

Poster for SLDS2018

Author: Yingjun Guan

Title: IMDB Review Mining and Movie Recommendation

Abstract

The world of movies contains enormous information and is worth digging for prediction and recommendation. IMDB database [1] is one good example for extracting the relevant features, including both the item-based information (only related to the movies), for example, the title, the poster, the trailer, the genre, the cast, the director, etc., and the user-related information, such as name, gender, age, occupation, review, number of watching (the same movie), etc.

Our project plans to benefit both the movie production corporations and the audience. For movie production corporations, our goals include better predicting the movie ratings given the item-based information and the user-related information. Data mining and supervised machine learning techniques are applied for labelling and classification to the nominal features. Text mining and sentimental analysis techniques are used for the review context in the database. For the audience, the hybrid recommendation system algorithm is introduced. The missing patterns [2] of the features in the dataset are researched via visualizations and analyses. And special analysis on the “cold start” phenomenon (recommendation for new items or from insufficient information) [3] on the database is also conducted.

Keywords: Recommendation System; Text Mining; Data Mining; Supervised Machine Learning

References:

- [1] *The Internet Movie Database*. IMDB.com, Inc, 2009. Web. 29 Oct. 2009. <<http://www.imdb.com>>
- [2] Bi, Xuan, et al. "A Group-Specific Recommender System." *Journal of the American Statistical Association* 112.519 (2017): 1344-1353.
- [3] Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002, August). Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 253-260). ACM.