SSN-SPARKS at SemEval-2019 Task 9: Mining Suggestions from Online Reviews using Deep Learning Techniques on Augmented Data

Rajalakshmi S, Angel Deborah S, S Milton Rajendram, Mirnalinee T T
Department of Computer Science and Engineering
SSN College of Engineering
Chennai 603 110, Tamil Nadu, India
rajalakshmis@ssn.edu.in, angeldeborahs@ssn.edu.in
miltonrs@ssn.edu.in, mirnalineett@ssn.edu.in

Abstract

This paper describes the work on mining the suggestions from online reviews and forums. Opinion mining detects whether the comments are positive, negative or neutral, while suggestion mining explores the review content for the possible tips or advice. The system developed by SSN-SPARKS team in SemEval-2019 for task 9 (suggestion mining) uses a rule-based approach for feature selection, SMOTE technique for data augmentation and deep learning technique (Convolutional Neural Network) for classification. We have compared the results with Random Forest classifier (RF) and Multi-Layer Perceptron (MLP) model. Results show that the CNN model performs better than other models for both the subtasks.

1 Introduction

Sentiment analysis is a process of computationally identifying and categorizing the opinions from unstructured data. This can be used to identify a user's perspective of a product — positive, negative or neutral. Opinion mining is used to identify whether the product is a success in the market or not. Suggestion mining finds out ways to enhance the product to satisfy the customers.

Review texts are mainly used to identify the sentiments of the user. Besides sentiments, review texts also contain valuable information such as advice, recommendations, tips and suggestions on a variety of points of interest (Negi and Buitelaar, 2017a). These suggestions will help other customers make their choices, on the one hand, and the sellers improve their products, on the other hand.

Suggestion mining is relatively a young field of research compared to sentiment analysis. While mining for suggestions, the propositional aspects like mood, modality, sarcasm, and compound statements have to be considered. It is observed that, in some cases, grammatical properties of the sentence alone can be used to identify the label, while in other cases semantics play a significant role in label classification (Negi and Buitelaar, 2017b).

"Task 9 – Suggestion mining from online reviews and forums" has two subtasks (Negi et al., 2019). Subtask A is to classify a sentence into a suggestion or a non-suggestion. Subtask B is a cross-domain testing in which the model learned from a domain-specific dataset is used to classify dataset from a new domain. We have built classifiers using MultiLayer Perceptron (MLP), Random Forest (RF) and Convolutional Neural Network (CNN) models. However, due to the imbalance in the data, we have augmented it using Synthetic Minority Over-sampling TEchnique (SMOTE). We found that the CNN model performs better compared to RF and MLP classifiers for both the subtasks.

2 Related Work

Ramanand et al. (2010) discusses the rule-based method to find out the suggestions from the reviews. They have identified two kinds of 'wishes' viz the desire to improve the product and the desire to purchase the product. They have formulated the rules using modal verbs and certain sentence patterns. Viswanathan et al. (2011) develops an ontology-based knowledge representation for suggestion mining.

Customer-to-customer (CTC) suggestions are extracted by Negi and Buitelaar (2015) using keywords, POS tags, and imperative mood patterns. Customers feedback are analyzed using CNN and GRU network by Gupta et al. (2017). Long Short-Term Memory (LSTM) and CNN are used for sentence classification by Negi and Buitelaar (2017a).

A Linguistic-based approach is used to analyze

Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019), pages 1237–1241
Minneapolis, Minnesota, USA, June 6–7, 2019.

customer experience feedback by Ordenes et al. (2014) and Brun and Hagege (2013). We have used MultiLayer Perceptron for performing task 1 and task 3 in SemEval 2018. Task 1 was to identify the affect in tweets (Angel Deborah et al., 2018) and task 3 was to identify the irony in English tweets (Rajalakshmi et al., 2018).

3 System Description

Suggestions are mined mainly for business people to improve the product or for fellow customers to detect advice (Negi et al., 2016). We have used a linguistic rule-based method for feature extraction. Data is augmented to balance the imbalance in the data. The extracted Bag of Words (BOW) features are used in MLP, RF and CNN classifier for suggestion mining.

