

Understanding the Users Input of IMDB Movie Reviews

Xu, Zheng
xzheng@tcd.ie

YuZhou, Peng
ypeng@tcd.id

Abstract

In this digital age, people all over the world can communicate and share opinions with each other through the internet. Text mining and summarisation techniques have been widely applied to understand people's opinions online, but many of them only analyse the overall document rather than taking care of its topic. In this paper, we will focus on the field of movie reviews with considering the movie genres and sentiments. We think that users tend to use the same words to describe different opinions in different types of movies. We first mine the movie review text and get the keywords for each of the reviews according to the movie types and review sentiments. RAKE and TF algorithms will be used to extract the keywords. Finally, we evaluate these keywords and get a classification accuracy reached above 80%. Based the keywords, we then analyze the words and understand the differences in the word usage in movie reviews of different movie types.

Keywords: Opinion, Topic, Movie Review, IMDB, RAKE, TF.

1 Introduction

The Internet has significantly changed the human's experience during past few years. Many social media services like Twitter, Facebook and IMDB, are served as a medium to help people share their opinions. Users can also easily give their feedback on the products from online stores like Amazon and eBay. As per the survey, three-quarters of American including 93% of adults have been online and used the web for communication and giving the opinion online. Internet hence becomes a good source of information for the service providers to improve their services according to people's feedback, so it is of great importance to utilise the online data and detect useful information from users input. A tremendous amount of user-generated text can now be found and analysed to gain a better understanding of people's thoughts and discover the hidden knowledge behind it.

Bibliography