Asian Journal of Information Technology 15 (21): 4250-4254, 2016
ISSN: 1682-3915


# Music Recommendation System

D. Jayashree, S. Goutham Manian and C. Pranav Srivatsav
Department of Computer Science and Engineering, Rajalakshmi Institute of Technology,
Kuthampakkam, Banglore Highway, 600123 Chennai, India

**Abstract:** The music recommendation systems, currently in use rely upon ratings, likes and generic generalization of genres. This produces far from ideal recommendations and possibly constricts the exploration span of the listener, thus limiting one's library to a set of popular artists and titles. In order to refine the recommendation system, scientific attributes of the track must be taken into account. These attributes can be represented in the form of vector parameters. These vector parameters can be meaningfully defined using a specially designed FFT algorithm and the derived data is sent to the main server which will serve as an open ended system. Thus resulting in a system which leads to an indirect recommendation performed by another user client.

**Key words:** Music recommendation system, exploration span, listener, library, vector parameters

## INTRODUCTION

The music industry is currently in a metamorphic stage where the shift towards digital distribution is becoming more evident with the sudden boom of streaming services like apple's itunes, Spotify and Pandora. As a result, automatic music recommendation has become an increasingly relevant problem, allowing listeners to explore new music and likewise targeting their wares to the right audience.

Recommendation can be defined as "the process of utilizing the opinions of a community of customers to help individuals in that community more effectively identify content of interest from a potentially overwhelming set of choices" (Resnick and Varian, 1997). Effective recommendation reduces the user's effort and time in making decisions. The majority of the existing researches solely focus on the transaction and preference data. The time has come for the recommender systems to evolve by involving contexts surrounding the attributes pertaining to the musical data.

Even with the ever increasing study on recommender systems, the problem of music recommendation continues to remain complicated because of the multiple parameters that influence the listener's preferences. The parameters range from genres to social and geographical factors.This makes the count of recommendable items very large. This number can be reduced by recommending albums or artists instead but this is not always compatible with the intended use of the system (e.g., automatic playlist generation) and it disregards the fact that the repertoire of an artist is rarely homogeneous and also add to the fact that listeners may enjoy particular songs more than others.

Most of the recommender systems have heavy reliance on usage patterns, listener ratings and meta data. This type of collaborative filtering mostly outperforms content-based filtering but exclusively with the inclusion of usage data. This makes it vulnerable to a cold start problem. In order to counter this drawback, we take into account the scientific parameters derived from the track which alleviates the dependency over usage data during the filtering process.

**Context awareness:** Context awareness is a term that describes the ability of the computer to sense and act upon information about its environment, such as location, time, temperature or user identity (Lee and Lee, 2007). Context is an important factor when one provides services such as music recommendation to the users sice user preferences could vary due to context where the user is (Park *et al*., 2006). Context can be classified into four types namely location, time, activity and identity. A context framework can be created by defining values to these four types. The additional advantage with recommendations made using context awareness is it alleviates the problem of information overload.

Schluter and Osendorfer (2011) presents a system that generates recommendations using content-based music similarity estimation. The distinction with his approach is that, he carefully by passes time-intensive features and still manages to extract good performance by

**Corresponding Author:** D. Jayashree, Department of Computer Science and Engineering, Rajalakshmi Institute of Technology, Kuthampakkam, Banglore Highway, 600123 Chennai India

pre-processing the data. Additionally, he employs unsupervised learning for local feature extraction and incepts an exciting approach for evolution of recommender systems.

Resnick and Varian (1997) states that the kinds of items being recommended and the people among whom evaluations are shared. Consider, first, the domain of items. The sheer volume is an important variable: Detailed textual reviews, although a rarity, may be practical but applying the same approach to thousands of daily subscribers would not be practical. First, recommenders may not explicitly collaborate with recipients who may be unknown to each other. Second, recommendations may suggest particularly interesting items in addition to indicating those that should be filtered out.

