Modeling Complex Clickstream Data by Stochastic Models: Theory and Methods

Choudur Lakshminarayan
Hewlett Packard Enterprises
choudur.lakshminarayan@hpe.com

Ram Kosuru
Hewlett Packard Enterprises
ram.kosuru@hpe.com

Meichun Hsu
Hewlett Packard Enterprises
meichun.hsu@hpe.com

ABSTRACT

As the website is a primary customer touch-point, millions are spent to gather web data about customer visits. Sadly, the trove of data and corresponding analytics have not lived up to the promise. Current marketing practice relies on ambiguous summary statistics or small-sample usability studies. Idiosyncratic browsing and low conversion (browser-to-buyer) make modeling hard. In this paper, we model browsing patterns (sequence of clicks) via Markov chain theory to predict users' propensity to buy within a session. We focus on model complexity, imputing missing values, data augmentation, and other attendant issues that impact performance. The paper addresses the following aspects; (1) Determine appropriate order of the Markov chain (assess the influence of prior history in prediction), (2) Impute missing transitions by exploiting the inherent link structure in the page sequences, (3) predict the likelihood of a purchase based on variable-length page sequences, and (4) Augment the *training* set of buyers (which is typically very small: $\sim 2\%$) by viewing the page transitions as a graph and exploiting its link structure to improve performance. The cocktail of solutions address important issues in practical digital marketing. Extensive analysis of data applied to a large commercial web-site shows that Markov chain based classifiers are useful predictors of user intent.

Keywords

Click Streams, Markov Chains, Link Analysis, Imputation, Prediction

1. INTRODUCTION

More recently, The Big Data movement of injecting analytics into business processes is driving the urgency to harness clickstream data for online analytics. It has been reported that the Chief Marketing Officer (CMO) may have a higher budget than the CIO and more particularly, the online analytics market is projected to be approximately 4 billion USD with a compounded annual growth rate (CAGR) of 22.5% by 2017(source Frost and Sullivan). In this context, the cutting edge is moving to digital marketing that utilizes data generated by customer interactions with websites. In spite of the steady growth of Internet commerce, and investment in website infrastructures, the average global conversion (turning a visitor into a buyer) rate is approximately $2 - 3\%$, an abandonment (browsers dropping off websites) rate of 75%, and for every dollar invested on converting a customer, 92 cents is spent on acquiring the customer (source: Gartner/marketing). Therefore, improving the conversion rate by a small fraction translates to billions in revenue.

Marketers have elaborate infrastructures to collect, store, manage and analyze vast amounts of customer data via web content management, reporting, multivariate testing and rudimentary analytics. However, ability to describe browser actions on websites are sadly confined to counting discrete web events such number of visits, number of pages clicked in a session, session duration, average page duration, abandonment, conversion, etc., followed by rudimentary predictive analytics. This paper is premised on the idea that a commercial website is a coherent network of related pages connected by a set of hyperlinks. A visitor's session-level sequence of clicks is viewed as navigation through a set of relevant pages (augmenting her/his knowledge about products and services rendered by the vendor) to eventually make a purchasing decision. It is therefore assumed that web navigation is an appropriate abstraction of browser intent and can be used to predict the likelihood of a buying-event. The presumed inter-relationships between pages in visitor sessions lend to modeling by a class of stochastic processes known as Markov chains (MC). Markov chains are studied extensively in web usage mining [1, 2, 3, 4]. Unlike other efforts in the literature, however, we will apply Markov chains to predict the likelihood of conversion based on variable-length page sequences within a session, while providing new techniques to enrich the Markov graph structures. Web sessions evolve as visitors traverse the website by clicking on links successively on pages. Thus, a session is a finite sequence of inter-linked pages. So, arrival at a given page is conditioned on being on a prior page. Markov chains postulate that the joint probability of a sequence of pages can be decomposed into a product of conditionally independent probabilities. Therefore, a Markov chain is specified by *conditional independence* and *order*. The notion of conditional independence assumes that the sequence of clicks is not a series of independent events. And conditional independence is parametrized by the degree of influence of past clicks (order). Treating the website

