The 2nd Faculty Annual Research Session
Faculty of Applied Science
University of Vavuniya, Sri Lanka
September 15, 2021

Fine-grained sentiment analysis on food recipe reviews

Madushani, P.P.E.
Faculty of Applied Sciences
Wayamba University of Sri Lanka
emadushani324@gmail.com

Arudchelvam, T.
Faculty of Applied Sciences
Wayamba University of Sri Lanka
arul@wyb.ac.lk

ABSTRACT

Among an oversized variety of food recipes available on the web, only some recipes are posted appropriately. The dish might not appear identical when preparing the dish by reading a specific food direction. It is imperative to find a suitable recipe to prepare a meal properly. Customer satisfaction is considered a good measure to find a suitable recipe. Therefore, recipes should have wise ratings from users. The ratings and comments facilitate users to search out suitable recipes. Sentiments, analysis techniques and classifier algorithms will be used to classify sentiments based primarily on words. Sentiment classifiers are used for binary classification (positive or negative). However, in this analysis, five separate categories of sentiments are used. They are strongly negative, weakly negative, neutral, weakly positive and strongly positive. This analysis used a supervised learning technique that categorized each new document by one or more class labels from a set of predefined classes. An improved multinomial naïve Bayes algorithmic program is employed to develop the model and compare the model accuracy and precision with fine-grained and binary classification. In this work, an algorithm was developed to choose the suitable recipe by analyzing user ratings.

Keywords: Multinomial naïve Bayes algorithm, Sentiment analysis, Supervised learning.

INTRODUCTION

The web could be a capable place to search out different kinds of food and cookery recipes in the current world. When considering some instructions and ingredients of the recipes, some are appropriately written, and some are not. Users will comment and share their expertise with alternative users regarding the dish that they made. There are more negative comments than positive comments in a wrong recipe, and there are more positive comments than negative comments in a correct recipe. Thus, alternative user comments provide additional helpful information for users to pick out an accurate and correct recipe among various food and cookery recipes. Sentiment analysis is a machine learning approach to mine text knowledge that analyses a message and establish whether the underlying sentiment is positive, negative or neutral. Fine-grained sentiment classification may be a more difficult task than binary sentiment classification (positive and negative). The fine-grained sentiment uses five distinct categories, as shown in Figure 1. A fine-grained analysis will offer additional precise results, while binary analysis will result in incorrect category predictions.

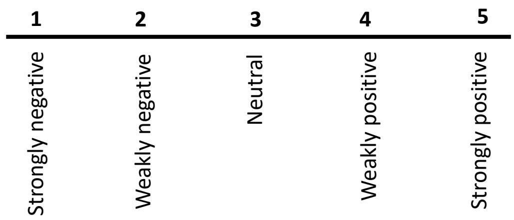

Figure 1: Sentiment Classification

This research is based on a supervised learning technique that categories each new document by assigning one or more category labels from a predefined class. The bag of words model is used wherever the order of the words is neglected, while the frequency of every word is employed as a feature for the trainset.

Different sentiment analysis techniques are accessible for analyzing text data. These techniques are primarily categorized as Machine learning techniques, lexicon-based techniques and rule-based techniques. Pugsee and Niyomvanich (2015a; 2015b) have analyzed food commentary recipes using polarity lexicon. It includes a list of negative and positive words and captures those words in comments using this polarity lexicon. Although it does not want any training data set, the main disadvantage of this process is that a more extensive range of words and expressions do not seem to be enclosed in sentiment lexicons. Rule-based approaches are used tokenization of each sentence and define various rules for opinions. Every sentiment started with an impartial score of 0 and became considered positive. If the very last polarity rating turned into greater than zero, or if the rating became less than 0 after the output of this technique, it would take a look at or ask whether the output was accurate or no longer (Devika et al., 2016). However, the efficiency and accuracy were more dependent on defining rules in this approach. Sentiment analysis can also be performed using deep learning algorithms. Deep Learning is a subtype of machine learning which learns features and tasks directly from data. Although this approach automatically deduced and modified the features to achieve the result, it requires substantial data to perform better than other techniques (Goularasv and Kamis, 2019). Also, there is no standard theory to aid in choosing the correct deep learning tools because it necessitates knowledge of topology, training method, and other characteristics. Machine learning techniques are used to construct a classifier that will determine sentiment text (Patil and Potdar, 2019). We train the associate machine learning model to acknowledge the sentiment based on words and their order employing a sentiment-labelled set (Saranya et al., 2019).

Fine-grained sentiment analysis: It is a sub-sentence level analyzing process and looking at small groupings of words. It uses ways like stemming, a bag of words, bi-gram: tri-gram and sentence level options to know the sentiment polarity. Fine-grained sentiment analysis refers to the acquisition of emotions, not on the document or post level, rather than on the sentence, sub-sentence (phrase/clause), and perhaps aspect-based level (Wang et al., 2016). It is a benefit for the characteristic sentiment that can be captured by document-based emotional analysis models and includes carefully defined language rules (Jurafsky and Martin, 2020).