Lee and Lee (2007) considers and collects more features that can represent the user's context and then select the appropriate features by feature selection process. For selecting the music for recommendation, he utilizes not only the user's demographics and behavioral patterns but also his/her context at the time of making recommendation. He further employs step case-based reasoning i.e. the first step for retrieving similar contexts and the second step for retrieving the similar users.

Yoshii *et al.* (2008) hybrid approach overcomes the conventional tradeoff between recommendation accuracy and variety of recommended artists. Collaborative filtering which is used on e-commerce sites, cannot recommend nonrated pieces and provides a narrow variety of artists.

Yading Song gives an effective summary of a basic metadata based model and two popular music recommender approaches: collaborative filtering and content-based model. The effectiveness of a hybrid model is perfectly illustrated as it outperforms a single model, since it incorporates the advantages of various methods. The relatively unexplored models based on social information, emotion and context are explored and it points at a positive inference with respect to the generated recommendation.

The catch with making inferences based on context is that they cannot deal with diverse information effectively and also add to the fact that the usage of discrete data can lead to loss of information which might decrease the resolution of the context on the basis of which inferences are made.

**Parameter analysis:** A large feature vector has some advantages, but also many disadvantages, such as for example the parameter separability problem. A variety of descriptors may allow for an easier differentiating between classifies genres. However, an important aspect of the parametrization effectiveness analysis is reducing the feature vector redundancy.

One of the methods to handle this redundancy is correlation analysis which makes use of a correlation matrix along with the acquired vector parameters to interpret the individual coefficients based on the statistics. This makes it easier to identify redundant parameters. A smaller vector thus resulting makes it efficient while handling server requests over the database.

Next the effectiveness of the vector needs to be ascertained. The three statistical parameters: average value (arithmetic mean), variance and skewness provides us the optimized value. This number is the optimization result. In this way, each sub-vector is described by three numbers, resulting in a major data reduction, that is, a shortening of parameters.

The usage of another vector allows us to accurately place the songs into their respective genres. This vector is based on fuzzy logic. Thus external data signatures need not be used for genre placement which reduces the data. The effectiveness tests for this particular vector makes use of the decision algorithms implemented using fuzzy logic and kNN algorithm which acts as a minimum distance classifier.

## MATERIALS AND METHODS

**User modeling:** User modeling as the one of the key elements, it models the difference in profile. For example, the difference in geographic region or age, their music preferences might be different. Interestingly, other factors such as gender, life styles and interests could also determine their choices of music. By including the user's current emotional state, music recommendations can be improved. Our next efforts will be to investigate how people regulate their emotions with music. That is what kind of emotionally laden music people are listening when being in a specific emotional state. Additionally, emotion regulation with music is related to their personality

**First step:** User profile modeling (Celma, 2009) suggested that the user profile can be categorized into three domains: demographic, geographic and psycho graphic shown in Table 1. Based on the steadiness, psychological data has been further divided into stable attributes which are essential in making a long term prediction and fluid attributes which can change on an hour to hour basis.

**Second step:** User listening experience modeling depending on the level of music expertise, their expectations in music are varied accordingly. Table 2

Table 1: User profile category

| Data types | Example |
|---|---|
| Demographic | Age, marital status, gender, etc. |
| Geographic | Location, city, country, etc. |
| Psychographic | Stable: Interests, lifestyle, etc. |
| | Fluid: mood, attitude, opinions, etc. |

Table 2: Listeners category

| Types | Percentage | Features |
|---|---|---|
| Savants | 8 | Extensive music knowledge |
| Enthusiasts | 31 | Above average knowledge |
| Casuals | 21 | Average knowledge |
| Indifferent | 40 | Low knowledge |

analyses the different types of listeners whose age range from 16-45 and categorized the listeners into four groups: savant, enthusiasts, casuals indifferent.

