

**DRAFT**

# **Hybrid music recommender using content-based and social information**

**Paulo Esteban Chiliguano Torres**

Submitted for the degree of Master of Science

Queen Mary, University of London

2015

**DRAFT**

# **Hybrid music recommender using content-based and social information**

**Paulo Esteban Chiliguano Torres**

## **Abstract**

.

Submitted for the degree of Master of Science

Queen Mary, University of London

2015

## Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Outline of the thesis . . . . .	8
<b>2</b>	<b>Background</b>	<b>9</b>
2.1	Recommender Systems . . . . .	9
2.1.1	Content-based Recommender Systems . . . . .	9
2.1.2	Collaborative filtering Recommender System . . . . .	9
2.1.3	Hybrid Recommender Systems . . . . .	9
2.2	Online Social Networks . . . . .	10
2.2.1	APIs . . . . .	10
2.3	Data Fusion Techniques . . . . .	10
<b>3</b>	<b>Main contribution</b>	<b>11</b>
3.1	Methods . . . . .	11
3.1.1	Content based modelling . . . . .	11
3.1.2	Collaborative filtering . . . . .	11
3.2	Algorithms . . . . .	11
3.2.1	Deep Belief Networks . . . . .	11
<b>4</b>	<b>Experiments</b>	<b>12</b>
4.1	Evaluation for recommender systems . . . . .	12
4.1.1	Types of experiments . . . . .	12
4.2	Evaluation settings . . . . .	13
4.2.1	Dataset . . . . .	13
4.2.2	Evaluation measures . . . . .	13
4.2.3	Experimentation aims . . . . .	13
<b>5</b>	<b>Results</b>	<b>14</b>

**DRAFT**

**6 Conclusion 15**

**Bibliography 16**

**DRAFT**

## **List of Figures**

**DRAFT**

## **Acknowledgements**

.

# **Chapter 1**

## **Introduction**

---

Recommender systems can be described as systems 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 audio tracks that would be appealing and enjoyable to the user. This system will combine two recommendation techniques. The first one is content-based technique to obtain acoustical features and metadata (ID3) from a music tracks library, and the second one would be a collaborative filtering technique to recognize music preferences on the basis of users social media profiles e.g. Twitter, Facebook, Last.fm.

Deep learning algorithms, such as Convolutional Restricted Boltzmann Machines (CRBMs) or Convolutional Deep Belief Networks (CDBNs), will be considered to extract features from audio files and for music clustering. Social network information will be obtained by crawling user information via Application Program Interface (API) available in many online social networks. These two sources of related data will be integrated considering data fusion techniques.

The evaluation of the system will be assessed by comparing the results obtained from the hybrid recommender system and the results obtained from a single source based recommender system.

## 1.1 Outline of the thesis

**Chapter 2** covers an introduction to recommender systems, online social networks and information fusion.

**Chapter 3** addresses the design of the hybrid recommender system.

**Chapter 4** evaluates the performance of the hybrid recommender system.

**Chapter 5** describes the results about the application of the hybrid recommender system. The performance is compared to a non-hybrid recommender system based.

**Chapter 6** presents the conclusion and some thoughts for further research.



## Chapter 2

### Background

---

#### 2.1 Recommender Systems

Recommender systems are software or technical facilities that provide items suggestions or predict individual preferences. These systems play an important role in commercial applications to increase sales and user satisfaction.

Recommender systems can be classified in two groups: Content-based systems and Collaborative filtering systems.

##### 2.1.1 Content-based Recommender Systems

Content-based recommender systems build an user profile by analysing user rated items. This profile is then processed to be correlated with another item to compute the interest of the user on this object. [6]

##### 2.1.2 Collaborative filtering Recommender System

In collaborative filtering, recommendations are based on similarities between user's ratings. A model is built from a priori ratings to make predictions. [2]

##### 2.1.3 Hybrid Recommender Systems

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. This project integrates item ratings from users and spectral features of audio and is based on a three-way

aspect model [8]. Real item ratings are obtained through Last.fm API and spectral information are represented by convolutional deep belief networks (CDBN) features computed from items' spectrogram [5].

## **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. [1]

### **2.2.1 APIs**

The publicly available music related information can be collected from user profiles on social networks using Application Program Interface (API).

## **2.3 Data Fusion Techniques**

Combination of multiple sources of information to obtain more relevant parameters is known as data fusion. In this study, a cooperative data fusion technique is considered to augment information provided from social network source to content-based system features. [3]

## Chapter 3

### Main contribution

---

#### 3.1 Methods

##### 3.1.1 Content based modelling

Classifier creates a model for each user based on the acoustic features of the tracks that user has liked.

##### 3.1.2 Collaborative filtering

At this stage, similarities between users is calculated to form a neighbourhood and predict user rating based on combination of the ratings of selected users in the neighbourhood.

#### 3.2 Algorithms

##### 3.2.1 Deep Belief Networks

Deep belief network is a probabilistic model that has one observed layer and several hidden layers.

*Convolutional Deep Belief Network (CDBN)*

## 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 [7] 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 [4] 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 the Last.fm Dataset - 1K users because the data format includes timestamps and it is publicly available. A 10-fold cross validation is performed which splits the data set in 90% for training and 20% 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's behaviour, 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.

**DRAFT**

## **Chapter 5**

### **Results**

---

**DRAFT**

## **Chapter 6**

### **Conclusion**

---

## Bibliography

- [1] danah m. boyd and Nicole B. Ellison. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2007.
- [2] R. Burke. Hybrid recommender systems: Survey and experiments. *User Modelling and User-Adapted Interaction*, 12(4):331–370, 2002.
- [3] F. Castanedo. A review of data fusion techniques. *The Scientific World Journal*, 2013, 2013.
- [4] Ò. Celma. *Music Recommendation and Discovery in the Long Tail*. PhD thesis, Universitat Pompeu Fabra, Barcelona, 2008.
- [5] H. Lee, L. Yan, P. Pham, and A. Y. Ng. Unsupervised feature learning for audio classification using convolutional deep belief networks. In *Advances in Neural Information Processing Systems 22 - Proceedings of the 2009 Conference*, pages 1096–1104, 2009.
- [6] Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. *Content-based Recommender Systems: State of the Art and Trends*, pages 73–105. Springer US, Boston, MA, 2011.
- [7] Guy Shani and Asela Gunawardana. Evaluating recommender systems. Technical Report MSR-TR-2009-159, November 2009.
- [8] K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno. An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model. *IEEE Transactions on Audio, Speech and Language Processing*, 16(2):435–447, 2008.