Movie Genre Prediction using Metadata and Multimedia Content

Anonymous

1 Introduction

Nowadays, with the development of information technology, the digital movie becomes more and more popular in our daily life. Everyone has their own preference about movie genres, for instance, some people are fond of horror movies, while some people are scared of it. Before watching a movie, people can obtain various information about it through different types of perception, in particular, the textual, audio and visual channels. The textual metadata includes a movie's title, year of release, as well as the human-provided tags. Also, a movie can be described by a trailer through audio and visual channels. A lot of information about a movie can be extracted from these data and make good use of it. The problem (hypothesis) investigated in this research is to compare the performance of different machine learning models trained with hybrid combinations of features in predicting movie genre and propose the classifier with the best performance to predict the movie genre.

The remaining parts of this paper are structured as follow. Section 2 describes a review of previous works and the methods used in this research. Section 3 presents the processes of the experiments for this research. The experimental results and error analysis are discussed in Section 4. Section 5 concludes the paper.

2 Related work

2.1 Previous Work

In the domain of movie rating, the most widely used dataset is MovieLens (ML) (Harper et al., 2015). There were four ML datasets released which are 100k, 1m, 10m and 20m, each of them consisted of attributes including user, rating, timestamp and item. In ML 10m and 20m, tag information and genre were introduced to the dataset.

Based on ML 10m and 20m, Multifaceted Movie Trailer Feature (MMTF-14K) dataset (Deldjoo et al., 2018), the most powerful dataset for the movie genre classification tasks was proposed in 2018. MMTF-14K comes with each movie's metadata features, audio feature, visual feature and the corresponding genre. Metadata feature includes movie id, title, year of release, tags and YouTube link of trailer. Audio and visual feature are precomputed from the trailers into state-of-the-art descriptors with the vectorized format. Genres contain 18 labels such as comedy, horror, romance and so on.

Besides, a lot of excellent works have been done in the context of movie recommending with the use of above datasets. The MediaEval (Deldjoo et al, 2018) provided different approaches to predict movie ratings through the usage of hybrid combinations of metadata, movie clips, audio and visual feature. CNN-MoTion (Gabriel et al., 2016) took deep learning approaches to deal with movie genre classification problem with image, audio, and video recognition. The results of above studies have a significant inspiration for this paper.

2.2 Methods

2.2.1 TF-IDF

Term Frequency–Inverse Document Frequency (TF-IDF) is a numeric statistic technology that evaluates the relevance of a word in a corpus and it vectorizes the textual features. The TF-IDF value is acquired by multiplying the metric of the number of times a word appears in a corpus and the metric of the inverse document frequency of the word across a corpus. Even though some words such as "and", "you" and "this" have a high frequency in the English language, their TF-IDF values stay low because they are not that important in that corpus. The output of a corpus converted by TF-IDF is a sparse matrix stores the TF-IDF value for every single word.

2.2.2 Zero-R

Zero-R is served as a baseline approach in this research to perform initial baseline experiments. Zero-R is the most commonly used baseline method in the machine learning area. It classifies the instances in the validation set according to the most common class in the training set.

2.2.3 Random Forest

Random Forest (RF) is an ensemble machine learning model which contains multitudes of the Decision Tree. To train an RF classifier, random samples are selected repeatedly by bootstrap aggregating with replacement and fits into trees. Each individual decision tree leads to a label, the label which is the mode of labels is the output classification label. RF can deal with both discrete values using the ID3 algorithm and continuous values using the C4.5 algorithm. The major advantage of RF is that the relatively uncorrelated trees protect each other from their individual errors and make sure the tree grows towards the correct direction. Compare to other machine learning models, in particular, Decision Tree, RF avoids overfitting the training set. In other words, the more training data, the better the performance RF can provide.

2.2.3 Logistic Regression

In this research, a multinomial Logistic Regression (LR) is adopted since the movie genre classification task is a multiclass problem. LR is a regression model that uses a linear predictor function to construct a score from the feature values and corresponding weight factors given by softmax function: $p(y = c|x; \Theta) = \frac{exp(\Theta_c x)}{\sum_k exp(\Theta_k x)}$ where c is the predicted label, x is a set of feature values, Θ is the set of weight factors. Multinomial LR predicts the label with the highest score as an outcome. LR is a more informative classifier since it provides not only the relevance of a predictor but also the direction of association. Besides, LR is suitable for frequency-based features, that is, natural language features such as tag and title features in this research.

