Predicting Movie Ratings with Multimodal Data

Yichen Yang Ruoyun Ma Min Haeng Cho
yyang9@stanford.edu ruoyunma@stanford.edu minhaeng@stanford.edu

Abstract

- **Motivation** For customers: predict movie ratings before theatrical release
  For producers: offer insight on the determining factors
- **Experiments** Linear regression, ridge regression, decision tree, random forest, support vector regression, neural network
  Feature importance, classification
- **Results** Best model: random forest, $R^2 = 0.4253$ (test set)
  Most important feature: movie genre

Data

- Open-source data from Kaggle (CSV file with textual information about movies & JPG files of posters)
- For colored posters, transformed pixel dimensions from 900x600 to 224x224
- Filtered out movies released before 1980 and non-English movies
- In total, 19429 movies in sample with train-validation-test split at 70%-15%-15% ratio (13600 training, 2914 validation and 2915 test data points)

Features

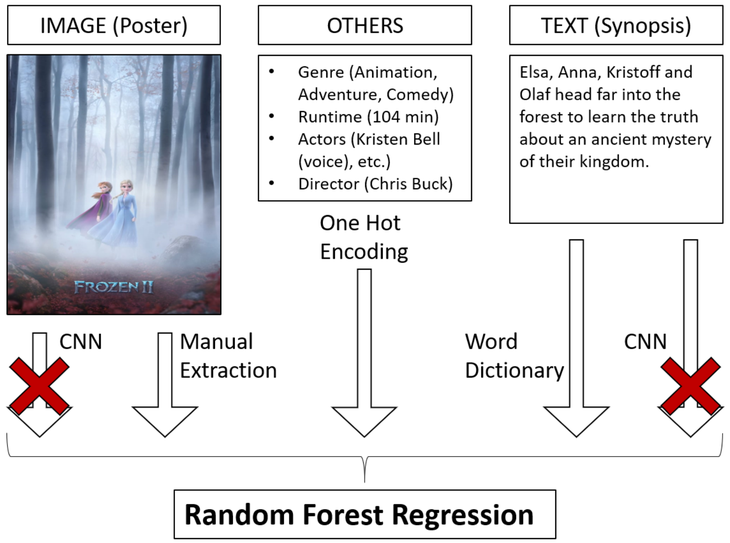
- Input feature categories: images (posters), text (synopses), others (cast, crew, runtime and genre)
- For posters, manually extracted 13 visual features including number of human faces and means and standard deviations of RGB and HSB
- For synopses, tokenized and kept words that appeared in at least 20 movies
- For cast and crew, extracted main director and top three leading actors for each movie; kept those involved in at least 5 movies in dataset
- These features are appropriate because before a movie is released, most people decide whether to watch the film based on these factors.

Data	Type	Dimension	Example
Genre	Categorical	23	Action
Runtime	Numerical	1	100 (minutes)
Actors	Categorical	1603	Robert Downey Jr.
Director	Categorical	492	Steven Spielberg
Poster	Numerical	13	Number of faces = 1
Synopses	Categorical	3884	“innocence”

Models

- Linear regression: $\theta^* = \arg\min_\theta \frac{1}{n}\sum_{i=1}^{n}(\theta^T x^{(i)} - y^{(i)})^2$
- Ridge regression: $\theta^* = \arg\min_\theta \frac{1}{n}\sum_{i=1}^{n}(\theta^T x^{(i)} - y^{(i)})^2 + \lambda||\theta||^2$
- Support vector regression: $\min \frac{1}{2}||\theta||^2$ s.t. $y^{(i)} - \theta^T x^{(i)} - b \leq \epsilon$ and $\theta^T x^{(i)} + b - y^{(i)} \leq \epsilon$
- Neural networks: CNN for synopses + non-textual features $\longrightarrow$ 5 ReLU layers
- Decision trees and random forest: our best model

Random forest is an ensemble learning method that aggregates outputs from a multitude of decision trees and is used for non-linear problems due to strong stability and overfitting reduction.



Results

Method	Train MSE	Train R^2	Test MSE	Test R^2
Linear Regression	0.7003	0.5314	0.9302	0.3745
Ridge Regression	0.7710	0.4842	0.8775	0.4099
Decision Tree Regression	0.4460	0.8281	0.8959	0.3975
Random Forest Regression	0.8819	0.1764	0.8546	0.4253
Support Vector Regression	0.5816	0.6253	0.8542	0.4256
Neural Network	0.4200	0.7290	0.8765	0.4109

- The models were trained on 13600 training samples, while hyperparameters were selected using 2914 validation samples and evaluated on 2915 test samples.
- Considering accuracy, efficiency, interpretability, random forest was best model.
- The optimal maximum depth was 32 and the optimal number of trees was 100.

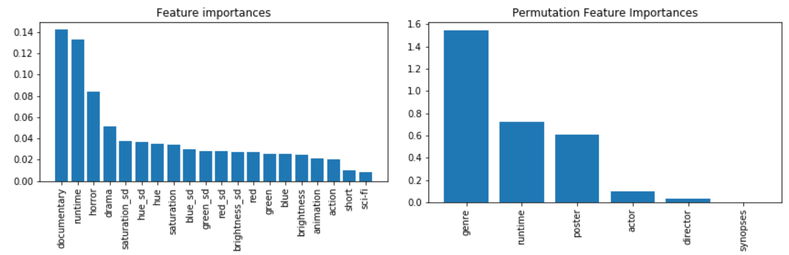
Discussion

Comparison with Previous Work:

- Compared to model that used movie trailers and genre to predict film ratings with MSE of 0.88 [1], we achieved smaller MSE as our model had more features.
- Compared to social media model which had MSE of 0.2 [2], our model performed worse because we used as features data available before theatrical release.

Feature Importance:

For both random forest and permutation feature importance (FI), genre, runtime and manually extracted visual features had the highest FI, while actors, directors and synopses had low FI.



(a) Random Forest Feature Importance **(b)** Permutation Feature Importance

Classification Perspective:

- We discretized movie ratings uniformly into 5, 10 and 20 classes, and chose a central value for each interval to represent each class.

Method	Test MSE	Test R^2
Classification (5 classes)	0.8722	0.4135
Classification (10 classes)	0.8649	0.4184
Classification (20 classes)	0.8608	0.4212
Regression	0.8546	0.4253

- The higher the number of classes, the higher the R^2 and lower the MSE on the test data. Regression outperformed classification because discretizing data renders some information missing from the original data.

Future Work

- Add new features such as movie trailers to improve prediction accuracy.
- Develop sophisticated nested model to combine data from all feature categories.

References

[1] F. B. Moghaddam, M. Elahi, R. Hosseini, C. Trattner, and M. Tkalčič. Predicting movie popularity and ratings with visual features. In *2019 14th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, pages 1–6. IEEE, 2019.

[2] A. Oghina, M. Breuss, M. Tsagkias, and M. De Rijke. Predicting imdb movie ratings using social media. In *European Conference on Information Retrieval*, pages 503–507. Springer, 2012.