3.1 Feature extraction

The dataset is preprocessed to remove the stop words and non-printable characters using the NLTK functions. The features from the unstructured text are extracted using parts-of-speech (POS) tag. Modal verbs (MD) and the base form of verbs (VB) are considered to be suggestion features. Since the dataset has an imbalance, we have fewer examples for the suggestion class. Hence we have added synonyms and certain keywords used in baseline to enhance the BOW feature set. Top 5 synonyms for a particular word is obtained using synsets function from wordnet and keywords such as 'suggest', 'recommend', 'add', 'extend', 'idea', 'enhance', 'helpful', and 'useful'are added to enhance the feature set.

3.2 Handling imbalanced data

Imbalance in data is a scenario where the number of observations of one class is significantly lower than those of other classes (Chawla, 2009). This problem is predominant in fraudulent transactions, rare disease identification, criminal detection and also in suggestion mining. The model developed on this dataset will be inaccurate and biased since the traditional machine learning algorithms do not consider the distribution of the classes.

Imbalance in data can be remedied using the following methods.

1. Data level resampling
   (a) Random undersampling
   (b) Random oversampling
   (c) Cluster based oversampling
   (d) Synthetic Minority Oversampling Technique (SMOTE)
   (e) Modified Synthetic Minority Oversampling technique (MSMOTE)
2. Algorithmic ensemble techniques
   (a) Bagging
   (b) Boosting: Ada boost, Gradient tree boosting, XG boost

Data augmentation can be done in data space or feature space. SMOTE algorithm is used to create augmented samples in feature space (Wong et al., 2016). A subset of minority data is used to generate similar instances, synthetically. Synthetic data are generated based on the k nearest neighbours. These synthetic data are added to the original dataset to balance it.

The procedure for balancing the minority class data (Chawla et al., 2002) is outlined in Algorithm 1. SMOTE algorithm is applied on the minority sample BOW feature vectors to generate the synthetic data.

Algorithm 1: SMOTE Algorithm
Input: Unbalanced dataset.
Output: Balanced dataset
begin
1. Set the balancing ratio as auto to balance the given dataset equally.
2. For each instance i in the minority sample
   (a) Compute k nearest neighbours.
   (b) For each instance n in neighbour list
      i. $\text{diff}_{i,n}$ = difference between i and n.
      ii. rand = Generate a random number between 0 and 1.
      iii. synthetic_sample = $i + \text{rand} * \text{diff}_{i,n}$
      iv. Add the synthetic_sample to original dataset

end

3.3 Classifier algorithms

MultiLayer Perceptron, Random Forest and Convolutional Neural Network algorithms are used to build models. The augmented data is given to each of these classifiers and the models are built in Python programming environment. The results show that the CNN model performs better than MLP and RF.

3.3.1 MultiLayer Perceptron

MLP is a feedforward neural network mainly used for classification. It comprises an input layer, one or more hidden layers, and an output layer. The number of neurons in the input layer is decided by the number of features in the feature vector. The number of neurons in the output layer depends upon the number of classes. In our network, we have 5048 input neurons and 2 output neurons for suggestion/non-suggestion classification. We have used two hidden layers with 512 and 256 neurons respectively. Relu activation function is used for the input and hidden layers, while the softmax function is used for the output layer. Nadam gradient descent algorithm is used for optimization of the model.

3.3.2 Random Forest

Random forest classifier is an ensemble learning technique that uses decision tree as a basic learning algorithm. This is used to overcome the overfitting problem present in decision tree. Random forest creates a set of decision trees for randomly selected subsets of training data. It then aggregates the results of all these decision trees to make the final prediction. We have used 100 decision trees to build the random forest classifier and information gain as the measure of split criteria.

