Aalto University
School of Science

Mat-2.4177 Seminar on Case Studies in Operations Research 2015

Estimation of consumer repurchase behavior

Project Plan

Client: Microsoft Mobile Oy

06/03/2015

Authors:

Niko Lietzén (Project Manager)

Alessandro Mancuso

Antti Melén

Pekka Alli

1. Background

The quickly changing nature of mobile phone market poses many challenges to companies selling mobile phones. The customer behavior varies by many different aspects such as the market, current trends and previous customer experience. Moreover, every customer is an individual and has his own preferences for the desirable phone. Customers also have different propensities to replace the phone that they are currently using. The knowledge of this propensity can be useful for direct marketing purposes. Too much direct marketing messages often annoy customers, but with right timing these messages have a change to activate the customers to replace their phones.

Improvements in information technology have increased the availability of customer transaction data. The marketing division would like to gather as much information as possible about the behavior of the customers. However, privacy laws constraint what data can be gathered. The size of these data sets is enormous, when we consider multinational firms like Microsoft Mobile Oy. The problem is to find as much relevant information from the data as possible, while keeping the model relatively simple. Too complex models can be computationally demanding and they are not necessarily as useful to the marketing specialist.

In this project, we analyze a customer activity data set collected by Microsoft Mobile. Our sample contains marketing information of approximately three million customers. We will perform the increasingly common customer-base analysis to the data set. We know that not all the markets in the world behave similarly. However, we hope to identify the markets that are similar to each other. Hence, we do not have to make an individual model for every separate market.

2. Objectives

Our primary goal of this project work is to create a forecasting model for a customer's next replacement purchase. We approach the problem by assuming that an underlying stochastic process determines the observed behavior of a single customer. We attempt to estimate the specific time, when the customer is most likely going to purchase a new phone. Additionally, we seek to determine the time intervals when it is unlikely that the customer is going to replace his phone. By decreasing the marketing activity in these time intervals, it is less likely that the customer will become annoyed with the constant direct marketing. Hereby, we can find the optimal time to focus the direct marketing.

We have decided to approach the problem using Bayesian techniques. The Markov chain Monte Carlo (MCMC) methods [5] allow us to try different probability distributions and choose the most suitable one. Furthermore, we need to perform some time series analysis to the data, to identify seasonal patterns and possibly implement these seasonal spikes into our model.

We have decided to perform the correspondence analysis (CA) [4] to the data. Correspondence analysis helps us identify the similar markets and their relations to the operating systems and the average purchase intervals. This allows us to limit the amount of different models.

We will focus on the mathematical model and therefore we leave the qualitative analysis of the customer behavior with little attention. The model will be based on an existing model found on a literature and possible changes to the model will be considered. Our goal is to implement the model with R language so that it is usable to Microsoft marketing team in the future.

3. Project Tasks

The following tasks are expected to be carried out during the project:

1. *Project planning (All)*
    - During this part our project leader Niko makes sure that our project objectives and team members' responsibilities are clear to every team member. A written project plan and presentation are prepared.
2. *Literature review and Model structure choice (Niko and others)*
    - Background familiarization and literature review is conducted to find the most suitable model structure for our project. Different models are evaluated to find the best one.
3. *Data elaboration (Pekka)*
    - The data received from Microsoft has to be prepared by eliminating errors and modifying it so that it is suitable for our model. The original data has is not grouped in any way so we need to define customer groups and organize the data in right way.
4. *Interim reporting (Antti and Alessandro)*
    - Project status and changes to project plan are reported in written form and a presentation is done.
5. *Implementation (Niko, Antti)*
    - The chosen model will be programmed with R. Programming code should be done as effective as possible to reduce the run time for large data used in the project.
6. *Verification & Validation (Pekka, Alessandro)*
    - We test the chosen model to ensure that it works with our data. We analyze the chosen model and reserve sufficient time to make changes to the chosen parameters.
7. *Final report (All)*
    - Written report and presentation are done to conclude the project. Final report is to include everything from the model to the final results.

The tasks are fairly allocated between all the team members. Main responsibility is on our project leader Niko. He makes the important decisions such as selecting our model.

