Norwegian University of Science and Technology

Master's thesis, spring 2014

User adaptation in an anonymous application setting

Author:
Jonas Myrlund

Supervisors:
Prof. Heri Ramampiaro
Prof. Helge Langseth

April 2014

Changelog

The latest changes are listed first.

2014-03-18 Tiny label fix. New version of PDF.

2014-03-18 Added condensed section on experimental setup.

2014-03-18 Added intro on similar domains and research.

2014-03-18 Expanded on how users are tracked in appear.in.

2014-03-14 Added changelog to compile pipeline and report.

2014-03-07 Some structural changes to report. Added ingestion pipeline figure.

2014-03-07 Updated motivation with a bit more on privacy concerns. Fixed chaptering.

2014-03-05 Fixed minor errors in report.

2014-03-05 Removed chapter 4 and moved higher ones to reflect.

2014-03-02 Updated report.

2014-02-17 Added preliminary version of the report source.

"I definitely need to find a timely quotation for my Master's thesis."

Jonas Myrlund, February 2014

NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Abstract

Faculty of Information Technology, Mathematics and Electrical Engineering
Department of Computer and Information Science

Master of Science in Computer Science

User adaptation in an anonymous application setting

by Jonas Myrlund

Identifying the different ways in which users use a product can be useful on several levels. In this project we will look at how identifying user classes allows for simple personalization of the product, and to what extent simple personalized treatments can alter user behavior.

It is often the case that some identified user classes are more desirable than others. Either because its associated users generate more revenue, use the product more, invite their friends, or similar. This project describes a system capable of not only identifying user classes, but more importantly a framework for identifying the most effective ways of driving users toward more desirable user classes.

More specifically, given a set of identified user classes and a set of predefined treatments, we want to find out how each treatment affects each user class. Although the project implementation will specifically target the video conferencing service appear.in, a major research question will be to what extent the results generalize.

We find that . . .

Acknowledgements

I would like to thank my supervisors, Heri Ramampiaro and Helge Langseth, for lots of good advice, while at the same time allowing me a great extent of freedom in planning and executing this project.

My supervisor at Telenor Digital AS – Svein Yngvar Willassen.

Contents

Changelog i

Abstract iii

Acknowledgements iv

Contents v

List of Figures vii

List of Tables viii

Abbreviations ix

1 Introduction 1
- 1.1 Background and motivation . . . 1
  - 1.1.1 Introducing appear.in . . . 2
  - 1.1.2 Usage patterns . . . 2
  - 1.1.3 Anonymity and privacy . . . 3
    - 1.1.3.1 Absence of identity . . . 3
    - 1.1.3.2 Tracking users . . . 4
  - 1.1.4 Why personalize? . . . 4
  - 1.1.5 Visualization of clustering . . . 4
- 1.2 Problem specification . . . 5
- 1.3 Organization of the thesis . . . 5

2 Survey 7
- 2.1 Similar problem domains . . . 7
- 2.2 Similar research . . . 7
  - 2.2.1 Relevant literature . . . 8

3 Approach 9
- 3.1 System overview . . . 9
- 3.2 Data ingestion and preprocessing . . . 9
- 3.3 User modeling and clustering . . . 10
- 3.4 Adaptation component . . . 10
  - 3.4.1 Evaluation metrics . . . 10

3.4.1.1 How to measure a positive user experience 10
3.4.2 Applying the personalized feature set 10
3.4.3 Tracking user treatments 10
3.4.4 Visualizing effects 10
3.4.5 Differentiating product features 10
3.5 Evolving the user models 10
3.5.1 Multi-arm bandits 10
3.5.2 Tracking individual treatment 10
3.6 Visualization requirements (?) 10

4 Evaluation 11
4.1 Description of the evaluated dataset 11
4.2 Prerequisites 11
4.3 Experimental setup 11
4.3.1 Clustering of users 12
4.3.2 A/B testing features 12
4.3.3 Evaluating the results 12
4.4 Results 13
4.5 Discussion 13

5 Conclusion 14
5.1 Summary 14
5.2 Generalizing the system 14
5.3 Suggestions for further work 14

A Evaluation Results 15

Bibliography 16

List of Figures

3.1 The ingestion pipeline broken into 4 steps. The color of each node indicates means of storage. 9

List of Tables

Abbreviations

Chapter 1

Introduction

1.1 Background and motivation

My motivation for doing a user adaptation project related to anonymous online video conferencing is comprised of two important factors. To understand how they are related and why they both are of equally big interest, some background on both the technological landscape of the web and the application case is needed.

