

**School of Electronic
Engineering and
Computer Science**

MSc Sound and Music Computing

Project Report 2015

Hybrid music recommender using content-based and social information

Paulo Esteban Chiliguano Torres

Acknowledgements

I wish to express my sincere gratitude to Dr. Georgy Fazekas, Lecturer in Digital Media at Queen Mary University of London, for providing me guidance and valuable suggestions during the planning and development of this project. I also wish to acknowledge the help provided by Dr. Mathieu Barthet and Mr. Tim Kay with the access to research servers.

I would also like to extend my thanks to Dr. Tony Stockman, my academic advisor, for the assistance provided during my time as a student at Queen Mary University of London.

I am particularly grateful with National Government of the Republic of Ecuador for awarding me with a scholarship to study a postgraduate degree at a high-quality research university in the United Kingdom of Great Britain and Northern Ireland.

Finally, I wish to thank my parents and brothers for their support and encouragement throughout my study.

Abstract

This is an abstract.

Contents

1	Introduction	1
1.1	Motivation	2
1.2	Aim	3
1.3	Outline of the thesis	4
2	Background	5
2.1	Online Social Networks	6
2.1.1	Last.fm	6
2.2	Music services platforms	7
2.2.1	Echonest	7
2.2.2	7Digital	7
2.3	Recommender Systems	7
2.3.1	Collaborative filtering	7
2.3.2	Content-based methods	8
2.4	Hybrid recommender methods	8
2.5	Music Information Retrieval	8
2.5.1	Musical genre classification	8
2.5.2	Deep Learning	9

2.5.3	Convolutional Neural Networks	9
2.6	Estimation of Distribution Algorithms	10
3	Methodology	11
3.1	Data collection	11
3.1.1	Taste Profile subset cleaning	12
3.1.2	Audio clips retrieval	12
3.1.3	Intermediate time-frequency representation for audio signals	13
3.2	Data preprocessing	13
3.3	Algorithms	14
3.3.1	CNN architecture	14
3.3.2	Continuous Bayesian EDA	14
3.3.3	EDA-based hybrid recommender	14
4	Experiments	15
4.1	Evaluation for recommender systems	15
4.1.1	Types of experiments	15
4.2	Evaluation method	17
4.2.1	Dataset description	17
4.2.2	Evaluation measures	17
5	Results	18
6	Conclusion	20
6.1	Future work	20

List of Figures

List of Tables

5.1	Genre classification Results	19
-----	--	----

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. Portable media players, mobile device applications or music streaming services enable us the access to a large volume of digital recorded music. This vast range of music tracks might include songs that are relevant or not to a listener, being necessary to develop facilities to bring out appropriate musical pieces to an user.

Recommender systems can be described as engines that guide the users to suitable objects from a large number of options in a particular domain such as books, films or music. The available information of users and items' attributes is analysed and exploited by the recommender systems to produce a list of previously unseen items that each user might find enjoyable. Depending on the analysed data, the design of a recommender can be focused on historical ratings given by users or similarities between the attributes of items that an user already rated.

1.1 Motivation

Due to the available information of relationship between users and items would be sparse, e.g., most part of the users tend to do not give enough ratings, the accuracy of predictions would decrease. Another disadvantage of traditional recommender systems, referred as *cold-start problem*, arises when a new item cannot be recommended until it gets enough ratings, or, equivalently, when a new user does not have any ratings (Melville & Sindhvani 2010). In order to alleviate the rating sparsity and cold-start problems, there is the motivation to combine two or more recommendation designs into hybrid approaches.

Deep learning is an approach to artificial intelligence for describing raw data as a nested hierarchy of concepts, with each abstract concept defined in terms of simpler representations. For example, deep learning can describe high-level features of an image of a car such as position, color or brightness of the object, in terms of contours, which are also represented in terms of edges. (Bengio et al. 2015)

Inspired in natural evolution of species, Estimation of Distribution Algorithms (EDAs) (Larranaga & Lozano 2002) are robust techniques developed during the last decade for optimisation in Statistics and Machine Learning fields. EDAs can capture the explicit structure of a population with a probability distribution estimated from the best individuals of that population.