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04
http://dx.doi.org/10.1145/2872518.2891070.

as a graph where the URLs serve as *nodes* and the *edges* as transition probabilities [10], the simplest Markov model is of the $1^{st} order$ in which the probability of being on the present URL page depends only on the previous page clicked (page transition probability). Higher order chains trace history to previously clicked pages $2, 3, \ldots$. While application of MC in clickstream analytics seems straight forward, idiosyncratic customer browsing, arbitrary session definitions and page instrumentation makes modeling difficult. It is further exacerbated by low conversion rates for certain consumer electronic products. Despite the challenges, as we shall see, MC models successfully predict customer *browsing* behavior.

Additionally, while the application of Markov models in clickstreams may be seen largely as web usage mining, it has implications in the idea of *web of data* and the *semantic web*. The predictions derived from web usage models can be connected to other web sources. For example, fundamental to clickstream data collection is the IP (Internet protocol) address and technology allows IP addresses to be mapped to Zip codes (zoned geographical areas in the US). The connection between IP addresses and ZIP codes paints a broader and richer picture of the browsers as clickstreams, and demographic features get combined. In machine learning terms, once the Markov model identifies a subset of browsers with high likelihood of conversion, the associated zipcode level features are leveraged to identify homogeneous subgroups by statistical clustering. The unique characteristics within the sub-groups contextualize the usage patterns which can then could be used as pointers to make specific product recommendations. The demographics and lifestyles features therefore can be construed as semantic terms defined for the agent (algorithm). Clearly, the combined data sources in conjunction with the algorithms create an ecosystem as enunciated by Berners-Lee et al [5] to bring meaning and structure to serve the end-users (marketers). Mabroukeh et al [6] propose using semantic information in sequential pattern mining. In this framework, if the website ontology is avaiable, it is exploited to prune the state-space and a lower order Markov chain is used to reduce computational complexity and obtain accurate predictions. In many commercial marketing optimization models, the webpage semantic information is used together with the clickstreams to contextualize user navigation and dynamically modify pages with appropriate offers to nudge customer into buying. In this paper however, we will focus on web navigations patterns for predictions.

2. RELATED WORK

Markov models were used for predicting online customer satisfaction by Lakshminarayan et al [9] in surveys of online browsers. Also, as we have pointed out in the Introduction, Markov chains were studied extensively in web usage mining [1, 2, 3, 4]. Borges and Levene [1] discuss predicting likelihood of the next link based on variable length Markov chains. Deshpande and Karypis [3] propose higher order Markov models and present different techniques for intelligently selecting parts of different order Markov models so that the resulting model has a reduced state complexity, while maintaining a high predictive accuracy. Eirinaki et al [4] incorporate the pagerank method to provide web recommendations. They propose the use of a Pagerank-style algorithm for assigning prior probabilities to the web pages based on their importance on the website. Mobasher [2] in an excellent review paper, outlines ways to predict the next user-action proposed in the literature. Srivastava et al investigated web data to discover usage patterns in order to understand and better serve the needs of web based applications [7]. In all these cases, the problem of on-line buyer behavior is not explicitly addressed. In the Digital Marketing arena, where customers have unlimited access to commercial websites, the volume of browsers is very high, while the conversion rate is but a tiny fraction. In this setting, understanding buyer behavior is a complicated problem. Unlike what was proposed in the literature, we will see in the consumer electronics space, higher order chains do not necessarily yield higher returns (predictions). We advance the literature by approaching the problem determining the probability of a next link using the Chapman-Kolmogorov equations [11] which we use as primary tool for missing value imputation as well. The Chapman-Kolmogorov equations go beyond merely predicting the next link. It enables to calculate the probability of future link n steps ahead. It is achieved by computing the product of the one-level transition probability matrix n times. Section 5 presents new techniques that are shown to be useful when conversion rates are very low (such as less than 0.2%).

