CLASSIFICATION OF REVIEWERS INTO PROMOTERS, DETRACTORS AND PASSIVES USING ARTIFICIAL NEURAL NETWORK

APURVA BHARGAVA BE/25022/15
PALLAVI JAIN BE/25001/15

Department of Computer Science, Birla Institute Of Technology, Mesra, Jaipur Campus

INTRODUCTION

Classification is one of the important areas of research in the field of data mining and neural network is one of the widely used techniques for classification.

Our objective is to classify product reviewers into promoters, detractors and passives using sentiment-analysis-based classification of their reviews using ANN.

Sentiment Analysis is the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral. ANNs are crudely based on the neuron network structure of human brain. They can be trained through input assignment via forward propagation and weight adjustment via backward propagation.

METHODOLOGY

We used a dense (5-core) subset of the Amazon product review dataset (cited: Modeling the visual evolution of fashion trends with one-class collaborative filtering; R. He, J. McAuley).

Text Preprocessing:

The reviews are in natural language that needs to be preprocessed to remove the grammatical structure and filler-words that are inessential to sentiment analysis.

- Tokenization: Tokenization is a way to split text into tokens. These tokens could be paragraphs, sentences, individual words or characters. nltk.tokenize.TreebankWordTokenizer()
- Stemming: Stemming usually refers to a crude heuristic process that chops off the ends of words and often includes the removal of derivational affixes to find the root word. PorterStemmer
- Lemmatization: Lemmatization usually aims to remove inflectional endings with the use of a vocabulary and morphological analysis of words to return the base or dictionary form of a word, which is known as the lemma.
- Case conversion, acronyms, named entities recognition
- POS tagging

Text Processing Example Flow

Input Text Review
" This is a wonderful product easily delivered."

↓

nltk.tokenize.TreebankWordTokenizer(text)
"This" "is" "a" "wonderful" "product" "easily" "delivered"

↓

nltk.PorterStemmer (text)
"This" "is" "a" "wonder" "product" "easily" "deliver"

↓

Wordnet.Lemmatizer(text)
"This" "is" "a" "wonder" "product" "easy" "deliver"

↓

nltk.pos_tag(text)
[('This',['DT']),('is',['VBZ']),('a',['DT']),('wonder',['JJ']),
('product',['NN']),('easy',['JJ']),('deliver',['VB'])]

Feature Selection/Extraction:

- Sentiment descriptors were selected (low informative words were removed)
- They were tokenized into N-grams following with the removal of high frequency and low frequency N-grams

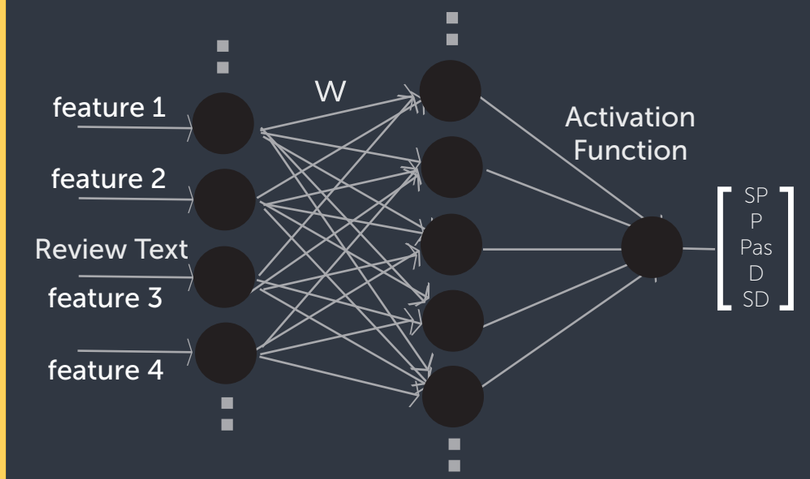
Feature Vector Generation:

There were two options to create the feature vector:
1. Occurrence (0/1)
2.Frequency/count can be replaced by TF-IDF to normalize
TF-IDF=TFi(t)*IDFi
TFi(t)=occurance count(t)/Total no of terms
IDFi=log(total no of reviews/total no of reviews containing t)

Employing ANN Model:

>Input: review
Input layer (preprocessed text reviews/ input matrix)
>Hidden layer (weight matrix for hidden layer, bias matrix for hidden layer)
>Output layer(output matrix, weight and bias matrices for output layer)
Output from output matrix: strengths assigned to each class,

ARTIFICIAL NEURAL NETWORK MODEL



RESULTS AND CONCLUSION

We have classified the reviewres into Strong Promoters, Promoters, Passives, Detractors and Strong Detractors. This powerful ANN technique may drastically improve classification of text data by increasing the number of hidden layers.This model shows better accuracy as compared against other models like NaiveBayes, Logistic Regression Model, Decision tree and more.

APPLICATIONS

It can be used for classiffication and gradually finding the NPS for a product or company.

Net Promoter or Net Promoter Score (NPS) is a management tool that can be used to gauge the loyalty of a firm's customer relationships. It serves as an alternative to traditional customer satisfaction research and claims to be correlated with revenue growth.