2.2.4 Multilayer Perceptron

Multilayer Perceptron (MLP) is a feedforward neural network composed of an input layer perceptron, an output layer perceptron and at least one hidden layer perceptron. Expect the input layer, each layer is a neuron using the non-linear activation function (sigmoid function): $f(x) = \frac{1}{1+e^{-x}}$. Input layer takes feature values as inputs and passes them to the first hidden layer. Then, the next hidden layer takes the values from the last hidden layer as inputs and keeps the above processes until it reaches the output layer and generates an outcome prediction label. In the training phase, there is a set of initial values of weights in every layer and backpropagation is adopted to modify each weight by applying gradient descent. MLP is a powerful classifier for the data which is not linearly separable, in particular, audio and visual feature in this research.

3 Experiments

3.1 Feature Engineering

In feature engineering, we need to choose the relevant features that have positive impacts on movie genre prediction, make sure there is no missing value in every row of the dataset and vectorize some features into a format that is suitable for fitting models. Firstly, we drop the irrelevant features, which are year of release and YouTube link. There are four remaining features, title, tag, audio and visual vector. Secondly, we drop the rows which have missing values. There are three rows which their titles are null. The three entire rows are dropped for keeping data integrity. Lastly, since the textual feature cannot be used directly as an input to fit machine learning models, TF-IDF is adopted to vectorize the tag and title features.

3.2 Data Splitting

The dataset, which composed of movies' metadata, audio and visual feature as well as the corresponding genre, is randomly split into a training set (5237 movies), a validation set (299 movies) and a test set (298 movies) using holdout strategy. The training set is used to fit models, the validation and test set are used to evaluate the models.

3.3 Classification

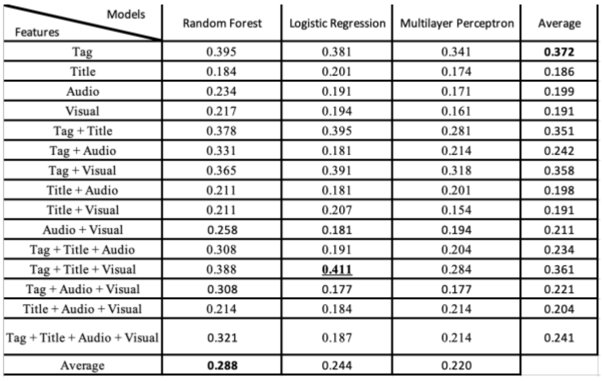
The movie genre prediction problem is treated as a classification in this research. After pre-processing the dataset, each movie has four features which are title, tag, audio and visual, as well as a label which is the movie genre. The dataset is split using the holdout strategy. Hybrid combinations of these four features (e.g. tag, tag + title, or tag + audio + visual) are used as inputs to train the models. The output is a movie genre predicted based on the above input(s). One baseline used in this research is Zero-R and there are three high-level machine learning classification models are adopted: Random Forest classifier with 100 trees in the forest, Logistic Regression classifier with l2 penalty and 0.01 tolerance for stopping criteria, and Multilayer Perceptron classifier with 100 hidden layers.

4 Results and Analysis

4.1 Experimental Results

The baseline method adopted in this research is Zero-R regardless of which feature(s) are involved. The most common label is 'Romance' with 791 movies support in the training set. We acquired a 0.171 accuracy by predicting all the instances in the validation set as romance.

The initial baseline experiment is not satisfactory, therefore, we decide to further investigate the performance of high-level models trained with various combinations of features.



Models / Features	Random Forest	Logistic Regression	Multilayer Perceptron	Average
Tag	0.395	0.381	0.341	**0.372**
Title	0.184	0.201	0.174	0.186
Audio	0.234	0.191	0.171	0.199
Visual	0.217	0.194	0.161	0.191
Tag + Title	0.378	0.395	0.281	0.351
Tag + Audio	0.331	0.181	0.214	0.242
Tag + Visual	0.365	0.391	0.318	0.358
Title + Audio	0.211	0.181	0.201	0.198
Title + Visual	0.211	0.207	0.154	0.191
Audio + Visual	0.258	0.181	0.194	0.211
Tag + Title + Audio	0.308	0.191	0.204	0.234
Tag + Title + Visual	0.388	**<u>0.411</u>**	0.284	0.361
Tag + Audio + Visual	0.308	0.177	0.177	0.221
Title + Audio + Visual	0.214	0.184	0.214	0.204
Tag + Title + Audio + Visual	0.321	0.187	0.214	0.241
Average	**0.288**	0.244	0.220	

Table 1 – Accuracy score of RF, LR and MLP classifier trained with various combination of features.

