A Novel Category Detection of Social Media Reviews in the Restaurant Industry

Mohib Ullah Khan · Abdul Rehman Javed · Mansoor Ihsan · Usman Tariq

Abstract Social media platforms have enabled users to share their thoughts, ideas, and opinions on different subject matters and meanwhile generate lots of information which can be adopted to understand people's emotion towards certain products. This information can be effectively applied for Aspect Category Detection (ACD). Similarly, people's emotions and recommendation based Artificial Intelligence (AI) powered systems are in trend to assist vendors and other customers to improve their standards. These systems have applications in all sorts of business available on multiple platforms. However, the current conventional approaches fail in providing promising results. Thus, in this paper, we propose novel convolutional attention based bidirectional modified LSTM by combining the techniques of the next word, next sequence, and pattern prediction with ACD. The proposed approach extracts significant features from public reviews to detect entity and attributes pair, which are treated as a sequence or pattern from a given opinion. Next, we trained our word vectors with the proposed model to strengthen the ACD process. Empirically, we compare the approach with the state-of-the-art ACD models that use SemEval-2015, SemEval-2016, and SentiHood datasets. Results show that the proposed approach effectively achieve 78.96% F1-Score on SemEval-2015, 79.10% F1-Score on SemEval-2016, and 79.03% F1-Score on SentiHood which is higher than the existing approaches.

Mohib ullah Khan
National University of Computer and Emerging Sciences, Islamabad, Pakistan
E-mail: mohib_khn@outlook.com

Abdul Rehman Javed
Department of Cyber Security, Air University, Islamabad, Pakistan
E-mail: abdulrehman.cs@au.edu.pk

Mansoor Ihsan
The University of Salford, Manchester, UK
E-mail: M.ihsan1@salford.ac.uk

Usman Tariq
College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Saudi Arabia
E-mail: u.tariq@psau.edu.sa

1 Introduction

The industrial internet of things (IIoT) is the use of smart interconnected sensors and instruments to enhance manufacturing and industrial processes. IIoT or industry 4.0 is characterized as machines, PCs, and people empowering intelligent modern tasks utilizing advanced data analytics for transformational business results Liao et al. (2019). A combination of social Media with IIoT has been progressively seen in industrial applications, for example, the manufacturing industry, education industry, smart devices industry, restaurant industry, automobile industry, marketing, and retail industry, etc Liao et al. (2019). IIoT uses the intensity of sensors, actuators, and constant examination to exploit the information that context-defined machines have delivered in modern settings for a considerable length of time Tripathi et al. (2020); Deep et al. (2019); Musaddiq et al. (2018); Iwendi et al. (2020a).

Fig. 1 depicts the major IoT roles in the modern world. The smart devices are connected by communication technologies to the internet that acts as a medium to connect any individual with other smart devices. This forms the internet of things (IoT) to interchange data or information based on which other devices can decide to execute a particular instance. IoT devices are widely deployed in various applications such as smart restaurants, vehicular ad-hoc networks, body sensor networks (BSN), and smart cities. Thus, IoT connectivity enables electronic systems to exchange real-world information in a digital form to perform autonomous features Gope et al. (2020). This establishes a need for innovation and automation in daily-performed typical manual tasks. In the diverse fields of business development and customer satisfaction, IoT and AI-automation play a vital role in providing comfort to both, the customer and vendor. Thus, if an adequately intelligent model is used for both, it can increase the comfort of customers and revenue for a vendor as the latest computer technique offers intelligent suggestions to the user according to their choice and vendor for profit maximization Hartanto and Utama (2020); Thanavisarnkajon and Jankaew (2020).

Machines use low-level language (i.e., Assembly) to understand real-world instruction and the same is with the human. Humans use language to communicate since their evolution, but this process is not as simple in the case of machines. Understanding language is a difficult task for machines which leads to ambiguity. This ambiguity of language makes Natural Language Processing (NLP) a challenging task for the computer to learn Kumar et al. (2020). Different NLP technique is being used by a machine to understand, analyze, and drive meaning from text sentences in a smart and useful way. We use NLP techniques in our work to address the problem of ACD and text classification. We use the NLP based techniques of the next word, next sequence, and pattern predictions with the combination of ACD.

Social media is expanding rapidly on the web. Every Social site contains bulks of text data that researchers and organizations use for building models Kiriu et al. (2019); Siriaraya et al. (2019); Mittal et al. (2019). Analysis of this vast amount
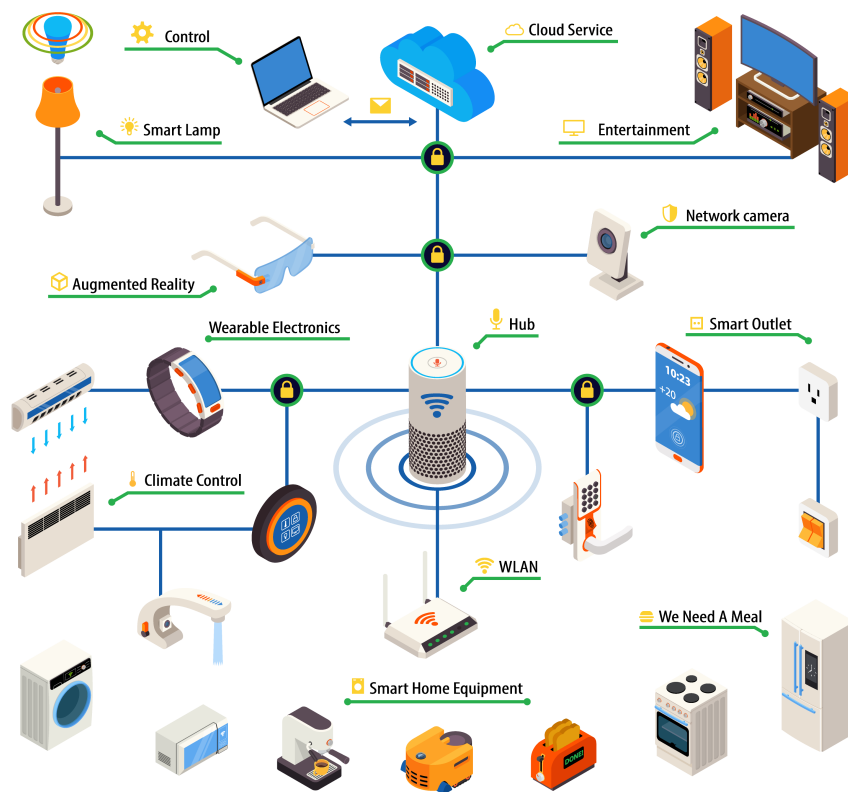
INTERNET of THINGS

Fig. 1: Graphical Representation of Modern IoT advancements and Usage Purposes.

of text data helps the researchers to extract people's ideas and reviews for decision making and the prediction of future decisions Wu et al. (2020); Kutia et al. (2019); Iwendi et al. (2020b). Also, understanding the demands of customers and their satisfaction level is one of the primary goals of tech consumer giants like Amazon, Yelp, and Apple to name a few. Keeping in view these things, we target opinion-mining and text classification problems to predict how people think about different things which include their perceptions, their expression of opinions, the way different people (agents) respond to different topics, and how their opinions about a subject change over time. These predictions are of great help for organizations and consumer-oriented businesses to enhance their sales and provide quality services according to the demand of the customers.

