

Aspect Extraction on Restaurant Reviews using Domain-Specific Word Embedding

1st Ahmad Satriamulya

School of Computing

Telkom University

Bandung, Indonesia

ahmadrayi@student.telkomuniversity.ac.id

2nd Ade Romadhony

School of Computing

Telkom University

Bandung, Indonesia

aderomadhony@telkomuniversity.ac.id

Abstract—Reviews on the internet can be an important part of a business and can influence owners or consumers for their decision making. Easy access to information in the form of opinions, experiences, and feedback from others can be used as a reference for taking an action. For businesses in the food and beverage sector, consumers usually provide reviews with negative or positive sentiments based on several aspects of the related business. The taste of the food, atmosphere, price, service are examples of aspects that are commonly written in a review. In this work, aspect extraction on consumer reviews of restaurants in Indonesia is going to be carried out. Reviews on the internet usually contains words that are informal and very domain specific. This is where Domain Specific Word embedding can be used to reduce the amount of out-of-vocabulary word (OOV) and give the model more context of the review text given. The model used is Deep Learning with Recurrent Neural Network architecture, using Domain Specific Embedding as Word Embedding, and several attempts to reduce out of vocabulary in the model. The model used is able to reduce OOV from 17.16% (based on previous research) to 3.62%, with an evaluation of the F1-Score model of 79.54% using the Bi-LSTM model.

Keywords—aspect extraction, Recurrent Neural Network, word embedding, out-of-vocabulary

I. INTRODUCTION

Reviews on the internet can be an important part of a business and can influence owners or consumers in making decisions. With easy access to information with the help of the internet and devices, information in the form of opinions, experiences, and feedback from others can be used as a reference for taking an action [1]. In the food and beverage business, customer satisfaction can describe the quality of a restaurant, and for consumers, reviews on the internet can be a source of information that helps reduce the risk of unwanted purchases that have been felt by other consumers [2]. Business owners is also able to use consumer opinions to learn more about their consumers views on their business so that they are able to develop their business based their consumer feedback [4].

In general, consumers provide reviews of restaurants in the form of opinions that describes aspects of a restaurant. A few example of restaurant aspects are food, price, place, atmosphere, and service. These reviews can also be classified into two categories, which is positive and negative, this classification is commonly referred to as sentiment [3].

Sentiment analysis is a field of study that includes the analysis of opinions, sentiments, evaluations, judgments, attitudes, and emotions of people from a text to entities such as products, services, organizations, individuals. Sentiment analysis has several levels, Document level which analyzes the sentiment of the whole text, Sentence level which analyzes sentiment in each sentence, and aspect level which analyzes sentiment directly based on the opinions contained in the text [3]. Sentiment analysis based on the aspect level (aspect-based sentiment analysis) can provide a more complex analysis than the other two levels [4]. An important process in aspect-based sentiment analysis is the extraction of aspects from the text, and building a relationship between these aspects and the opinion in question [5].

In this work, we propose an approach to reduce the OOV numbers on aspect extraction based on suggestion on previous research [6]. The goal of OOV numbers reduction is to improve the aspect extraction performance. To achieve this, domain specific word embedding and several word pre-processing techniques are used. Domain specific word embedding will be made using restaurant reviews data in Indonesia. We are going to compare the performance of aspect extraction from restaurant reviews in Indonesia before and after several processes that reduces OOV and do analysis how OOV impacts the result of this work. feature extraction will be done in a form of a tri-gram to get the context of the adjacent word, and word2vec is used to create the word embedding.

II. RELATED WORKS

In previous research conducted by Widhianto, Rachmansyah Adhi, and Ade Romadhony [6] they have done research called "aspect extraction on restaurant review data in Bandung, Indonesia and compared several Machine Learning and Deep Learning methods". The RNN models (LSTM and BiLSTM) are the two best models compared to the SVM and CNN models with f1 scores of 79.1 for 11 tags, and 55.1 for 10 tags while not counting O-tags in the dataset, because the majority of the tags are O-Tag. In this research, the word representation is made using skip-gram from Word2Vec trained with wikipedia article text. The data train is done by changing the sentence into a tri-gram with the 2nd word as a reference for the output, for example a trigram sequence of