Improving Marketing Interactions by Mining Sequences

Ritwik Sinha[1], Sanket Mehta[1](✉), Tapan Bohra[2], and Adit Krishnan[3]

[1] Big Data Intelligence Lab, Adobe Research, Bangalore, India
{risinha,samehta}@adobe.com
[2] Georgia Institute of Technology, Atlanta, GA, USA
tbohra3@gatech.edu
[3] IIT Madras, Chennai, India
adit@cse.iitm.ac.in

Abstract. The advent of digital marketing has revolutionized how a marketer reaches the organization's customers. Since each interaction with the customer is recorded today, the marketer can do a better job of measuring the effectiveness of marketing efforts. With multi-channel marketing data, comes a new set of challenges; those of measuring the effect of individual channels, understanding synergistic effects, and finally leveraging the information stored in the sequence of marketing activities. While there is some work addressing the first two challenges, we aim to shed light on the last question. The combinatorial explosion in the number of possible marketing sequences requires a systematic approach to address this problem. We propose an approach based on sequence mining to identify marketing touch sequences that are most likely to lead to a stated marketing goal. Our approach provides a rapid way of creating marketing campaigns with the highest chance of success. We test our proposed approach on a real world dataset of a retail chain (with web visits, digital marketing channels, email data, instore and online purchase data). We compare against baseline approaches, and observe interesting insights in the real data.

Keywords: Sequence mining · Association rules · Digital marketing · Marketing attribution

1 Introduction

Marketing is the process of communicating the value of products or services to customers. Marketing campaigns carried out through a range of channels play a paramount role in this process. Hence, the task of understanding the value of individual campaigns and marketing channels is important to the marketer. The area of Marketing Attribution is devoted to understanding the true value of different marketing mediums in a multivariate fashion. With the advent of

All authors contributed equally to this work.

J. Wang et al. (Eds.): WISE 2015, Part I, LNCS 9418, pp. 277–292, 2015.
DOI: 10.1007/978-3-319-26190-4_19

digital marketing, much of today's marketing data is recorded in a variety of databases (web logs, transaction databases, email marketing databases, social media sites, search marketing, and display networks). The area of marketing attribution has three aspects. First, algorithmically computing attribution to campaigns or marketing channels. Second, using this information for optimal marketing spend allocation. Finally, all algorithmic attribution needs the entire interaction history between an organization and customers; this data can be mined to design improved targeting strategies. This last aspect is what our work is related to.

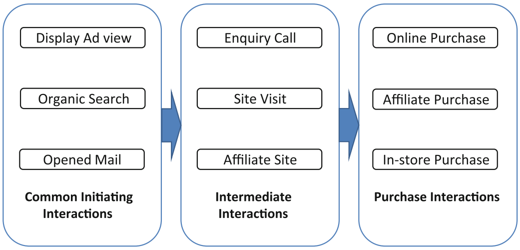


Fig. 1. A customer can follow many paths to a purchase. This schematic shows possible interactions a customer can have at different stages of the purchase process.

Marketing automation refers to a solution that is designed for marketing organization to more effectively manage multi-channel marketing. Many organizations that provide marketing automation solutions claim to reach the customers with the *right message* at the *right time* through the *right channel* [1]. But, most such solutions are not algorithmic in nature, rather, they depend on a marketer's understanding of the business. Hence, this is an area where significant advances can be made by designing data driven algorithmic solutions. We aim to address the question of identifying which channel to use at which point of a customer's journey.

An important source of information about a customer's preferences is stored in the sequence of interactions the customer has with the organization. Figure 1 is a schematic of how a customer may interact with different channels at different stages of the purchase and complete the purchase on different mediums. To test if the sequence of touches is indeed important, we conducted a hypothesis test on a large database of sequences from a major office and home supplies retailer. The hypothesis test, comparing future purchase propensity for certain observed sequences against complements of these sequences, was found to be highly statistically significant (details in Sect. 5). This indicates that the sequence of marketing activities plays an important role in determining marketing effectiveness. This motivates the need for our approach to find high impact marketing sequences which leads us to the primary contributions of our work. In addition to finding evidence for the importance of marketing sequences in real data, we propose an extension of sequential mining techniques to identify sequences that are most likely to lead to a pre-specified marketing goal. We show significant improvements over baseline targeting schemes. Finally, we build a tool that lets the end-user define a goal and retrieve the marketing sequences that have the highest likelihood of achieve the goal.

This paper is organized as follows. In Sect. 2, we describe how our work fits into existing literature. After that, we describe the methodology of our approach and algorithms in Sect. 3. Next,we describe the data on which we applied our work in Sect. 4. In Sect. 5 we describe some of the interesting results we obtained. Finally, we finish with our conclusions and future work in Sect. 6.