4. Resources

The project team is composed of four people: Niko Lietzén, Pekka Alli, Antti Melén and Alessandro Mancuso. Niko is the designed team leader and has an Operations Research background, as well as Pekka and Antti. Furthermore, Niko's current work is related to statistics, thus he has important experience related to the data analysis. Alessandro is a Master student in Industrial Engineering and Management from Politecnico di Milano. We believe that the integration of our competences will bring meaningful value for the project.

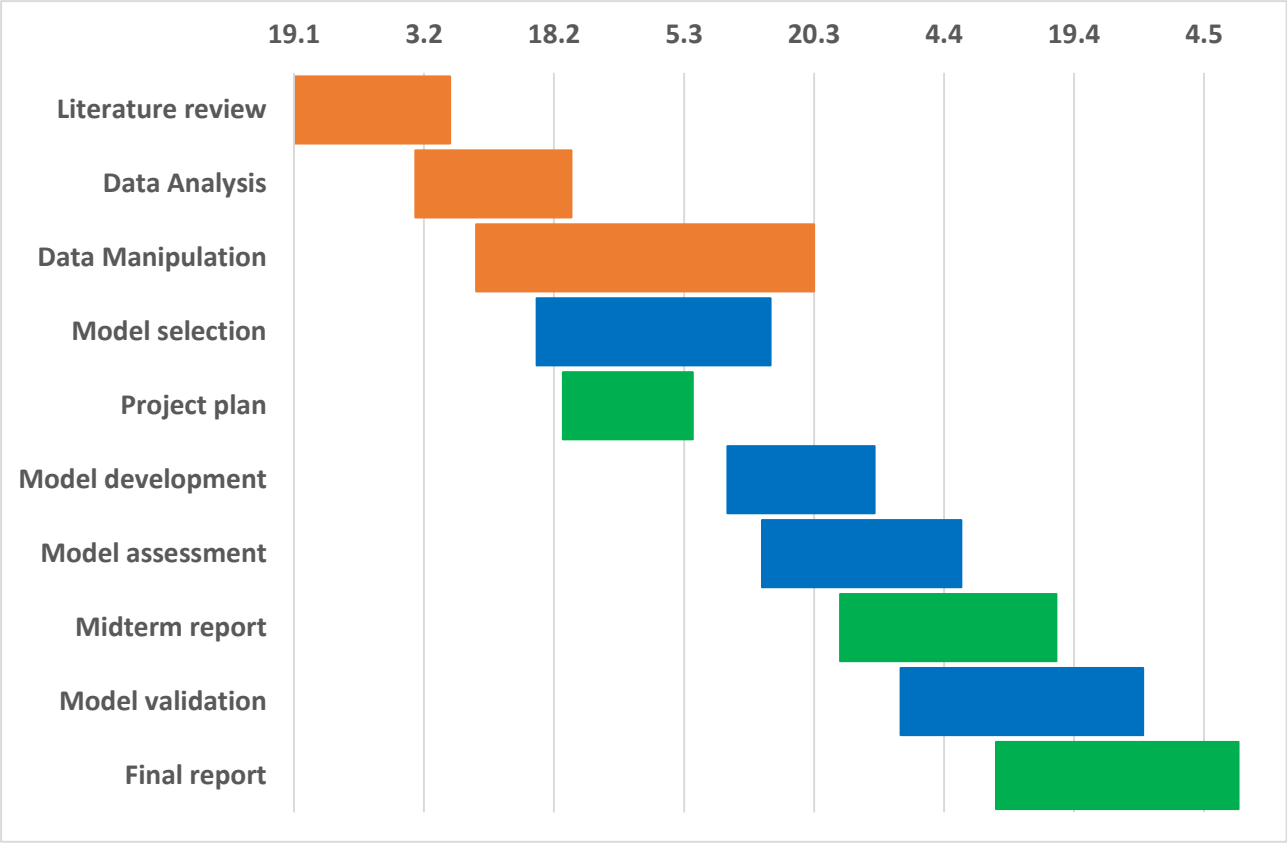
The steering group consists of Lauri Salminen (Microsoft Mobile Oy), professor Ahti Salo (Aalto University) and assistant Joonas Lanne (Aalto University), but we expect significant feedback and ideas also from the opponent team.