**TV item profiling:** Music meta-data is categorized into three categories: editorial meta-data, cultural meta-data and acoustic meta-data.

**Editorial meta-data:** Meta-data obtained by a single expert or group of experts. This is obtained literally by the editor and also it can be seen as the information provided by them. For example the cover name, composer, title or genre etc.

**Cultural meta-data:** Meta-data obtained from the analysis of corpora of textual information, usually from the Internet or other public sources. This information results from an analysis of emerging patterns, categories or associations from a source of documents, e.g. Similarity between music items.

**Acoustic meta-data**: Meta-data obtained from an analysis of the audio signal. This should be without any reference to a textual or prescribed information, eg., beat, tempo, pitch instrument, mood, etc.

**Semantic fissure in music:** Latent factor vectors form a compact description of the different facets of users' tastes and the corresponding characteristics of the items (Celma, 2009). In order to demonstrate this, latent factors computed over a small set of usage data and a small selection of artists whose songs have very positive or negative values for each factor in Table 3. Since, usage data is scarce for many songs, it is often impossible to reliably estimate these factor vectors. Therefore it would be useful to be able to predict them from music audio content.

This would require an acknowledgement towards the fact that there lies a large semantic fissure between the characteristics of a song that affect user preference and the corresponding audio signal. Extracting high-level properties such as genre, mood instrumentation and lyrical themes from audio signals requires powerful models that are capable of capturing the complex hierarchical structure of music. Additionally, some properties are impossible to obtain from audio signals alone, such as the popularity of the artist, their reputation and and their location. Various researches are concerned with the retrieval of these high level properties of music, with several modern approaches inclining towards the usage of neural networks.

**Dataset:** The usage of large scale datasets is limited due to the presence of licensing issues. Tracks which are open sourced partly for music information retrieval purposes are our primary source of dataset. One such source is the Million Song Dataset (MSD) which is a collection of meta data and precomputed audio features for one million contemporary songs. Several other datasets linked to the MSD are also available, featuring lyrics, cover songs, tags and user listening data (Hoffman *et al*., 2009). This makes the dataset suitable for a wide range of different music information retrieval tasks.

Due to its size, the MSD allows for the music recommendation problem to be studied in a more realistic setting than was previously possible. It is also worth noting that the taste profile subset is one of the largest collaborative filtering datasets that are publicly available today.

Another source that is being used are the 29 sec data clips provided by 7digital.com. The 7 digital offers music tracks in MP3 320, 256, M4A, 16-bit and 24-bit FLAC audio which provides us with some amount of flexibility in understanding the varying levels of detail that can be obtained from different formats.

**Automatic tags:** Automatic tags help us to capture acoustic similarity between tracks by marching semantic characteristics of the audio content.

The automatic tagging algorithm is used to define a track using various semantic tags. The tags include descriptors of the instruments, vocals, emotion and usages of the track.

The most common unsupervised machine learning model for this type of task is latent Dirichlet Allocation (LDA). This model automatically infers a collection of attributes over a corpus of data based on the bit elevation in the tracks. Running LDA on a set of music would assign matches which are relevant to that track in a probabilistic manner. This falls under the general category of 'clustering' algorithms. There are many possible choices of such algorithms but it still remains an active area of research.

Table 3: Examples of different genre category

| Genre | Artists with +ve values | Artists with -ve values |
|---|---|---|
| Electronic music | Katy Perry, John Legend, Taylor Swift | The Kills, Interpol, Man Man, Beirut, the bird and the bee |
| Classic rock | Bonobo, Flying Lotus, Cut Copy, Chromeo, Boys Noize | Shinedown, Rise Against, Avenged Sevenfold, Nickelback, Flyleaf |
| Indie rock | Phoenix, Crystal Castles, Muse, Royksopp, " Paramore | Traveling Wilburys, Cat Stevens, Creedence, The police |

For a given song, the output of this "auto-tagger" is a set of probabilities that indicate the relevance of each tag to the song. These probabilities may be interpreted as the parameters of a "semantic" multinomial distribution that characterizes the song, just as a human listener might use words to describe a song's acoustic content (e.g., "very jazzy, features a lot of saxophone and piano and good to listen to on a date"). The similarity is computed by taking the divergence between semantic distributions into account. The songs with minimum divergence with respect to the seed song are taken to feed into the recommender output.