2 Related Work

Measuring the effect of different marketing activities has long been an important problem. This problem is as old as the famous quote by John Wanamaker, "Half the money I spend on advertising is wasted; the trouble is I don't know which half". Much of the effort to address this question has been based on Market Mix Models (MMM). MMMs work on time series data aggregated at the marketing channel level. The goal is to find relationships between the time series of revenues, with those of spends for different marketing channels. For example, temporal analysis of marketing channel data is performed in [11,13]. However, with a significant shift of marketing to digital mediums, it is today possible to evaluate the value of advertising at an individual customer level. This aids to draw better causal interpretations of the role played by each advertisement.

Using individual level data for credit assignment to marketing channels falls under the area of Multi-Channel Marketing Attribution. In [2], the authors propose a Hidden Markov Model based approach to model a customer's journey across different states. While theoretically appealing, their model requires a large number of parameters to be estimated from the data. An attribution model aimed at estimating the incremental effect of individual channels has been proposed in [16]. A game theoretic formulation of the marketing attribution model aimed at estimating the causal effects is studied in [7]. While some of these approaches address the question of interaction between multiple channels, none of them look at the effect that the sequence of marketing activities has, on the potential for purchase.

The area of Association Rule Learning has been developed to find interesting relationships between variables in large databases. The literature of this field may be classified into two broad categories. In the first category, association rules are used as an exploratory tool [4], while in the other group they are used as a predictive tool [14]. The problem of mining a large collection of transaction baskets was introduced in [3]. While the authors address the question of finding patterns within transactions, the sequence the items appear in is not considered. In [10], the problem of finding sequential patterns within transactions is addressed where sequence in which items are purchased is considered. As [3,10] are concerned with intra-transactions, they assume that there is no repetition of items in the basket. The question of mining sequential patterns in large databases of transactions has been addressed in [4]. In this work, the authors look beyond situations where the entities may be considered a set (like a basket), but consider a temporal ordering of the items, thus leading to a sequence. In [4,17] a horizontal representation of the database is used for mining sequential patterns while in [8,9] a vertical representation of the database is used. A vertical representation leads to faster algorithms to computing frequencies compared to horizontal representations.

In this work, we have applied algorithms for fast computation of frequencies of different marketing touch sequences. In particular, we have explored SPADE [17], CM-SPADE, CM-ClaSP [8], and VMSP [9]. In our experience, the VMSP algorithm leads to fast computations, and has the beneficial property that it extracts maximal sequences thus avoiding short sequences which are obvious and do not reveal much information, for instance search followed by web visit. These are eliminated by maximality of our frequent sequences. In the next section, we provide the details of our work.

3 Method

Here we provide the foundation of our approach. First, we describe the statistical hypotheses tests that are conducted to show that the sequence of marketing activities is important, when it comes to informing what happens in the future. Next, we describe the exploratory part of our work, that is, how we compute the frequency of sequences, and their corresponding confidence for certain objectives.

3.1 Importance of Sequences

The whole premise of our work hinges on the notion that the sequence in which a set of marketing activities takes place is important to the future relationship an organization has with a customer. We test this hypothesis in a staged manner. In the first step, we conduct a hypothesis test for the importance of the interaction of marketing channels.

Interaction Effects Between Marketing Touches. Let x_{ij} denote the indicator variable of whether customer i $(i = 1, \cdots, N)$ has had an interaction of type j $(j = 1, \cdots, J)$ with the organization. Also, let y_i denote whether this customer made a purchase. Then, consider the logistic regression

$$\log\left(\frac{P(y_i = 1|\mathbf{x}_i)}{P(y_i = 0|\mathbf{x}_i)}\right) = \alpha + \sum_{j=1}^{J} \beta_j x_{ij} + \sum_{j=1}^{J}\sum_{k=1}^{J} \beta_{jk} x_{ij} \times x_{ik}.$$

Where $\mathbf{x}_i = (x_{i1}, \cdots, x_{iJ})$. Then the Wald test [5] statistic for the parameter β_{jk} may be used to test the hypothesis that the j^{th} and k^{th} channels have an interactive effect. A positive coefficient estimate of the interaction term denotes that the presence of both channels leads to a larger increase in the log-odds ratio of purchase, than just the sum of the effects of the two channels. Such a finding will show that channels have synergistic effects. With J marketing channels, we will be testing $\binom{J}{2}$ different hypothesis. Since we will be conducting multiple hypothesis tests, we use the Bonferroni correction [6] to control for it with a family wise error rate of 0.05.

Importance of Marketing Touch Sequence. If the above hypothesis is established, one may further be interested in understanding if, it is not just the presence of both channels j and k, but that their appearance in a particular order is also important. To test this hypothesis, we performed a formal evaluation.

