

Hybrid music recommender using content-based and social information

Paulo Esteban Chiliguano Torres

Project report 2015

School of Electronic Engineering and Computer Science

Queen Mary University of London

2015

Abstract

Contents

1	Introduction	7
1.1	Outline of the thesis	8
2	Background	9
2.1	Recommender Systems	9
2.1.1	Collaborative filtering (CF)	9
2.1.2	Content-based methods	10
2.1.3	Hybrid methods	10
2.2	Online Social Networks	10
2.3	Deep Learning	11
2.3.1	Convolutional Neural Networks (CNN)	11
2.4	Estimation of Distribution Algorithms (EDAs)	11
3	Methodology	12
3.1	Data collection	12
3.1.1	Taste profile subset filtering	12
3.1.2	Audio samples collection	12
3.1.3	Log-mel spectrograms	12

3.2	Algorithms	12
3.2.1	CNN implementation	12
3.2.2	Continuous Bayesian EDA	12
3.2.3	EDA-based hybrid recommender	12
4	Experiments	13
4.1	Evaluation for recommender systems	13
4.1.1	Types of experiments	13
4.2	Evaluation settings	14
4.2.1	Dataset	14
4.2.2	Evaluation measures	15
4.2.3	Experimentation aims	15
5	Results	16
6	Conclusion	17
6.1	Future work	17

List of Figures

List of Tables

Chapter 1

Introduction

Recommender systems can be described as facilities that guide users to interesting objects in a huge space of information. In order to achieve high performance, there is the need of hybridization of two or more recommendation techniques.

This project is going to examine a different approach to develop a hybrid music recommender system in order to suggest new items that would be appealing and enjoyable to the users. This system will combine two recommendation techniques. The first technique is collaborative filtering to predict music preferences on the basis of users' information from an online social network (OSN) such as Last.fm¹, and the second technique is content-based filtering in which acoustical features from audio tracks are correlated to compute their similarities.

Users' information will be obtained from the complementary Taste Profile subset, which is a part of the Million Song Dataset². The music library will

¹<http://last.fm/>

²<http://labrosa.ee.columbia.edu/millionsong/>

be consolidated by crawling songs' information via 7digital API³.

A convolutional neural network (CNN), which is a deep learning model, will be employed for describing the audio files of the music library. Estimation of distribution algorithms (EDA), which are optimization methods in statistics and machine learning, will be investigated to model user profiles that will be comparable with the features of the audio files to predict ratings and produce new item recommendations.

The evaluation of the hybrid recommender system will be assessed by comparing the results with a purely content-based system.

1.1 Outline of the thesis

The rest of the report is organised as follows:

Chapter 2 reviews related work with deep learning techniques and Estimation of Distribution Algorithms on recommendation systems.

Chapter 3 explains the proposed approach of the hybrid system for recommending new music items.

Chapter 4 addresses the experiments and the evaluation scenarios of the performance for the hybrid recommender system.

Chapter 5 discusses and analyses the results from the conducted experiments to evaluate the performance of the proposed hybrid music recommender system approach.

Chapter 6 presents the conclusions and some thoughts for further research.

³<http://developer.7digital.com>

Chapter 2

Background

2.1 Recommender Systems

Recommender systems are software or technical facilities to provide items suggestions or predict customer preferences. These systems play an important role in commercial applications to increase items sales and user satisfaction. Depending on the application, recommender systems can be categorised in the following groups: collaborative filtering, content-based methods and hybrid methods.

2.1.1 Collaborative filtering (CF)

In collaborative filtering (CF), recommendations are based on correlation between users' ratings or they can be predicted from historical user data. The strength of CF is that the recommendation process is independent from the item features. On the other hand, CF would not be suitable when the user-item matrix is sparse. (Burke 2002)

2.1.2 Content-based methods

Content-based methods build user profiles by analysing the users' rated items. Each profile is then processed to be correlated with another item, which has not been rated, to compute the interest of the user on this object. (Lops et al. 2011)

2.1.3 Hybrid methods

Hybrid recommendation is based on the combination of techniques mentioned above, by using the advantages of one system to compensate the disadvantages of the other system.