3. DATASETS, DESCRIPTION, AND CLASSIFICATION METHODOLOGY

The products we studied consumer are electronic goods such as desktops, notebooks, printers and supplies. The dataset for a given product is a table (T) where each row consists of a Session-id, Page name (URL), a class label ($'Buy,'$ $'No\ Buy'$). The objective is to identify prospects who will convert based on their session-level page navigation. To do this, we divide the dataset into two classes ($C_i, i = 1, 2$): $'buyers,'$ and $'non-buyers.'$ The *training* data is composed by obtaining 70% from ($C_i, i = 1, 2$) and the *testing* set is obtained by combining the remaining 30%. In the next step we collect all URL sequences from the *training* set to create transition probability matrices (TPM) for the two classes C_1 and C_2. Each session-level URL sequence in C_1 is decomposed into pair-wise URLs $i \to j$. Note that the transitions $i \to j$ are the $'From'$ and $'To'$ pages occurring in the sessions contiguously. For example, if a session consisted of page transitions $U_1 \to U_2 \to U_3$, the pair-wise transitions would be, $U_1 \to U_2$ and $U_2 \to U_3$. From the set of all $(i, j) \in C_1$, we compute transition probabilities p_{ij}. The pair-wise probabilities p_{ij} form the TPM ($\mathbb{P}_1$) . Similarly, we generate the matrix $\mathbb{P}_2$ for class C_2. When a session from the *testing* set is presented to the MC, it is classified based on following reasoning. Let a new session denoted by U consist of the sequence of pages $(U_1, U_2, \ldots U_k)$. We compute class-conditional joint probabilities;

$$P(U_1, U_2, \ldots, U_k|C_1) \ , \ P(U_1, U_2, \ldots, U_k|C_2)$$

denoted by $\mathcal{L}_1$, $\mathcal{L}_2$ respectively. The joint probability by the application of Markov chains, decomposes into conditionally independent one-level page transition probabilities(p_{ij}) stored in $\mathbb{P}_1$ and $\mathbb{P}_2$. For example, the joint probability, $P(U_1, U_2, \ldots, U_k|C_1)$ is decomposed as follows:

Session Name	Page	Class
320611406779612746869641750-4	us:welcome-home	0
320611406779612746869641750-4	us:sale:static:springsale	0
320611406779612746869641750-4	us:en-us:laptops notebook pc	0
320611406779612746869641750-4	us:laptops pavilion 15t-n200 notebook pc with windows 7	1

Table 1: A sample user session

$$\begin{aligned} P(U_1, U_2, \dots U_k) = P(U_k|U_{k-1} \dots, U_3, U_2, U_1) \\ * P(U_{k-1}|U_{k-2}, \dots, U_3, U_2, U_1) \dots \\ * P(U_3|U_2, U_1) * P(U_2|U_1) * P(U_1) \end{aligned} \quad (1)$$

$$= P(U_1) \prod_{i=2}^{k} P(U_i|U_{i-1}) \quad (2)$$

As we can see, Eq. 1 is simply the decomposition of the joint probability of the page sequence into product of conditionally independent events. Eq. 2 is a result of invoking the property of the first-order Markov chain. We dropped the class label C_1 for simplicity of notation.

Figure 1 is an overview of the modeling methodology including the decision rule. The integers $'1'$ and $'0'$ in the flow chart represent the $'Buy'$ and $'No\ Buy'$ classes. The decision rule for classifying sequences is determined by the ratio, $\mathcal{D} = \frac{\mathcal{L}_1}{\mathcal{L}_2}$. If $\mathcal{D} \leq 1$ then $U \in C_2$, else $U \in C_1$. Repeating the procedure for each session in the *testing set*, the performance metrics; recall (re)$=\frac{TP}{TP+FN}$ and false positive rate (FPR) $=\frac{FP}{TN+FP}$ are evaluated. The quantities TP, TN, FN, FP are respectively true positives, true negatives, false negatives, and false positives. In practice the model ought to predict conversion based on a user's current session-depth (length). Next section summarizes model performance (recall, and FPR) as a function of session length (l).