Before we describe our approach, let's introduce some notation. Let Γ be the set of possible interactions $\{C_1, C_2, \cdots, C_J\}$ (for example, opened email, viewed display ad, or clicked paid search). Define a marketing interaction sequence of a customer as the chronologically ordered list of interactions. We denote a sequence S_i of length i_k by $< I_1, I_2, \cdots, I_{i_k} >$, where $I_j \in \Gamma$ for all $j = 1, \cdots, i_k$. Further, the sequence S_l denoted by $< I^l_1, I^l_2, \cdots, I^l_{l_k} >$ is *contained in* sequence S_m denoted by $< I^m_1, I^m_2, \cdots, I^m_{m_k} >$ if $I^l_1 = I^m_{m_1}, I^l_2 = I^m_{m_2}, \cdots, I^1_{l_k} = I^m_{m_{l_k}}$ for $m_1 < m_2 < \cdots < m_{l_k} \leq m_k$. For example, the sequence $S_a = < S, A >$ is contained in the sequence $S_b = < S, W, A, G >$, however it is not contained in the sequence $S_c = < S, T, U, D >$.

For a sequence $S_i = < I_1, I_2, \cdots, I_{i_k} >$, define its complement $\Omega(S^c_i)$ as the set of all sequences $S_m = < I^m_1, I^m_2, \cdots, I^m_{m_k} >$ such that all singleton sequences $< I^m_j >$ are contained in S_i, for all $j = 1, \cdots, m_k$, but, S_i is not contained in S_m. In other words, the set of sequences that contain all the individual marketing channels, but not in the exact same order. For example, if the sequence whose effect is being tested is $S_d = < W, A, S >$, then $\Omega(S^c_d)$ has the five sequences $< W, S, A >$, $< A, S, W >$, $< A, S, W >$, $< S, W, A >$, and $< S, A, W >$.

To test the effect of a sequence, we performed the following exercise. For a sequence S_i, we can compute how often the sequence ends with a purchase, denoted by $p_S = P(S_i \rightarrow 1)$ (S_i leads to a purchase). Similarly, define $p_{S^c} = P(\Omega(S^c_i) \rightarrow 1)$ (for sequences in the complement of S_i, how often do they lead to a purchase). The hypothesis $H_0 : p_S = p_{S^c}$ may be tested using the asymptotic normality test [12]. To ensure we do not bias our hypothesis tests, we selected a number of sequences ensuring sufficient presence in the data, but without looking at its association with purchase. As in the earlier section, we will use the Bonferroni correction to control for multiple tests with a family wise error rate of 0.05. Further, since we are working with large sample sizes, traditional statistical significance may be associated with small effects, hence we also consider the effect size given by $(p_S - p_{S^c})$ when deciding whether there is an effect of business importance.

3.2 Confidence and Support

Before we describe our approach, let's define a few terms. Let D denote $\{S_1, S_2, \cdots, S_n\}$, as a database of user activity sequences. A sequence S_l *supports* a sequence S_m if sequence S_m *is contained in* sequence S_l. The *support for a* *sequence* is defined as the fraction of total sequences which support this sequence. For example, in Table 1, the sequence $< P, A, W >$ has a support of 3/6.

We define a *frequent sequence* as a sequence that has support greater than or equal to a user-specified minimum support (*minsup*). A *frequent maximal sequence* is one which is not strictly contained in any another frequent sequence. In Table 1, if $minsup = 0.2$, the sequence $< S, A >$ fails to meet the minimum support. The sequence $< P, A >$ has a support of 4/6. However, $< P, A >$ is not a frequent maximal sequence because it is contained in $< P, A, W >$ which is also a frequent sequence.

Table 1. Some example user activity sequences, each letter denotes a different marketing interaction.

Customer ID	User activity sequence
1	O → W → A → D → 1
2	D → P → A → M → W → 1
3	D → P → A → 1
4	P → M → A → O → D → W → 0
5	S → W → A → G → 1
6	O → P → M → A → M → W → 0

(1 – Purchase, 0 – Non Purchase)

Note that some sequences may end with a purchase (call them purchase sequences). In other words, $S_i = < I_1, I_2, \cdots, I_{i_k} >$, such that, I_{i_k} is a purchase event (a sequence cannot have a purchase in its interior, if they do, such sequences are broken into separate sequences, each with at most one purchase event). A purchase may have additional attributes like revenue, product type, mode of purchase, and so on. Let us denote a mapping of purchase events to classes of *objectives*, $F(I_{i_k}) \in \{O_1, \cdots, O_K\}$. An objective is a marketing goal that the marketer may envision for customer sequences, for example, one objective could be purchase of product type "furniture", with a "high" revenue that happen "online". Also, for a purchase sequence S_i, define the function $antecedent(S_i) = < I_1, I_2, \cdots, I_{i_k - 1} >$, the sequence S_i without the purchase event.