Multinomial Naïve Bayes algorithm: Two Naïve Bayes event models are usually used; multivariate Bernoulli event model and multivariate event model . Multivariate Bernoulli event model is often called as multinomial Naïve Bayes model. The samples in the multinomial Naïve Bayes model represent frequencies when certain events are produced by multinomial $(p_1, ..., p)$ where the p is that the chance (probability) that event i happens (or such multinomials are in multiclass mode) (Xu et al., 2017). The feature vector $V = (v_1, ..., v)$ is then a histogram, with v enumeration the quantity of times event i used to be determined during an explicit instance. This can be the event model generally used for document classification, with events representing the prevalence of a word during a single document. Multinomial Naïve Bayes algorithm may easily handle large datasets, and the forecasts made by this algorithm are fast (Jurafsky and Martin, 2020).

N-gram model: The N-gram language model predicts the N-gram possibilities provided within any word order in the language. N-gram is identified as a sequence of N words. For a uni-gram model, N=1; for a bi-gram model, N=2; for a tri-gram model, N=3; and so on. As an example, "delicious" could be a uni-gram, "so delicious" could be a bi-gram, and "simple and delicious" could be a tri-gram. In this research, uni-gram, bi-gram and tri-gram are used to tokenize the words in food recipe reviews.

Term Frequency – Inverse Document Frequency (TF-IDF): The scoring method is named TF-IDF, where, Term Frequency (TF) is the frequency of a word within the current document, and the Inverse Document Frequency (IDF) is the frequency of how rare a word is across entire documents. The prominence of the extremely frequent vocabularies which do not belong to more informational content can be a crux when evaluating the scores of frequencies. Standard terms in every document such as 'the', 'and', 'if' do not give a higher meaning to specify documents. The frequency increases and approaches 0 when rare the word in a document and vice versa.

METHODOLOGY

The used public dataset acquired from Kaggle belongs to 1132367 records with two specific columns named review and rating. Preprocessing steps are applied, and then words in the reviews are tokenized using 3 N-grams (uni-gram, bi-gram, tri-gram). The text is then weighted using TF-IDF, and an improved multinomial event model is applied to analyze the sentiments of food recipes.

Data preprocessing: Text preprocessing is an essential step for sentiment classification. Inconsistent effects can occur when not applying for preprocessing techniques or when use incorrect text preprocessing steps. Some data preprocessing steps are applied for this text dataset, such as removing duplicate rows, removing empty rows, removing some unwanted characters, removing stop words, converting numbers to string, removing multiple spaces with single space, and removing URLs (HTML tags), and lemmatization. Lemmatization and stemming are two text normalization techniques within the field of text preprocessing. Lemmatization is incredibly the same as stemming. However, the distinction is that lemmatization cannot cut things out; it turns words into a particular root. Therefore, lemmatization is used for text standardization of food recipe reviews instead of stemming.

Balancing the dataset: After preprocessing steps, records are labelled into five different classes as weakly negative, weakly positive, neutral, strongly positive and strongly negative. The chosen dataset cannot be balanced; distribution to all known classes may discriminate against them. Therefore, it has to be balanced. There are some strategies for balancing, such as upsampling, downsampling and boosting. Among them, upsampling is used to increases the sampling rate.

Proposed method: The algorithm needs to pursue high dimensional problems because the review document contains many words usually identified by the high dimension of feature space. A feature vector wherever a given term represents the number of times it seems (frequency) is adopted by multinomial Naïve Bayes (Xu et al., 2017).

$$p\left(D_i|C=k\right) = \frac{\sum_{tx_{it}}!}{\prod_{t=1}^{d} x_{it}!}\prod_{t=1}^{d} p(w_t|C)^{x_{it}} \alpha \prod_{t=1}^{d} p(w_t|C)^{x_{it}} \tag{1}$$

where $x_i t$ is the frequency of word (number of words) w_t in document D_i and $\sum_{tx_{it}}!$ total number of words in a document D_i. The probability $p(w_t|C)$ (probability of a word w_t containing in a document of class C) is calculable using the frequency information of the words from the document feature vectors. The product of word likelihoods for every word within the document is taken by $\prod_{t=1}^{d} p(w_t|C)^{x_{it}}$.

The multivariate event model has two main problems. The frequency-based likelihood estimate will be zero if a given category and feature value never appear within the

Table 1: Results of Accuracy

No of classes	Uni-gram	Bi-gram	Trigram
Fine-grained (strongly negative, weakly negative, neutral, weakly positive and strongly positive)	63.6%	90.3%	92.8%
Binary (positive, and negative)	75.8%	90.3%	95.7%

training data. Then the whole product becomes zero because the likelihood estimate is directly proportional to the number of feature value events. Therefore, Laplace smoothing (equation 2) is used to handle that downside in this research.

$$p(w_t|C) = \frac{(N_{w_tC} + 1)}{(N_C + N_f)} \tag{2}$$

where N_{w_tC} is the number of word w_t in class C, N_C is the total number of words in class C, and N_f is the number of features.

