PERSONALIZED CONTENT BASED MOVIE RECOMMENDER SYSTEM USING NATURAL LANGUAGE PROCESSING

LOCHANA BIMANTHA SAMARAWICKRAMA RANATUNGA

A dissertation submitted in partial fulfillment of the requirement for Bachelor of Science (Honours) degree in Software Engineering

Department of Computing

Informatics Institute of Technology, Sri Lanka in collaboration with University of Westminster, UK

2021

Abstract

Introduction : Modern recommendation systems are built by a combination of content based and collaborative filtering methods. Although sometimes the quality of the recommendations given is somewhat lacking. This is mainly due to two reasons, the misuse of certain movie metadata and automated systems that forgo any personalization or transparency. In this dissertation a system is built that makes up for these issues by making use of a movies most underused attribute, it's plot and combining it alongside it's other attributes as well as allowing the user to select the bias filters on which the recommendations are based on.

Method : The NLP technique of topic modeling was used to breakdown the movie plot and subject it to a similarity check. This will be done through the LDA (Latent Dirichlet allocation) statistical model alongside cosine similarity. A corpus of nearly 30,000 movies scraped from Wikipedia was used to train, evaluate and test the model. A basic GUI was also created to showcase the main functions of the system as well as it give an idea how the end user will interact with it.

Results : Through testing and evaluation it was revealed that the with big plots the results given out by this system were either on par or in some cases offered better quality recommendations than mainstream recommenders. However with smaller plots the results were inconsistent.

Discussion : This leads to the conclusion that a plot included recommendation system is not only viable but with the added bonus of it being customizable leads the user to have an overall better experience as they have more control over the recommendation process.

Keywords : Movie, Film, Plot, Synopsis, Recommendation, NLP, LDA, Python, LSA, Topic Modeling, Feature Extraction, Gensim, Wikipedia, IMDB, Movielens