Automated Movie Genre Classification with LDA-based Topic Modeling

International Journal of Computer Applications
Foundation of Computer Science (FCS), NY, USA

Volume 145
Number 13

Year of Publication: 2016

Authors:
Brandon Chao, Ankit Sirmorya

10.5120/ijca2016910822
2016910822.bib

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

2. B. T. Truong, S. Venkatesh and C. Dorai, Automatic Genre Identification for
Content-Based Video Categorization, IEEE International Conference on Pattern Recognition, 2000.
3. Z. Rasheed, Y. Sheikh, and M. Shah, On the Use of Computable Features for Film
5. R. S. Jasinschi and J. Louie, Automatic TV program genre classification based on audio
7. S. Oger, M. Rouvier and G. Linares, Transcription Based Video Genre Classification,
8. M. Blosseville, G Hebrail, M. Monteil, N. Penot, Automatic Document Classification:
Natural Language Processing, Statistical Analysis, and Expert System Techniques Used
models and the nested Chinese restaurant process, Advances in Neural Information Processing
Conference on Uncertainty in Artificial Intelligence, 1999.

Index Terms

Computer Science
Automated Systems

Keywords

Video Genre Identification, Latent Dirichlet Allocation, LDA