

A Novel Hybrid K-Means PLSA Technique for Music Recommender

Gurpreet Singh^{1*}, Rajdavinder Singh Boparai² and Mohit Kathpal³

Computer Science and Engineering Department, Chandigarh University, Mohali - 140413, Punjab; India;
Gur92reet@gmail.com, boparai_raji@yahoo.co.in, mohitcu92@gmail.com

Abstract

We propose a hybrid approach of music recommendation. According to this approach, user will be recommended with the songs searched by the other users similar to the current user on the basis of content and context of the songs listened by the set of similar users. Here we will try to find the similar users on the basis of context using k-means clustering technique and among those users we will find the most frequent or most listened songs on the basis of content using PLSA technique. The proposed hybrid approach is expected to achieve the higher accuracy and efficiency as compared to other approaches for music recommendation.

Keywords: Context Based, Content Based, Hybrid, K-Means, Music Recommendations, PLSA Technique

1. Introduction

With the dramatic increase of network in the past decades, internet has become the major origin of retrieving multimedia information such as video, books, and music etc. People have regarded music as a special aspect in their lives and they listen to music, an activity they engaged in frequently. Lesser hardware costs and up gradations in technology have led to a diverse growth in digital music. However, the vast amount of music available has made it more difficult for users to find the music they enjoy. Now, it has become more important than ever to filter out music that are relevant to each user. Thus, recommendation systems have become vital because of their capacity to result only the suitable products from the large gamut of accessible data.

However, the concern is about the management of huge music data that the population produces. Music information retrieval schemes already developed to handle various problems like classification of genre, identification of artist, and acoustic features recognition. Moreover, music recommendation systems are there that helps to filter out proper songs according to the tastes of

users. A successful music recommendation system is one that automatically finds preferences and results relevant playlists. Meanwhile, recommender system provides a big chance for music industry to group the users who are found of listening music.

Therefore, the thing that is important to know and model is the music tastes of the listeners. Presently, there exist various music recommendation approaches like on the basis of users' listening behavior and his ratings given to the music items in the past, collaborative approach is performing well. Content-based music recommendation approach is also there in which low level acoustic features like pitch and high-level features like genre of the song, instruments played etc., are considered for the process of recommendation. Also, such websites also provides information that can be used by the researchers.

In¹ proposed a technique for Music recommendation system. This system works using text analysis on various song requests. In this work, user gives the input in the textual form to the system. In order to evaluate the system; user evaluation test using conventional metrics has been performed. Results indicates that there is some connection between the textual input similarity and song

*Author for correspondence

similarity, showing the capability of using text as main source to recommendation system.

Author² in his paper described about music preference structure base model to present the content based music recommendation system. This system is depends on set of attributes which is derived from different music preferences. The result indicates that the recommendation system is effective in music preference estimation.

Author³ in his paper described a new approach that employed data gathered from mobile for fulfilling users' temporary music needs. Also, probabilistic model has been presented in order to integrate the contextual information to provide music recommendation for regular activities. The performance results show that model has good accuracy.

In⁴ proposed recommending system based on the collaborative filtering for twitter user. In other words, recommender system based on neighborhood which suggested URLs to twitter users has been presented. In this work, tweets, hash tags, consider as main component of the system. The results show the precision of URL recommendation in Twitter.

Author in⁵ presented a personalized e-Learning Material Recommendation System. E-learning environments are relying on interactive services. It provides the e-learning material to the students. The main objective of this paper is to help the students to fined learning material. Furthermore, two new approaches have also been proposed. One is a multi-attribute evaluation method and other is fuzzy matching method.

Author in⁶ proposed an Incrementally Trainable Probabilistic Generative Model. This model is based on An Efficient Hybrid Music Recommender System. The performance of the system indicates that the system is very precise and provides the high degree of accuracy.

Author⁷ in this paper, latent factor model for recommendation is proposed. This proposed approach is compared with other traditional approach in this work, when usage data cannot attained then latent factor from music audio is predicted and this prediction gives sensible recommendations.