Researchers focused on understanding and classifying opinionated text contents that are scattered in the form of unstructured data (i.e., Twitter, Facebook) Babu et al. (2017); Cliche (2017). Sentiment Analysis (SA) in this aspect, is gain-

ing sufficient importance to understand the opinions expressed in a piece of text Koumpouri et al. (2015). It is a method that classifies the review sentences in binaries (positive or negative). This might lead to the other thousands of possible ways and dimensions through which opinions might be classified into. Consider a review that "Subway's Teriyaki Chicken was delicious but expensive". If we apply Sentiment Analysis on this particular review, it will be placed in the category of conflict sentiments because it is reflecting the positive as well as negative sentiments at the same time. However, if we observe, we can conclude that the sentiment expressed for the food "Subway Teriyaki Chicken" is positive, but the price is making the product unattractive for the customer. So whenever the customer joins any online site, first he analyzes the other people's reviews about that particular product in which he is interested and then decides whether he wants to buy it or not. Organizations and companies are using the same techniques to improve their services and solve their issues related to customer care. Before a person decides to visit a restaurant, he is particularly interested to know about their quality of food, the environment, the ambiance, and their service quality. To get this information, the candidate looks for external (unbiased) reviews. People give their opinions, mostly in the form of comments or star ratings. Usually, it is quite tricky for a computer system to understand these reviews. Moreover, manually analyzing the vast number of reviews is also not possible.

To our best knowledge, existing work focused on domain-specific aspect-based sentiment analysis by targeting aspect only, use pre-trained word vectors (i.e., glove), and also lack in providing efficient accuracy to detect the aspect Cliche (2017); Marzieh Saeidi and Riedel (2016). Thus, in this research work, we focus on generalized aspect based category detection rather than sentiment analysis and use our custom word vectors for the detection of customer opinion. We develop a Common sense model to predict the next word sequence by semantic aspect and contextual information to extract the category from customer reviews.

The main contributions of this work can be summarized as follows:

- Propose a novel Convolutional Attention Based Bidirectional Modified LSTM for Aspect Category Detection (ACD) of online English restaurant reviews in the restaurant industry.
- Proposed approach combines the method of next word, next sequence, and pattern prediction with ACD.
- Reformulate the sequence to label the problem into a sequence to sequence classification by using our custom words vectors.
- Proposed approach effectively enhances the detection rate of aspect category with consistent performance in comparison with existing studies.

The remainder of the paper is structured as follows. Section 2 provides the background of our approach consisting of Opinion Mining (OM), Aspect Category Detection (ACD), Sentiment Analysis (SA), and presents an overview of the related work on the background. Our technical contribution is presented in Section 3 and demonstrates the running example of our approach. Section 4 discusses the experimental setup and empirical results. Finally, in Section 5 concludes the paper by outlining the contributions of the proposed approach and stating future directions.

2 Related Work

This section provides the background of Aspect Category detection (ACD) and Long term Short Memory (LSTM). It also presents the state-of-the-art studies related to Aspect based category detection in which we proposed the methodology to predict if the sequence of the next word based on semantic aspect and contextual information using attention-based modified LSTM model.

2.1 Aspect Category Detection (ACD)

In ACD, the task is to detect every entity (E) and attribute (A) pair represented as (E#A) to which a user review is being expressed in a sentence. Categories for an entity (E) and attribute (A) should be assigned from a set (e.g. Laptop, Restaurant) and attribute types (e.g. price, quality). The Fig. 2 represents how to extract entity attribute pair from any sentence.(food=entity) and (Cost & Quality=attribute). Opinion mining is used to describe the user's opinion, emotions,

"**Big Mac 'n' Cheese Burger** is so
Tasty but **Expensive**."

{**Food#Cost**}
{**Food#Quality**}

Fig. 2: Aspect detection of a customer opinion

and attitude. A Social application like Facebook, Stack Exchange, and Twitter are the sites where a large amount of raw data is available. This data can be beneficial for ACD.

2.2 Sum of Word Vectors (SWV)

We analyze all the word vector methods Like Normalized Average Word Vectors, Normalized word vectors, Average Sum of the vectors (Average-SUM), L2-Normalized average Word Vectors (L2-NAV), a difference of word vectors (DWV), Concatenate both sentence vectors that is (CONCT-Vec), etc and compare them with a sum of word vector (SWV) method. The word vectors are summed up and used as sentence features for the LSTM classifier. This approach is presented in Mikolov et al. (2013) sum of word vector is used to represent the sentence using this Equation 1 and Graphical representation of SWV is shown in Fig. 3.

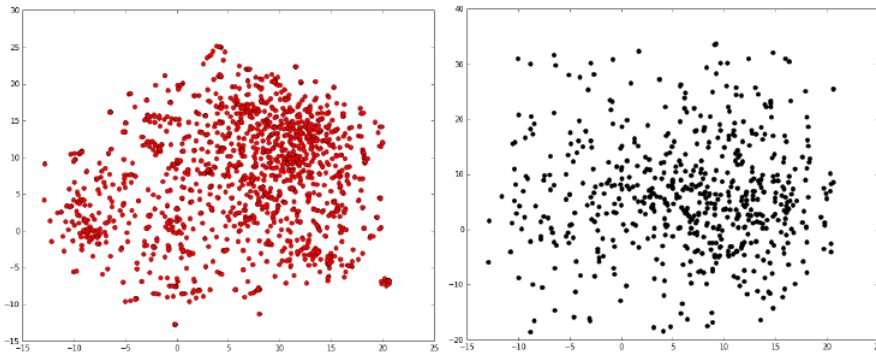$$SWV = \frac{\sum_{i=2}^{N} \overrightarrow{w_i}}{N} \tag{1}$$

Fig. 3: Red plot on the left shows the sentence vector training set. The black plot on the right shows the test set. These plots of sentence vectors are plotted using t-Distributed Stochastic Neighbor Embedding (t-SNE) technique to reduce dimension.

2.3 Long term Short Memory (LSTM)

Long Short Term Memory (LSTM) networks are introduced by Hochreiter & Schmidhuber (1997). LSTM is explicitly designed for targeting long term dependencies. LSTM networks can remember various kinds of information for a long time. This works amazingly well in various kinds of sequential problems without encountering the problems like exploding gradient or vanishing gradient. Their default behavior is to remember information for long periods. LSTM internally decides what to keep in memory and what to discard and then merge the current memory, previous state, and the input. Like all recurring neural networks that have repeating modules of the neural network, RNN has a similar but simplified structure such as a single Hyperbolic tangent (tanh) layer.

