Paper 121-28

A Simple Bayesian Approach in Mining the Touch Point Data

Aaron Lai[1], San Francisco, CA

Abstract

In any CRM cycle, we often adopt a "multiple touch point" strategy by using phone, mail, email or other communication channel. It is important to know how will a customer react to different touch points, or different combination of touch points. The analytical work could easily become unrealistic as the combinatorial computation could soon beyond reach. Just a 30 touch points system and consider only 5 stages of interaction; the total cases will be 30 to the power of 5 (24.3 million cases!). This is also a path dependent case, as the customer will behave differently if you email first and then call versus call first and then email. I am proposing a simple Bayesian approach to analyze the touch point data.

Introduction

Customer Relationship Management, or CRM, is an approach in improving the customer experience. The idea behind CRM is to provide a comprehensive strategy in customer solicitation, acquisition, retention and collection. It is particularly important to sales force automation (SFA). Many companies, such as Siebel and Oracle, have rushed to develop sophisticated software to help other achieving this goal. The most basic requirement for the system is to record every interaction between the customer or prospect and the company. Since the customer will have many interactions before a successful or a fail sale (e.g. not returning call), it would be useful to know how does the interaction affect the chance of outcome. Since the order of interactions will play a significant role in affecting the outcome (i.e. mail and then call will be different from call and then mail), it will be difficult to estimate the probability of success due to the fact that there are so many different kinds of interaction combinations. In this paper, I am going to propose a simple Bayesian prediction method that could tackle this problem.

Situation

The most challenging aspect of data mining is seldom the lack of data but the opposite of it. The goal of this exercise is to find out what kind of action will lead to success. In the traditional framework of analysis such as market basket analysis, the goal is “to find groups of items that tend to occur together in transactions, typically supermarket checkout data.”[2] The most common approach to this problem is the implementation of association rule. The basic idea of association rule is to find out how would a product or service be purchased together. A typical example is how does a supermarket find out the beer customers tend to buy chips together. From the point of sales (POS) data, there is a high correlation that both products are bought together. This result leads to a more efficient allocation and arrangement of space for the goods on the shelves. Since the order of buying is not relevant (i.e. buying beer and then chips, or buying chips and then beer contain the same information), it is a commutative relationship problem.

The sales force interaction is a different situation because it is an non-commutative relationship problem. In another words, we do care whether the customers buy beer first and then buy chips, versus buy chips first and then buy beer. Put it into a CRM perspective, the impact is different if you call the customer first and then email, versus email the customer first and then call. We would like to know what kind of action, or what series of actions, will leads to a higher conversion rate.

Literature Review

In statistics, if two variables were related to each other, we would say that they are *correlated* and we would calculate the correlation coefficient (or covariance) between them. A stronger relationship will lead to a higher correlation coefficient. For more than two variables, the relationship could be established using multiple regression and partial correlation. This is a standard topic in any statistics textbook. However, it becomes difficult when other factors are considered. Hence, many econometrics techniques such as Granger causality, Hausman specification test, latent variables, SUR (Seemingly Unrelated Regression), VAR (Vector Autoregression), … etc. have been developed to handle different aspects of inter- or intra- equations issues. However, this is a computationally expensive operation. It will become incomprehensible if the numbers of variable are over 20 and the numbers of observation become large.

Association Rule is a well-researched topic in data mining and machine learning. Almost all of the data mining or knowledge discovery books have contained some kind of discussion on this topic.[3] One of the most widely cited paper by Agrawal et al (1993) discussed how the association rule be modified for a large database. The goal was to discover the relationship among itemsets that had support (predictive power) above a user-defined threshold. Others had proposed a number of variations, including the use of cache memory on a specific data structure called ADtree (Moore and Lee (1997)). The ADtree used a probability table to cache sufficient statistics given a certain amount of memory. Subsequent improvements of ADtree could be found in Anderson and Moore (1998). In this paper, they introduced a new ADtree algorithm for quickly counting the number of records that match a precondition.

Another variation of the association rule was to investigate the indirect relationship. In Tan et al (2000), they investigated the situation that a pair of variables had low support (prediction) but has high dependence on another itemset.