Author in⁸ presented Auralist recommendation framework, Moreover, Auralist is evaluated over a metrics set this Auralist recommendation system is for the user to study on music recommendation,

Author⁹ in this work, Singular value decomposition has been presented. Singular value decomposition likes things appear more like and unlike thing more unlike.

This can only be described by the help of vector in the reduced version of U and these vectors are short.

Author in¹⁰ gives the review of various algorithms of power recommender systems. Also social and monetary relevance of recommender system has been presented.

Author¹¹ in this paper, Carpooling Recommendation System for Taxicab Services has been presented. It is based on a data-driven approach. Trip distribution for efficient data processing and reciprocal price reduce mechanism.

Author in¹² has explained various music recommendation models. User modeling, item profiling, and match algorithms are the three main elements of the system has discussed.

Author¹³ in this paper therefore reviews advanced application developments of recommender systems, some important topics are found and listed as new directions. This review will help the researcher in understanding the development of application in recommender system.

2. Proposed Methodology

We propose a hybrid *K-Means-p-LSA* technique for music recommender. Here we will find the similar set of users with the K-Means cluster procedure on the context of the songs listened by all the users. After determining the set of similar filtering the relatively close set of users to the current user among all the music listeners from the database, system will find the most appropriate songs on the basis of content of the songs among all the songs listened by the similar users. The songs listened by the users will be categorized proportionally after which the most set of songs listened by other users will be recommended to the current user.

Generally, we have n data points, $x_i, i = 1 \dots n$ which is divided into clusters. The purpose is the assignment of a cluster to every data point. *K-means* is a clustering process with the intention of finding the positions $\mu_i, i = 1 \dots k$ of the clusters that decrease the square of the distance from the data points to the cluster. *K-means* clustering solve

$$\arg \min_c \sum \sum d(x_i - \mu_i)^2 = \arg \min_c \sum \sum \|x_i - \mu_i\|^2$$

where c_i is the all the points that corresponds to cluster i. It uses the Euclidean distance formula $d(x, \mu_i) = \|x_i - \mu_i\|^2$. As the problem comes under NP-hard (non-trivial), the K-means process simply finds the total minimum, sometimes get trapped in a diverse solution.

2.1 K-Means Process Working

The k-means process (The Lloyd’s algorithm) solves the k-means cluster problem. It works as follow. First, choose how many clusters we want to construct let’s say k after that:

1. Initialize the center of the clusters $\mu_i = \text{some value}, i = 1 \dots k$
2. Attribute the closest cluster for each data point $c_i = \{j : d(x_j, \mu^i) \leq d(x_j, \mu^l), l \neq i, j = 1, \dots, n\}$
3. Place the location of each cluster to the average of all data points belonging to concerend cluster $\mu^i = 1/|c_i| \sum_{j \in c_i} x_j, \forall i$
4. Reiterate steps 2-3 until convergence. $|c| = \text{number of elements in } c$

2.2 Probabilistic Latent Search Algorithm

PLSA is a statistic’s procedure to evaluating the two-mode and co-occurrence data. In Probabilistic latent semantic, one can get a low-dimensional illustration of the variables from which Probabilistic latent semantic analysis is involved. In comparison with typical latent semantic analysis which originates from linear algebra and step down the occurrence tables (generally by means of SVD), probabilistic latent semantic analysis depends on a combination break down resulting from a latent class model. Taking observations as co-occurrences of terms and papers, PLSA models the likelihood of each co-occurrence as a combination of provisionally independent multinomial distributions:

$$P(t, p) = \sum_c P(c)P(p|c)P(t|c) = P(p) \sum_c P(c|p)P(t|c)$$

Being the terms’ topic. The first formulation which is said to be symmetric formulation, where and are produced from the latent class in similar manners (using $P(p|c)$ and $P(t|c)$), whereas the second formulation is said to be the asymmetric formulation, in which, for each paper, a latent class is chosen provisionally to the paper corresponding to $P(c|p)$, and a word is then generated from that class corresponding to $P(t|c)$. Although we have used terms and papers in this example, the co-occurrence of any couple of discrete variables may be modeled in exactly the same way.