Let the database of sequences $D(O)$ denote all observed sequences that satisfy objective O, for example, all sequences ending in high revenue online furniture purchases. Algorithm 1 outlines the steps to compute the support and confidence of sequences. In the first step, the VMSP sub-routine [9] is called on the sequence database $D(O)$, this finds all frequent maximal sequences with a specified minimum support. All such sequences end with the objective O, next take the antecedent of the sequence, and compute the support for this sequence in the entire data (D). The ratio of the two frequencies gives us the confidence of the sequence leading to the objective of interest. This approach lets us find the best sequences likely to lead to a certain objective. It also gives a measure of the chance of this objective being satisfied by the sequence.

Algorithm 1. Mining High Confidence sequences

```
Require: Two database of sequences D and D(O)
Supp ← VMSP(D(O),minsup,maxlen)
Count ← empty dictionary
for seq in Supp do
    antseq ← antecedent(seq)
    for x in D do
        if antseq is contained in x then
            Count[seq]++
        end if
    end for
end for
Conf ← empty dictionary
for seq in Supp do
    Conf[seq] = Supp[seq] / Count[seq]
end for
```

Table 2. Details of the marketing interactions along with frequencies (Frequency is the total count of sequences where that particular channel is reflected).

Channel	Symbol	Source	Frequency	Percentage
Direct Web Visit	T	Web	9,088,525	36.28
Display Ad View	d	Display Ad	4,167,055	16.64
Display Ad Clicked	D	Web	218,604	0.87
Opened Email	o	Mail	13,955,334	55.71
Clicked Email	O	Web	1,699,826	6.79
Social	S	Web	10,537	0.04
Paid Search	P	Web	682,837	2.73
Organic Search	G	Web	1,933,748	7.72
Owned Ads	A	Web	1,408,392	5.62
Money Saving Sites	M	Web	295,668	1.18
Instore	I	Instore	16,770,764	66.95
Online	E	Web	4,422,140	17.65

4 Data Description

While our approach can be reasonably easily applied to any marketing interaction data, in this section, we provide details of the data-set on which we applied our approach.

We had access to the marketing (email and display), web interaction, as well as online and in-store purchase data for a large office and home supplies retailer. The data ranged over a 100 day period during the summer of 2013. The data contained about 54 Million interactions with 18 Million customers. For marketing activities to be valuable, it is important for organizations to stitch customer data across different sources. The data from different sources like email marketing, display marketing, web analytics and transaction databases were stitched to have a single key with the use of email addresses, loyalty cards, and web cookies. While there has been some work on probabilistically stitched customer data [15], we assumed that the data was correctly stitched in a rule based fashion.

Table 2 provides details of the channels from which we had data, and the relative frequencies of them. We had the following interactions: Direct Web Visit (typing of url or bookmark visit), Display Ad View, Display Ad Click, Opened Email, Clicked Email, Visit from a Social Site, Click on a Paid Search, Click on an Organic Search, Click on an Owned Ad (ads shown on affiliate sites), and Click on a link on a Money Saving Site (bargain hunting sites). Lastly, we had two kinds of purchase events possible, either an In-Store purchase or an Online purchase. There were more in-store purchases than online purchases in this dataset. Among the interactions, email opens, direct web visits, and display ad views were the more common interaction channels.

Table 3. Distribution of the sequences across multiple-channels. The diagonal value denotes the percentage of sequences containing that channel. The off-diagonal value denote the percentage of sequences containing the two channels corresponding to the row and column. Please refer to Table 2 for what the row and column names stand for.

	Channels											
	T	d	D	o	O	S	P	G	A	M	I	E
T	36.284	11.415	0.632	22.208	3.095	0.031	1.305	3.698	3.980	0.725	15.313	14.147
d		16.636	0.854	9.536	1.948	0.018	0.953	2.557	1.714	0.476	7.427	6.838
D			0.873	0.487	0.141	0.003	0.060	0.170	0.276	0.076	0.312	0.434
o				55.714	5.669	0.028	1.220	3.728	3.402	0.675	37.207	5.442
O					6.786	0.009	0.237	0.951	0.734	0.191	4.156	0.747
S						0.042	0.003	0.009	0.009	0.004	0.023	0.008
P							2.726	0.523	0.352	0.075	0.998	1.29
G								7.720	0.983	0.243	2.868	3.445
A									5.623	0.258	2.239	2.315
M										1.180	0.503	0.420
I											66.954	0.000
E												17.654

Since our problem looks at sequences, we explore the interaction between channels and the importance of sequences, we looked at how often multiple channels appear in the same sequence. Table 3 describes these interactions. We see that there is significant cross-channel interactions. Additionally, we find in our data, 75 % of the customers had at interactions with at least 2 channels, 30 % of them interacted with at least 3 channels, and 24 % of them interact with 4 or more channels.