For a long time, developing video conferencing services was an extremely challenging discipline, and as a result the market has consisted of a correspondingly low number of actors. However, this trend is currently in the process of being shaken and turned on its head with the introduction of HTML5; more specifically, with the introduction of the WebRTC[1] specification.

As the name implies, WebRTC handles real-time communication, but an important aspect of the technology is that it is designed to do so peer-to-peer. Although it per design is a protocol for exchanging arbitrary data between peers, it is particularly geared toward multimedia streams. For instance, the traditionally cumbersome task of setting up a two-way audiovisual connection is now a matter of dropping around 40 lines of boilerplate Javascript into a web page[2].

[1] Web Real-Time Communication.

[2] For an excellent introduction, see: http://www.html5rocks.com/en/tutorials/webrtc/basics/.

Although the WebRTC specification is still officially a *working draft* in the W3C[3], several large browser vendors have already implemented it, and applications previously unseen on the web pop up every week.

1.1.1 Introducing appear.in

One of these applications is called appear.in, and it will serve as the main use case in this thesis. Like many other WebRTC applications, appear.in concerns itself with video conferencing. The idea is simple enough: a conversation happens between users who are in the same room at the same time. The central idea, though, is that the room is identified solely by the URL in use, and not in any way by the peers connecting. As an example, if any two people are visiting `https://appear.in/ntnu` at the same time, they will see and hear each other and can start chatting away without any further call setup or configuration.

This view of a conversation as not really being an enitity in its own right, but rather an *effect* of people being in the same room at the same time, breaks with the traditional model of audiovisual communication. Traditionally, talking to someone not present has been a process of one person *calling* the other one, with group conversations usually being nothing more than an extension of this concept. appear.in doesn't concern itself with distinguishing between callers and callees, has no simple concept of a "conversation", and generally does not enforce any particular way of using the service – apart from requiring the conversation venue to be identifiable by a string of characters.

1.1.2 Usage patterns

Until the arrival of WebRTC, this particular way of thinking about conversations for anything but textual conversations hasn't been a big thing. However, the simplicity of the room concept opens the service up for a wide variety of uses[4]. These wildly varied use cases are where the motivation for this project stems from:

1. If the users' behaviors are quantified, will any clear and distinct usage patterns emerge?

[3] The latest specification can be found here: `http://dev.w3.org/2011/webrtc/editor/webrtc.html`.

[4] In addition to traditional video calls, we've already seen it used for everything from virtual offices and team meeting rooms to baby monitoring and remote tutoring, just to name a few.

2. If so, can the different uses be better served by dynamically adapting the product to fit each of them?

1.1.3 Anonymity and privacy

appear.in is an anonymous communication service. No personal information is ever collected about the users, and not even IP-addresses or geolocational data is logged on an individual level. By tracking individual *browsers* using cookies, then logging behavioral events along with a cookie identifier, we can measure user behavior over time.

This all opens a wide series of questions bordering to sociological aspects of web usage:

1. To what extent can an anonymous web service be personalized? Is user behavior enough to provide a satisfactory personalized user experience?
2. Will a personalized user experience go against the users' expectations of appear.in as an anonymous web service?
3. How can we measure any of this?

1.1.3.1 Absence of identity

Several others have written about how privacy constraints impact personalized systems [1, 2]. In many ways, the very nature of anonymity is an extension of privacy. However, the absence of *identity* presents an entirely different challenge.

There are ways of coping with the absence of user identification. Kobsa, for instance, describes an approach that makes heavy use of pseudonyms designed around this problem [3]. However, appear.in is not only a pseudonymous service – it is a fully anonymous service; there are no user accounts, and there is no login. Tracking users, then, is obviously a bit of a struggle.

1.1.3.2 Tracking users

appear.in runs in the browser. The first time a user enters the site, a cookie is set with a random value uniquely identifying the browser over time. This value is sent with every event to the instrumentation service used for user analytics.

