**Music Recommender system**

Perform Exploratory Data Analysis and Build a Music Recommend system using Python.

The Dataset for building this music recommend system is provided by KKBOX.COM (KKbox is one of the famous website for music, its available on both mobile and computer). It has more than 10 million users worldwide.

The dataset consist of information of 1st observable listening event or each unique user song pair with in it specific time duration.

The metadata of each unique song and user pair is also provided training set and test set are selected from the users listening history in the given time period.

The 3 main important files for this project

1. Train.csv
2. Songs.csv
3. Members file

Train.csv🡪Here in train.csv we have different column like source system tab, song id, source screen name source type

Test.csv is uses for submission of prediction.

Songs.csv🡪song\_id etc….

KKbox’s Music Recommendation Challenge Solution with

Feature engineering

Jianyu Zhang Françoise Fogelman-Soulié

School of Computer Software

Tianjin University

(edzhang, soulie) @tju.edu.cn

ABSTRACT

Recommendation is widely used in our daily life. Especially in

the e-commerce area, a good recommendation system can help

users a lot. In this paper, we introduce our approach for the

KKbox’s Music Recommendation Challenge. In this challenge,

we were asked to build a recommendation system that can

predict whether a user will listen again to a song within one

month after the user’s very first observable listening event in

KKbox. Our solution was mostly based upon systematic and

extensive feature engineering and an ensemble of simple

boosting tree classification algorithms, both of which could

easily be used in industry. However, we did not use timestamp

of user-song interactions here, since this was hidden by Kaggle

to avoid leakage.

KEYWORDS

Feature engineering, recommender systems, gradient boosting

tree, SVD.

1 INTRODUCTION

In the WSDM Cup KKBox’s Music Recommendation

Challenge [6], we had to build a recommendation system that

can predict whether a user will listen to a song again within one

month after the user’s very first observable listening event in

KKbox. If the user did not listen to the song again within one

month, the target variable will be 0, and 1 otherwise. The

training and test sets consist in unique user/song pairs selected

from users’ listening history in a certain period, split by time.

The test set is split 50/50 between public and private

leaderboards; obviously, targets are unknown in the

leaderboards test set, but the data distribution is, and we can

thus use it as we will show later. At the time of the competition,

submissions were evaluated on the public leaderboard, while

KKbox’s Music Recommendation Challenge Solution with

Feature engineering

Jianyu Zhang Françoise Fogelman-Soulié

School of Computer Software

Tianjin University

(edzhang, soulie) @tju.edu.cn

ABSTRACT

Recommendation is widely used in our daily life. Especially in

the e-commerce area, a good recommendation system can help

users a lot. In this paper, we introduce our approach for the

KKbox’s Music Recommendation Challenge. In this challenge,

we were asked to build a recommendation system that can

predict whether a user will listen again to a song within one

month after the user’s very first observable listening event in

KKbox. Our solution was mostly based upon systematic and

extensive feature engineering and an ensemble of simple

boosting tree classification algorithms, both of which could

easily be used in industry. However, we did not use timestamp

of user-song interactions here, since this was hidden by Kaggle

to avoid leakage.

KEYWORDS

Feature engineering, recommender systems, gradient boosting

tree, SVD.

1 INTRODUCTION

In the WSDM Cup KKBox’s Music Recommendation

Challenge [6], we had to build a recommendation system that

can predict whether a user will listen to a song again within one

month after the user’s very first observable listening event in

KKbox. If the user did not listen to the song again within one

month, the target variable will be 0, and 1 otherwise. The

training and test sets consist in unique user/song pairs selected

from users’ listening history in a certain period, split by time.

The test set is split 50/50 between public and private

leaderboards; obviously, targets are unknown in the

leaderboards test set, but the data distribution is, and we can

thus use it as we will show later. At the time of the competition,

submissions were evaluated on the public leaderboard, while

KKbox’s Music Recommendation Challenge Solution with

Feature engineering

Jianyu Zhang Françoise Fogelman-Soulié

School of Computer Software

Tianjin University

(edzhang, soulie) @tju.edu.cn

ABSTRACT

Recommendation is widely used in our daily life. Especially in

the e-commerce area, a good recommendation system can help

users a lot. In this paper, we introduce our approach for the

KKbox’s Music Recommendation Challenge. In this challenge,

we were asked to build a recommendation system that can

predict whether a user will listen again to a song within one

month after the user’s very first observable listening event in

KKbox. Our solution was mostly based upon systematic and

extensive feature engineering and an ensemble of simple

boosting tree classification algorithms, both of which could

easily be used in industry. However, we did not use timestamp

of user-song interactions here, since this was hidden by Kaggle

to avoid leakage.