2.4 Literature of Natural Language Processing and Machine Learning Models

Authors inMikolov et al. (2013) showed how Word2Vec works with unsupervised dataset. They used skip-gram (SG) and a continuous bag of words (CBOW) to catch surrounding words. They used a model named NLANGP to extract Opinion Target Expression (OPE). However, this paper uses unsupervised datasets, and predictions are based on similarities. Also, they only extracted the target words and neglected the aspects. In Manning et al. (2014), the author solves this issue by working on detecting OTE by using named entity recognition (NER). For Sentiment analysis on aspects-categories, the author used Stanford CoreNLP for pre-processing, Part of Speech (POS) tagging. The POS tagging method is, however, not favorable in aspect category detection. They also haven't worked much on aspect categories. For this reason, Saias (2015) proposed a system Sentinel, in which the author worked on extracting features and syntactic features. His proposed model is divided into two parts 1) assigning sentences to multiple entities 2) making the correct pair of E#A. They used a machine learning methodology for this task. The limitation for this is that they used machine learning for

feature extraction and neglected word vectors. To use vector-based features, authors in Alghunaim (2015) targeted the vector-based features for extracting aspect categories and aspect terms. However, it is seen that feature extraction from machine learning is not as effective in category detection so for better accuracy. The authors in Maddikunta et al. (2020) proposed a machine-learning algorithm to locate the optimum venue to initialize the business based on the customer availability and demand of the particular item. The algorithm considers the quality of recommendation in terms of average service time, and customer-business ratio while providing recommendations to the user. The authors in Gadekallu et al. (2019) proposed their experimental work domain-specific feature-based heuristic which provides movie reviews on aspect-level. The aspect-oriented system examines word-based evaluations from actors of movies and tag aspects. The authors claim that integrating syntactical information in the models is dynamic to the sentiment analysis method.

Deep learning techniques for the extraction of semantic and contextual information. Authors in Wang and Liu (2015) focused on extracting aspect category and sentiment analysis using deep learning for feature extraction. To improve this, the author in Siwei Lai and Zhao (2015) proposed a model to extract the contextual information to learn the word representations. The proposed model uses a Recursive NN structure to capture the contextual information and builds the representation of text using CNN. However, the semantic and contextual information is lacking in this proposed model so to improve this, Liu et al. (2016) is introduced. In this paper, the author proposed to multitask learning framework based on RNN architecture (text sequence model with multitask learning). The author targeted two different mechanisms for the model, shared information, and shared layers. The entire RNN is trained jointly to perform different tasks. This framework is used to learn arbitrary mapping text into SVR (semantic vector representations) with shared layers. As this paper is RNN based so the handling of contextual information is limited so to solve this issue and get better accuracy. The authors in Asghar et al. (2020) proposed a deep learning model incorporated with a hybridized Fuzzy model for computing customer recommendations for vendors which assists them to improve their service quality. The proposed model initially uses Bidirectional Long Short Term Memory followed by Fuzzy logic with an attention mechanism to perform the computation over customer choices.

Attention architecture models for aspect detection. Authors in Yequan Wang and Zhu (2016) introduced in which the author revealed that the sentence polarity is not only related to the content, but it is also based upon the essential aspect. The proposed work revolved around Attention LSTM to classify the Aspect based Sentiment Analysis (ABSA). This attention architecture concentrates on various parts of a sentence when aspects are extracted as input data. The results of proposed attention mechanisms (AE-LSTM and ATAE-LSTM) are very prominent for aspect level sentiment analysis. This work continued in this paper Fei Liu and Baldwin (2018), where the author proposed a novel architecture to keep track of numerous aspect entities independently. This architecture is based upon an Update mechanism known as delayed memory (external "memory chains"). The proposed model is based upon GRU architecture. This model is trained to track and update the entity states at the right period with an external memory mechanism.

Cliche (2017) presented the study about the Twitter sentiment classifier. In this paper, the author proposed a model using LSTM networks and CNN which

are trained on the SemEval 2017 Twitter sentiment dataset. Aspect category detection is also added with sentiment analysis in this paper Marzieh Saeidi and Riedel (2016), where the author introduced the aspect-base sentiment analysis (ABSA) task. The main goal of this paper is to extract the fine-grained information w.r.t entities in a user review. To handle all the issues and limitations suffered by the papers mentioned above in aspect category detection, this Paper Yukun Ma and Cambria (2018), proposes a novel method to target ABSA by exploiting knowledge-based common sense. The author incorporated the LSTM network with an attention-based mechanism and sentence level attention. Here, knowledge-based common sense is used for sentiment classification. A hierarchical attention model is proposed to target ABSA. Effective commonsense knowledge is incorporated into a deep neural network.

The authors in Yadollahi et al. (2017) focused on sentiment analysis which plays a crucial part in Opinion Mining (OM). Some major OM tasks are also covered in this paper like Subjectivity Detection (Detection of objectivity and subjectivity in the given text), Opinion Polarity Classification, Opinion Summarization. The techniques and methodologies of OM describe in Liu (2012) are machine learning and NLP. For text classification, SVM performance is much better than any other classifier. It is because SVM works on the principle of structural risk minimization as compare to the Naive Bayes classifier which performs efficiently where features are highly dependent. K-mean algorithm is used where opinions are related to each other, and there is a need to cluster them, as done in paper Zhou et al. (2017) describes how OM is combined with topic modeling. The paper Babu et al. (2017) targets the problems related to noisy data, teaches how to make a cluster of related tweets, and handles unstructured data. In this paper, the author identified the groups of people with similar reviews about product features and understand the sentiment. In Koumpouri et al. (2015), the author combined different techniques for sentiment classification including Bag of words, lexicon-based approach, statistical-based, and content-based. The author in Wu and Wang (2017)developed a feature-based sentiment analysis model using ReviewMiner. It uses the LARA analyzer algorithm which analyzes the reviews from each category.

Authors in Touati et al. (2016) the author focuses on OM in Arabic and proposes a new discriminative model called Conditional Random Fields (CRF). This model works in 4 phases of data extraction from news articles then create Lexicon, it extracts sentences from opinion and tokenizes them using AROA Tool, then by applying CRF for OM and SA. In paperIsmail et al. (2016) author develops a model for sentiment analysis on Arabic reviews. The proposed model not only works as sentiment analysis but also summarizes the opinions. Teerapong Sungsri (2017) In this paper, the author focused on non-English language (Thai) sentiment analysis. Feature-based opinion mining is used to analyze the opinion in different aspects. In Che et al. (2015) paper, the Author adds a sentiment sentence compression in the proposed framework, before feature extraction data pass off from sentiment sentence compression (Sent Comp). It compresses the complicated sentiment sentence into a shorter sentence.