Unfortunately, using only cookies for identification has its clear downsides. Should the user clear the browser cache, switch browsers, use the browser "incognito"[5], or simply use multiple machines, then different user ids will be generated for each case.

Without tracking more user information – IPs and locations, for instance – there is no easy way of tying these user ids together[6], and to reason about them as a single user. However, collecting this information about users goes against appear.in's privacy policy, and it has been deemed more interesting to see what can be done without crossing this line.

1.1.4 Why personalize?

@TODO: Write out in full-text.

- Premise: Users use service in different ways, and want a simple UI targeted at their specific use case.
- Goal: Provide an interface which stimulates activation and retention.
- Result: Increase general user base, as well as their happiness (as measured by metrics such as recurring visits and time on site).

1.1.5 Visualization of clustering

@TODO: What is the state of the art regarding visualization of spacio-temporal clusters. (Dimensionality reduction etc.)

Package into clustering tool to improve business intelligence. Challenges from lack of demographics.

[5] Google lingo for private browsing, where previously set cookies are not available (among other things).

[6] Some suggestions on how to solve this is found in the suggestions for further work, in section 5.3.

1.2 Problem specification

Main research question: Can users of anonymous video conferencing services be clearly divided into user classes based on their behavior, and if so, to what effect can personalization improve their activity level?

1. Are users of video conferencing services such as appear.in clearly dividable into separate groups based only on their behavior within the service? Do these patterns reflect those seen elsewhere – in other types of internet services or even in real life?
2. Is it feasible to personalize treatments to these user classes? Does it stimulate users into becoming more active users?
   (a) Is this something these users want?
   (b) Do the inferred preferences of the detected user classes significantly differ from each other?
   (c) Can the personalized treatments be devised in such a way as to stimulate the moving of users in the direction of any desired user class?
3. How can a toolkit be devised to handle the following?
   (a) User classification based on behavior.
   (b) Product personalization based on a relevant user's class.
   (c) Tracking of each treatment's effect on each user class.
   (d) Prioritize using the most effective treatments without introducing statistical bias (see multi-armed bandit).
   (e) Allow product developers to easily access results to improve future feature prioritization.

1.3 Organization of the thesis

This paper is organized as follows.

Chapter 2 surveys relevant literature, similar applications, provides an in-depth analysis of the available data. Chapter 3 describes the system design and the reasoning behind central design choices. In chapter 4 we'll look at the execution results, and see how they evaluate. Chapter 5 summarizes the most important takeaways, and suggests further work.

Chapter 2

Survey

2.1 Similar problem domains

Although the case in question is a specific service, the techniques in use and the limitations needing to be dealt with should apply to many kinds of applications.

When demographics are absent and identity is highly unreliable, this will unevitably lead to some highly sparse usage data and quite a bit of noise in the user modeling data. Assuming that the usage patterns of the users actually differ, will these patterns be clearly and reliably identifiable? Will it be possible to base an adaptive personalization system on this kind of data? When using an anonymous system, will users welcome a personalized product, or will they regard this as breaking with their perceived concept of anonymity?

These are all questions that apply especially to appear.in, but that may also apply to many other anonymous, web-based systems.

2.2 Similar research

The approach taken to adaptive personalization is based heavily on the work by Vrieze in "Fundaments of Adaptive Personalization" [4].

2.2.1 Relevant literature

Teltzrow and Kobsa's work on privacy-driven personalization systems provides insight into a lot of the issues surrounding the lack of demographic information in personalization [1, 2]. However, the main focus of their work seems to be that users are more willing to provide demographic information if that information is not backtracable to themselves, through pseudonymous personalization. This matches the application in question quite poorly, as the goal is not to obtain demographics – rather to attempt to cope without it.

@TODO Explore and include literature on:

- Clustering and segmenting users, choice of algorithms etc.
- Visualizing and evaluating clusters.
- A/B testing, multivariate testing, multi-arm bandits.
- Motivations for adaptive personalization. Online business models?
- Anonymity: do users experience personalization as an overstep?

