Aalborg University

Multimedia Recommendations

P3-Project

Group ds305e12:

Frederik Daniel Bank Gundersen

Frederik Meyer Bønneland

Michael Jensen

Naburan Amirthalingam

Peter Gurnæs Havbro Nielsen

Rasmus Bak Christiansen

AALBORG UNIVERSITET

Det Teknisk-Naturvidenskabelige
Computer Science and Software
Selma Lagerlöfs Vej 300
Telefon 96 35 97 31
Fax 98 13 63 93
http://tnb.aau.dk

Title:
Elektroniske Lommepenge

Theme:
Programmering og problemløsning

Projectperiod:
P2, 2013

Projectgroup:
ds305e12

Participants:
Frederik Daniel Bank Gundersen
Frederik Meyer Bønneland
Michael Jensen
Naburan Amirthalingam
Peter Gurnæs Havbro Nielsen
Rasmus Bak Christiansen

Advisor:
Emmanouil Valsomatzis

Oplagstal: 10

Sidetal: 60

Bilagsantal og –art: 2 + 1 CD

Afsluttet den: 22. maj 2013

Synopsis:

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Contents

Chapter 1

Introduction

Entertainment media is a part of almost all people's lives in some shape or form. It is something we seek out and consume almost every single day. Here is things like books, films, video games, music, etc included. For that, there exist various kinds of way to spread a piece of media to the public, like websites, application, advertising, and social groups. Here can a system which generates recommendations based on various kinds of data, like personal information, media data, and social connections, also be included. Either directly or indirectly. This kind of recommendation is tailored towards a certain person, and is there to make them aware that there might be other products which he could be interested in. Like a movie that is similar in genre to what a person previously have watched, or an add-on product to a previous purchase on a retail website.

Recommendation systems have various problems before them hindering its effectiveness, like data shortage. It is also very crucial that proper weights can be generated for the various kinds of data, so the most important aspect of a piece of media is highlighted for the individual person. And interesting feature could be if it were able to generate recommendations across different kinds of media, E.g. if you liked these books, maybe you would like this movie, then it would also require a set of suitable connections between them, to create the recommendation. Since these kinds of systems is centered around the user, it is also important to include some kind of survey to study people's habits regarding media. These surveys could also provide weights for how important certain kinds of data is. These challenges is what creates the base problem for this project.

1.1 Initiating Problem

What challenges exist for recommender systems, what is different peoples habits regarding recommendations from various sources, and to what degree is it possible for recommending media across different kinds of media?

Chapter 2

Problem Analysis

2.1 Recommender Systems

2.1.1 Collaborative Recommendations

2.1.2 Content Based Recommendations

2.2 Recommendation Statistics

In a study conducted by Harris Interactive (Footer with link) among 2.311 U.S. adults ages 18 and over, on how they use and feel about interaction on the internet. It was conducted between April 28 and 30 2010. They asked different question and some of them are quite interesting. Although it was conducted in America, the information can still be used in our project. The idea behind our project is to create a website that can be accessed from across the world. So different studies in different countries are useful for the information needed to determine the target group. That is why this information is useful for our project.

The first graph asks the question about what the asked person shares through social media. They were to select all that applies to them. If we take a look on what the different options was, one stands out and that is: TV and movie recommendations. In the range from 18 to 34 year old, 36 percent of them says that they share TV and movie recommendations on social media.

The first interesting question and answer is, that in short terms asks what the person share on social media and select all that applies. If we take a look on what the different options were one strikes out and that is; TV and movie recommendations. In the range from 18 to 34 year olds, 36 percent of them

