

# Hybrid music recommender using content-based and social information

Paulo Esteban Chiliguano Torres

Project Report 2015

MSc Sound and Music Computing

School of Electronic Engineering and Computer Science

Queen Mary University of London

August 2015

# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Outline of the thesis . . . . .	6
<b>2</b>	<b>Background research</b>	<b>8</b>
2.1	Online Social Networks . . . . .	9
2.1.1	Last.fm . . . . .	9
2.2	Music services platforms . . . . .	10
2.2.1	Echonest . . . . .	10
2.2.2	7Digital . . . . .	10
2.3	Recommender Systems . . . . .	10
2.3.1	Collaborative filtering . . . . .	10
2.3.2	Content-based methods . . . . .	11
2.4	Hybrid recommender methods . . . . .	11
2.5	Music Information Retrieval . . . . .	11
2.5.1	Musical genre classification . . . . .	11
2.5.2	Deep Learning . . . . .	12
2.5.3	Convolutional Neural Networks . . . . .	12
2.6	Estimation of Distribution Algorithms . . . . .	13

<b>3</b>	<b>Methodology</b>	<b>14</b>
3.1	Data collection . . . . .	14
3.1.1	Taste Profile subset cleaning . . . . .	15
3.1.2	Audio clips retrieval . . . . .	15
3.1.3	Intermediate time-frequency representation for audio signals . . . . .	16
3.2	Algorithms . . . . .	16
3.2.1	CNN architecture . . . . .	16
3.2.2	Continuous Bayesian EDA . . . . .	17
3.2.3	EDA-based hybrid recommender . . . . .	17
<b>4</b>	<b>Experiments</b>	<b>18</b>
4.1	Evaluation for recommender systems . . . . .	18
4.1.1	Types of experiments . . . . .	18
4.2	Evaluation method . . . . .	20
4.2.1	Dataset description . . . . .	20
4.2.2	Evaluation measures . . . . .	20
<b>5</b>	<b>Results</b>	<b>21</b>
<b>6</b>	<b>Conclusion</b>	<b>22</b>
6.1	Future work . . . . .	22

# Chapter 1

## Introduction

Music has accompanied social activities on our daily lives and has influenced the shape of the technology landscape that we have today, such as portable media players, mobile device management applications and music stream services.

Recommender systems can be described as facilities that guide users to interesting objects in a huge space of information. In order to enhance performance, there is the motivations 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.fm, 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 Dataset. The music library will be consolidated by crawling songs' information via 7digital API.

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 prediction accuracy and performance comparison with a typical 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 recom-

mender system approach.

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

## Chapter 2

# Background research

Recommender systems set up opportunities and challenges for industry to understand consumption behaviour of users. In particular, for music industry, the develop of recommender systems could improve sales for artists and labels, and the discovery of new songs for listeners. However, regarding that music tastes vary from one person to another person, an advantageous music recommender system should be able to infer listeners needs through their historical listening preference information, similarities with another listeners, and audio signal features from their music collections.

In the following sections, the importance of online social networks for retrieving user-item information among with previous work on music recommender systems are presented. Subsequently, a novel approach of an hybrid recommender system based on Estimation of Distribution Algorithm (EDA) is introduced and examined.

## 2.1 Online Social Networks

boyd & Ellison (2007) describe social network sites (SNSs) as:

“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.”

During the last decade, online social networks have become the outstanding source of multimedia information.

### 2.1.1 Last.fm

Last.fm is a social network system that accumulate a list of played audio tracks from registered users through *scrobbling* to provide to any user a detail about listening preference and taste similarities between connected friends in the network. Last.fm also uses scrobbling to feed its music recommendation service to help to users to discover new artists.

Users' information such as recently played tracks, loved tracks, or top songs over a time period e.g. weeks, months, can be retrieved by using Last.fm API<sup>1</sup> methods.

---

<sup>1</sup><http://www.last.fm/api>



## **2.2 Music services platforms**

### **2.2.1 Echonest**

### **2.2.2 7Digital**

Both Echo Nest and 7digital require to sign up to their API to get unique keys for OAuth authentication in order to retrieve desired information. As well, free account has limited number of calls, in the case of Echo Nest is limited to 20 request per minute and in the case of 7digital is limited to 4000 request per day.

## **2.3 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. In general, recommender systems can be categorised in the following groups: collaborative filtering and content-based methods.

### **2.3.1 Collaborative filtering**

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 if the user-item matrix is sparse. (Burke 2002)

### **2.3.2 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)

## **2.4 Hybrid recommender 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.

## **2.5 Music Information Retrieval**

Music Information Retrieval (MIR) is an extend of audio signal processing for understanding the usefulness and applications of music data by using time-frequency representations or low-level features. Applications of MIR include artist identification, genre classification and music recommendation.

### **2.5.1 Musical genre classification**

Music classification is one of the principal components for clustering audio tracks based on similarities between features of pieces of music. Automatic

musical genre classification approach proposed by Tzanetakis & Cook (2002), which uses GTZAN genre dataset<sup>2</sup>, has been widely used in the past decade. Nonetheless, the GTZAN dataset has inaccuracies (Sturm 2012), it still provides an useful baseline to compare musical genre classification systems.