We define a sequence as a chronologically ordered collection of interactions for an individual. Thus, the entire sequence for the individual can be $< A, B, C, P_1, D, E, P_2, F, G, H >$ (alphabets A–H denote different interaction channels, while P_1 and P_2 denote purchase events), such an sequence is broken into the following 3 sequences $< A, B, C, P_1 >$, $< D, E, P_2 >$, and $< F, G, H >$. The first two are purchase sequences, but the last one is a non-purchase sequence.

Table 4. Sequences distribution across Product Category and Mode of Purchase.

Product category	Purchase mode	Total sequences	Unique sequences
Furniture	Online	115,987	34,998
	Instore	265,662	46,758
Machines	Online	901,906	214,633
	Instore	3,769,764	467,098
Electronics	Online	272,500	83,443
	Instore	1,597,079	231,740
Home	Online	2,275,956	485,701
	Instore	5,248,325	638,993
Services A	Online	0	0
	Instore	1,304,517	106,761
Office	Online	736,024	184,215
	Instore	3,837,277	480,918
Services B	Online	5,542	2,593
	Instore	228,657	34,964
Store	Online	114,225	49,697
	Instore	519,483	81,429
Non purchase		3,855,469	277,762
Total		25,048,373	808,254

Breaking an individual's entire sequence by purchase events was necessary because we wanted to profile sequences by various attributes of the purchase. We had additional information for purchase interactions, in particular, the product purchased, the revenue of the purchase and the mode of purchase. Table 4 shows the distribution of purchases from the different product categories. In all, there were about 25 Million sequences considered in our analysis. Each of which was one of $808, 254$ unique sequences. This table also provides the distribution for different product categories and mode of purchase. Apart from "Services A" and "Services B" being only (or mostly) sold in-store, all other product categories have a large number of sequences.

Figure 2 shows the sequence for three customers. The x-axis of the plot is the date of interaction, on the y-axis, we have the sequence number. In the first panel, we see a customer who may be dubbed an "impulsive customer". This

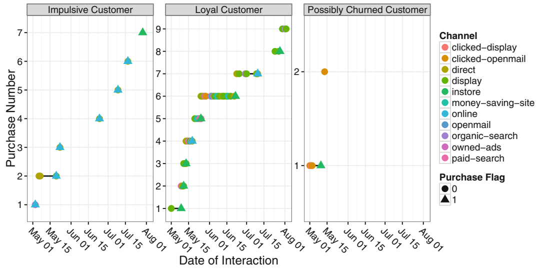


Fig. 2. The marketing interaction sequences for 3 customers. The x-axis has the date of interaction, on the y-axis are the resulting sequences from the customer, after breaking them by purchases. The three panels denote three different kinds of customers. The color and symbol denote the type of interaction and whether the interaction included a purchase or not (Color figure online).

customer had 7 purchases (denoted by seven rows of data in the panel), in each case making the purchases soon after the interaction. The second customer may be called a "loyal customer" because of the large number of interactions with the organization, as well as large number of purchases. The last customer is possibly a churned customer, after a single purchase. This customer's last sequence is a non-purchase sequence. Apart from such customers there were also customers who had no purchases over the entire 100 day period.

5 Results

In this section, we describe our results. We first provide the findings of the hypotheses tests that prove the importance of marketing sequences, next we describe some exploratory data analysis. After that, we describe some results contrasting high confidence sequences for different marketing objectives.

Table 5. Test of the importance of interactions between channels. Only selected interaction coefficients are displayed.

Variable interaction	Estimate	Std. error	z value	Pr(>|z|)
Direct Web Visit × Display Ad	2.85	0.02	125.37	0.00
Direct Web Visit × Openmail	1.61	0.04	45.79	0.00
Display Ad × Paid Search	0.87	0.04	20.95	0.00
Display Ad × Organic Search	0.83	0.03	31.56	0.00
Openmail × Clicked Openmail	1.34	0.04	35.71	0.00
Openmail × Owned Ad	0.71	0.03	22.71	0.00

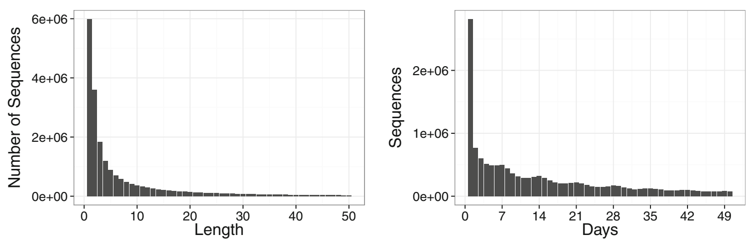
5.1 Importance of Marketing Sequence

As described in Sect. 3, we first test two hypotheses. The first looks at the importance of interaction between channels. Table 5 shows selected coefficients from the logistic regression, along with the standard error, z–values and p–values. Of the 45 interaction effects in the model, 28 interactions are statistically significant at the Bonferroni corrected level of 0.001 (0.05/45), also 20 of these are positive. Thus, in our data, there is a strong evidence that interaction between marketing channels is important to encourage a positive likelihood of purchase.