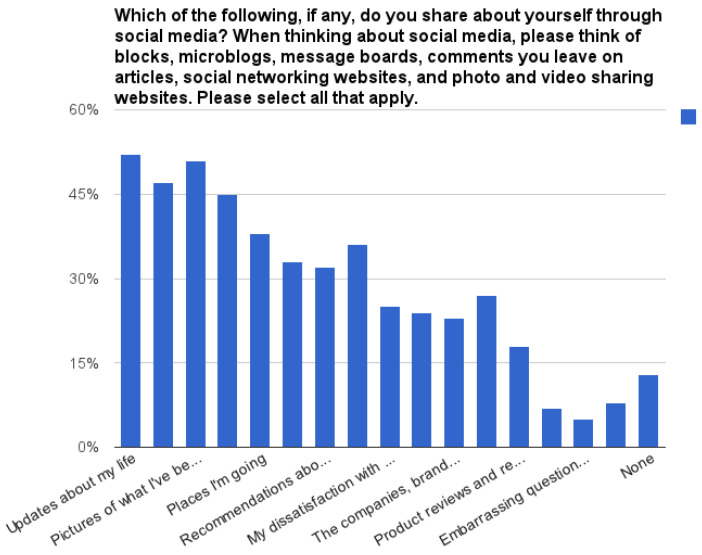

Figure 2.1: 1

says that they share TV and movie recommendations on social media.

The second graph pose the question about how much different reviews from different people influence them. It shows that 71 percent of those asked replied that they are influenced a great deal or a fair amount by reviews from family members or friends.

The second graph asks the question about how much different things influence them. It shows that a staggering 71 percent of those asked, replied that they are influenced a great deal or a fair amount by reviews from family members or friends. That shows that social recommendation is something that a lot of people have in mind when it comes to buy or watch something.

The third and last graph shows the age group of those asked the previous question. That graph shows that the 18 to 34 year olds are more influenced by blog and social media, than the older ages were. At the same time, the older people are more influenced by their family and friends.

The third and last graph shows the age group of those asked the previous question. It shows that 18-34 year old are more influenced by blogs and social media than the older ages. While the older people have are more influenced by their family and friends.

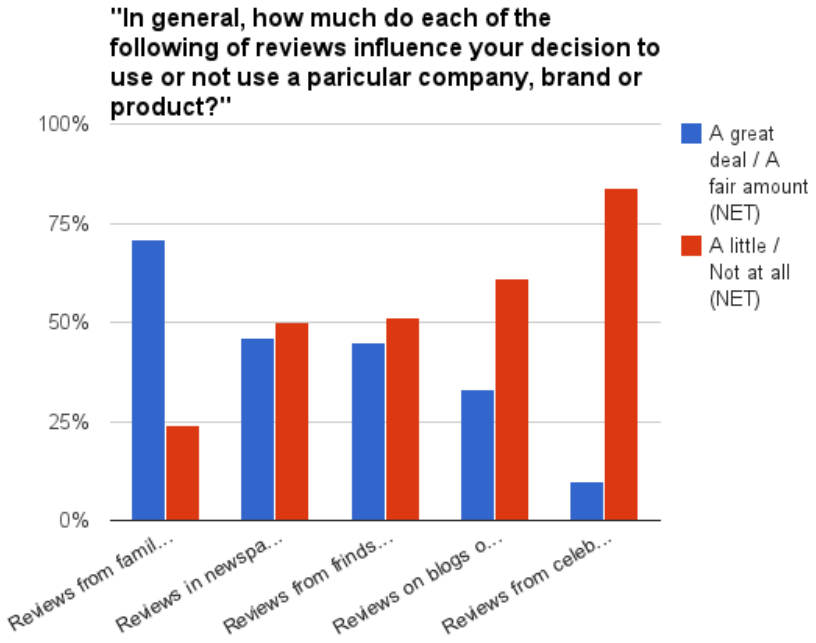

Figure 2.2: 1

2.3 Existing Solutions

According to Alex Iskold, from readwrite.com, there are four approaches for recommendation. The first approach is personalized recommendation which is based on the individual's past behavior. The second approach is social recommendation which is based on the past behavior of similar users. The third approach is item recommendation which is based on the item itself. The last approach is a combination of the all three approaches mentioned above. With these four approaches this section will look at solutions on existing recommendation systems.

The four approaches are:

- Personalized recommendation - recommend things based on the individual's past behavior.
- Social recommendation - recommend things based on the past behavior of similar users.
- Item recommendation - recommend things based on the item itself.
- A combination of the three approaches.