We show the schema of the dataset in the Table 1. During a session, a user visits many pages. The class column contains a 1 or 0 depending on whether user made a purchase at the end of the session. The total number of sessions is more than 2 million; the number of distinct pages is 5800; the collected data recorded purchases of desktops, notebooks, printers, and supplies, with conversion rates of 0.2 %, 0.2%, 0.1%, 1% respectively. The conversion rate is about 1.5% when all purchases are consolidated.

4. RESULTS

We will examine results for the product categories "supplies," and "notebooks." We chose supplies and notebooks because, the product category $'supplies'$ has the highest conversion among all the products, and notebooks have multiple product lines giving a richer sequential patterns. Also, we only report results for Markov chains up to 3 *orders* since performance decays at higher *orders*. More specifically, for the product category $'supplies'$, recall is in the range of (0.55-0.66) range, while FPR is dismally high in the range of (0.5, 0.66) over the session depth (5-35). Thus the decision to limit analyses to chains up to third order Markov chains. The metrics recall and FPR shown are averages calculated from multiple datasets randomly generated from the original table (T). Repetition over multiple sets provides an estimate of the standard error (se) to assess algorithm stability. Figures 2 and 3 summarize the results for the product category "supplies." Note that the X-axis is session depth (length). Clearly, there are no significant differences between Markov chains of various *orders*. At smaller session lengths the models deliver impressive recall (85%) and false positive ($\leq$ 10%). So a 1^{st} order MC is preferable due to low complexity, manageable bookkeeping and ease of implementation. We notice likelihood of purchase is high when the session (l) is between 5 and 30 pages. The decline in recall has a slightly steeper gradient for session lengths > 30. So we call attention to length (l) in the interval [5, 30]. In order to study *repeatability* of the classifier, we generate the *training* and *testing* sets by resampling in 70/30 proportions multiple times (6). The purpose is to evaluate the stability of the classifier in the presence of randomness. The variability from sample to sample is measured by the standard error(s) (s.e.) of the metric(s). The standard error of recall increases being in the interval $[0.001, 0.04]$ at session-length ($5 \leq l \leq 30$). Similarly, the standard error of FPR is in the interval [0.01-0.02] when ($5 \leq l \leq 30$). This analysis confirms algorithm stability. Note that we report standard errors for the $1^{st} - order$ Markov chain only. The actionable insight is that in a real-time setting, marketers should target sessions ($l \leq 30$) when customers are likely to buy. The product category notebooks whose conversion rate is 0.2% is analyzed next. The performance summary is depicted in Figures 4 and 5. The average Recall is lower compared to "supplies" obviously due to low conversion in this product category. However, FPR is small and decreases with increasing session-length. The standard errors are in the ranges [0.06, 0.11], and [0.02, 0.04] for recall and FPR respectively. The small bump at session length 40 (Figure 3) is an artifact of the dataset (unexplainable variation). Needless to say, higher conversion rate will yield better performance. The product categories Desktops and printers are excluded due to space constraints. The results for the products desktops (DT) and printers (Pr) are given in Table 2. The columns in the Table are self explanatory. In summary, the recall declines gradually as the session length increases. The FPR rate shows similar behavior as well. Due to low conversion, the recall is poor, but the algorithm is able to identify the true negatives (non-buyers). The results for the product category printers follow the same pattern. In spite the sub optimal performance, the standard error (se) is small, demonstrating the stability of the algorithm. Another common problem in MC modeling is missing transition probabilities p_{ij} in the *training* set, but appear in the testing phase. Missing transitions (values) is a vexing research issue and can not be understated. We impute missing probabilities by the Chapman and Kolmogorov (C-K) equations[11]. C-K equations is a tool to estimate $(n-level, n \geq 2)$ transi-