Table 1: Summary of existing studies focusing sentient analysis and aspect based sentiment analysis

Paper	Focus	Limitation
Cliche (2017)	Author proposed GRU model for twitter sentiments analysis	Short sentences are used for the job. For context relation information GRU perform average with attention Low F1-score
Christos Baziotis and Doulkeridis (2017)	Authors proposed Attention module and Siamese bi-directional LSTM for message level and topic-based sentiment analysis	Did not focus on aspects based sentiment analysis binary classification for significant feature extraction model rely on attention module and LSTM rather than CNN layers.
Yukun Ma and Cambria (2018)	Author targeted the Aspect base sentiment analysis using the sentence level attention mode	Focused on aspects of dependent sentence attention while ignoring the context for significant feature extraction model rely on attention module and LSTM rather than CNN layers.

3 Proposed Approach for Aspect Category Detection

In this section, we provide a brief overview of the proposed approach to implement an efficient system that increases the performance of ACD for Sentence Level Aspect-Based Sentiment Analysis. We transform the problem statement into a sequence-2-sequence classification task such that each sentence is represented by a vector that succeeds by a classifier to predict the aspect-categories. Fig. 4 present the steps of our proposed approach. The proposed approach comprise of: data collection, pre-processing, Word2Vec, sentence representation, and MLSTM model.
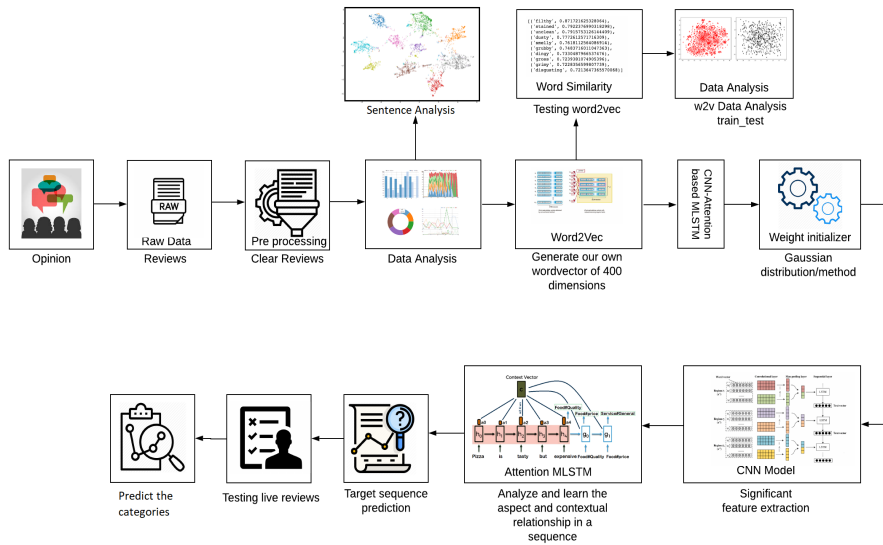

Fig. 4: Overall taxonomy of the proposed approach

3.1 Data Prepossessing and Word Embedding

Data Preprocessing is that segment in which the data gets Encoded into a state at which it can be easily parsed in the machine learning process. First, to Clean the dataset by removing the stop words, extra white spaces, special characters, and punctuation. In the next phase, the dataset is split into testing and training data. Then we train our Word2Vec model on 6 lac training sentences and get 231,778 unique words from these sentences. We take a training sentence and concatenate it with its label to create our training example. This is a demo sentence "For the price, you cannot eat this well in Manhattan" and these are its target labels RESTAURANT#PRICES, FOOD#QUALITY. What we did is we concatenate the demo sentence with its label and create a new sequence like this "For the price, you cannot eat this well in Manhattan RESTAURANT#PRICES, FOOD#QUALITY. Thus this is how we generate a training sentence that is used to predict the next sequence.

3.2 Word2Vec Model

Word2Vec is a technique to learn word embeddings using *shallow neural network*. To train Word2Vec, we use a skip-gram (SG) model on 10,00K Yelp and 4000 sentences from the given training set using the Gensim Rahim library in python Language. We set different hyperparameters (e.g., context window, min-word count, number of word dimensions) to train the Word2Vec model. The value of these parameters highly influences the quality of word vectors. We tune parameters to train Word2Vec so that they suit best to our required task. The following are the 5 parameters that we currently use to control the word vector's quality.

- Word Dimensions: To control the size of word-embedding, we use 400 word-dimensions.
- Min word Count: To control the word count, minimum numbers of the word allowed in the Word2Vec model.
- Number of workers: For the training process no threads used to manage to the desired parallelism.
- Context: Size of the context window.
- Down Sampling: To control the frequent-words in text corpus while training. Our down sampling ranges are from 1e-5 to 1e-3.

3.3 Attention Mechanism Base Bi-directional MLSTM

Our work is based on Aspect-based category detection in which we propose the methodology of sequence-to-sequence prediction. The architecture which we use for this task is based on attention base MLSTM Networks which is a modified version of LSTM where we add PS_{t-1} previous cell state information in each memory cell.

In the modified LSTM, we embed the information from the previous cell into the current cell. The current cell state, using the information (forget gate g_f information as well as input gate g_i information) from the previous cell state PS_{t-1} this

can make better decisions and can tune its weights more efficiently and faster as shown in Equation 2, 3, 6. In this way, our sequence to sequence model can learn our input sequences in a better way as shown in Equation 4, 5. We also embed the attention mechanism with our MLSTM model which can predict a better output sequence.

$$g_f = \sigma(W_f[PS_{t-1}, h_{t-1}, x_t] + b_f) \tag{2}$$

$$g_i = \sigma(W_i[PS_{t-1}, h_{t-1}, x_t] + b_i) \tag{3}$$

$$\tilde{S}_t = tanh(W_s[h_t - 1, x_t] + b_s) \tag{4}$$

$$S_t = g_f * PS_{t-1} + g_i * \tilde{S}_t \tag{5}$$

$$g_o = \sigma(W_o[S_t, h_{t-1}, x_t] + b_o) \tag{6}$$

Where h_{t-1} represents Equation 7:

$$h_t = g_o * tanh(S_t) \tag{7}$$

We solve the text classification and sentence Aspect Categories Detecting (ACD) problem by concatenating the given labels with a sentence and then treat it as a training example or by treating labels as the last words in a sentence. We mold the sequence-to-label classification problem to sequence-to-sequence classification problems but concatenate the labels with the given sentence. This type of method allows us to apply the attention mechanism with the bi-directional MLSTM network for ACD Task in which we aim to predict the next words in the sequence by the previous significant words given. To make our system more efficient, fast, and to extract better features for accurate prediction of the series, we integrate CNN (Convolutional Neural Network) with the proposed model. We use word vectors for the representation of each word in the sentence sequence and combine the CNN and attention mechanism with MLSTM to predict the sequence of next words. For the Attention Mechanism, we use these equations by JK.Chorowski and Bengio (2015).