The aforementioned works were targeted at commutative relationships. With the growing interest in Internet, a different set of approaches on path problems have been developed. In Chen et al (1996), they "explore a new data mining capability which involves mining path traversal patterns in a distributed information providing environment like world-wide-web." Firstly, they defined the maximum forward paths, and then developed an algorithm for determining longest reference, as well as hashing and pruning. In Parsa (2000), he talked about the specific topics on "web-log based" path analysis, with particular attention to data collection and incomplete information issues. Our touch point problem could not be formulated into the web analysis framework as we had a non-reversible path. Even though Srikant and Agrawal (1995) had studied the data mining issues in sequential pattern, their studies were transaction focus and they concentrated on how people buy two goods together, in one shopping or multiple shopping.

The marketing researchers were attacking similar problems from different angles. They tried to find out how the people behave in different situations and then tried to find out why. Mittal and Kamakura (2001) investigated how did the satisfactory rating of a customer affect his repurchase behavior. A lot of other literatures were on how to improve the channel effectiveness by market segmentation, instead of adopting a unified approach.

Lai (2001) discussed the analytical issues around analyzing touch point data. Horvitz et al (1998) used a Bayesian network approach to anticipate user interactions.

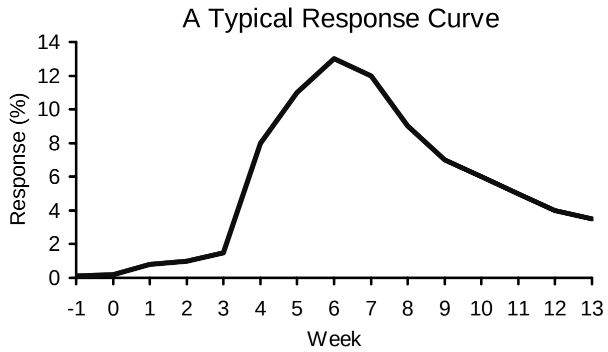
Approach

Codify Interactions

The first step is to define the time and period of study, as well as the meaning of "success" and "failure". The study period cannot be too short that it is not representative; but it also cannot be too long that it becomes incomprehensible. "Success" and "failure" are used to denote the end of the customer interaction process.

In the acquisition phase of marketing promotion, "success" is usually defined as "a successful conversion". However, the word "successful" is also ambiguous enough to require further elaboration. It is not uncommon for a company to provide some special gifts to attract new customers. If a customer closes his or her account after one month, should we still consider it "a successful conversion"?

"Failure" is also an ambiguous concept. How do we know that the prospect is not interested in the product anymore? If a non-responsive prospect in direct mail campaign is a "failure", how long should we wait for he or she to respond? All direct marketers know that the direct mail responses will more or less follow a logistic shape curve. (as shown below) Where should the cut-off point be set? When should we give up on a prospect?



Sequence Interaction and Counting

The next step is to put the data into a usable form. We have to define what activity would we consider in the analysis and codify them into a single character code. For example, if an outbound call is coded as "`A`" and a mail piece is coded as "`B`", then an interaction that involve an outbound call and then a mail will be coded as "`AB`"; while a mail followed by an outbound call will be coded as "`BA`". Please note that "`AB`" is different from "`BA`". Following this procedure, the data for analysis will be similar to the one shown below.

Account ID	Interactions	Other Information
0012251151	ABABBCCDEQGTA**S**	...
5528811441	ABCDDW**F**	...
5587110014	BBEDCEWD**S**	...
8811111487	HHYGREE**S**	...

Since SAS is a very strong tool for data manipulation, we can easily rearrange the data using the **`RETAIN`** statement or **`LAG`** function in data step to append the new codes to a string.

After transforming the account information, we will count the number of cases for each kind of interactions. We would like to know how many accounts have interaction "AABBCCS", for example. This can be done by using **PROC SQL** statement in a program as shown below:

```
PROC SQL;
  CREATE TABLE RESULT
  AS
  SELECT
    ACTION,
    COUNT(*) AS N
  FROM
    INDATA
  GROUP BY
    ACTION
  ;
```

Intelligent Collapsing

This is the most important part of the analysis. Since there are practically infinite number of interaction combinations, the analyst would need to simplify the interaction pattern intelligently. The analyst would need to apply his/her domain knowledge to help facilitating this process. For example, it is very common for a salesperson to play "phone tags" with the customers. Is it reasonable to consider it as "one phone call" if there are more than one call within 24 hours? Two days? One week?

We also need to limit the number of interactions. To facilitate the analysis, we have to set a limit on how long the series of interaction could be. Suppose that over 90% of the sales process were completed within ten interactions, we could restrict our analysis to the first ten interactions (after we have intelligently collapsed some) and assign the final outcome (success/failure) as its 11th interaction.