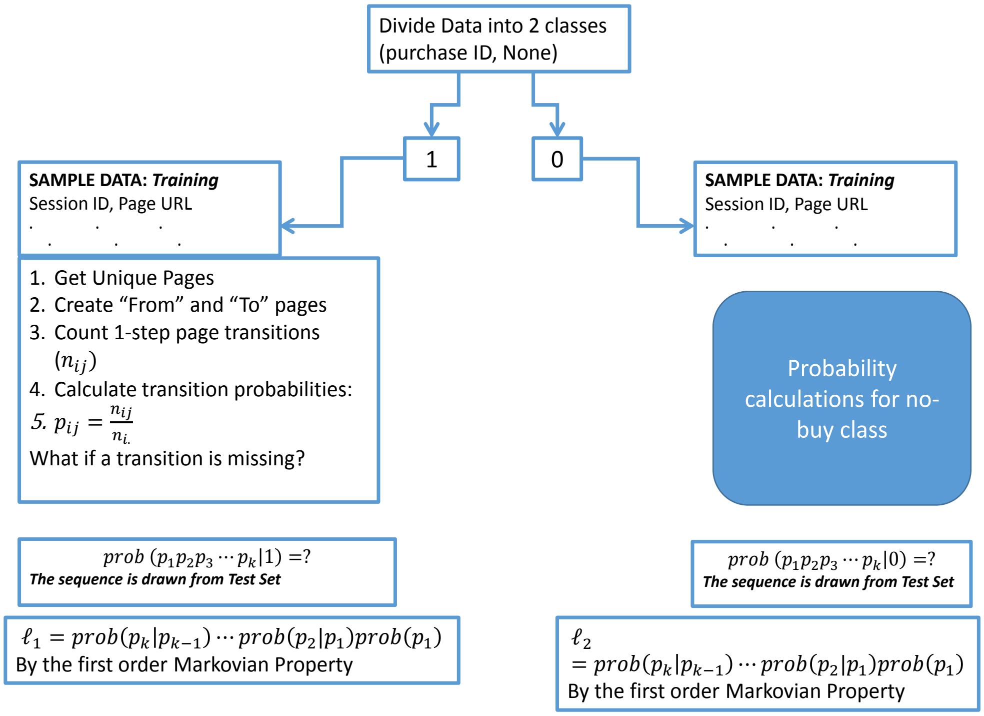

Figure 1: A flowchart describing the proposed classifier

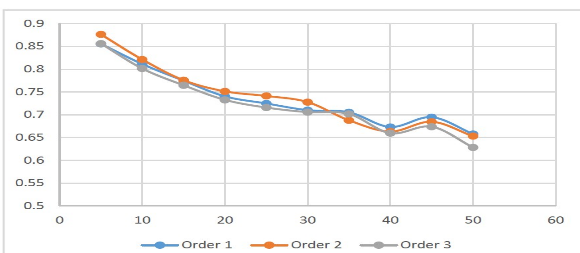

Figure 2: Recall(Supplies): Performance is $\sim$ equal across $1^{st}, 2^{nd}, 3^{rd}$ orders. X-axis is the partial length of sequence within a session

Figure 3: FPR(Supplies): 1^{st} order MC is within range of higher order chains. X-axis is the partial length of sequence within a session

tion probabilities. If pages $i \rightarrow k$ are linked and pages $k \rightarrow j$ are linked, C-K imputes p_{ij} from $i \rightarrow k \rightarrow j$ by marginalizing over all $k \in \mathbb{S}$, $\mathbb{S}$ is the collection of all pages on the website. In conclusion, Markov models are practical tools to identify prime prospects. Despite low conversions for consumer electronic products products, idiosyncratic browser behavior, similarity of product offerings (laptops, desktops yielding similar navigation paths), missing probabilities, and large number of pages hosted (5800), results show Markov chains are viable. As a cautionary note, issues cited require a deliberative approach; modifying, evaluating until a suitable model is found.

