

# Data Mining for Music Recommendation

Yue Guo  
Indiana University  
107 S. Indiana Avenue  
Bloomington, IN 47405-7000  
yueguo@iu.edu

XXXX  
Indiana University  
107 S. Indiana Avenue  
Bloomington, IN 47405-7000  
XXXXXX@iu.edu

## ABSTRACT

This paper provides a sample of a  $\LaTeX$  document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings.

## KEYWORDS

music recommendation, use up to 5 keywords

## 1 OBJECTS AND SIGNIFICANCE

Our goal is to build a better music recommendation system, “the dataset is from KKBOX, Asia’s leading music streaming service, holding the world’s most comprehensive Asia-Pop music library with over 30 million tracks.”[? ]. Although KKBOX currently has their music recommendation system, we try to find new techniques to lead better results.

Nowadays, large amount of companies have their own recommendation system. For example, when you buy things on Amazon or search some words on Google or follow somebody on Facebook, they will recommend you what to eat, what to listen, what to buy, what to watch and whom you can connect with. Data mining skills can be used to those recommendation systems. Although, there are lots of skills to build recommendation systems now, it is hard to create new tools with new skills, especially with large data. Having a good recommendation systems is very important to a company, since their customer will feel the company really “understand” them and know exactly what they like and need, then customers will show much loyalty to this company.

## 2 BACKGROUND

KKbox “currently use a collaborative filtering based algorithm with matrix factorization and word embedding in their recommendation system”[? ]. The goal of collaborative filtering algorithm is to predict and recommend new music to users with the data from user former preferences. And In a collaborative filtering based algorithm, there is a list of users and a list of items. Each users shows different preferences to each items, and it is possible that a user has very few records and even has no preferences. There are two kinds of Collaborative Filtering algorithms, one is Memory-based and the other is Model-based .[? ]

Memory-based Collaborative Filtering Algorithms is to find a group of users as neighbors, which all of them have similar taste of music. Since we believe personal tastes are correlated, we can produce top-N recommendations for users with the knowledge of their neighbors’ preferences. This techniques, also called as nearest-neighbor or used-based collaborative filtering.

Model-based Collaborative Filtering Algorithms is to calculate the expected value of a vote, given what we know about the user.[? ]

And we can use different kinds of machine learning models for collaborative filtering, like clustering models and Bayesian network. In Cluster Models, the idea is users can be grouped as a class and they have common preferences and tastes. And given a class, the preferences for different items are independent. Bayesian network formulates a probabilistic model with a node corresponding to each item in the domain.

However, there are two main challenge for collaborative filtering recommender systems.[? ] The first is the scalability. Nearest-neighbor algorithms need computation that grows with the scale of users and items at the same time. With millions of users and items, the recommendation system will have scalability problem. The second challenge is accuracy. And in some ways, those two challenges are in conflict, if we reduce the time to compute neighbors and system can be more scalable, but the accuracy may become lower.

## 3 PORPOSE APPROACH

## 4 INDIVIDUAL TASKS

Yue Guo and Yinghua will figure out how to build a better recommendation system together. We both need to research and identify different algorithms and skills to apply on this project, and imply those independently.

## 5 REFERENCES

## ACKNOWLEDGMENTS

The authors would like to thank Yue Guo for his support and suggestions to write this paper.