Recommender Systems for Blog Websites

Spring 2016

Yuki Nagae
Student ID: 11945627

Introduction

This document proposes a data mining project in order to increase the number of users and users' satisfactions for the blog website, named JamBlog (name changed for confidentiality reasons) by using recommender systems. The blog website provides the platform where any users to post their blogs on the website and the monthly page views (PV) are around 1 million PVs with 50 thousands users. The project will develop a data mining model based on existing users' access histories of each blog page.

It first looks at the aims and objectives of this project, following the background of the website and recommender systems. In the section of research project, the significance of the problem and the innovation of our approach will be described. In the next section, our proposed approaches and the process will be explained. Finally, the schedule and budget of the project will be explained in order to make the project feasible, following personnel will be shown briefly.

Aims and Objectives

The key aim of this project is:

- To increase the number of users and users' satisfactions by using recommender systems for the JamBlog website using analysis of users' access histories

The key objectives of this project are:

1. To extract patterns of users access patterns from users' access histories
2. To develop a data mining model which recommends blog posts to users
3. To deploy an implementation of the recommender systems

Background

Problem

JamBlog website is the most well-known blog website in Sydney, having that the monthly page views (PV) are around 1 million PVs with 50 thousands users. However, in recent years, JamBlog website has been facing the decrease in the number of both the existing users and new users, also seeing the gradual decrease in the user's' satisfactions. There might be several reasons of the problem but the most crucial issue is that the rise of the new blog website called BlogGo (name changed for confidentiality reasons). In order to solve the problem, our analytics team offers a data mining approach which is about implementing recommender systems for JamBlog website, differentiating JamBlog website from other blog services.

Recommender systems

Recommender Systems are software tools and techniques making recommendations of items to a user, resulting in better decision making for users such as what items to buy, what music to listen to (Ricci, Rokach & Shapira 2011).

There are various techniques for recommender systems, therefore the algorithms of each recommender system are different each other. The below is the list of popular recommendation techniques by Ricci, Rokach & Shapira (2011).

- Collaborative filtering: recommending to the active user the items that other users with similar tastes liked in the past
- Content-based: recommending items that are similar to the items that the user liked in the past
- Demographic: recommending items based on the demographic profile of the user
- Knowledge-based: recommending items based on specific domain knowledge about how certain item features meet users needs and preferences
- Community-based: recommending items based on the preferences of the user's' friends.
- Hybrid recommender systems: recommending items based on the combination of the other recommendation techniques

In the project, collaborative filtering approach will be used to implement the recommender systems, considering following advantages and disadvantages of the approach.

Pros

The prominent advantages of collaborative filtering are below by Hu, Koren & Volinsky (2008):

- Domain free: only past user behaviours are required
- Accuracy: more accurate than other approaches such as content-based filtering

Cons

While there are powerful advantages of collaborative filtering, there are also several disadvantages underneath in collaborative filtering (Thorat, Goudar & Barve 2015):

- Cold-Start problem: require a huge amount of existing data on which user can make exact recommendations
- Scalability: a huge amount of computation power is often essential to compute Recommendations.
- Sparsity: only a small subset of the entire database is rated by most active users

Fortunately, there are already enough data about past user behaviours such as the view histories of each user's accessing each blog post, therefore collaborative filtering approach will be the best option in the project.

Research Project

Significance

Since JamBlog website has been facing the decrease in the number of both the existing users and new users, also seeing the gradual decrease in the user's' satisfactions, any action to recapture users and make the difference from other blog services should be taken, otherwise JamBlog website may lose its position in the near future.

Innovation

Importing recommender systems into the blog website would be fairly new and unique, comparing from other blog services. As Ricci, Rokach & Shapira (2011) point out, there are several reasons that companies started using recommender systems in order to meet their business needs. The below are the reasons and benefits why companies make use of recommender systems.