Chapter 3

Approach

3.1 System overview

3.2 Data ingestion and preprocessing

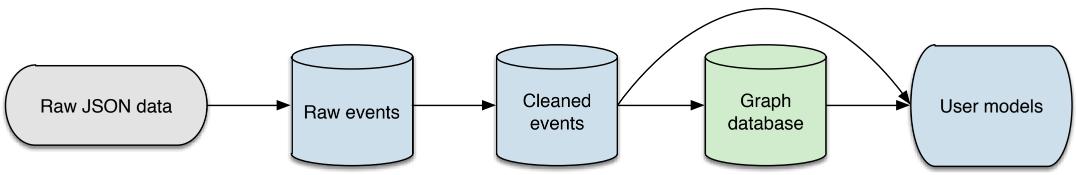

FIGURE 3.1: The ingestion pipeline broken into 4 steps. The color of each node indicates means of storage.

Blue indicates a RDBMS, *green* indicates a graph database, whereas *gray* is used to indicate flat file storage.

3.3 User modeling and clustering

3.4 Adaptation component

3.4.1 Evaluation metrics

3.4.1.1 How to measure a positive user experience

3.4.2 Applying the personalized feature set

Description of the FlagService model.

3.4.3 Tracking user treatments

3.4.4 Visualizing effects

3.4.5 Differentiating product features

3.5 Evolving the user models

3.5.1 Multi-arm bandits

3.5.2 Tracking individual treatment

3.6 Visualization requirements (?)

Chapter 4

Evaluation

4.1 Description of the evaluated dataset

4.2 Prerequisites

For any of this user adaptation to have any actual use, we will need to find out whether there is actually significant variation in feature adoption between clusters; it matters little whether users exhibit differing behavior with regard to existing functionality, if this behavior yields no basis for predicting feature adoption in the future. We need an inductive bias.

An easy way of discovering whether this inductive bias is present in the user base is to simply log whether there is significant variance across the user groups in their users' adoption of new features. Thus, this question of whether the predictive bias actually exists will be included as a central part of the actual experiment itself, and be thoroughly discussed through the next sections.

4.3 Experimental setup

Condensed version.

The main experiment determines whether an alteration of the service affects different kinds of users differently. More specifically, do users in clusters $C_1, \ldots, C_n$ employ function f in significantly differing ways?

We break the process into three steps, each elaborated on in 4.3.1, 4.3.2, and 4.3.3:

1. Divide users into clusters $C_1, \ldots, C_n$.
2. A/B test feature f independently of clusters, noting the outcomes for each user.
3. Analyze results: did the test results vary significantly between clusters?

If the final answer is "yes", we can store our result and use it to toggle feature f for users in the relevant clusters.

4.3.1 Clustering of users

@TODO: Describe methods and algorithms used. What kind of results are expected?

4.3.2 A/B testing features

@TODO: Brief description of how to fit the actual testing regime into the production system.

4.3.3 Evaluating the results

Seeing as the application does not yield any direct and real payoff, we will evaluate the performance of each variation relatively, using some general key metrics. This approach is described in [5].

For appear.in, the most important key metrics are "time on site" and "number of visits in the last n days". These are combined in providing a relative success metric for each variation.

@TODO: Verify key metrics.

4.4 Results

4.5 Discussion

Chapter 5

Conclusion

5.1 Summary

5.2 Generalizing the system

5.3 Suggestions for further work

Appendix A

Evaluation Results

Bibliography

[1] Maximilian Teltzrow and Alfred Kobsa. Impacts of user privacy preferences on personalized systems. *...personalized user experiences in eCommerce*, (0308277), 2004. URL http://link.springer.com/chapter/10.1007/1-4020-2148-8_17.

[2] Alfred Kobsa. Privacy-Enhanced Web Personalization. pages 628–670, 2007.

[3] Alfred Kobsa and Jörg Schreck. Privacy through pseudonymity in user-adaptive systems. *ACM Transactions on Internet Technology*, 3(2):149–183, May 2003. ISSN 15335399. doi: 10.1145/767193.767196. URL http://portal.acm.org/citation.cfm?doid=767193.767196.

[4] Paul De Vrieze. *Fundaments of Adaptive Personalisation*. ISBN 9789090211138.

[5] Yisong Yue, Josef Broder, Robert Kleinberg, and Thorsten Joachims. The K-armed dueling bandits problem. *Journal of Computer and System Sciences*, 78(5):1538–1556, September 2012. ISSN 00220000. doi: 10.1016/j.jcss.2011.12.028. URL http://linkinghub.elsevier.com/retrieve/pii/S0022000012000281.