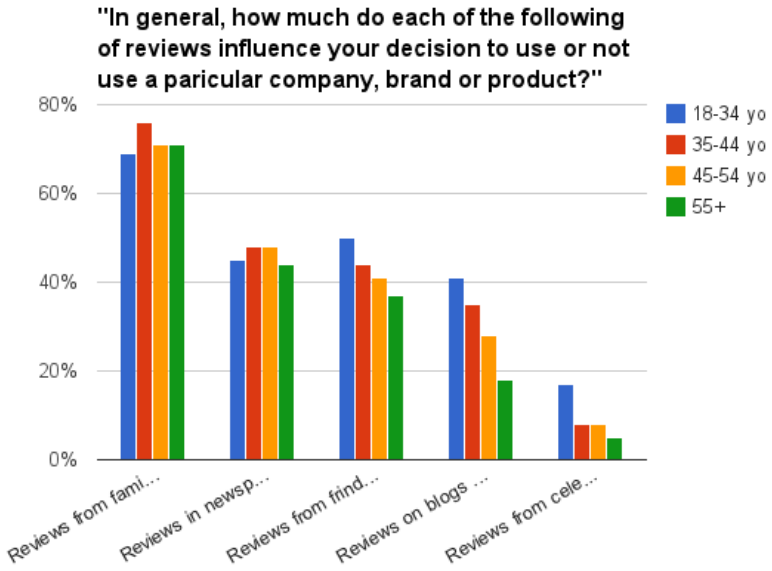

Figure 2.3: 1

2.3.1 Solutions

Amazon

Amazon is the world's largest online retailer. It started as an online bookstore but then it soon got every kind of things like movies, electronics, clothes.

Amazon system uses the combination of the three approaches. Which means the system is based on individual behavior, and either the item itself or behavior of similiar users. Amazon have a "what other customer also bought" under the item you are viewing, which is a social and personalized recommendation. A customer reviews is also possible to be view for the item, where the specific customer is available to rate and comment on the item. And based on the reviews it shows the top 3 similiar statements. Amazon gives you other recommendation based on what your history, which is personalized recommendation

YouTube

YouTube is a popular video sharing website which is under control by Google. YouTube system uses the personalized recommendation. YouTube takes your history, and other activity on the site, to give you recommendations, on which it gives recommendations for channels and videos. So basically whatever video you click on will be placed in your personal history list and based on that it will give you some recommendation.

The problem with YouTube is that even if you only watched one second

of a certain video, it will be included in the recommendations to come. Unwanted videos can easily be recommended because of this. Channels also gets recommended even after only watching a single video from the channel.

IMDB

Internet Movie Database(IMDB) is database with old to new movies. It gives information about new movie releases and you can be able to watch trailers from the movies before you decide to watch it. IMDB is a kind of personalized recommendation and social recommendation system. The recommendation system goes by the image X

It also has some based on what you've previously seen and rated. The system is to a certain degree, flawed. Despite being sort of a social recommendation, because it is based on what other people have previously seen, there is no friend list or similar, so there is no control regarding what people these social recommendations is based on. The personalized recommendations also has some flaws. As the project groups members tried seeing what it would generate, as a member and not a member of the site, it gave the same recommendations, and showed movies that have already been seen and rated.

LinkedIn

LinkedIn is a social networking website for professional occupations. It is way for people to find, and be found, for projects and/or work based on ones skills, previous experience, and descriptions from other users. It uses text analysis to find certain keywords, like "trustworthy" or "dedicated", to highlight a person's abilities. LinkedIn also has a "apply from LinkedIn" button, where users can apply for a position in a company, through their LinkedIn profile.

The system is based on item recommendation, as it is the descriptions, skills, and keywords extracted about the user, that is used for the recommendation. When the user itself search for a person or a company that will be a personalized recommendation. The system could also be some of hybrid recommendation system There can be some problems, if someone begins to write wrong descriptions about another user, both if it intentionally better or worse than it in reality is.