1. **Increase the number of items sold:** To sell an additional set of items compared to those usually sold without any kind of recommendation

2. **Sell more diverse items:** To enable the user to select items that might be hard to find without a recommendation.
3. **Increase the user satisfaction:** To improve the experience of the user with the site or the application. The user will find the recommendations interesting, relevant.
4. **Increase user fidelity:** To foster the loyalty of a user, leveraging the information acquired from the user in previous interactions. Consequently, the longer the user interacts with the site, the more the recommender output can be effectively customized to match the user's preferences

Proposed Approaches

Tasks

The below table is the basic task list of the project, following the CRISP-DM methodology which is one of the standard processes in the field of data mining.

Phase	Task	Description	Deliverables
Business Understanding (optional)	Collect requirements	This task may be unnecessary since the business problem has already been identified.	Reports
	Assess situation		
	Determine data mining goals		
	Produce project plan		
Data Understanding	Collect initial data	This task is involved in understanding the existing data which has already been obtained in the website application.	Reports
	Describe data		
	Explore data		
	Verify data quality		
Data Preparation	Select data	This task aims to format and clean the existing data. If the data is already clean enough, the task may not be necessary.	Datasets
	Clean data		
	Construct data		
	Integrate data		
	Format data		
Modeling	Select modeling technique	This task is to develop a model	Datasets, Reports

	Generate test design	based on collaborative filtering approach, with several parameter tunings.	
	Build model		
	Assess model		
Evaluation	Evaluate results	This task is to evaluate the results of the modeling phase, and review the results.	Reports
	Review process		
	Determine next step		
Deployment	Plan deployment	This task is to deploy actual implementation of the recommender systems.	Reports, Deployment plan, Implementation of the recommender systems
	Plan monitoring and maintenance		
	Produce final report		
	Review project		

Table 1. Brief project task list

Schedule

The project is to be completed and implemented by the 1st of December 2017. Following the approval of this proposal on the 15th of November 2016, a more detailed project plan will be drafted for approval and the submission of this plan. The commencement of the project following the approval is 15th of November to 1st of December 2016.

From the provided time estimates of an optimistic 8 months, a pessimistic 16 months and a most likely estimate of 12 months, an accurate assessment of the project time frame from the Program Evaluation and Review Technique (PERT) is 12 months.

PERT Time estimate = [Optimistic + (4 * Most likely) + Pessimistic] / 6
= [6 + (4 * 12) + 16] / 6
= 12 months

From the PERT evaluation, the project will likely meet the deadline of December 2017.

Outcomes

The outcomes of this project will be:

- The patterns of user behaviours
- The data mining model
- The implementation of recommender systems

Benefits

The benefits of this project will be:

- The increase of users derived from uniqueness of the blog website
- The increase in the number of page views (PV) because each user would see more diverse blog posts
- More user satisfaction by recommending preferable blogs posts to users

Budget

The Project has been allocated a budget of AUD $550,000 and Table 2 provides an estimated breakdown of the distribution of the budget.

Suggested Component	$AUD
Resources Cost for 1 year	150,000
Recommender Systems Implementation	250,000
Hardware and network infrastructure and security components	50,000
Database (purchase & implementation)	50,000
Contingency	50,000
Total	550,000

Table 2. Breakdown of the project budget

Personnel

The following personnel will be required on full time and part time.

Role	Number	Full Time / Part Time	Contribution
Project Manager	1	Full Time	Managing the project with regards of resources, schedule and budget overall
Business Analyst	1	Part Time	Providing insights and perspectives
Software Developer	1	Full Time	Implementing recommender systems

Accountant	1	Part Time	Overseeing the budget of the project
Data Analyst	1	Full Time	Analysing the data and develop a data mining model

Table 3. Personnel list

References

Hu, Y., Koren, Y. & Volinsky, C. 2008, 'Collaborative filtering for implicit feedback datasets',

2008 Eighth IEEE International Conference on Data Mining, Ieee, pp. 263-72.

Ricci, F., Rokach, L. & Shapira, B. 2011, *Introduction to recommender systems handbook*, Springer.

Thorat, P.B., Goudar, R. & Barve, S. 2015, 'Survey on collaborative filtering, content-based filtering and hybrid recommendation system', *International Journal of Computer Applications*, vol. 110, no. 4.