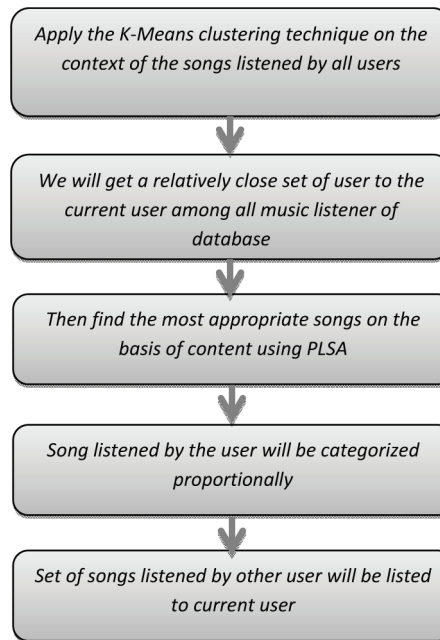


Figure 1. Flow chart of proposed methodology.

So, the total parameters are, and total parameters expanding linearly with the total documents.

3. Future Scope

Finding patterns and similarities of users in the field of recommender system is becoming the major part so as to recommend the users accurately. In the field of music recommender it become quite challenging to get the reference or meaning out of some song or music. For that an hybrid approach has been proposed to achieve the higher accuracy for music recommender. This technique is quite useful in other recommender system where the content of the topic is available so that some meaning can be extracted from it and used to recommend users. We will implement our proposed methodology to find out the results.

4. References

1. Hyung Z, Lee K, Lee K. Music recommendation using text analysis on song requests to radio stations. Expert Systems with Applications. 2014; 41(5):2608–18.
2. Soleymani M, Aljanaki A, Wiering F, Veltkamp RC. Content-based music recommendation using underlying music preference structure. In IEEE International Conference on Multimedia and Expo (ICME); Turin. 2015. p. 1–6.

3. Wang X, Rosenblum D, Wang Y. Context-aware mobile music recommendation for daily activities. In Proceedings of the 20th ACM International Conference on Multimedia; 2012. p. 99–108.
4. Yazdanfar N, Thomo A. Link recommender: Collaborative-filtering for recommending URLs to Twitter users. *Procedia Computer Science*. 2013; 19:412–9.
5. Lu J. Personalized e-learning material recommender system. *International Conference on Information Technology for Application*; 2004. p. 374–9.
6. Yoshii K, Goto M, Komatani K, Ogata T, Okuno HG. An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model. *IEEE Transactions on Audio, Speech, and Language Processing*. 2008; 16(2):435–47.
7. Van den Oord A, Dieleman S, Schrauwen B. Deep content-based music recommendation. In *Advances in Neural Information Processing Systems*. 2013; 2643–51.
8. Zhang YC, Seaghdha DO, Quercia D, Jambor T. Auralist: Introducing serendipity into music recommendation. In *Proceedings of the 5th ACM International Conference on Web Search and Data Mining*; 2012. p. 13–22.
9. Baker K. Singular value decomposition tutorial. The Ohio State University; 2005. p. 24. Desheng Zhang et. al. Carpooling Recommendation System for Taxicab Services. *IEEE Transactions on* , pp. 2014.
10. Huttner J. From tapestry to SVD: A survey of the algorithms that power recommender systems [BA thesis]. Haverford College; 2009.
11. Zhang D, He T, Liu Y, Lin S, Stankovic JA. A carpooling recommendation system for taxicab services. *IEEE Transactions on Emerging Topics in Computing*. 2014; 2(3):254–66.
12. Song Y, Dixon S, Pearce M. A survey of music recommendation systems and future perspectives. In *9th International Symposium on Computer Music Modeling and Retrieval*; 2012.
13. Lu J, Wu D, Mao M, Wang W, Zhang G. Recommender system application developments: A survey. *Decision Support Systems*. 2015; 74:12–32.