In the second hypothesis, we test the effect a sequence has on the likelihood of a future purchase. Table 6 displays the results from a selection of this analysis. Of the 219 sequences tested against their complements (please refer to Sect. 3 for the definition), we found 212 of them to have significantly different purchase propensities (at the Bonferroni corrected level of 0.0002). Also, as displayed in the table, the magnitude of this difference was often large. In fact, for 208 of these hypotheses, the effect size was 1 % or more. The effect may have been either positive or negative, but since we did not select the sequences to test based on their propensity of purchase, this is to be expected. Also note that the effect sizes were often large.

Table 6. Test showing the confidence of sequences and their complements. Please refer to Table 2 for the sequence name legend.

Sequence	Sequence confidence (%)	Complement confidence (%)	P-value	Difference (%)
dodT	83.41	79.19	0.00	4.22
TdT	85.89	85.26	0.00	0.63
Todo	80.95	85.14	0.00	−4.19
dGdT	81.55	86.27	0.00	−4.72
odoT	82.94	77.53	0.00	5.41
GdT	82.16	89.78	0.00	−7.62



(a) Distribution of Sequences by Length of Sequence. (b) Distribution of Sequences by Duration of Sequence.

Fig. 3. Number of sequences plotted against the length of the sequence (in number of touches or duration in days).

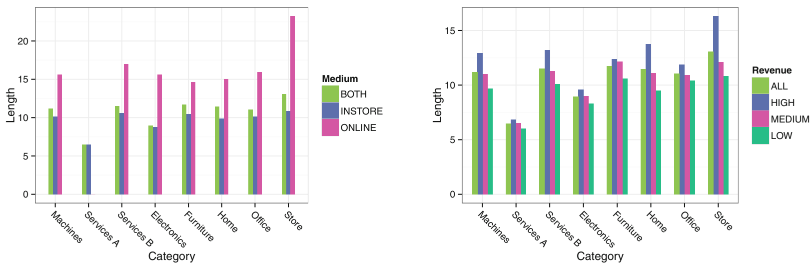
5.2 Exploratory Data Analysis

Our data contains a number of interesting facets on which we performed exploratory data analysis. In particular, in addition to the sequence data, we had the timing of interactions, the types of products purchased (if any), the mode of purchase (online or in-store) as well as the revenue of the purchase.

We first looked at the distribution of the number of sequences by the number of touches in the sequence (Fig. 3(a)). As expected, we see a highly positively skewed distribution, with the largest number of sequences being short (of length 1 or 2), but there is still a significant number of sequences longer than 2. When looking at the duration of these sequences (Fig. 3(b)), we again see a positive skew, but with a large proportion of sequences lasting a week or more. We also see an interesting seasonal pattern, with a periodicity of a week (due to a weekly activity cycle).

Next, we explored how the length of the sequence associates with the mode of purchase (Fig. 4(a)) and the revenue of the purchase (Fig. 4(b)). We see that the online sales tend to have a larger number of touches, across all product categories. We also see that the higher revenue sequences require a higher number of touches, indicating that customers take longer to decide on higher revenue purchases. As before, we again see that "Services A" and "Services B" purchases require fewer touches.

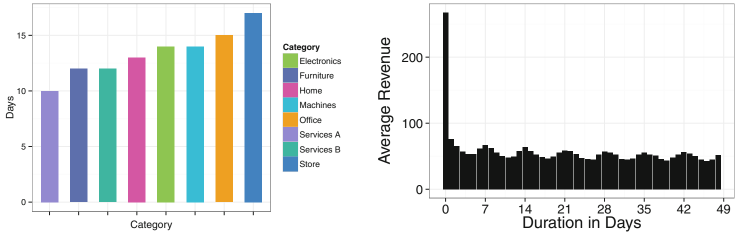
To explore the effect of duration further, we looked at the purchase sequences by different product categories (Fig. 5(a)) and at the average revenue of the purchase against the duration of the purchase (Fig. 5(b)). We see that different products take different number of days for the purchase to be completed. We also see that, on the average, the shortest (in terms of days) sequences lead to the highest revenues. On further exploration, we found a lot of high revenue business purchases (in product category "Services A" and "Services B") happening rapidly. The weekly recurring pattern is more apparent when plotting the average revenue against duration of the sequence.