Figure 4: Recall(Notebooks): 1^{st} order MC delivers superior performance. X-axis is the partial length of sequence within a session

5. AUGMENTED GRAPHS WHEN CONVERSION IS LOW

In practice, the *training* set for the class $'Buy'$ is much smaller than that of the class $'No\ Buy'$; a phenomenon that

Length	Re(DT)	FPR(DT)	se(Re:DT)	se(FPR:DT)	Re(Pr)	FPR(Pr)	se(Re:Pr)	se(FPR:Pr)	Order
5	0.65	0.11	0.04	0.01	0.60	0.11	0.04	0.01	1
10	0.49	0.06	0.05	0.01	0.52	0.06	0.03	0.01	1
15	0.36	0.04	0.08	0.01	0.39	0.04	0.03	0.01	1
20	0.29	0.03	0.05	0.01	0.30	0.03	0.06	0.01	1
25	0.24	0.02	0.09	0.02	0.23	0.02	0.07	0.02	1
30	0.25	0.02	0.07	0.02	0.19	0.01	0.02	0.02	1

Table 2: Performance Summary of Desktops(DT) and Printers (Pr).

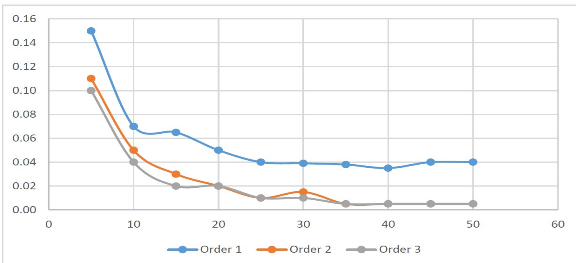

Figure 5: FPR(Notebooks): 1^{st} order MC has slightly higher false positives, but within range. X-axis is the partial length of sequence within a session.

we call *information asymmetry*. Information asymmetry known in the machine learning literature as *skewed* data and affects the predictive power of the classifier resulting in an inflation of false positives and false negatives. Therefore, the classifier-influenced business processes might be incorrect with a negative impact on the desired business outcome; For example, The Wall Street Journal [12] claims that recommendations from algorithms on e-commerce websites are often unreliable. There may be multiple reasons for this, but this condition could be ameliorated by augmenting the link structure of the $'Buy'$ class. In other words by adding new links to the graph structure associated with the $'Buy'$ class, we seek to enrich the graph to improve the classification. As most sequences are in $'No\ Buy'$ class we do not augment the corresponding graphs. We augment the *training* set of the class "buy" to improve the classification and apply Markov chains of various orders (complexity) to distinguish buyers from non-buyers. The techniques are also used for multi-class classification and are thus extensible. Establishing new edges (augmenting a graph) and consequently estimating transition probabilities is demonstrably beneficial. When conversion rates are low (<0.5%), our experimental results show improved performance.

The idea is to build new graphs from the existing URL links in the website by identifying paths from $i \to j$ via many intermediate layers (pages), i.e., $i \to k \to l \to j$ (where k and l are pages in the intermediate nodes on the path from i to j). In a G_3^a graph (the index "a" denotes augmented graph), path $i \to k \to l \to j$ is computed, using a G_2 path $i \to k \to l$ and an existing edge $l \to j$ from G_1. So by linking $i \to k \to l$ and $l \to j$ and marginalizing over nodes k and l with edge $k \to l$ we obtain p_{ij}. Most sequences belong to the "no buy" class (about 98% of the sequences), therefore the corresponding graphs are not augmented. These graphs are labeled nG_i where $i = 1, 2, 3, \ldots$ They are constructed using sampling from the *training* data set of the "no buy" class; sampling is used so that number of sessions from the *training* dataset is about 10 times that of the number of entries from the *training* data set of the buy class.