3.3.3 Convolutional Neural Network

Convolutional neural network is a deep learning technique that has already achieved remarkable results in computer vision. In text processing, deep learning techniques are used to learn the word vector representation through various neural models (Kim, 2014) and (Zhang and Wallace, 2015). We have used input embedding layer, convolutional one dimension layer with 32 filters, max-pooling layer with pool-size as 2, flatten layer, fully connected dense layer and output layer. We have added the dropout layer as 0.2 to regularize the network (Srivastava et al., 2014). The batch size of the model is set as 128 and the learning rate as 0.01. Softmax activation function is used in output layer and relu activation function is used in all other layers. Nadam algorithm is used for optimization.

3.4 Algorithm

The procedure for suggestion mining is outlined in Algorithm 2:

Algorithm 2: Suggestion Classification

Input: Augmented dataset.
Output: Suggestion/Non-Suggestion class labels
begin

1. Preprocess the dataset
   (a) Separate labels and sentences.
   (b) Remove the stop words and non-ascii characters from the sentences using NLTK functions.
   (c) Perform tokenization and Parts-of-Speech tagging using functions of the NLTK toolkit.
2. Feature selection
   (a) Identify the features using MD (Modal verbs) and basic form of verbs (VB).
   (b) Add the features and their synonyms into BoW.
   (c) Encode the features of sentences as a one-hot vector.
   (d) Represent the labels as a one-hot encoding of binary class in target vector.
3. Balance the dataset using SMOTE technique.
4. Build models (MLP, RF, CNN) with BoW feature vectors and target vectors.
5. Predict the labels for the test dataset.
   (a) Preprocess the test dataset
   (b) Represent the sentences as one hot vector with the help of BoW features of the training set.
   (c) Predict the labels of the test sentences by giving the BoW feature vector as input to the built model.
6. Calculate the accuracy and F1 score.

end

4 Dataset

The dataset given for suggestion mining task is prepared by a study of suggestions which appeared in different domains (Negi et al., 2018). For subtask-A, suggestion forum dataset is used for training and testing. For subtask-B, suggestion forum dataset is used for training and hotel review dataset (Wachsmuth et al., 2014) is used for testing purposes. The suggestion forum training dataset has 2085 instances of suggestion class and 6415 instances of non-suggestion class. The trial test set for suggestion forum dataset has equal number of instances (296) for both classes. The trial test set for hotel review dataset has an equal number of instances (404) for both the classes.

5 Performance Evaluation

The performance of the system is measured using precision, recall and F1-score for suggestion examples alone, using formulas shown in Equations 1 to 3.

$$\text{Precision } (\text{P}_{sugg}) = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{Recall } (\text{R}_{sugg}) = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{F1 score}_{sugg} = 2 \times \frac{\text{P}_{sugg} \times \text{R}_{sugg}}{\text{P}_{sugg} + \text{R}_{sugg}} \quad (3)$$

where TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative.

The F1 score for subtask A and subtask B using various models are shown in Table 1 and 2 respectively. Results show that CNN performs slightly better than MLP and RF models. The performance of the CNN model depends on the dataset size. Hence on increasing the data, we can get better results in CNN.

F1-Score	**Value**
MLP	0.45
RF	0.4
CNN	**0.49**
Baseline	0.2676

Table 1: Performance for Subtask A

F1-Score	**Value**
MLP	0.154
CNN	**0.155**

Table 2: Performance for Subtask B

We also worked with original data as such (without balancing) and created models using MLP, RF, and CNN. The F1 score for those models are very low, as almost all the samples are classified to non-suggestion class. CNN model with hand-selected features converges in less time with the same accuracy when compared to the CNN model with pre-trained Word2Vec embeddings. We intend to further investigate the model behavior using the variations of SMOTE such as borderline SMOTE, ADASYN and MSMOTE. For subtask B, the results are very low, since we have used the model built for suggestion forum dataset to make the prediction on hotel reviews dataset. The performance can be increased by incorporating transfer learning.

