

Chapter 1 : INTRODUCTION

This chapter presents an overview of the thesis. Section 1.1 discusses motivation, section 1.2 discusses problem statement and research questions, section 1.3 discusses objectives, section 1.4 explains the hypotheses, section 1.5 explains the scope and limitation, section 1.6 discusses contribution, and section 1.7 discusses thesis overview.

1.1 MOTIVATION

Online marketing media usually provide a review feature for users to fill out reviews on products or services purchased. Mining information and analyzing sentiment from reviews can find out the market response to the product. Knowing the market response, the owner can make the right decisions to increase sales. However, many reviews are difficult to analyze manually, and to overcome that can use sentiment analysis.

Sentiment analysis is one of the techniques for social analysis, also called opinion mining, opinion analysis, or subjectivity analysis [1]. Sentiment analysis is only a classification of sentiment polarity values, namely positive or negative in a review [2]. To improve the information, it is necessary to know the context contained in the review. To get a sentiment that is oriented to a particular context, it can use the aspect-based sentiment analysis (ABSA). ABSA is a text analysis technique that categorizes aspects and classifies sentiments that are oriented towards certain aspects [3,4]. An example of ABSA for example “the food is delicious, but the price of the food is expensive” from the review it can be concluded that the aspect is “quality of food” with the sentiment being “positive” and in the second sentence the aspect is “the price of the food” with the sentiment being “negative”.

An ontology in computer science is an explicit specification of a conceptualization. Ontology can be said to represent a domain of knowledge explicitly about a concept by giving meaning, nature, and relationships to the concept, so that it is collected in the knowledge domain and forms a knowledge base [5].

There are several ABSA studies that have some differences in the classification method. Research [6] uses the maximum entropy method by using one feature or a combination of several features, while the features are Bag of N-gram, Bag of Cluster CBOW, Bag of Cluster GloVe, Bag of Head Words, Bag of k-skip-bigram, in the restaurant domain with Indonesian language, the evaluation compares using only one feature and a combination of several features, the best result for sentiment classification is using a combination of Bag of N-gram features, Bag of Cluster CBOW, Bag of GloVe with F1-measure 64.2 %. Research [7], using the Naïve Bayes Method with Chi Square in English restaurant reviews, Chi Square is used to reduce the number of features used

by Naive Bayes, the experimental results using Chi Square can speed up the process from 3 to 7 times, but the use of Chi Square reduces F1-measure value, the best F1-measure result without Chi Square is 78.12%. The SVM method [8,9,10] is used for aspect-based sentiment analysis on reviews of hotels, restaurants, laptops, cell phones, digital cameras, museums, and telecommunications with the greatest accuracy for SVM is 82.93%. Research [3] uses SVM for classification, Ontology, Synonyms of Ontology and TF-IDF, using synonyms aimed at expanding the terms of Ontology, so that terms generated from ontology and from synonyms ontology, evaluation compares between methods using Ontology and synonyms with methods without ontology and without synonyms, the results show the use of Ontology and Synonyms produces the best F1-measure is 81.19%, so it can be said that using Ontology and Synonyms is effective to improve F1-measure. Research [11] uses Coupled Multi-Layer Attention (CMLA) and multiple embedding mechanisms in the hotel domain in Indonesian, evaluates by comparing several variations of RNN for CMLA. The variations of RNN are GRU, LSTM, B-GRU, B-LSTM. The best method is using CMLA with the variation of RNN B-LSTM resulting in an F1-measure of 91.4%. Research [12] using the Recursive Neural Tensor Network (RNTN) in the hotel domain in English, resulted in an F1-measure of 77%, the experimental results of the RNTN model were the best in categorizing aspects that were positive. Research [13] using Latent Dirichlet Allocation (LDA) for aspect extraction and Naïve Bayes for sentiment classification in the hotel domain with English yields an F1-measure of 72%. Research [14] uses Conditional Random Field (CRF) with IOB notation and several features such as syntax, lexical, semantic features, sentiment lexicon for opinion target extraction (OTE) or commonly called aspect extraction, and uses Logistic Regression for sentiment classification, domain restaurant, hotel, laptop in English, the accuracy results from the 3 domains show that it is always above the baseline so this model is good for use in several domains, the accuracy result from the restaurant is 75.5%. Research [15] uses Ontology-based and fuzzy Product Review Miners in several evaluation domains comparing the Association rule, SVM, Ontology-based Product Review Miner (OBPRM) and OBPRM-light methods, the results show OBPRM produces the highest F1-measure is 80.82%, so it can be said that the use of ontology is quite effective in increasing the value of F1-measure. Research [16] recommends products by extracting information with ontology and SVM, ontology is used to determine aspects and polarities of the hotel domain in English, evaluation compares type-2 fuzzy ontology, fuzzy ontology with SVM, merged ontology with SVM, the best results use the method Merged ontology with SVM is 85%, so it can be said that merging SVM with ontology is better than without merging. Research [17] uses Ontology with SVM for aspect classification and sentiment classification, Bag of Words for feature extraction, domain in restaurant with English, experiments by comparing Ontology, SVM with BoW, SVM with BoW with Ontology, the best F1-measure results using Ontology with SVM with BoW is 91.7%, so it can be concluded that ontology alone is not good enough in classification, so ontology

needs to be combined with other methods such as SVM and BoW. From several research references, the best accuracy result using SVM with Ontology with BoW is 91.7%.