$$C_i = \sum_{i=1} a_{ji} h_i \tag{8}$$

$$a_{ji} = softmax(f_{attn}(h_i, s_j)) \tag{9}$$

$$f_{attn}(h_i, s_j) = v_a tanh([W_a h_i, W_a s_{j-1}]) \tag{10}$$

Where C_i represents context vector, a_i represents attention scores, h_i represents present hidden state, s_{j-1} represents previous hidden state and $f_{attn}(h_i, s_j)$ represents attention function used to capture the alignment between input that is i and output that is j . This mechanism works by generating the weights of significant words which represent in Equation 9 and 10. These weights are multiplied with their respective word vectors. The results are then added up to form a single
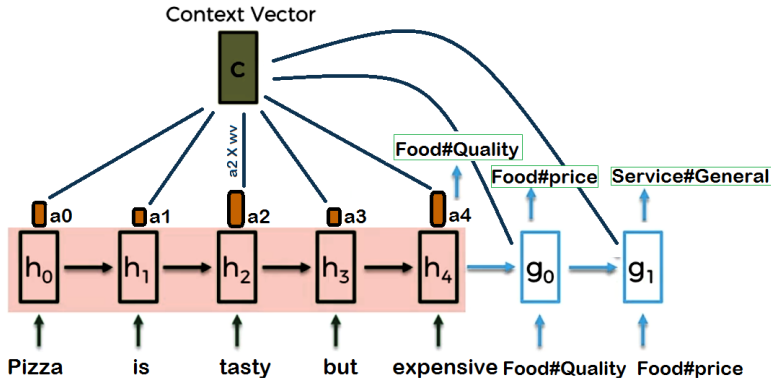
Fig. 5: Attention mechanism with MLSTM

context vector as represent in Equation 8. This context vectorC_i is then used for the prediction of the sequence. Fig. 5 represents the attention mechanism with MLSTM.

In Fig. 5 $h = 0, 1, 2...$ represents the MLSTM cells, $g(i)$ represent the output sequence and $a = 0, 1, 2..$ represents the weights of the words given by attention mechanism. Our proposed approach comprises of four parts: 1) Convolutional Neural Network (Conv NN), 2) MLSTM model, 3) Attention Mechanism (AM), and 4) Prediction.

1. Data is fed to CNN which returns significant words sequence of the given sentence.
2. To check the context and aspect of the data, we pass it to MLSTM Model. The model will also learn from the sequence w.r.t context and aspect of the words and will tune its weights accordingly.
3. The Attention Mechanism will assign weights to the words in the sentence according to their significance. After this, it will multiply the weights of the words with corresponding word vectors and will generate a context vector after adding them up.
4. The context vector is used to predict the sequence of the next words.

Algorithm 1 Opinion Mining And Text Classification

 1: **procedure** Text_Labeling($Data$)
 2:     $Sent\_Seq \leftarrow Sent.\textbf{concat}(Labels)$
 3:     $Word\_Vec \leftarrow Embedding(Sent\_Seq)$
 4:     **return** $Word\_Vec$
 5:     $Significant\_Words \leftarrow CNN(Word\_Vec)$
 6:     $Context\_Aspect\_Check \leftarrow MLSTM(Significant\_Words)$
 7:     $Word\_Weight \leftarrow Attention\_Mechanism$
 8:     $Context\_Vec \leftarrow \sum Word\_Weight \times Word\_Vec$
 9:     $Next\_Word \leftarrow Predict\_Class(Context\_Vec)$
10: **return** $Next\_Word$

3.4 Training and Testing Phase

The proposed CNN attention base bi-direction MLSTM model comprises 3 layers. We use 300 memory units in the first layer and 250 memory units in the third and second layers. The bi-direction MLSTM network uses dropout with a probability of 0.2. The dense layer of our model is fully connected to the hidden layer of MLSTM. We add CNN layers before MLSTM layers to efficiently extract significant words from sequences. The output layer predicts the next word. We use the Softmax activation function to ensure the character of normalized probability in our output. To optimize the log-loss (cross-entropy), we use the ADAM optimization algorithm to speed up the process. Bi-direction LSTM network is slow to train (it's about 200 sec per epoch on our CPU). Because of our optimization requirements, we use model check-pointing to record all the network weights to file every time progress in loss is detected at the end of the epoch.

To test the proposed model, the attention mechanism is used to predict the required sequence. A function is created that takes the input which includes our model, Word2Vec dictionary, seed text, and several categories. Then, the test sentence was converted into a vector sequence. When we execute the function, it takes a test sentence as an input, predicts the sequence of the next words, and returns the vectors of the sequence of the words with the highest probability. Like ("Pizza is tasty but expensive") and returns "Food#taste Food#Quality Service#General" as output words sequence.

3.5 System Demonstration

In this section, we provide a short demo of our proposed work. In Fig. 6, we present a test sentence as an input to our ACD model, where it returns the predicted categories against the input sentence sequence. In this demo, we use the SWV method to represent sentences for training our model for making sequence predictions. Our model returned sequence of categories is the same as it was provided in the test examples. Such type of results represents that the SWV method is good enough to represent long sentences as well.

4 Experimental Setup and Results

In this section, we explain the experimental setup of our approach. We evaluate the Aspect Category Detection approach by computing F1-scores (micro-averaging). These scores are based on the ratio between correctly classified labels, the set of predictions and the gold-standard Pontiki et al. (2016) and Androutsopoulos (2015) and Pontiki et al. (2016). We use Precision (P) and Recall (R), and F1-Score (F1) to evaluate the performance of our proposed approach.

4.1 Dataset Selection

We use Pontiki et al. (2016) restaurant review dataset(SemEval-2015, SemEval-2016). The dataset SemEval-2016 is annotated just like it was done previously for
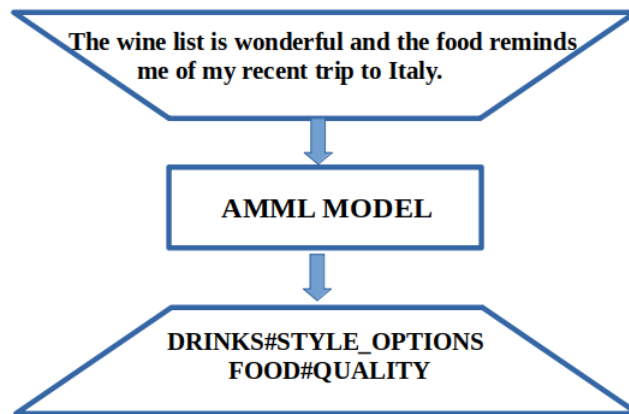
Fig. 6: ACD MLSTM model demonstration