We plan to have regular meetings throughout the project development and several meetings with Lauri Salminen from Microsoft in order to ensure that the focus of the project is in line with the client's wishes. Meetings will be decided in agreement with the team or called by the project leader.

Communication among team members and with Lauri Salminen will occur by e-mails, phone calls and sharing documents through Dropbox folders.

5. Task Scheduling

In the following Gantt chart, we present three different macro activities: the first one (orange) refers to acquiring accurate knowledge and information related to Microsoft case study, the second one (blue) regards all the activities concerning the definition of an appropriate model, finally the last one (green) refers to the development of periodic project reports.



In particular, January is approximately dedicated to literature review and data preparation for future model development. Literature comes both from Lauri Salminen suggestions and from our team research. At the same time, data manipulation needs some effort given the initial dataset structure.

Several models will be evaluated before making the final selection. After that we are going to write the first project plan for future activities and risk prevention, while on the other hand we will start developing a dynamic model for purchase timing.

After developing a consistent section of the model, assessment and validation will be the next steps. In the meanwhile we are going to write a midterm report related to the performed analysis and eventually the first results.

Final report will be carried forward throughout the project development, but we are going to really focus after the midterm submission.

6. Risks

Here we present the most significant risk factors, the relative impact and our intentions to prevent them.

Risk factor	Probability	Risk outcome	Impact	Prevention
1. Poor data quality	Likely	Significance of results could be seriously undermined	Intermediate	Sufficient data preparation and manipulation. Cooperation with Microsoft Mobile
2.Wrong model selection	Unlikely	Inconclusive project results	High	Accurate literature review and reserved time for eventual modifications
3. Overly complex model	Unlikely	Model is not implemented	High	Accurate model selection
4. Alessandro's departure	Certain	Some troubles for meetings and decisions	Low	Skype meetings

We identified four risks that might affect our project. First of the risks is poor data quality, which is something beyond our group, but may still affect the result. However, we can try to prevent this risk from being realized with accurate data preparation and manipulating. It is estimated to only have intermediate impact. Three other risks are related to our own work efforts. The second risk is that we choose an inappropriate model. We may notice our mistakes too late and the results would not be desirable. As a matter of fact, model selection is a very delicate stage of our project, given that it will have a strong impact on the rest of our work.

The third risk is the possibility of selecting an overly complex model. This will lead to useless work especially when estimating the parameters. Prevention is done by keeping it simple. Last identified risk is Alessandro’s departure from Finland. This will make communication within the team more difficult as we cannot have normal meetings anymore. This will also lead to Alessandro being unable to attend the seminars. This risk can be prevented with good project planning and fair task allocation among the team members. The impact is considered to be only intermediate as this will only effect the last part of the project.

References

[1] Allenby, G. M., Leone, R. P. and Lichung, J.: "A Dynamic Model of Purchase Timing with Application to Direct Marketing." *Journal of the American Statistical Association* 94, no. 446 (1999): 365–74.

[2] Fader, P. S. and Hardie, B. G. S.: "Probability Models for Customer-Base Analysis." *Journal of Interactive Marketing* 23, no. 1 (2009): 61–69.

[3] Karvanen, J., Rantanen, A. and Luoma, L.: "Survey Data and Bayesian Analysis: A Cost-Efficient Way to Estimate Customer Equity." *Quantitative Marketing and Economics* 12, no. 3 (2014): 305–29.

[4] Mardia, K. V., Kent, J. T. and Bibby, J. M.: "*Multivariate Analysis*." Academic press, London, 1979.

[5] Martin, A. D., Quinn, K. M. and Park., J. H.: "Mcmcpack: Markov Chain Monte Carlo in R." *Journal of Statistical Software* 42, no. 9 (2011): 1–21.

[6] Rossi, P. E. and Allenby, G. M.: "Bayesian Statistics and Marketing*." Marketing Science* 22, no. 3 (2003): 304-328.