Table 1 shows the accuracy score evaluated by three different machine learning models on the validation set trained with hybrid combinations of features. First of all, from the perspective of the comparison of learning models, RF is the best predictor while MLP is the worst. Secondly, regarding every single feature, tag feature has the best performance while the title feature has the worst performance. For the hybrid combination of features, a combination of tag, title and visual feature outperforms. Last but not least, the highest accuracy score on the validation set is using LR with tag, title and visual feature, which achieves 0.411 and highlighted in bold and underlined in Table 1.

As a result, LR classifier trained with tag, title and visual feature is selected to predict the movie genre on the test set and analyse the errors.

4.2 Error Analysis

The fine-grained performance is analysed by observing the three classifiers' learning curve on the training set and the confusion matrix on the validation set.

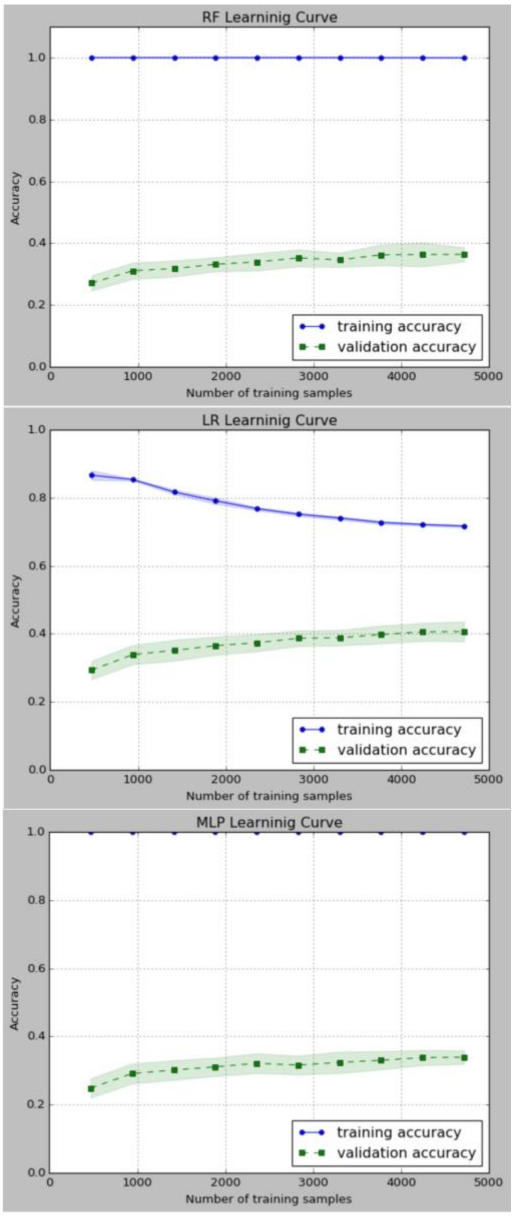


Figure 1 – The learning curve of RF, LR and MLP trained with tag, title, and visual feature

The learning curves are generated with 10-fold cross-validation and the classifiers are trained with tag, title, and visual feature using training set. Figure 1 demonstrates that RF and MLP classifier are overfitting on this training set with high variance and low bias, while LR seems fits better but the accuracy is low. There are some remedies to improve the fitting performance of LR and MLP classifier. Firstly, we can add more training data by adopting the Bagging strategy. Bagging constructs new datasets by randomly selecting the training data with replacement and it can reduce variance. Secondly, some features with low irrelevance can be dropped to in the training set. Finally, the model complexity can be reduced since the complex models are prone to high variance.

