, no. 3, pp. 515-524

[13] Shao Jinbo, Li Xiu, and Liu Wenhuang (2007), "The Application ofAdaBoost in Customer Churn Prediction," In *Proceedings of Service Systems and Service Management, 2007 International Conference on*, pp.1-6

[14] Gary King and Langsche Zeng (2001), "Explaining Rare Events in International Relations," *International Organizations*, vol.55, pp. 693-715

[15] Gary King and Langsche Zeng (2001), "Logistic Regression in Rare Events Data," *Political Analysis*, vol.9, pp. 137-163

[16] B. Lariviere, and D. Van den Poel (2004), "Investigating the role of product features in preventing customer churn by using survival analysis and choice modeling: The case of financial services," *Expert Systems with Applications*, 27, 277-285

[17] B. Lariviere, and D. Van den Poel (2005), "Predicting customer retention and profitability by using random forests and regression forests techniques" *Expert Systems with Applications*, 29, 472-484

[18] Aurelie Lemmens and Chistophe Croux (2006), "Bagging and Boosting Classification Trees To Predict Churn," *Journal of Marketing Research,* vol. 43, no 2, pp. 276-286

[19] , Michael C. Mozer, Robert Dodier, Michael D. Colagrosso, Csar Guerra-salcedo, and Richard Wolniewicz (2002), "Prodding the ROC Curve: Constrained Optimization of Classifier Performance", *Neural Information Processing Systems*, 14, MIT Press, Cambridge, 1409-1415

[20] Michael C. Mozer, Richard Wolniewicz, David B. Grimes, Eric Johnson, and Howard Kaushansky (2000), "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry," *IEEE Transactions on Neural Networks*, vol.11, no.3, pp.690-696

[21] Junfeng Pan, Qiang Yang, Yiming Yang, Lei Li, Frances Tianyi Li, and George Wenmin Li (2007), "Cost-Sensitive-Data Preprocessing for Mining Customer Relationship Management Databases," *Intelligent Systems, IEEE* , vol.22, no.1, pp.46-51

[22] Greg Ridgeway (2007), "Genearlized Boosted Models: A guide to the gbm package," available at http://cran.r-project.org/web/packages/gbm/vignettes/gbm.pdf

[23] Jiayin Qi, Yangming Zhang, Yingying Zhang, and Shuang Shi (2006), "TreeLogit Model for Customer Churn Prediction," In *Proceedings of IEEE Asia-Pacific Conference on Services Computing, 2006*, pp.70-75

[24] K.A. Smith, R.J. Willis, and M. Brooks (2000), "An analysis of customer retention and insurance claim patterns using data mining: A case study," *Journal of the Operational Research Society*, 51, 532-541

[25] Yaya Xie, and Xiu Li (2008), "Churn prediction with Linear Discriminant Boosting algorithm," In *Proceedings of Machine Learning and Cybernetics, 2008 International Conference on* , vol.1,.228-233

[26] Guo-en Xia, and Wei-dong Jin (2008), "Model of Customer Churn Prediction on Support Vector Machine," *Systems Engineering - Theory and Practice*, vol. 28, Issue 1, pp. 71 - 77

[27] Lian Yan, R.H. Wolniewicz, and R. Dodier (2004), "Predicting customer behavior in telecommunications," *Intelligent Systems, IEEE*, vol.19, no.2, pp. 50-58

[28] Weiyun Ying, Xiu Li, Yaya Xie and Johnshon Ellis (2008), "Preventing Customer Churn by Using Random Forests Modeling," in *Proceedings of Information Reuse and Integration, IEEE International Conference On,* pp. 429 - 434