From some of the ABSA studies above, there are studies [3,17] that do not only use SVM for classification but use Ontology for classification with string matching, so that classification is 2 times with Ontology and SVM. there are several string matching studies that compare 2 ontologies such as research [18] which is a survey of methods for ontology matching, there are several techniques for ontology matching, one way is by string matching. Research [19] uses a string matching algorithm to equate 2 ontologies, the algorithm used is Levenstein, Sub-String, Jaro-Winkler, Monge-Elkan, Q-Gram, Smith-Waterman, Needleman-Wunsch, Sim. There are several best algorithms, one of which is Q-Gram.

From several previous ABSA studies, the best method at ABSA Restaurant is SVM with Ontology with BoW producing the highest F1-measure of 91.7%, but the features used are only based on Ontology so there are a large group of aspects whose sentiment expression cannot be found and there are several words in the review that are not can be detected by Ontology [17]. Research [3] uses SVM, Ontology, TF-IDF and WordNet Synonyms, WordNet Synonyms are used to expand terms in Ontology, experiments compare without and SVM with Ontology with Synonyms, the best results use SVM with Ontology with synonyms to produce F1-measure 81.19%, the use of Ontology and WordNet Synonyms is a factor to increase accuracy, the use of synonyms can increase accuracy because words in the review that could not be found in ontology can be found in synonyms [3]. So that the potential for expanding Ontology is combining SVM with Ontology with BoW in paper [17] and combined with Wordnet Synonyms in paper [3].

1.2 Problem Statement and Research Question

From research [17] using Ontology, SVM, Bag of Words, but the features used are only based on Ontology so that there are a large group of aspects whose sentiment expression cannot be found and there are several words in the review which cannot be detected by Ontology. Research [3] can expand Ontology by looking for synonyms for Ontology so that the number of terms becomes larger. To overcome the lack of paper [17], it is necessary to expand the ontology. The main question for this research is whether extending Ontology by adding synonyms combined with SVM, Bag of Words, Ontology can produce better F1-measures than without synonyms?

1.3 Objectives

The purpose of the research is

1. Working on ABSA with aspect classification and sentiment classification using a combination of Ontology, SVM, Bag of Words and Wordnet Synonyms.

2. Knowing the effect of adding WordNet synonyms is better or not by calculating the F1-measure.

1.4 Hypothesis

Research [17] uses ontology, SVM, BoW for ABSA which produces the highest F1-measure of 91.7% in the restaurant domain. However, the features used are only based on Ontology, so there are a large group of aspects whose sentiment expression cannot be found and there are several words in the review that are not can be detected by Ontology. Research [3] can expand Ontology by looking for synonyms of Ontology so that the number of terms in ontology is more, the evaluation proves that the use of ontology and synonyms is effective in increasing F1-measure from 0.8020 to 0.8119. The hypothesis of this study is to use Ontology, SVM, BoW combined with WordNet Synonyms to expand the terms in Ontology so that the use of synonyms is effective in increasing F1-measure.

1.5 Scope and Limitation

To ensure that the scope does not extend to unrelated topics, the scope of the problem needs to be limited. The scope and limitations of this study are using ABSA Restaurant data from SemEval 2016, which totals 2000 sentences. The SemEval 2016 data already contains a text review, aspect category labels and sentiment labels so there is no need to do labeling manually. This study conducts sentiment classification and aspect classification in the restaurant domain in English. Sentiment classification is classifying text reviews into positive or negative labels. Aspect classification is classifying text reviews into 12 labels namely Ambience#General, Drinks#Prices, Drinks#Quality, Drinks#Style_Options, Location#General, Food#Prices, Food#Quality, Food#Style_Options, Restaurant#Miscellaneous, Restaurant#Prices, Restaurant#General, Service#General. The Classification method cannot handle sentences that contain negations.

1.6 Contribution

Contribution to this research is the improvement of the method for ABSA Restaurants. Recent research on ABSA uses SVM, Ontology, BoW. This research combines SVM, Ontology, BoW with WordNet synonyms. The addition of WordNet synonyms can expand the terms in Ontology so that it can increase the number of features used in BoW. To find out the performance of this method, see F1-measure which compares 3 simulations.

1.7 Thesis Overview

This thesis is organized into five chapters consisting of introduction, literature review, algorithm design, and implementation, experiment and data analysis, and conclusions. The explanation of each chapter is as follows:

a. Introduction

This chapter discusses the background of the problems raised in this thesis, the definition of problems and research questions, research questions, objectives to be achieved, an overview of the research methodology and an overview of the thesis.

b. Literature Review

This chapter discusses the theoretical basis of the concepts and theories that support and are applied in this final project. Literature Review refers to the theoretical basis used as a basic reference for the guidelines for the final project.

c. Research Method

This chapter describes the general description of the system and the proposed model for solving the problems in this final project. In this design, an overview of the system is described, both in terms of detailed system structure, proposed model, data flow, and system usage scheme.

d. Experiment and Analysis

This chapter explains the purpose of testing and test scenarios for solving the problems that exist in this final project. At the analysis stage of the test results, it is explained about the data derived from the test results, the graph of the test results and the analysis.

e. Conclusion

This chapter describes the conclusions of this study. The conclusion includes the results of the final analysis that has been carried out and explains the answers to the problem formulation formulated in this thesis. At the recommendation points, suggestions are outlined regarding the possibilities that can be developed from the results of this research.