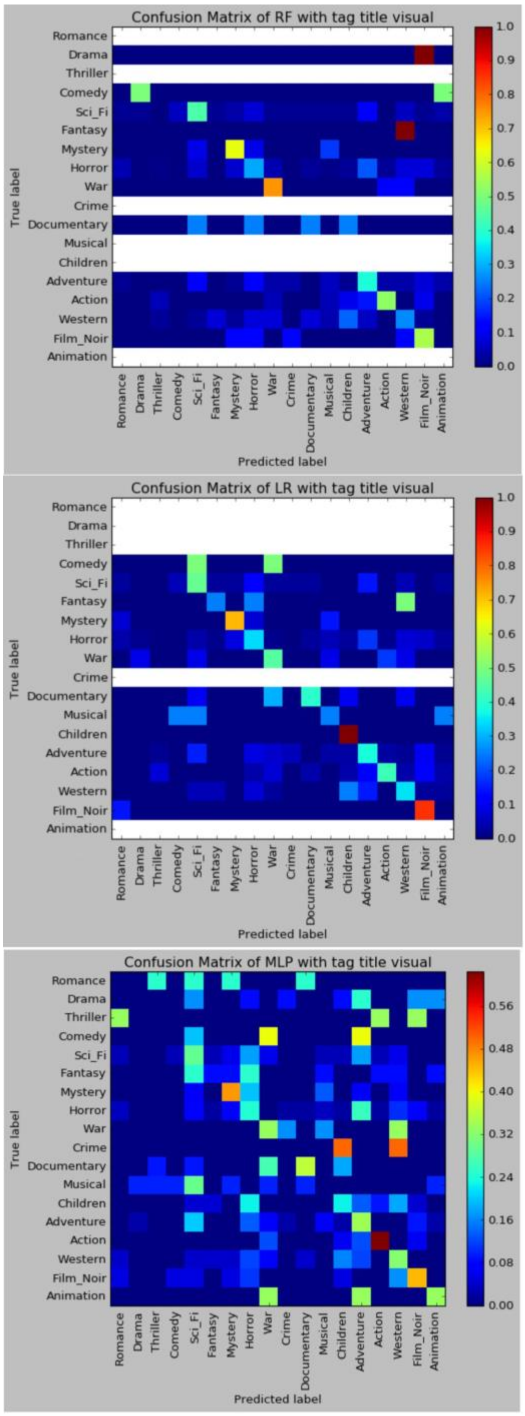


Figure 2 – Confusion Matrix of RF, LR and MLP classifier trained with tag, title, and visual feature and evaluated on validation set

Figure 2 shows the Confusion Matrix of RF, LR and MLP trained with tag, title, and visual feature and evaluated on the validation set. From the confusion matrix of RF, there is a high chance that RF classifier wrongly predicts the genre drama and fantasy as film noir and western respectively. Besides, the RF classifier doesn't make any correct prediction to romance, thriller, crime, musical, children and animation genre. The same errors also appear in LR classifier. The LR classifier is not available to predict romance, drama, thriller, crime and animation genre and mistakenly predicts comedy as war and science fiction. MLP classifier outperforms in predicting action and mystery genre while doesn't perform well in predicting other genres, especially, it predicts romance as thriller, science fiction, mystery and documentary.

5 Conclusions

In this research, we compared the performance of learning models RF, LR and MLP trained with hybrid combinations of features in predicting movie genre. Based on the experimental results, we proposed a movie genre classifier trained with tag, title and visual feature using logistic regression. The learning curves and confusion matrixes are used to evaluate the model and analyse the errors for every specific genre.

References

F. M. Harper, J. A. Konstan, The MovieLens Dataset: History and Context. In *ACM Transactions on Interactive Intelligent Systems (TiiS), Vol. 5, No. 4, Article 19. New York, NY, USA, 2015*

Y. Deldjoo, M. G. Constantin, B. Ionescu, M. Schedl, and P. Cremonesi. MMTF-14K: A Multifaceted Movie Trailer Dataset for Recommendation and Retrieval. *In Proceedings of the 9th ACM Multimedia Systems Conference (MMSys 2018), pages 450-455. Amsterdam, the Netherlands, 2018*

Y. Deldjoo, M. G. Constantin, H. Eghbal-Zadeh, M. Schedl, B. Ionescu, and P. Cremonesi. Audio-Visual Encoding of Multimedia Content to Enhance Movie Recommendations. In *Proceedings of the Twelfth ACM Conference on Recommender Systems, ACM, pages 455-459, Vancouver, BC, Canada, 2018.*

G. S. Simões, J. Wehrmann, R. C. Barros, D. D. Ruiz. Movie genre classification with Convolutional Neural Networks. In *2016 International Joint Conference on Neural Networks (IJCNN), pages 259-266, Vancouver, BC, Canada, 2016.*