2.3.2 Overview

The most common recommendation type between the four existing solutions is the personalized kind of recommendations. This is followed by the social recommendation type. Item recommendation is visible on some of the solutions, but is less used compared to the other recommendation types.

Personalized and social recommendations seems to fit into collaborative

filtering, while item recommendations falls into content-based filtering recommender systems. To cover all possibilities, a hybrid recommendations seems to be the best option, considering Amazon's success with this kind of recommendation system.

It seems like the personalized and social recommendations is the most important, considering what this project is aiming for. Item recommendation is based solely on data regarding a specific item, rather than a recommendation based on the person's interests, past behavior, and similarities with other people. This will be further investigated in the survey section to find out what people weigh the most, when they receive new media they might consume, or find themselves.

2.4 Connections Between Media

This section is going to be about what possible similarities in attributes there can be between different kinds of entertainment media, as a way to recommend content across them. This is going to look at the content-based recommendation possibilities, which is purely based on data and attributes about the content, like genre and involved people, rather than user and friend ratings.

First we have books, movies, and videogames. These three have a clear connection as they can, and often is, adapted into each other. There exist many movies which is based on a book, like the 'Lord of the Rings' trilogy, and various videogames based on books, like 'The Witcher' and 'Metro 2033' series. For videogames, the other way around is usually a result of rising popularity, and uses novels and books for world-building and secondary plot-lines. Genre is also a clear connection, as all three has the same genres, like crime and science fiction, prevalent through them. Associated people can also provide another connection, like a script writer or director, who previously have worked on both movies and videogames. Television shows can also be fitted into these three, especially movies, who shares many of the same similarities.

For something like music, there isn't any clear connections to any of the previously mentioned entertainment media. Music does appear in the visual entertainment media, like movies and videogames, but mainstream music usually only appear in some movies, and more often than not, movies and videogames has their own scores. Genre connections doesn't apply either, as music has its own set of genres, like rock and pop. For associated people there can be some connections, like if a piece of music shares a composer with a score from a movie. When it comes to books, which doesn't have any music attached

to its form of entertainment media, there is even less suitable connections to be made. It could be argued that there does exist connections between music and books, because for every topic in existence, there will be books about it. Music could appear in books as a story element, or in educational music books. There could also exist connections with biographies depicting the lives of musicians. This is still problematic as these connections are quite niche. This project's topic is also about entertainment media, so something like educational media does not apply here.

A majority of people do listen to some kind of music though, so it could possible, or more suitable, to link users with similar taste in music. Through that music can be incorporated, together with the connections that do exists. This also applies to any other which has already been mentioned, and so, their recommendations can further be reinforced and improved together with their own connections.

Another thing that could provide usable connections can be indirect connections. For example, if you have a certain book, and it has a movie adaption. This movie uses a certain piece of music, which turns out to have variable that match with other media items which the same user likes. This can provide a whole new dimension of connections and recommendations, but also make the recommendation process much more complex.

There exist numerous ways to make connections between the books, movies, and video games, which can be used as parameters for generating content-based recommendations. For music though, there is a clear lack of connections suitable for generating proper recommendations, as it shares minor similarities with movies, videogames, and books. The connections that does exist are quite niche, and may serve more of an educational purpose, than entertainment. It is quite clear though, that music is something a majority of people consumes. So if it can't get as many connections to other kinds of media, it could work in a more social kind of way, where it takes personal information into a higher consideration, and links people who have a similar taste in music, and though that might even recommend a movie or a book.

2.5 Surveys

To find out how people generally perceive media recommendations, it was decided that there had to be collected some data from potential users. For that reason, both an interview and a questionnaire were constructed and executed. This was also done for the purpose of approximating a more precise target

audience, and create possible weights based on real data, once the project's product is going to be designed. This section will look at and discuss the data collected from the interview and questionnaire, and what was learned from doing these activities.