1.2 Aim

We aim to design and implement a hybrid music recommender to suggest new music tracks that an user would find them appealing and enjoyable. The architecture of our hybrid recommender combines two recommendation techniques.

the second technique is *content-based filtering* where recommendations are produced by computing similarities between representations of content of items that an user

in which are correlated to compute similarities between them.

Users' information is obtained from the Taste Profile dataset, which is a complementary subset of the Million Song Dataset¹. The music library that contains sample audio clips of the rated songs in the Taste Profile dataset is consolidated by fetching audio files using 7digital API.

A convolutional neural network (CNN), which is a deep learning model, is employed to describe each audio file of the music library with a n-dimensional vector, whose dimensions represent music genres.

An Estimation of Distribution Algorithm (EDA) technique is implemented to model user profiles in terms of music genres in order to compare each profile with the vector representation of the audio clips to compute similarities between them. Recommendation is achieved by choosing the clips with highest similarity values.

The evaluation of our hybrid music recommender will be assessed by comparing the prediction accuracy with a traditional content-based recommender

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

p.

1.3 Outline of the thesis

The rest of the report is organised as follows: Chapter 2 provides an overview in recommender systems. Recommendation process, associated challenges, and related work based on state-of-the-art techniques are discussed. In Chapter 3, we present our proposed hybrid recommendation approach and describe the stages and algorithms in detail. The experiments and evaluation protocols are to assess the performance of the hybrid recommender presented in Chapter 4. We proceed to discuss and analyse the results from the conducted experiments to evaluate the proposed hybrid music recommender. In Chapter 6, we present the conclusions and some thoughts for further research.

Chapter 2

Background

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.

social net info edges between user

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 API¹ methods.

¹<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 dataset², 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.

²<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.

These algorithms were applied in complex problems such as load balancing for mobile networks (Hejazi & Stapleton 2015) or software reliability prediction

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 ¹ which provides 48,373,586 triplets that consists of Last.fm user ID, Echo Nest song ID and

¹<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 Pyechonest² package which allow us to acquire track ID with *get_tracks* method through Echo Nest API³ requests. The reason behind obtaining track IDs is because for each ID we can retrieve a 30-60 seconds preview audio clips through 7digital API⁴.

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.

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

³<http://developer.echonest.com>

⁴<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 LibROSA⁵ 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 Data preprocessing

- Rating complementary cumulative distribution
- Flattenning spectrogram Sidsig

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

3.3 Algorithms

3.3.1 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 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.

3.3.2 Continuous Bayesian EDA

3.3.3 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,

Chapter 5

Results

fadslkfjdsalfjdsalf el mindsafa se va a cabarakl;dflakdjfl;akds dflk;djflkajflkajf

jlk;d;fjlk;ajdlf;kajsl;d

jkdl;fkaj

Table 5.1: Genre classification Results

Trial	Validation error (%)	Test error (%)	Iterations	Time elapsed (min.)
1	58.0	65.2	650	7.00
2	37.6	46.0	2150	13.07
3	39.6	46.0	700	7.54
4	35.6	36.8	550	6.01
5	36.4	40.0	250	5.47
6	40.4	44.8	150	5.41
7	32.4	40.4	800	8.64
8	36.0	38.8	250	5.42
9	34.0	38.8	850	9.14

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

Hejazi, S. A. & Stapleton, S. P. (2015), ‘A self-organized network for load balancing using intelligent distributed antenna system’, *Canadian Journal of Electrical and Computer Engineering* **38**(2), 89–99.

URL: *www.scopus.com*

Larranaga, P. & Lozano, J. A. (2002), *Estimation of distribution algorithms: A new tool for evolutionary computation*, Vol. 2, Springer Science & Business Media.

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*

Melville, P. & Sindhvani, V. (2010), Recommender systems, in ‘Encyclopedia of machine learning’, Springer, pp. 829–838.

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. Ghahra-

mani & 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.