In this project, CF, that provides song ratings, is integrated with a content-based method, that compare spectral features of song to achieve hybridisation.

2.2 Online Social Networks

Social network sites (SNSs) are web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. (boyd & Ellison 2007)

2.3 Deep Learning

2.3.1 Convolutional Neural Networks (CNN)

2.4 Estimation of Distribution Algorithms (EDAs)■

Chapter 3

Methodology

3.1 Data collection

3.1.1 Taste profile subset filtering

3.1.2 Audio samples collection

3.1.3 Log-mel spectrograms

3.2 Algorithms

3.2.1 CNN implementation

Genre classification

3.2.2 Continuous Bayesian EDA

3.2.3 EDA-based hybrid recommender

Chapter 4

Experiments

4.1 Evaluation for recommender systems

4.1.1 Types of experiments

The scenarios for experiments requires to define an hypothesis, controlling variables and generalization of the results. Three types of experiments (Shani & Gunawardana 2009) can be used to compare and evaluate recommender algorithms:

- **Offline experiments:** where recorded historic data of users' ratings are used to simulate online users behaviour. The aim of this type of experiment is to refine approaches before testing with real users. On the other hand, results may have biases due to distribution of users.
- **User studies:** where test subjects interact with the recommendation system and its behaviour is recorded giving a large sets of quantitative measurements. One disadvantage of this type of experiment is to

recruit subjects that represent the population of the users of the real recommendation system.

- **Online evaluation:** where the designer of the recommender application expect to influence the users' behaviour. Usually, this type of evaluation are run after extensive offline studies.

Also, evaluation for recommender systems can be classified (Celma 2008) in:

- **System-centric evaluation:** The accuracy is based only on users' dataset.
- **Network-centric evaluation:** Other components of the recommendation system such as diversity of recommendations are measured as a complement of the metrics of system-centric evaluation.
- **User-centric evaluation:** The perceived quality and usefulness of recommendations for the users are measured via provided feedback.

4.2 Evaluation settings

The hybrid recommender system of this project is evaluated with an offline experiment and system-centric metrics.

4.2.1 Dataset

For the purpose of evaluation of the hybrid recommender system, a part from the Taste Profile subset is used because the data format includes user-item

ratings and it is publicly available. A 10-fold cross validation is performed which splits the data set in 90% for training and 10% for testing.

4.2.2 Evaluation measures

Because the data set does not include explicit ratings, hence, the number of plays of tracks are used as users' behaviours, decision-based metrics are considered.

4.2.3 Experimentation aims

In order to evaluate the performance of the hybrid recommender, the prediction ratings are compared with a model-based collaborative filtering.

Chapter 5

Results

Chapter 6

Conclusion

6.1 Future work

Bibliography

boyd, d. m. & Ellison, N. B. (2007), ‘Social network sites: Definition, history, and scholarship’, *Journal of Computer-Mediated Communication* **13**(1), 210–230.

URL: <http://dx.doi.org/10.1111/j.1083-6101.2007.00393.x>

Burke, R. (2002), ‘Hybrid recommender systems: Survey and experiments’, *User Modelling and User-Adapted Interaction* **12**(4), 331–370. cited By 0.

URL: <http://www.scopus.com/inward/record.url?eid=2-s2.0-0036959356&partnerID=40&md5=28885a102109be826507abc2435117a7>

Celma, O. (2008), Music Recommendation and Discovery in the Long Tail, PhD thesis, Universitat Pompeu Fabra, Barcelona.

URL: http://mtg.upf.edu/static/media/PhD_ocelma.pdf

Lops, P., de Gemmis, M. & Semeraro, G. (2011), Content-based recommender systems: State of the art and trends, in F. Ricci, L. Rokach, B. Shapira & P. B. Kantor, eds, ‘Recommender Systems Handbook’, Springer US, pp. 73–105.

URL: http://dx.doi.org/10.1007/978-0-387-85820-3_3

Shani, G. & Gunawardana, A. (2009), Evaluating recommender systems,

Technical Report MSR-TR-2009-159.

URL: *<http://research.microsoft.com/apps/pubs/default.aspx?id=115396>*