KEYWORDS

Feature engineering, recommender systems, gradient boosting

tree, SVD.

1 INTRODUCTION

In the WSDM Cup KKBox’s Music Recommendation

Challenge [6], we had to build a recommendation system that

can predict whether a user will listen to a song again within one

month after the user’s very first observable listening event in

KKbox. If the user did not listen to the song again within one

month, the target variable will be 0, and 1 otherwise. The

training and test sets consist in unique user/song pairs selected

from users’ listening history in a certain period, split by time.

The test set is split 50/50 between public and private

leaderboards; obviously, targets are unknown in the

leaderboards test set, but the data distribution is, and we can

thus use it as we will show later. At the time of the competition,

submissions were evaluated on the public leaderboard, while

KKbox’s Music Recommendation Challenge Solution with

Feature engineering

Jianyu Zhang Françoise Fogelman-Soulié

School of Computer Software

Tianjin University

(edzhang, soulie) @tju.edu.cn

ABSTRACT

Recommendation is widely used in our daily life. Especially in

the e-commerce area, a good recommendation system can help

users a lot. In this paper, we introduce our approach for the

KKbox’s Music Recommendation Challenge. In this challenge,

we were asked to build a recommendation system that can

predict whether a user will listen again to a song within one

month after the user’s very first observable listening event in

KKbox. Our solution was mostly based upon systematic and

extensive feature engineering and an ensemble of simple

boosting tree classification algorithms, both of which could

easily be used in industry. However, we did not use timestamp

of user-song interactions here, since this was hidden by Kaggle

to avoid leakage.

KEYWORDS

Feature engineering, recommender systems, gradient boosting

tree, SVD.

1 INTRODUCTION

In the WSDM Cup KKBox’s Music Recommendation

Challenge [6], we had to build a recommendation system that

can predict whether a user will listen to a song again within one

month after the user’s very first observable listening event in

KKbox. If the user did not listen to the song again within one

month, the target variable will be 0, and 1 otherwise. The

training and test sets consist in unique user/song pairs selected

from users’ listening history in a certain period, split by time.

The test set is split 50/50 between public and private

leaderboards; obviously, targets are unknown in the

leaderboards test set, but the data distribution is, and we can

thus use it as we will show later. At the time of the competition,

submissions were evaluated on the public leaderboard, while

KKbox’s Music Recommendation Challenge Solution with

Feature engineering

Jianyu Zhang Françoise Fogelman-Soulié

School of Computer Software

Tianjin University

(edzhang, soulie) @tju.edu.cn

ABSTRACT

Recommendation is widely used in our daily life. Especially in

the e-commerce area, a good recommendation system can help

users a lot. In this paper, we introduce our approach for the

KKbox’s Music Recommendation Challenge. In this challenge,

we were asked to build a recommendation system that can

predict whether a user will listen again to a song within one

month after the user’s very first observable listening event in

KKbox. Our solution was mostly based upon systematic and

extensive feature engineering and an ensemble of simple

boosting tree classification algorithms, both of which could

easily be used in industry. However, we did not use timestamp

of user-song interactions here, since this was hidden by Kaggle

to avoid leakage.

KEYWORDS

Feature engineering, recommender systems, gradient boosting

tree, SVD.

1 INTRODUCTION

In the WSDM Cup KKBox’s Music Recommendation

Challenge [6], we had to build a recommendation system that

can predict whether a user will listen to a song again within one

month after the user’s very first observable listening event in

KKbox. If the user did not listen to the song again within one

month, the target variable will be 0, and 1 otherwise. The

training and test sets consist in unique user/song pairs selected

from users’ listening history in a certain period, split by time.

The test set is split 50/50 between public and private

leaderboards; obviously, targets are unknown in the

leaderboards test set, but the data distribution is, and we can

thus use it as we will show later. At the time of the competition,

submissions were evaluated on the public leaderboard, while