(a) Average sequence length by Purchase Mode and Categories.

(b) Average sequence length by Revenue Levels and Categories.

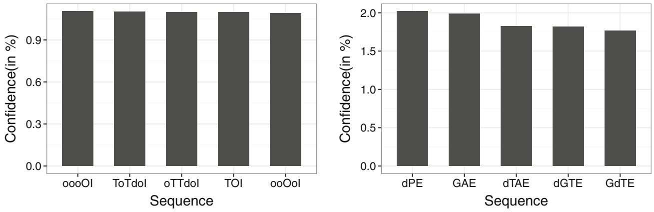
Fig. 4. Plot of Sequence length by Category, Revenue Level and Purchase Mode.



(a) Average conversion time (in days) by product category.

(b) Total Revenue by Duration of sequence.

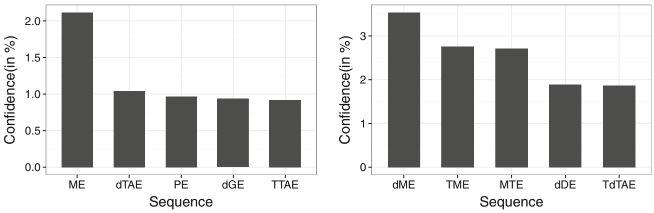
Fig. 5. Duration of the sequence plotted against the product category and Average Revenue of the transaction.



(a) In-store furniture purchases.

(b) Online furniture purchases.

Fig. 6. Comparing Furniture buys across online and in-store modes across all revenue categories. The x-axis has the sequence names and the height of the bar denotes the confidence of the sequence. Please refer to Table 2 for the sequence name legend. We present the top five sequences, in terms of their confidence toward the marketing goal.



(a) Low revenue category electronics purchases.

(b) High revenue category electronics purchases.

Fig. 7. Online Electronic purchases across low and high revenue categories. Please refer to Table 2 for the sequence name legend.

5.3 Confidence and Support

The primary goal of our work is to give the marketing practitioner the ability to find the best sequences for a given objective. Let's say a marketer wants to know the kind of sequences that have the highest confidence for furniture purchases that happen online, we provide a tool for the marketer to identify these sequences readily. Figure 6(b) shows the most promising sequences for such purchases. The marketer can additionally contrast the in-store furniture purchases (Fig. 6(a)) against the online furniture purchases (Fig. 6(b)), and find vastly different sequences with the highest confidence. We see that display followed by paid search leads to the highest confidence for an online furniture purchase, where as the high confidence sequences for in-store furniture sales tend to be more influenced by emails and direct web visits.

When we compare high revenue and low online Electronic purchases (Fig. 7(b) and (a)), we see that the presence of a display ad prior to the money saving sites visit can leads to higher confidence in the high category purchases. This could be an indication for the marketer to target customers with display ads, to try to get them to money saving sites. Many similar such insights can be gleaned by the marketer when he explores marketing sequences using our tool.

Figure 8 provides a view of the demo tool we have designed help the marketer identify top performing sequences for a particular marketing goal. At the top, the marketer specifies the goal of a particular campaign. For example, in this view, the marketer is interested in selling furniture online of a low revenue category. The best marketing sequence for this goal is Display ads followed by Paid Search. At the bottom are the confidence and support for the five sequences with the highest confidences for the specified marketing goal. It is easy for the marketer to use this tool to rapidly create marketing campaigns for any given marketing goal. It is also possible to contrast sequences for different marketing goals.

Fig. 8. A tool created to help marketer to retrieve the best sequence of interactions based on a marketing goal. The marketer specifies the goal, the best sequence to target for this goal is identified. At the bottom, on the left are the confidences of the different sequences, and on the right are their corresponding supports.

5.4 Comparison to Baseline

We compare the results of our proposed approach to two baseline approaches. The first is the random strategy. The random strategy is the one where we

Table 7. Comparison of high confidence sequences, in their confidence to lead to orders of "Machines". The last two columns provide multiplicative lifts compared to two baseline strategies.

Marketing sequence	Confidence (%)	Lift over random	Lift over 100^{th} sequence
odMAE	4.00	3.33	2.01
dodGE	3.20	2.67	1.61
MPE	3.17	2.64	1.59
GdE	2.76	2.30	1.38
GTE	2.69	2.24	1.35
TTAE	2.60	2.17	1.31

consider the confidence of a randomly selected sequence to lead to a certain marketing goal. The second comparison is the following. For a given marketing goal, we look at the ranked list of sequences, ranked by their confidence to achieve the given marketing objective. We then compare the highest ranked sequences with the hundredth ranked sequence. Table 7 provides these details. The last two columns of the table provides the lift achieved by using the highest ranked sequences as a multiplicative factor compared to the two baselines. We see that the top six sequences have between 2 and 3.3 times increase over the random approach. When we compare them to the 100^{th} most successful sequence, we see lifts of 1.3 to 2 times.