Now the classification using augmented graphs follows by slightly modifying the logic used in Figure 1: The 3^{rd} order Markov chain (G_3^a) enables computation of conditional probability $P(j|l,k,i)$ for the path $i \to k \to l \to j$. Using G_3^a, we have $P(i \to k \to l \to m \to j) = P(j|m,l,k) * P(m|l,k,i) * P(l|k,i) * P(k|i) * P(i)$. Using G_2^a, we have $P(i \to k \to l \to m \to j) = P(j|m,l) * P(m|l,k) * P(l|k,i) * P(k|i) * P(i)$. By viewing a session as a walk on the buy graph G_i^a or the no-buy graph nG_i, we compute joint probability of the walk from both the graphs G_i^a or nG_i where i=2,3,...; we choose the higher of the two joint probabilities to classify the session. We use sessions from the *testing* dataset to query the graphs to obtain joint probability. It is clear that there is no augmented G_1^a graph and hence we use G_i^a for $i = 2, 3 \ldots$.

5.1 Results

Experiment 1: We show that the metric FPR is much better when using G_2^a over G_2. We combined all buy sessions into one class. Using the training data, we constructed graphs G_i, nG_i for $i = 1, 2, 3$ to model the i^{th} order Markov chain. Similarly we constructed augmented G_i^a for i=2,3. Figures 6 and 7 show performance of the 2nd order Markov chain using G_2, nG_2 and augmented G_2^a. Figure 6 compares how performance of FPR changes over session length: FPR(TC) shows the false positive ratio when G_2^a, nG_2 are used for classification, that is, TC is for the augmented case. FPR(MC) shows false positive ratio in the normal case, that is, MC is for normal case. FPR is much better in the augmented scenario compared to normal scenario across all session lengths. Recall is slightly better in the standard scenario when session length ≥ 30. This is under analysis.

Experiment 2: We consider the notebook purchases that have a conversion rate $\leq$ 0.2%. This experiment clearly shows the power of augmentation where both metrics perform better when using augmented graphs. We trained a 3^{rd} order Markov chain by constructing the three graphs G_3, G_3^a and nG_3. Using the *testing* data we queried G_3 and nG_3 to classify our testing data to report FPR (MC). We queried the augmented G_3^a and the no-buy nG_3 to classify our testing data to report FPR (TC). Figure 8 shows that FPRs from the augmented case outperform the normal case. We also report the performance of metric Recall in Figure 9. For notebooks, we used 3866 sessions in the training set for the buy class; 314400 sessions are used to construct the graphs for the no-buy class. The size of *testing* sets are 383, 30991 for the "buy'' class and the "$No\ Buy''$ class respectively.

6. CONCLUSION

Maintaining large transition probability matrices requires bookkeeping and in real time may affect speedy response.

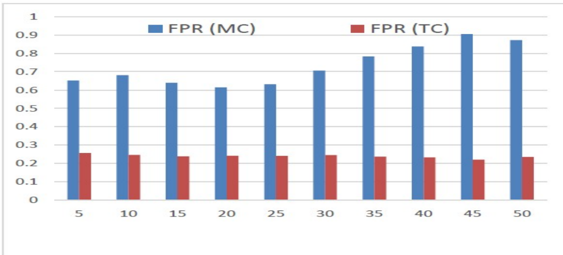

Figure 6: False positive rates are much better with TC. X-axis is the partial length of sequence within a session.

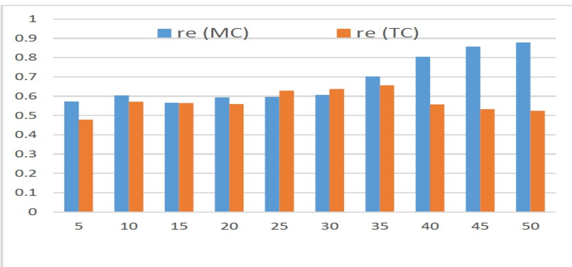

Figure 7: Recall could be better with TC. X-axis is the partial length of sequence within a session.