**Content v/s metadata:** There always exists a competition between the influences of acoustic and artist similarity. These rivaling influences are taken into account while measuring the divergence from the seed song. These measurements, obtained from divergence also give bad songs. Thus, a better than random level artist similarity is captured.

But, the suggestion that a simple minimization of divergence can lead to good recommendations is unhealthy. Likewise, recommending similar artists is not sufficient as many bad songs had high artist similarity.

**Hybrid recommender systems:** A hybrid recommender system that unites theoretically the content-based data and the collaborative data is the most elegant approach to make the recommender systems much more effective (Yoshii *et al*., 2008). There lies two problems in designing a hybrid recommender.

**Reliable integration:** The initial problem is to figure out a way to reflect the collaborative and content-based data when making recommendations. Two contrasting ways to solve this problem is to either use collaborative and content based methods in parallel or in cascade. The drawback with these solutions is that although meta recommender systems have been proposed to select a recommender system among conventional ones on the basis of certain quality measures, the disadvantages of the selected system are inherited. Moreover, the heuristics-based integration dealt with in other studies lacks a principled justification.

**Efficient calculation:** The second problem is to make the recommender efficiently adaptable to the increase in rating scores and users. Here a memory-based approach looks feasible as the whole data is used to make recommendations, however, this results in delayed responses. A more efficient method is to train an aspect model for model-based collaborative filtering. The joint probability distribution over users, pieces and features is decomposed into three independent distributions which are respectively conditioned by genres. These distributions are statistically estimated so that the probability of generating the observed data is maximized. This allows us to incrementally train the aspect model according to the increase in users and rating scores at low computational cost, i.e., we only need to partially update the parameters.

**Mapping of latent factors:** The regression problem that is encountered while mapping latent factors can only be faced by learning a function that maps a time series to a vector of real numbers. We evaluate two methods to achieve this: one follows the conventional approach by extracting local features from audio signals and aggregating them into a Bag-of-Words (BoW) representation. Any traditional regression technique can then be used to map this feature representation to the factors. The other method is to use a deep convolutional network. Latent factors can be used to train the prediction models and the compatibility for large implicit feedback datasets.

Many latent factors can be interpreted, for instance in a music domain as the amount of bass or treble in a track, bitrate, record quality and so on.

Dimensionality reduction of the latent factors is achieved by defining a reverse mapping from data space onto latent space, so that every data point is assigned a representative in latent space. In latent variable modeling, once the parameters are fixed, Bayes' theorem gives the posterior distribution in latent space given a data vector, i.e. the distribution of the probability that a point in latent space was responsible for generating the data point.

## RESULTS AND DISCUSSION

### Evaluation of latent factors

**Quantitative evaluation:** The predictions provided by the prediction models are used to make quantitative assessments over the latent factors. For every use 'u' and for every song 'i' in the test set we compute a score and recommend the songs with the highest scores first. As mentioned before, we also learn a song's similarity metric on the bag-of-words representation using metric learning to rank. In this case, scores for a given user are computed

Table 4: Qualityn evalation

| Query | Most similar (WMF) | Most similar (Predicted) |
|---|---|---|
| Daft punk-touch | Daft Punk-Shortvision | Savage garden-truly madly deeply |
| | Daft Punk-Short Circuit | Gotye-making mirrors |
| Coldplay-paradise | Coldplay-viva la vida hasley-ghost | Coldplay-yellow lorde-rule the world |

by averaging similarity scores across all the songs that the user has listened to. Finite mixtures of latent variable models can be constructed in the usual way using a linear combination of component models.

