Abstract

Movie genre classification is a challenging problem with many potential applications. Whereas many prior approaches rely on image, audio, or motion features to classify movies, we consider using textual content analysis instead, which is a comparatively less computationally expensive and time consuming process. In this paper, we present a novel system for movie genre classification that uses probabilistic topic modeling of the movie’s script as its main component. Our approach uses latent Dirichlet allocation, a topic modeling algorithm, to train our model and discover common themes present in movie scripts of the same genre. We then compute the cosine similarity of the feature vectors from our trained and test models and use this value to identify the movies’ genres.

References


**Index Terms**

Computer Science

Automated Systems

**Keywords**

Video Genre Identification, Latent Dirichlet Allocation, LDA