Estimating Success/Failure Probability

To estimate the probability of success, we could use a backward recursive techniques. If we assume that there are at most three interactions:

$P(S|ABA) = P(ABAS)/P(ABA)$

$P(S|AB) = P(S|ABA) + P(S|ABB) + P(S|ABC) + \ldots$
and it could be calculated by successive substitutions.

Interpretation and Application

After we have found out all necessary cases, we could start interpreting the result. Suppose that

$P(S|ABA) = 0.60$
$P(S|ABB) = 0.25$
$P(S|AB) = 0.40$

If we do nothing after the interactions "AB", the probability of success is 0.4. However, if we do "A", we have improved our probability to 0.6 — a 0.2 increase in success probability!

Cost-Benefit Analysis

If we assume that our profit of a successful sale is $100, our expected profit after "AB" is $40. We could increase our expected profit to $60 if we do "A". If doing "A" will cost us $10, we still have a $10 net gain by doing so. We can then choose our course of action with the highest expected profit.

Business Implication

Recency, Frequency and Monetary (RFM) is the most important ruler to marketer. We could find out how will the prospect behave when different interactions have occurred between the company and the prospect. By adjusting the RFM, company could optimize its multi-channel marketing effort.

By knowing the consumer behavior, the company could determine what kind of information should be disclosed to the customer at different stage of interaction. Kivetz and Simonson (2000) have shown that customer will *perceive* products differently just by withholding information in presentation.

Conclusion

This is a preliminary study on the touch point interaction strategy. Just by knowing what kind of interactions are effective could provide the company an edge in promotion. The proposed analytical approach is easy to use and intuitive, future research could be done to understand why the prospect will behave in different situation.

Reference

Agrawal, R. Imielinski, T. and Swami, A. (1993), *"Mining association rules between sets of items in large databases". Proceeding ACM SIGMOD International Conference on Management of Data*, Washington D.C., U.S.A.

Anderson, B. and Moore, A. (1998), *"ADtree for Fast Counting and for Fast Learning of Association Rules". Proceedings Fourth International Conference on Knowledge Discovery and Data Mining*

Chen, M. Park, J.S. and Yu, P.S. (1996), *"Data Mining for Path Traversal Patterns in a Web Environment", Proceeding of the 16th International Conference on Distributed Computing Systems*

Horvitz, E., Breese J., Heckerman D., Hovel D. and Rommelse K. (1998) *"The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users"* in *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, WI, 1998.

Kivetz, R. and Simonson, I. (2000), "*The Effect of Incomplete Information on Consumer Choice*", *Journal of Marketing Research*, November 2000

Lai, A. (2001), *"Mining the Touch Point Data and the Acquisition Strategy of CRM*', *WUSS (Western Users of SAS Software) Conference*, San Francisco, CA, U.S.A.

Mittal, V. and Kamakura, W.A. (2001), *"Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the Moderating Effect of Customer Characteristics."* in *Journal of Marketing Research*, February 2001

Moore, A.W. and Lee, M.S. (1997), *"Cached Sufficient Statistics for Efficient Machine Learning with Large Datasets"*, CMU Robotics Institute Tech Report TR CMU-RI-TR-97-27

Parsa, I. (2000), *Mining Web Data for Personalization Solutions Course Note*, Epsilon Corporation and SAS Institute

Srikant, R. and Agrawal, R. (1995), *"Mining Sequential Patterns"* in *Proceeding on 11th International Conference on Data Engineering, ICDE*

Tan, P., Kumar, V. and Srivastava, J. (2000), *"Indirect Association: Mining Higher Order Dependencies in Data", PDKK*

Weiss, S.M. and Indurkhya, N. (1998), *Predictive Data Mining, A Practical Guide*

Witten, I.H. and Frank, E. (1999), *Data Mining, Practical Machine Learning Tools and Techniques with Java Implementations*

Endnotes

[1] Aaron Lai CFA FSS MIEEE is a Database Marketing Manager of a major brokerage firm. The opinion and information expressed in this article does not reflect the company opinion nor his employer endorses it. It also does not imply that the company is considering or has adopted the approach presented. This article is an extended version of Lai (2001). All errors are mine and all comments are welcomed. Please send to aaron-lai@attbi.com.

[2] Witten and Frank (1999) p.25

[3] Witten and Frank (1999) and Weiss and Indurkhya (1998)