**Qualitative evaluation:** The complexity involved in evaluating recommender systems is the fact that the accuracy metrics by themselves do not provide enough insight into whether the recommendations are sound. The qualitative assessment is done by cross correlating the predictions provided by the convolutional neural network. Qualitative analysis identifies three variables that are manifested when indicators of management of exception are observed. A few songs and their closest matches according to both models are shown in Table 4. When the predicted latent factors are used, the matches are mostly different but the results are quite reasonable in the sense that the matched songs are likely to appeal to the same audience. Furthermore, the variance that is generated is a useful property for recommender systems.

## CONCLUSION

The primary conclusion arrived is that accretion of the feature vectors and optimization of latent factors is essential for resolving recommendations from larger data sets. High performance tests (>80% efficiency), thus far have arrived only from using small scale data sets, none >1000 tracks. Therefore, in order to implement the proposed solutions effectiveness, one needs to consider the length of the audio speck, quantum of the test set and the ability of the recommender to assimilate, observe and refine its recommendations.

The music database used in the study contains 20-30 sec recordings fragments. The observed studies have indicated that the most effective portion for conducting trials over the efficacy of the classifier is the portion between the bridge and the chorus of the track. The ideal number of parameters is given by the degree of narrowness in the recommended spectrum which with the help of the current studies is ascertained as 28.

Removal of bias resulting from artist names, song familiarity and other meta-data based factors from the evaluator is important to distribute concentration between all the contexts equally.

Collaborative filters form the basis of state-of-the-art recommendation systems but cannot directly form recommendations or answer queries for items which have not yet been consumed or rated. By optimizing content-based similarity from a collaborative filter this provides a simple mechanism for alleviating the cold-start problem and extending music recommendation to novel or less known songs

It is particularly critical for the recommender to be furnished with lossless and optimized input. The effectiveness in the genre classification is further aided by the weights assigned to the parameters, with respect to the amount of data that needs to be monitored for obtaining them. The resultant effectiveness approaches an impressive 85%. The prevalent results arrived from the terminal portion of this study is acceptable and further research is required to increase the classification effectiveness for substantially larger datasets. The proposed methods are quite general and may apply to a wide variety of applications involving content-based similarity such as nearest-neighbor classification of audio signals

## REFERENCES

Celma, H.O., 2009. Music recommendation and discovery in the long tail. Ph.D Thesis, Universitat Pompeu Fabra, Barcelona, Spain. http://repositori.upf.edu/handle/10230/12223

Hoffman, M.D., D.M. Blei and P.R. Cook, 2009. Easy as CBA: A simple probabilistic model for tagging music. ISMIR., 9: 369-374.

Lee, J.S. and J.C. Lee, 2007. Context Awareness by Case-Based Reasoning in a Music Recommendation System. Springer, Heidelberg, pp: 45-58.

Park, H.S., J.O. Yoo and S.B. Cho, 2006. A Context-Aware Music Recommendation System using Fuzzy Bayesian Networks with Utility Theory. In: Fuzzy Systems and Knowledge Discovery, Lipo, W., L. Jiao, G. Shi, X. Li and L. Jing (Eds.). Springer, Berlin, Germany, ISBN:978-3-540-45916-3, pp: 970-979.

Resnick, P. and H.R. Varian, 1997. Recommender systems. Commun. ACM., 40: 56-58.

Schluter, J. and C. Osendorfer, 2011. Music similarity estimation with the mean-covariance restricted boltzmann machine. Proceeding of the 2011 10th International Conference on Machine Learning and Applications and Workshops ICMLA, December 18-21, 2011, IEEE, Munich, Germany, ISBN:978-1-4577-2134-2, pp: 118-123.

Yoshii, K., M. Goto, K. Komatani, T. Ogata and H.G. Okuno, 2008. An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model. IEEE Trans. Audio Speech Language Process., 16: 435-447.