SemEval-2015 so that the task should stay almost the same. The dataset contains 2000 unique training sentences and 676 test sentences. Out of these, 1708 reviews in the training data and 576 reviews in the test data which are correctly labeled with correct categories. Thus the left out sentences cannot be used in the SemEval-2016 Task 5 final evaluation. In this way, we have 576 reviews for testing purposes and 1708 reviews for training. The reviews are labeled with coarse categories of aspects with their attributes listed in the form of (E#A) pairs. Each review is in the form of a text sentence. Thus, it can have more than one category with each sentence. Unique categories are 12 as shown in Fig. 7.
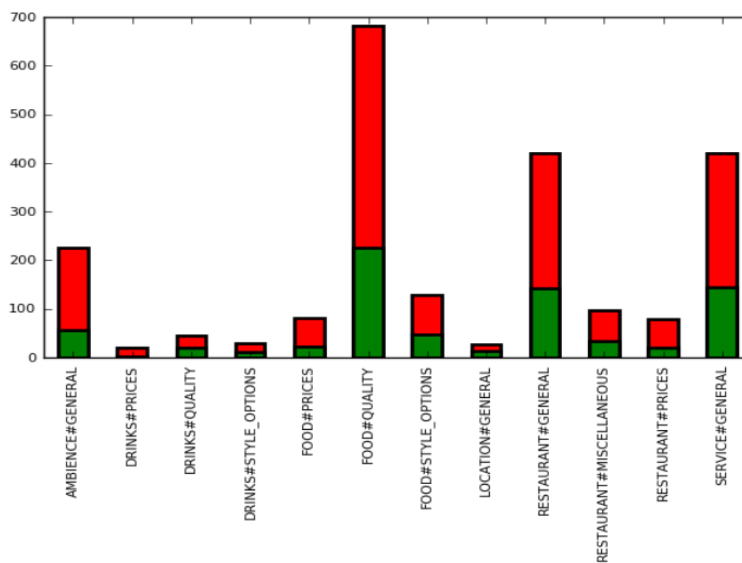

Fig. 7: Histogram of aspect categories distribution across restaurant reviews

We use three aspect base data sets, namely SemEval-2015, SemEval-2016, and SentiHood to evaluate the performance of the aforementioned approaches. Yelp dataset is used for computing word embedding. In the following, a description of each data set is provided.

- SemEval-2015[1] : SemEval-2015 Task 12 (SE-ABSA15) provided datasets annotated with aspect terms (e.g., "hard disk", "pizza") and their polarity for restaurant and laptop reviews, as well as coarser aspect categories (e.g., price) and their polarity only for restaurants. It contains 2200 Training Examples and 685 Test examples.
- SemEval-2016[2] : The dataset of SemEval-2016 is annotated just like it was done previously for SemEval-2015 so that the task should stay almost the same. It contains 2258 Training Examples and 783 Test Examples.
- SentiHood[3]: SentiHood is a dataset for the task of aspect-based sentiment analysis (ABSA).It's based on the text taken from the question-answering platform of Yahoo Answers. There are 5,215 sentences in this dataset. 3,862 sentences have a single target while the remaining 1,353 have multiple targets. We use Sentihood because it can be used as a benchmark for comparing our model with existing models (i.e.,).
- Yelp reviews[4]: The Yelp dataset is obtained from the Yelp-Dataset Challenge. This dataset contains 1,569,264 samples that review texts.

4.2 Results

A Three-layered MLSTM network model is used with signal layer CNN for Aspect Category Detection. First, we use our pre-trained Word2Vec to convert all the sentences to word vectors. Then we use 85% of sentences (reviews) from the training set to train the model. We use 1452 sentences (reviews) for training and 256 sentences (reviews) for validation from a total of 1708 labeled sentences. The performance of the model can be highly affected by many hyperparameters in the MLSTM network. To overcome this, we tune those parameters to make our model more efficient for the task of ACD. The detail of the parameter is shown in Table 2.

We use Word Vectors (WVs) to represent sentences. These vector representations use simple arithmetic operations like sum, subtract, average. This aggregating operation is then performed on both un-normalized and normalized WVs. Our research has shown that the sum of WVs methods for sentence vector representation performs better in the ACD task. Instead of normalizing representation methods of a sentence-vector.

Table 3 presents the results of the normalized representation of sentence-vectors in ACD (Simple Word Averaging Method). In this given category, we use 2 methods to represent feature WVs of a sentence. First by taking Average $\sum WVs$ and 2nd one is the Difference of WVs. But both methods perform less than the best score given in Sentic LSTM. Average $\sum WVs$ method score 5th position and the

---

[1] Available for download from http://alt.qcri.org/semeval2015/task12/

[2] Available for download from http://alt.qcri.org/semeval2016/task5/

[3] http://annotate-neighborhood.com/download/download.html

[4] Available for download from https://www.kaggle.com/yelp-dataset/yelp-dataset

Table 2: Details of the proposed MLSTM model

Parameter	Value
Batch size	128
Epochs	120
Decay rate	0.086
Internal layer	3
No of hidden layers	3
Hidden units at first layer	300
Hidden units at second layer	250
Hidden units at Third layer	250
Activation function (non-linearity)	RELU
Learning Optimizer	Adam (Stochastic)
Regularization	L2-regularization= 0.003

Average difference of WVs score 7th position in un-constraint systems as shown in Table 3.

Table 3: Normalized representation of sentence-vectors in ACD (Simple Word Averaging Method)

Sentence/word Features	Precision	Recall	F1 Score
Average sum of the WVs	66.68	77.81	71.81 (Rank 5)
Average difference of the WVs	62.89	79.82	70.35 (Rank 7)

In Table 4 we use the count of sentence-words to normalize the sum of WVs known as the average word vector. In this method, we focus on average WVs for normalization rather than the word counts. So for this experiment, we focus on 2 methods using normalized average WVs. The first one is normalized average word-vectors (Normalized average WVs L1) in which we use L1-norm to Normalized average WVs L1 and we use L2-norm to Normalized 2nd average WVs L2. Experiment results show Normalized average WVs L1 performs better than Normalized average WVs L2 and score 4th in an un-constraint system.

Table 4: Normalized average word vectors

Sentence/word Features	Precision	Recall	F1 Score
Normalized average WVs L1	74.66	73.31	73.31 (Rank 4)
Normalized average WVs L2	65.89	77.56	71.25 (Rank 5)

Table 5 shows in this experiment, we use 2 simplified versions of a sentence vector method which are non-averaging techniques used to represent the semantics. In the first method, we just sum up all the WVs of a sentence to obtain a one big sentence vector and then normalized it with L1-norm to obtain the ultimate sentence vector. The second method we use L2-norm to normalize the resultant sum of the WVs. These are the pretty simple method, but when we train our model using such type of sentence representations, then we figured out that norm vector L1 and norm vector L2 is performing better than Normalized average WVs.

Our normalized vector L1 method and normalized norm vector L2 scored 3rd and 4th rank in the category of un-constraint systems.

Table 5: Normalized word vectors

Sentence/word Features	Precision	Recall	F1 Score
Normalized WVs L1	77.54	72.50	74.93 (Rank 3)
Normalized WVs L2	69.56	77.56	73.79 (Rank 4)

Table 6 shows the results of the Unnormalized representation of sentence-vectors in ACD (Addition and Subtraction of Word Vectors) on SemEval-2016. In this section, we discuss the methodology for representing our sentence vectors. we use 3 different strategies: the first one is Sum of the WVs (SWV), in which we just sum up all the WVs of a sentence, the second one is to make the difference by subtracting all the WVs and then represent a sentence with a final resultant vector, known as the difference of WVs (DWV) and the third one is the Concatenation of sentence vectors that is known as (CONCT-VEC). Its 2 times large vector as compared to other methods used to accommodate the features from both SWV and DWV methods. We trained our model individually on all 3 methods and the results were outperformed all of the previous methods which we discussed before.