6 Conclusions

In this work, we have shown on a large retail data-set that the sequence of marketing touches has a significant effect on the propensity of achieving a stated marketing goal. Next, we propose a novel application of sequence mining to marketing touch data. While association rule learning has been applied to the market basket problem in the realm of marketing, this is, to our knowledge, the only application of sequence mining to marketing interaction data. In doing so, we look beyond the individual effect of marketing channels toward combinations of marketing channels in the form of sequences. Our approach aids better measurement of marketing effectiveness, and aids in better targeting. We also provide a tool to the marketer which can be used to rapidly mine this data to find high confidence sequences for a particular marketing goal.

Our approach has a number of advantages, we were able to mine large amounts of data in a scalable fashion. We handle in excess of 25 Million marketing interaction sequences without difficulty, a size that is likely to be bigger than most organizations' customer base. We also consider all possible interactions between marketing channels and capture the most important of these. We show the marketing lift achieved by our approach over random targeting.

There are a number of possible extensions to our work. While this paper talks about finding an optimal sequence to launch a particular campaign, it does not consider the attributes of an individual customer. We also aim to overlay sequence mining with the stage of a customer in the purchase life cycle. This will aid us in understanding if certain channels play a more important role in different stages of a customer's purchase cycle. We also aim to explore the effect that channels like television have on sequences. Such marketing channels are harder to capture in sequence mining because confirmed touch information is not available.

References

1. Gartner names salesforce.com a leader in the CRM customer engagement center magic quadrant. sforce.co/1z5Lesd. Accessed: 22 July 2014
2. Abhishek, V., Fader, P., Hosanagar, K.: Media exposure through the funnel: a model of multi-stage attribution (2012). SSRN 2158421
3. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. ACM SIGMOD Rec. **22**, 207–216 (1993). ACM
4. Agrawal, R., Srikant, R.: Mining sequential patterns. In: Proceedings of the Eleventh International Conference on Data Engineering, pp. 3–14. IEEE (1995)
5. Agresti, A., Kateri, M.: Categorical Data Analysis. Springer, Heidelberg (2011)
6. Bland, J.M., Altman, D.G.: Multiple significance tests: the bonferroni method. BMJ **310**(6973), 170 (1995)
7. Dalessandro, B., Perlich, C., Stitelman, O., Provost, F.: Causally motivated attribution for online advertising. In: Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy, p. 7. ACM (2012)
8. Fournier-Viger, P., Gomariz, A., Campos, M., Thomas, R.: Fast vertical mining of sequential patterns using co-occurrence information. In: Tseng, V.S., Ho, T.B., Zhou, Z.-H., Chen, A.L.P., Kao, H.-Y. (eds.) PAKDD 2014, Part I. LNCS, vol. 8443, pp. 40–52. Springer, Heidelberg (2014)
9. Fournier-Viger, P., Wu, C.-W., Gomariz, A., Tseng, V.S.: VMSP: efficient vertical mining of maximal sequential patterns. In: Sokolova, M., van Beek, P. (eds.) Canadian AI 2014. LNCS, vol. 8436, pp. 83–94. Springer, Heidelberg (2014)
10. Kamakura, W.A.: Sequential market basket analysis. Mark. Lett. **23**(3), 505–516 (2012)
11. Naik, P.A., Raman, K., Winer, R.S.: Planning marketing-mix strategies in the presence of interaction effects. Mark. Sci. **24**(1), 25–34 (2005)
12. Newcombe, R.G.: Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat. Med. **17**(8), 873–890 (1998)
13. Ramaswamy, V., DeSarbo, W.S., Reibstein, D.J., Robinson, W.T.: An empirical pooling approach for estimating marketing mix elasticities with pims data. Mark. Sci. **12**(1), 103–124 (1993)
14. Rudin, C., Letham, B., Salleb-Aouissi, A., Kogan, E., Madigan, D.: Sequential event prediction with association rules. In: Conference on Learning Theory (2011)
15. Saha Roy, R., Sinha, R., Chhaya, N., Saini, S.: Probabilistic deduplication of anonymous web traffic. In: Proceedings of the 24th International Conference on World Wide Web Companion, pp. 103–104. International World Wide Web Conferences Steering Committee (2015)
16. Sinha, R., Saini, S.K., Natarajan, A.: Estimating the incremental effects of interactions for marketing attribution. In: 2014 International Conference on Behavior, Economic and Social Computing (BESC), pp. 1–6. IEEE (2014)
17. Zaki, M.J.: SPADE: an efficient algorithm for mining frequent sequences. Mach. Learn. **42**(1–2), 31–60 (2001)