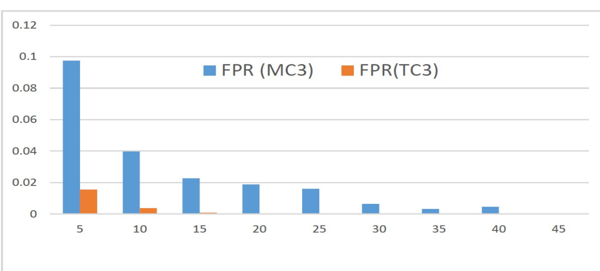

Figure 8: False positive rates are much better with TC. X-axis is the partial length of sequence within a session.

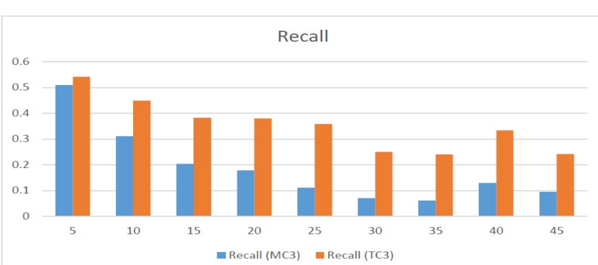

Figure 9: Recall rates are much better with TC. X-axis is the partial length of sequence within a session.

We are working on equivalent, sufficient statistics for transition probabilities to overcome the potential bottleneck. When conversion rates are low, augmenting the higher order Markov chains results in better classification. A possible next step is to look for a way to parallelize the construction of graphs so that we can build higher order Markov chains; some applications may benefit from more history on the random walk for a given query. In conclusion, Markov chain models due to their simplicity and elegance are a suitable alternative approach. As we could not cover several aspects, detailed reports are available upon request.

7. REFERENCES

[1] J. Borges, M. Levene, Evaluating Variable-Length Markov Chain Models for Analysis of Web Navigation Sessions IEEE Transactions on Knowledge and Data Engineering, Vol. 19, No. 4, April 2007.

[2] Bamshad, Mobasher, Data Mining for Web personalization, P. Brusilovsky, A. Kobsa, and W.Nejdl (Eds.): The Adaptive Web, LNCS 4321, pp 90-135, 2007.

[3] M. Deshpande and G. Karypis, Selective Markov Models for Predicting Web Page Accesses, ACM Transactions Internet Technology, Vol. 4, pp. 163-184, May 2004.

[4] M. Eirinaki, M. Vazirgiannis, and D. Kapogiannis, Web Path Recommendations Based on page Ranking and Markov Models, Proc. Seventh Annual ACM int'l workshop Web Information and Data Management (WIDM), pp. 2-9, 2005.

[5] Tim Berners-Lee, James Hendler, Ora Lassila Scientific Merican, 2001

[6] N. R. Mabroukeh, C.I. Ezeife, Using Domain Ontology for Semantic Web Usage Mining and Next Page Prediction CIKM, 2009

[7] Srivastava, R. Cooley, M. Deshpande and P. Tan Discovery and Application of Usage Patterns from Web Data ACM SIGKDD Explorations, 1(2):12-23, 2000

[8] Phillip E. Pfeifer, Robert L. Carraway Modeling Customer Relationships as Markov chains Journal of Interactive Marketing, Vol 14 / Number 2 / Spring 2000

[9] Choudur K. Lakshminarayan, Alen E. Benson, Qingfeng Yu Website Analysis combining Quantitative and Qualitative Data 2013

[10] Pierre Baldi, Paolo Frasconi, Padhraic Smith, Modeling the Internet and the Web, Probabilistic Methods and Algorithms, Wiley, 2003.

[11] Sheldon, M. Ross, Introduction to Probability Models, 11^{th} Edition, Academic Press, 2014.

[12] Wall Street Journal, "These Product recommendations do not compute", `http://www.wsj.com/articles/those-product-recommendations-just-dont-compute-1434641400`, as of October 2015.