USING WEB SCRAPING IN SOCIAL MEDIA TO DETERMINE MARKET TRENDS WITH PRODUCT FEATURE-BASED SENTIMENT ANALYSIS

Thamarasi Nanayakkara

199480A

Thesis/Dissertation submitted in partial fulfillment of the requirements for the Degree Master of Science

Department of Computational Mathematics

University of Moratuwa

Sri Lanka

July 2022

Declaration

I declare that this is my own work and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Name of Student: Thamarasi Nanayakkara

Signature of Student: Date: 25.07.2022

The above candidate has carried out research for the Masters dissertation under my supervision.

Name of Supervisor(s): Dr. Sagara Sumathipala

Signature of Supervisor(s):   Date: 25.07.2022

Acknowledgement

Abstract

Customer product reviews are openly available online and they are now widely used for deciding quality of product or service and to determine market trends and influence decision making of users. Due to the availability of a massive number of customer reviews on the web, summarizing them requires a fast classification system. Compared to supervised and unsupervised machine learning techniques for binary classification of reviews, fuzzy logic can provide a simple and comparatively faster way to model the fuzziness existing between the sentiment polarities classes due to the uncertainty present in most of the natural languages. But the fuzzy logic techniques are not much considered in this domain. This thesis proposes a model which measures product market value by using sentiment analysis conducted on the reviews of online products which are collected from a well known ecommerce website "Amazon". Fuzzy logic approach is used in calculating the final product market demand.

Hence, in this paper we propose a fine grained classification of customer reviews into weak positive, average positive, strong positive, weak negative, average negative and strong negative classes using a fuzzy logic model based on the most popularly known sentiment based lexicon SentiWordNet. By creating rules and relationships between fuzzy membership functions and linguistic variables, we can analyze the customer opinions towards online products. This proposed model provides the most reasonable sentiment analysis because we try to reduce all the problems from the related past researches. The outcomes can allow the business organization to understand their customer's sentiments and improve customer loyalty and customer retention techniques in order to increase customer values and profits result. Fine grained classification accuracy approximately in the range of 74% to 77% has been obtained by experiments conducted on datasets of electronic products containing reviews of smart phones, TV and laptops.

Keyword- Sentiment analysis, Fine grained classification, Fuzzy Logic, SentiWordNet, Online reviews

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Description
POS	Part of Speech
NLP	Natural Language Processing