2.5.1 Interviews

2.5.2 Questionnaires

2.6 Target Audience

To find our target audience we first want to examine who might be interested in our potential solution. To do this we have taken several different approaches, first we created personas and use cases to try and get an idea of what groups of people we want to target our program towards. To try and confirm our assumptions, further understand who our target audience is and also define them more accurately, we did interviews.

2.6.1 Personas and Use Cases

Through doing our personas and use cases we got a deeper understanding of who is going to be our target audience. We started by creating personas for three of what seemed like the most obvious users, while constructing situation where our project could might be useful. Along with helping form our interviews it also helped form our program, through showing what kinds of features different audiences might want.

It also helped showing us what age groups we might want to ignore. Because children and people with certain disabilities, particularly bad eyesight, impose certain restrictions on a potential solution, restrictions that is not immediately necessary for our program to function, we have decided to exclude these groups from our initial demands for the program. Note that we consider kids to be everyone younger than 17 years, because that is the oldest age restriction you can find on media(in EU at least). Based on that we have decided to focus our solution towards people in their late teens and people without disabilities.

2.6.2 Interviews

We can see from the interviews that younger people seem to have more interest in our proposed project. One reason this might be the case is because younger people already use the internet to find new media to consume, and most also use other recommendation systems. While our sample size isn't big enough to make a definite conclusion it helped us define questions for our questionnaire, such at the what age groups we should split people into.

The interviews also hint that peoples' current employment status doesn't have a significant effect on their interest in our project. There was one person that outright said she had no interest in our project, and had only one relevant difference from the other people we interviewed which was her age. So that doesn't signify a trend based on employment. And the other 10 interviewees were a mix of unemployed, employed and had various educational levels.

To note about privacy which was a minor focus of the interviews we see that age doesn't really impact concerns on privacy. While younger people were a bit more specific on the technicals and on the different forms of data, they all agreed that they wanted to choose what data the company keeps on them. They also agreed that the data, which the company does not need to run their service, should never be stored.

The interviews not only helped us confirm that there is an interest in the project but also to specify our target audience. We now know that our focus should be towards the age group 18-50, based on the responses we got.

2.7 User Privacy and Rights

For a recommendation system it is required to collect data about the intended recipients of the recommendations. The recommendations become better and more precise, as the amount of data becomes bulkier, so a recommendation system will usually try to gather as much it possibly can. This can also be called data profiling, where the intended system for this project would categorize people depending on their taste in media, personal information like age and sex, and their common friend connections. In other instances it could might also include variables like previous purchases, view history, tags, and keywords extracted through text analysis. All of these might be more revealing, and raises privacy concerns.

Even with this amount of data gathering this project plans to handle, there will be some data privacy concerns. Users will most likely be required to

register a profile or account to utilize the system, and will have to hand over personal information for verification and recommendation purposes. This is of course not any different from many other systems who does the same thing. But in this case, we're talking about a recommendation system, which poses some new challenges in this aspect. Because recommendations can be based on the data of other people, it would make it possible to deduce connections to other people. Especially if it is through some uncommon element, like an obscure cambodian film, where fewer people is connected to. This can ultimately lead to personal information being found out. This risk is further heightened if the person also has access to the database, and can place queries.

This is a glaring challenge for any recommendation system, but is also a complex and difficult problem to solve, or even answer to. It also depends on exactly how this projects product is going to be constructed, which is not yet formalized, to properly answer these questions. At this point it might be out of this project's scope.

We conducted some interviews where we asked potential users several questions regarding the project, including questions regarding how much information they would be okay with being available, either to a company or the entire public. The questions was asked in a rising degree, to see exactly where people's threshold would be. Most people were okay with their contact information being available, as it was most likely already available in some form of way. When it came to their more personal information, like interests and age, they were more uncertain. Almost all was okay with having their very personal information, like emails, chats, and pictures, available, as expected. An interesting feature a lot of them mentioned were an option to choose whether or these information was shared, for the user themselves. This could be a more user-engaging way to collect information for the recommendation system, and solves most problems regarding user rights, as the user decides for themselves what they want to share.

2.8 Project Boundary

2.9 Problem Formulation

2.10 Product Requirements