### 2.5.2 Deep Learning

One of the aims of learning algorithms is to identify high-level features that help us make sense of an observed data., e.g. genre, mood or release time in a music library. However, it could be difficult to compute these abstract features directly from audio waveforms. Deep learning can solve the difficulty of extracting high-level representations by expressing them in terms of simpler features, e.g. spectrograms. Deep learning allows the computer to build complex concepts out of simpler concepts. (Bengio et al. 2015)

Sigtia & Dixon (2014) examined and compared three implementations of deep neural networks to learn features for music genre classification, using Rectifier Linear Units (ReLUs), dropout regularisation and Hessian Free optimization.

### 2.5.3 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are type of neural networks that uses convolution operation instead of matrix multiplication for processing data that has grid-like topology (Bengio et al. 2015) such as images collection.

---

<sup>2</sup><http://marsyas.info/downloads/datasets.html>

van den Oord et al. (2013) used a convolutional network approach to predict latent factors from music audio in a content-based recommendation system.

## 2.6 Estimation of Distribution Algorithms

Estimation of distribution algorithms (EDAs) (Pelikan et al. 2015) are optimisation techniques by constructing a probabilistic model from a sample of solutions, generating a new population and leading to an optimal solution (Santana et al. 2010).

Liang et al. (2014) exploited an EDA to model user profiles by using weighted featured vectors of keywords from a set of items that the user had rated above a threshold.

In this chapter, previous work on recommender systems has been reviewed and novelty techniques for representing acoustical features and for modelling user profiles has been presented. The next step is to implement the algorithms to collect the dataset by crawling online social information, to extract the acoustical features of a collection of songs for representing them as vectors, to model the user profiles by an EDA, and therefore, to return predicted recommendations.

# Chapter 3

## Methodology

The methodology used to develop the hybrid music recommender consists of three main stages. First, the collection of real users' data corresponding to the number of playings of specific songs and the retrieval of audio samples of the identified songs in the users' data. Secondly, the implementation of the deep learning algorithm to represent the songs as vectors and the EDA to model the user profiles

### 3.1 Data collection

The Million Song Dataset (Bertin-Mahieux et al. 2011) is a collection of audio features and metadata for a million contemporary popular music tracks which purpose in MIR is to provide a ground truth for evaluation research. This collection is also complemented by the Taste Profile subset <sup>1</sup> which provides 48,373,586 triplets that consists of Last.fm user ID, Echo Nest song ID and

---

<sup>1</sup><http://labrosa.ee.columbia.edu/millionsong/tasteprofile>

play count of the song.

### 3.1.1 Taste Profile subset cleaning

Due to potential mismatches between song IDs and track IDs on the Echo Nest database, it is required to filter out the wrong matches in the Taste Profile subset. A Python script is implemented to discard the triplets that contain the song ID values from the mismatches list available also on the Million Song Dataset webpage. The resulting triplets are stored in a new CSV file.

### 3.1.2 Audio clips retrieval

The list of songs IDs from the triplets obtained in the last step are used to retrieve the track IDs through a Python script that includes the Pyechonest<sup>2</sup> package which allow us to acquire track ID with *get\_tracks* method through Echo Nest API<sup>3</sup> requests. The reason behind obtaining track IDs is because for each ID we can retrieve a 30-60 seconds preview audio clips through 7digital API<sup>4</sup>.

Additionally, the Python script accumulates the song ID, the URL, artist and song metadata of each track available in a text file. If the track for a song ID is not available, the script skips to the next song ID to retrieve information of it. The generated text file can be used to reduce more the triplets dataset from the last section.

---

<sup>2</sup><http://echonest.github.io/pyechonest/>

<sup>3</sup><http://developer.echonest.com>

<sup>4</sup><http://developer.7digital.com>

### 3.1.3 Intermediate time-frequency representation for audio signals

For representing audio waveforms of the song collection obtained through 7digital API, a similar procedure suggested by van den Oord et al. (2013) is followed:

- Read 3 seconds of each song at a sampling rate of 22050 Hz and mono channel.
- Compute log-mel spectrograms with 128 components from windows of 1024 frames and a hop size of 512 samples.

The Python script for feature extraction implemented by Sigtia & Dixon (2014) is modified to return the log-mel spectrograms by using the LibROSA<sup>5</sup> package.

“Representations of music directly from the temporal or spectral domain can be very sensitive to small time and frequency deformations”. (Zhang et al. 2014)

## 3.2 Algorithms

### 3.2.1 CNN architecture

The input of the CNN consist of the 128-component spectrograms obtained in feature extraction. The batch size considered is 20 frames. Each convolutional layer consists of 10 kernels and ReLUs activation units. In the first

---

<sup>5</sup><https://bmcfee.github.io/librosa/index.html>

convolutional layer the pooling size is 4 and in the second layer the pooling size is 2. The filters analyses the frames along the frequency axis to consider every Mel components with a hop size of 4 frames in the time axis. Additionally, there is a hidden multi perceptron layer with 513 units.