6 Conclusion and Future Scope

Customers generally express their opinions about an item through online reviews, blogs, discussion forums, or social media platforms. These opinions not only contain positive or negative sentiments but also contain suggestions to improve the item or advice to other customers. We have used CNN for suggestion mining. Dataset is augmented using SMOTE technique to handle the imbalance. Rule-based approach is used for feature extraction.

The performance can be improved by extracting the features using lexicons and increasing the number of convolutional layers in CNN structure. We intend to work with variations of SMOTE algorithm for balancing data and compare the results. We would also like to investigate the performance of Recurrent Neural Network (RNN) for mining suggestions from unstructured data.

References

S Angel Deborah, S Rajalakshmi, S Milton Rajendram, and TT Mirnalinee. 2018. Ssn mlrg1 at semeval-2018 task 1: Emotion and sentiment intensity detection using rule based feature selection. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 324–328.

Caroline Brun and Caroline Hagege. 2013. Suggestion mining: Detecting suggestions for improvement in users' comments. *Research in Computing Science*, 70(79.7179):5379–62.

Nitesh V Chawla. 2009. Data mining for imbalanced datasets: An overview. In *Data mining and knowledge discovery handbook*, pages 875–886. Springer.

Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.

Deepak Gupta, Pabitra Lenka, Harsimran Bedi, Asif Ekbal, and Pushpak Bhattacharyya. 2017. Iitp at ijcnlp-2017 task 4: Auto analysis of customer feedback using cnn and gru network. *Proceedings of the IJCNLP 2017, Shared Tasks*, pages 184–193.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Sapna Negi, Kartik Asooja, Shubham Mehrotra, and Paul Buitelaar. 2016. A study of suggestions in opinionated texts and their automatic detection. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 170–178.

Sapna Negi and P Buitelaar. 2017a. Suggestion mining from opinionated text. *Sentiment Analysis in Social Networks*, pages 129–139.

Sapna Negi and Paul Buitelaar. 2015. Towards the extraction of customer-to-customer suggestions from reviews. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2159–2167.

Sapna Negi and Paul Buitelaar. 2017b. Inducing distant supervision in suggestion mining through part-of-speech embeddings. *arXiv preprint arXiv:1709.07403*.

Sapna Negi, Tobias Daudert, and Paul Buitelaar. 2019. Semeval-2019 task 9: Suggestion mining from online reviews and forums. In *Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019)*.

Sapna Negi, Maarten de Rijke, and Paul Buitelaar. 2018. Open domain suggestion mining: Problem definition and datasets. *arXiv preprint arXiv:1806.02179*.

Francisco Villarroel Ordenes, Babis Theodoulidis, Jamie Burton, Thorsten Gruber, and Mohamed Zaki. 2014. Analyzing customer experience feedback using text mining: A linguistics-based approach. *Journal of Service Research*, 17(3):278–295.

S Rajalakshmi, S Milton Rajendram, TT Mirnalinee, and S Angel Deborah. 2018. Ssn mlrg1 at semeval-2018 task 3: Irony detection in english tweets using multilayer perceptron. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 633–637.

Janardhanan Ramanand, Krishna Bhavsar, and Niranjan Pedanekar. 2010. Wishful thinking: finding suggestions and'buy'wishes from product reviews. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, pages 54–61. Association for Computational Linguistics.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958.

Amar Viswanathan, Prasanna Venkatesh, Bintu G Vasudevan, Rajesh Balakrishnan, and Lokendra Shastri. 2011. Suggestion mining from customer reviews. In *AMCIS*.

Henning Wachsmuth, Martin Trenkmann, Benno Stein, Gregor Engels, and Tsvetomira Palakarska. 2014. A review corpus for argumentation analysis. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 115–127. Springer.

Sebastien C Wong, Adam Gatt, Victor Stamatescu, and Mark D McDonnell. 2016. Understanding data augmentation for classification: when to warp? In *2016 international conference on digital image computing: techniques and applications (DICTA)*, pages 1–6. IEEE.

Ye Zhang and Byron Wallace. 2015. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*.