Table 6: Unnormalized representation of sentence-vectors in ACD (Addition and Subtraction of Word Vectors)

Sentence/word Features	Precision	Recall	F1 Score
Sum of the word vectors (SVW)	79.88	78.49	79.10 (Rank 1)
Difference of word vectors (DWV)	77.90	75.88	76.87 (Rank 1)
Concatenate sentence vectors (CONCT-VEC)	78.08	77.29	77.68 (Rank 1)

Table 7 shows the results of the proposed approach on three different datasets. We get good results in all 3 datasets. However, It can be seen that the F1 score of SenticLSTM of the Sentihood dataset is closer to ours, but the SenticLSTM is domain-dependent. However, the score of our proposed model is better overall in all 3 datasets. This proves that our model, unlike SenticLSTM, is more generic than domain-dependent.

Table 7: Comparison of results with state-of-the-art approaches

Datasest	Models	Strict Acc	Macro F1-Score	Micro F1-Score
Semeval-2015	Conv Bi-LSTM	63.16	70.01	68.07
	Conv Bi-LSTM+AM	65.73	74.36	73.67
	Sentic LSTM	67.34	76.44	73.82
	MLSTM+AM	**68.08**	**78.96**	**77.87**
Semeval-2016	Conv Bi-LSTM	65.05	71.71	70.21
	Conv Bi-LSTM+AM	65.78	75.91	75.11
	Sentic LSTM	66.08	76.66	76.17
	MLSTM+AM	**68.08**	**79.10**	**78.54**
Sentihood	Conv Bi-LSTM	60.04	69.28	67.76
	Conv Bi-LSTM+AM	64.76	75.61	74.08
	Sentic LSTM	67.63	78.18	77.66
	MLSTM+AM	**69.95**	**79.03**	**78.31**

Fig. 8 and 9 represent the loss and accuracy plot showed the results of our model after every 5 epochs. The experiments show that the proposed model achieves a promising F1-Score of 79.10%. Results showed that the proposed model is more generic than existing models. The loss curve starts from 3.2% and goes down to 2.35% and then it becomes consistent.
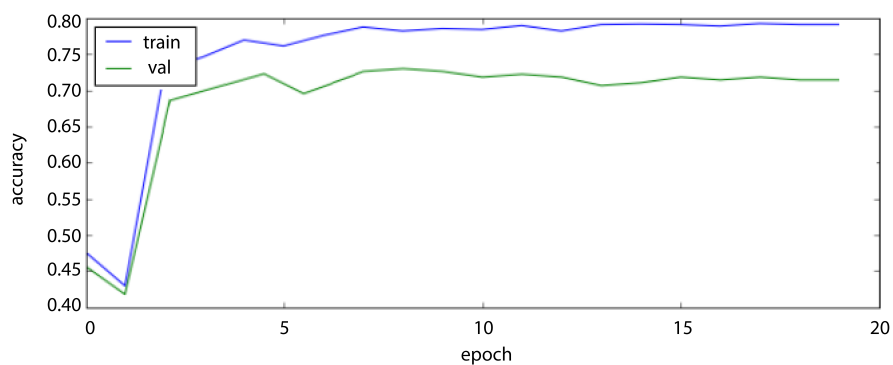

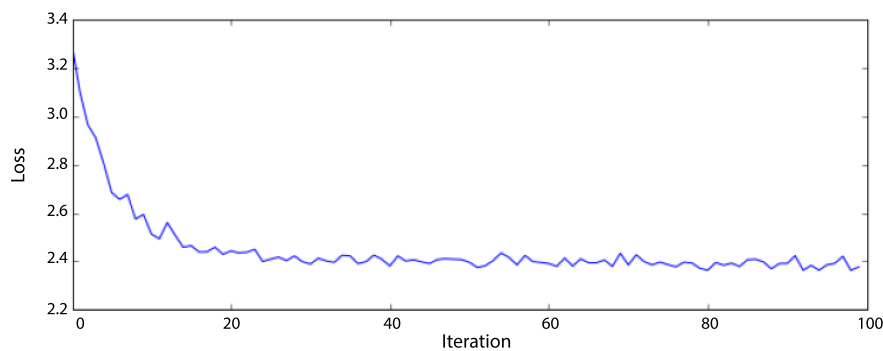Fig. 8: Trained MLSTM network model using Soft-max cross entropy accuracy function

Fig. 9: Trained MLSTM network model using Soft-max cross-entropy loss function

5 Conclusion and Future Work

In this research, a novel social media analytic based approach is proposed for Aspect Category Detection of user reviews about restaurants. This approach evaluated the significance of the prediction of the sequence of the next words based on the semantic aspect and contextual information for Aspect Category Detection (ACD). Our approach convolutional attention based bidirectional modified LSTM combines the techniques of the next word, next sequence, and pattern prediction with ACD. Our proposed approach outperforms state-of-the-art ACD models over the publicly available datasets: SemEval-2015, SemEval-2016 datasets, and Senti-Hood. Our approach achieved 78.96% F1-Score on SemEval-2015, 79.10% F1-Score on SemEval-2016, and 79.03% F1-Score on SentiHood which is higher than the existing ACD models. Results suggest that our proposed approach is best for Aspect Category Detection and sentiment analysis. This study will help those organizations that mostly rely on public opinion. In the future, we intend to work on the following research problems: A) Associate aspect categories with sentiments. Sentences (reviews) will be represented as attributes and aspect words. Such types of sentences will be passed to a Model to detect aspect categories. After detecting the aspect-categories, the dependency of a parse tree of a given sentence will be generated. B) Sentence Words that are associated with the aspect are detected according to the attribute words. Those words will then classified as positive words or negative words by using a classifier model Also the model will be able to perform all the operations based on the live feed according to customer satisfaction.