After plenty of distinctive words are observed, there is a tendency to produce little value by computing the merchandise of the many likelihood terms. The computer can be run out of memory to represent those small values or be rounded to zero. Adapting the logarithm (equation 3) to the model is better to avoid that problem (Xu et al., 2017).

$$p(C|D) = \log\left(p(C)\prod_{t=1}^{d} p(w_t|C)^{x_{it}}\right) \tag{3}$$

The classifier is trained with five discrete classes of reviews using a multinomial naïve Bayes algorithm and binary classification. A model is defined to find term frequencies and vectorize and classify the sentiments. N-grams ranges, IDF and TF-IDF boundaries were utilized and estimated on the test dataset. Laplace smoothing constant $p(w_t|C)$ is added to evade wrong results when a term has not appeared at all in the training data.

RESULTS AND DISCUSSIONS

About 70% of the data is used to train the model, while 30% is kept to test the model. The classification uses three types of n-gram and multinomial Naïve Bayes algorithm with Laplace smoothing and logarithm calculations. Accuracy results are observed for five discrete classes (strongly negative, weak negative, neutral, weak positive and strongly positive) and two discrete classes (positive, negative).

As shown in Table 1, accuracy is increased gradually with uni-gram, bi-gram and tri-gram, respectively. In unigram, we assume that every word's incidence is freelance of its previous word. For example, 'it', 'was', and 'worst' assume that every three words are independent. However, it is not the case in real languages. In bi-gram and tri-gram, we assume that every incidence of every word depends solely on its previous word or 2-words. As N-gram is increased, it can be including more information about words.

Figure 2 indicates that the confusion matrix was plotted using *scikit-learn* and *matplotlib*; it tabulates the correct predictions versus incorrect predictions for five categories. Most of the sentiments were predicted as their actual class, but 54748 strongly positive sentiments were also false weakly positive. Also, none of the strongly negative sentiments is predicted as false, weakly positive or strongly positive.

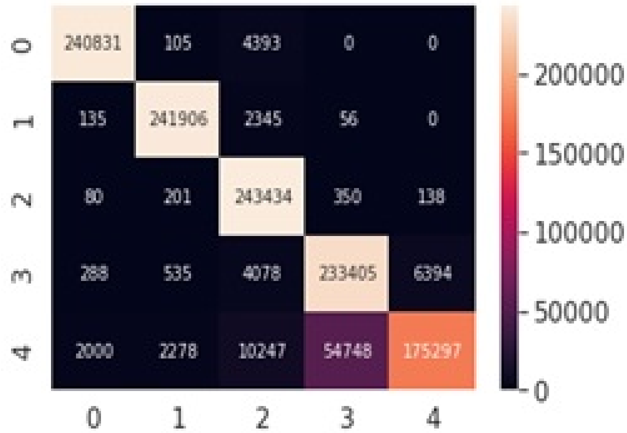

Figure 2: Confusion matrix of the model for five discrete classes

CONCLUSION

This study used the proposed model to analyze user opinions using fine-grained sentiment analysis techniques. The method in the proposed model used a multinomial Naïve Bayes algorithm to determine sentiment polarities. Term frequencies and TF-IDF were calculated, and three types of n-gram ranges were predicted to tokenize words within text segments. It recompensed the accuracy up to 90% with bi-gram and tri-gram and 63% with uni-gram. It was noted that the accuracy was increased with the value of n in n-gram as increased from one to three. However, although the accuracy of the results was high, some reviewer comments could not be analyzed because they had used stickers and symbols to describe their opinion.

REFERENCES

Devika, M.D., Sunitha, C. and Amal, G. (2016) Sentiment Analysis:A Comparative Study On Different Approaches. Fourth International Conference on Recent Trends in Computer Science & Engineering, Chennai, Tamil Nadu, India, 44–49.

Goularas, D. and Kamis, S. (2019) Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data. International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), Yeditepe University of Turkey, 12–17.

Jurafsky, D. and Martin, J.H. (2020) *Speech and Language Processing*, web.stanford.edu, Available: https://web.stanford.edu/ jurafsky/slp3/4.pdf

Patil, S.T. and Potdar, R.D. (2019) Recipe recommendation systems: a review, OSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, p-ISSN: 2278-8727, 21(3): 17–21.

Pugsee, P. and Niyomvanich, M. (2015a) Comment Analysis for Food Recipe Preferences, 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Chulalongkorn University of Thailand, 1–4.

Pugsee, P. and Niyomvanich, M. (2015b) Sentiment analysis of food recipe comments. ECTI Transactions on Computer and Information Technology (ECTI-CIT), 9(2): 182–193.

Saranya, K., Roslin, R.L. and Priyan, L.H.H. (2019) Sentiment Based Comments Rating Approach, International Journal of Innovative Research in Computer and Communication Engineering, 7(2): 1102–1106.

Wang, Z., Chong, C.S., Lan, L., Yang, Y., Beng Ho, S. and Tong, J.C. (2016) Fine-grained Sentiment Analysis of Social Media with Emotion Sensing, Future Technologies Conference (FTC). Institute of High Performance Computing (IHPC) of Singapore, 1361–1364.

Xu, S., Li, Y. and Wang, Z. (2017) Bayesian Multinomial Naïve Bayes Classifier to Text Classification. International Conference on Future Information Technology. Institute of Scientific and Technical Information of China, 347–-352.