### **Genre classification**

The classification of genre for each frame is returned by negative log likelihood estimation of a logistic stochastic gradient descent (SGD) layer.

### **3.2.2 Continuous Bayesian EDA**

### **3.2.3 EDA-based hybrid recommender**



# Chapter 4

## Experiments

In order to evaluate the performance of a recommender system, there are several scenarios to be considered depending on the structure of the dataset and the prediction accuracy. It is therefore necessary to determine a suitable experiment for evaluation of the proposed hybrid music recommendation system that employs an user-item matrix and vector representation for songs as inputs to predict ratings of items that an user has not previously listened to. In addition, the performance of the hybrid approach is compared with a pure content-based recommender algorithm.

### 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.

Besides, evaluation of recommender systems can be classified (Celma 2008) in:

- **System-centric** process has been extensively exploited in CF systems. The accuracy of recommendations is based exclusively on users' dataset.
- **Network-centric** process examines other components of the recommendation system, such as diversity of recommendations, and they are measured as a complement of the metrics of system-centric evaluation.
- **User-centric:** The perceived quality and usefulness of recommendations for the users are measured via provided feedback.

## **4.2 Evaluation method**

The hybrid music recommender system proposed in this project is evaluated through an offline experiment and the results are presented with system-centric metrics.

### **4.2.1 Dataset description**

For the purpose of evaluation of the hybrid recommender system, a sample 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 dataset does not include explicit ratings, hence, the number of plays of tracks are considered as users' behaviours,  
decision-based metrics are considered.

## Chapter 5

### Results

## Chapter 6

## Conclusion

### 6.1 Future work

# Bibliography

Bengio, Y., Goodfellow, I. J. & Courville, A. (2015), Deep learning. Book in preparation for MIT Press.

**URL:** <http://www.iro.umontreal.ca/~bengioy/dlbook>

Bertin-Mahieux, T., Ellis, D. P., Whitman, B. & Lamere, P. (2011), The million song dataset, *in* ‘Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)’.

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](http://mtg.upf.edu/static/media/PhD_ocelma.pdf)

Liang, T., Liang, Y., Fan, J. & Zhao, J. (2014), ‘A hybrid recommendation model based on estimation of distribution algorithms’, *Journal of Computational Information Systems* **10**(2), 781–788. cited By 0.

**URL:** <http://www.scopus.com/inward/record.url?eid=2-s2.0-84892865461&partnerID=40&md5=a2927d36b493e8ef4d1cdab3055fa68b>

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](http://dx.doi.org/10.1007/978-0-387-85820-3_3)

Pelikan, M., Hauschild, M. W. & Lobo, F. G. (2015), Estimation of distribution algorithms, *in* ‘Springer Handbook of Computational Intelligence’, Springer, pp. 899–928.

Santana, R., Bielza, C., Larraaga, P., Lozano, J. A., Echegoyen, C., Mendiburu, A., Armaanzas, R. & Shakya, S. (2010), ‘Mateda-2.0: A matlab package for the implementation and analysis of estimation of distribution algorithms’, *Journal of Statistical Software* **35**(7), 1–30.

**URL:** <http://www.jstatsoft.org/v35/i07>

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>

Sigtia, S. & Dixon, S. (2014), Improved music feature learning with deep neural networks, pp. 6959–6963. cited By 0.

**URL:** <http://www.scopus.com/inward/record.url?eid=2-s2.0-84905259152&partnerID=40&md5=3441dfa8c7998a8eb39f668d43efb8a1>

Sturm, B. (2012), An analysis of the gtzan music genre dataset, pp. 7–12.  
cited By 0.

**URL:** <http://www.scopus.com/inward/record.url?eid=2-s2.0-84870497334&partnerID=40&md5=40a48c1c9d787308dd315694b54b64ec>

Tzanetakis, G. & Cook, P. (2002), ‘Musical genre classification of audio signals’, *IEEE Transactions on Speech and Audio Processing* **10**(5), 293–302.  
cited By 976.

**URL:** <http://www.scopus.com/inward/record.url?eid=2-s2.0-0036648502&partnerID=40&md5=72d2fee186b42c9998f13415cbb79eea>

van den Oord, A., Dieleman, S. & Schrauwen, B. (2013), Deep content-based music recommendation, in C. Burges, L. Bottou, M. Welling, Z. Ghahramani & K. Weinberger, eds, ‘Advances in Neural Information Processing Systems 26’, Curran Associates, Inc., pp. 2643–2651.

**URL:** <http://papers.nips.cc/paper/5004-deep-content-based-music-recommendation.pdf>

Zhang, C., Evangelopoulos, G., Voinea, S., Rosasco, L. & Poggio, T. (2014), A deep representation for invariance and music classification, in ‘Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on’, IEEE, pp. 6984–6988.