References

Alghunaim A (2015) A vector space approach for aspect-based sentiment analysis. PhD thesis, Massachusetts Institute of Technology

Asghar MZ, Subhan F, Ahmad H, Khan WZ, Hakak S, Gadekallu TR, Alazab M (2020) Senti-esystem: A sentiment-based esystem-using hybridized fuzzy and deep neural network for measuring customer satisfaction. Software: Practice and Experience

Babu AG, Kumari SS, Kamakshaiah K (2017) An experimental analysis of clustering sentiments for opinion mining. In: Proceedings of the 2017 International Conference on Machine Learning and Soft Computing, pp 53–57

Che W, Zhao Y, Guo H, Su Z, Liu T (2015) Sentence compression for aspect-based sentiment analysis. IEEE/ACM Transactions on audio, speech, and language processing 23(12):2111–2124

Christos Baziotis NP, Doulkeridis C (2017) Datastories at semeval-2017 task 4: Deep lstm with attention for message-level and topic-based sentiment analysis. In: Proceedings of the 11th International Workshop on Semantic Evaluations(SemEval-2017), pp 747–754

Cliche M (2017) Bb twtr at semeval-2017 task 4: Twitter sentiment analysis with cnns and lstms. In: Proceedings of the International Workshop on Semantic Evaluation (SemEval-2017), pp 573–580

Deep S, Zheng X, Hamey L (2019) A survey of security and privacy issues in the internet of things from the layered context. arXiv preprint arXiv:190300846

Fei Liu TC, Baldwin T (2018) Recurrent entity networks with delayed memory update for targeted aspect-based sentiment analysis. In: Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2018), pp 278–283

Gadekallu TR, Soni A, Sarkar D, Kuruva L (2019) Application of sentiment analysis in movie reviews. In: Sentiment Analysis and Knowledge Discovery in Contemporary Business, IGI Global, pp 77–90

Gope P, Gheraibia Y, Kabir S, Sikdar B (2020) A secure iot-based modern healthcare system with fault-tolerant decision making process. IEEE Journal of Biomedical and Health Informatics

Hartanto M, Utama DN (2020) Intelligent decision support model for recommending restaurant. Cogent Engineering 7(1):1763888

Ismail S, Alsammak A, Elshishtawy T (2016) A generic approach for extracting aspects and opinions of arabic reviews. In: Proceedings of the 10th international conference on informatics and systems, pp 173–179

Iwendi C, Jalil Z, Javed AR, Reddy T, Kaluri R, Srivastava G, Jo O (2020a) Keysplitwatermark: Zero watermarking algorithm for software protection against cyber-attacks. IEEE Access 8:72650–72660

Iwendi C, Moqurrab SA, Anjum A, Khan S, Mohan S, Srivastava G (2020b) N-sanitization: A semantic privacy-preserving framework for unstructured medical datasets. Computer Communications

JKChorowski JKDSKC Dzmitry Bahdanau, Bengio Y (2015) Attention-based models for speech recognition. In: In Advances in neural information processing systems, pp 577–585

Kiriu T, Mittal M, Siriaraya P, Kawai Y, Nakajima S (2019) Development of an acoustic ar gamification system to support physical exercise. In: Proceedings of the 27th ACM International Conference on Multimedia, pp 1056–1058

Koumpouri A, Mporas I, Megalooikonomou V (2015) Evaluation of four approaches for" sentiment analysis on movie reviews" the kaggle competition. In: Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS), pp 1–5

Kumar A, Saini M, Sharan A (2020) Aspect category detection using statistical and semantic association. Computational Intelligence

Kutia S, Chauhdary SH, Iwendi C, Liu L, Yong W, Bashir AK (2019) Socio-technological factors affecting user's adoption of ehealth functionalities: A case study of china and ukraine ehealth systems. IEEE Access 7:90777–90788

Liao H, Zhou Z, Zhao X, Zhang L, Mumtaz S, Jolfaei A, Ahmed SH, Bashir AK (2019) Learning-based context-aware resource allocation for edge computing-empowered industrial iot. IEEE Internet of Things Journal

Liu B (2012) Sentiment analysis and opinion mining. Synthesis lectures on human language technologies 5(1):1–167

Liu P, Qiu X, Huang X (2016) Recurrent neural network for text classification with multi-task learning. arXiv preprint arXiv:160505101

Maddikunta PKR, Gadekallu TR, Al-Ahmari A, Abidi MH, et al. (2020) Location based business recommendation using spatial demand. Sustainability 12(10):4124

Manning C, Surdeanu M, Bauer J, Finkel J, Bethard S, McClosky D (2014) The stanford corenlp natural language processing toolkit. In: Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, pp 55–60

Marzieh Saeidi ML Guillaume Bouchard, Riedel S (2016) Sentihood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods. In: Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016), pp 1546–1556

Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. CoRR abs/1301.3781, URL http://arxiv.org/abs/1301.3781, 1301.3781

Mittal M, Siriaraya P, Lee C, Kawai Y, Yoshikawa T, Shimojo S (2019) Accurate spatial mapping of social media data with physical locations. In: 2019 IEEE International Conference on Big Data (Big Data), IEEE, pp 4113–4116

Musaddiq A, Zikria YB, Hahm O, Yu H, Bashir AK, Kim SW (2018) A survey on resource management in iot operating systems. IEEE Access 6:8459–8482

Pontiki M, Galanis D, Papageorgiou H, Androutsopoulos I, Manandhar S, Mohammad AS, Al-Ayyoub M, Zhao Y, Qin B, De Clercq O, et al. (2016) Semeval-2016 task 5: Aspect based sentiment analysis. In: Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016), pp 19–30

Saias J (2015) Sentiue: Target and aspect based sentiment analysis in semeval-2015 task 12. Association for Computational Linguistics

Siriaraya P, Zhang Y, Wang Y, Kawai Y, Mittal M, Jeszenszky P, Jatowt A (2019) Witnessing crime through tweets: A crime investigation tool based on social media. In: Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp 568–571

Siwei Lai KL Liheng Xu, Zhao J (2015) Recurrent convolutional neural networks for text classification. In: Proceedings of AAAI'15 Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, pp 2267–2273

Teerapong Sungsri UUa (2017) The analysis and summarizing system of thai hotel reviews using opinion mining technique. In: ACM. ICIET'17, pp 167–170

Thanavisarnkajon W, Jankaew W (2020) Structural equation model for factors affecting customer choice to japanese restaurant in bangkok. International Journal of Development Administration Research 4(1):41–49

Touati I, Graja M, Ellouze M, Belguith LH (2016) Towards arabic semantic opinion mining: identifying opinion, polarity and intensity. In: Proceedings of the

Mediterranean Conference on Pattern Recognition and Artificial Intelligence, pp 131–136

Tripathi AK, Sharma K, Bala M, Kumar A, Menon VG, Bashir AK (2020) A parallel military dog based algorithm for clustering big data in cognitive industrial internet of things. IEEE Transactions on Industrial Informatics

Wang B, Liu M (2015) Deep learning for aspect-based sentiment analysis

Wu D, Wang H (2017) Reviewminer: An aspect-based review analytics system. In: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, pp 1285–1288

Wu P, Li X, Shen S, He D (2020) Social media opinion summarization using emotion cognition and convolutional neural networks. International Journal of Information Management 51:101978

Yadollahi A, Shahraki AG, Zaiane OR (2017) Current state of text sentiment analysis from opinion to emotion mining. ACM Computing Surveys (CSUR) 50(2):1–33

Yequan Wang LZ Minlie Huang, Zhu X (2016) Attention-based lstm for aspect-level sentiment classification. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp 606–615

Yukun Ma HP, Cambria E (2018) Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI-18), pp 5876–5883

Zhou X, Tao X, Rahman MM, Zhang J (2017) Coupling topic modelling in opinion mining for social media analysis. In: Proceedings of the International Conference on Web Intelligence, pp 533–540