Thesis Outline

for Scientific Method (TDDD89)

Pontus Persson, ponpe412@student.liu.se

December 14, 2016

Thesis Title

The working title of my thesis is; "Evaluating association analysis algorithms for finding factors affecting user retention in mobile browsers".

Initial Problem Description

When developing a mobile browser, or any software where the revenue is in direct correlation to the number of users, it is important to continuously improve and maintain a good experience the end users. Especially important is to convert new users to stay in your product instead of using a competitors product.

To improve the user conversion rate the goal of this thesis is to test and evaluate different clustering and association analysis algorithms to find rules or indicators of behaviour that if observed for a new user within its first few sessions greatly improves the probability of that user converting.

Initial Approach

To achieve the goal of this thesis I have been given access to data collected from new users of a mobile browser. The data is collected from sessions during the first week and contains what interactions a user makes with the browser during a session along with time stamps. The methods that are to be tested and evaluated are different algorithms used for association analysis. Whether a particular user has converted or not is stated in the available data, so the goal is to find association rules on the form $\{indicator_1, \cdots, indicator_n\} \rightarrow converts$.

1 Time Plan

Figure 1 shows a preliminary time plan for my thesis work.

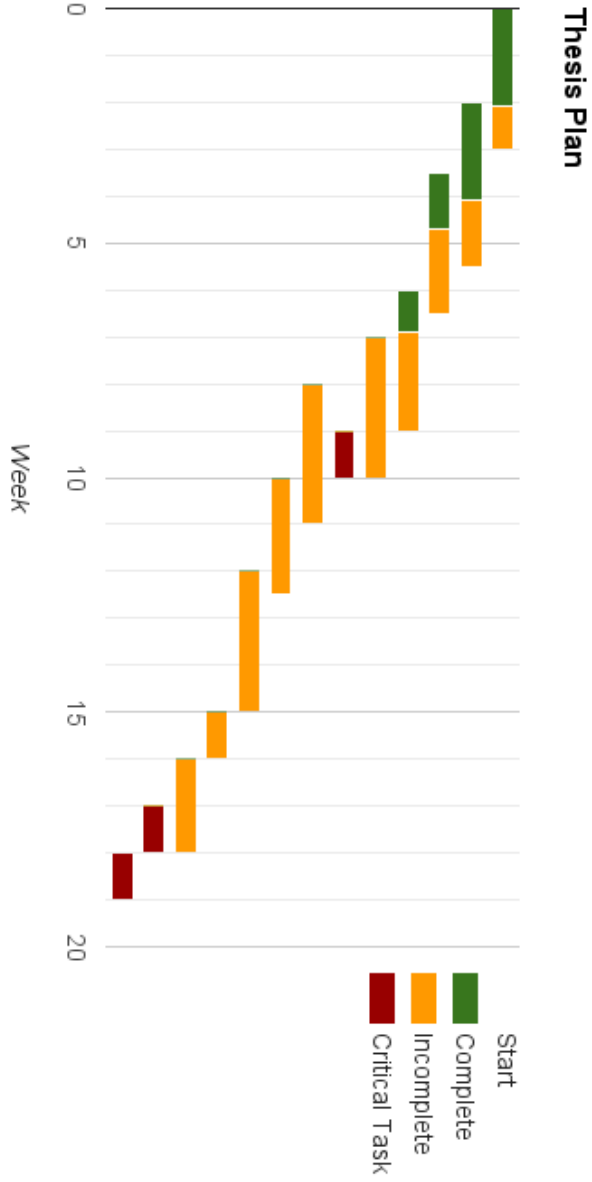

Figure 1: Thesis time plan

The tasks are described below:

Write introduction Set the context for the thesis and motivate why there is a problem that needs solving. Describe the aim and formulate questions. Also write delimitations section

Research Main research part. Search for similar work and other papers with theory that might be of use. This is needed to formulate theory and method chapters.

Write theory Using the research, write a theory paper. Should include relevant theories to the problem and similar work.

Write method Using the theory written. Formulate a method that will be used to answer the questions from the introduction. Make sure the method has a high level of reliability, validity and replicability.

Implementation Implement the method. The method might be modified during this time. Focus on creating a minimum viable product (MVP), to be able to generate results as soon as possible.

Experimentation As soon as possible, start generating results. The initial results might open up new questions and force you to reevaluate your questions and method choices.

Write results Try to write down your findings and present them such that they answer the questions.

Discussion Evaluate the method and results. Discuss the meaning of the results and what the consequences of them are.

Conclusions Summarize to what extent the questions and aims have been answered/met. What are the consequences of this thesis?

Finalize Paper Cut superfluous sections, make sure there is a read thread throughout the whole paper. Write abstract, acknowledgements.

There are three critical tasks in my plan, mid term evaluation, final presentation and the final hand in. The mid term evaluation is scheduled after 9 weeks, just before the half way mark. At this point I should have implemented most of my method and started getting some results. This is a very critical time, since if I need to change my method and or questions, this might be my last chance.

The final presentation and final hand in of the report are scheduled the 17th and 18th week. When presenting the paper should be finished, with only minor tweaks to be fixed after the presentation.

1.1 Decision points and risks

The main decision point is the choice of method. If chosen poorly this can cause me to start implementing a method that will not give me the results I need. Hopefully this will not happen, and if it does I will find out as early as possible. One other risk is if the data is not available to me. A data mining thesis report is not very interesting to read if there is no data.

Possible literature

- R Agrawal and R Srikant. Fast algorithms for mining association rules. *Proc. 20th int. conf. very large data bases, VLDB*, 1994
- Jian Pei and Jiawei Han. Can we push more constraints into frequent pattern mining? In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '00*, pages 350–354, New York, New York, USA, 2000. ACM Press
- Jiawei Han, Jian Pei, Yiwen Yin, Jiawei Han, Jian Pei, and Yiwen Yin. Mining frequent patterns without candidate generation. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data - SIGMOD '00*, volume 29, pages 1–12, New York, New York, USA, 2000. ACM Press
- Gaming App User Retention: Only 22% Return After One Month
- User Onboarding — A frequently-updated compendium of web app first-run experiences. https://www.useronboard.com/

Listed below are a number of keywords that might help me find relevant information in order to write this thesis.

data mining, association analysis, association rules, user retention, customer retention, mobile browser behaviour, user conversion...

I will primarily focus my information gathering on finding relevant papers and studies through IEEE Xplore Digital Library, ACM Digital Library and Google Scholar.

Relevant Courses

The list of past courses that I think will be of most relevance to me when working on this thesis is quite short. The two courses that come to mind are Artificial Intelligence (TDDC17) and Data Mining (TDDD41).

References

[1] Gaming App User Retention: Only 22% Return After One Month.

[2] User Onboarding — A frequently-updated compendium of web app first-run experiences. https://www.useronboard.com/.

[3] R Agrawal and R Srikant. Fast algorithms for mining association rules. *Proc. 20th int. conf. very large data bases, VLDB*, 1994.

[4] Jiawei Han, Jian Pei, Yiwen Yin, Jiawei Han, Jian Pei, and Yiwen Yin. Mining frequent patterns without candidate generation. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data - SIGMOD '00*, volume 29, pages 1–12, New York, New York, USA, 2000. ACM Press.

[5] Jian Pei and Jiawei Han. Can we push more constraints into frequent pattern mining? In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '00*, pages 350–354, New York, New York, USA, 2000. ACM Press.