IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 26, 2019, accepted July 9, 2019, date of publication July 15, 2019, date of current version August 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2928872

Aspect-Based Opinion Mining on Student's Feedback for Faculty Teaching Performance Evaluation

IRUM SINDHU[1], SHER MUHAMMAD DAUDPOTA[1], KAMAL BADAR[2], MAHEEN BAKHTYAR[3], JUNAID BABER[3], AND MOHAMMAD NURUNNABI[2]

[1]Sukkur IBA University, Sukkur 65200, Pakistan
[2]College of Business Administration, Prince Sultan University, Riyadh 11586, Saudi Arabia
[3]Department of Computer Science and Information Technology, University of Balochistan, Quetta 87650, Pakistan

Corresponding author: Irum Sindhu (irum.sindhu@iba-suk.edu.pk)

ABSTRACT Students' feedback is crucial for academic institutions in order to evaluate faculty performance. Handling the qualitative opinions of students efficiently while automatic report generation is a challenging task. Indeed, most organizations deal with quantitative feedback effectively, whereas qualitative feedback is either processed manually or ignored altogether. This study proposes a supervised aspect based opinion mining system based on two-layered LSTM model. The first layer predicts the aspects described within the feedback and later specifies the orientation (positive, negative, and neutral) of those predicted aspects. The model was tested on a manually tagged data set constructed from the last five years students' comments from Sukkur IBA University as well as on a standard SemEval-2014 data set. Unlike many other LSTM models proposed for other domains, the proposed model is quite simple in terms of architecture which results in less complexity. The system attains a good accuracy using the domain embedding layer in both tasks: aspect extraction (91%) and sentiment polarity detection (93%). To the best of our knowledge, this study is a first attempt that uses deep learning approach for performing aspect based sentiment analysis on students' feedback for evaluating faculty teaching performance.

INDEX TERMS Aspect extraction, deep learning, long short term memory network, opinion mining, polarity detection, student feedback.

I. INTRODUCTION

In this era of digital world, mining people's opinion is very crucial in order to dig out the useful insights and their sentiments regarding any specific entity. It is a common practice that when it comes to decision making, individuals or organizations prefer to seek out others' opinions [1]. Similarly, in academia, faculty teaching performance is evaluated through student's feedback provided at the end of each course. The quantitative responses are then aggregated and used as a measure to gauge the teaching quality of concerned faculty members. Besides, this is also considered as one of the key factors in the annual appraisal process. Though student feedback form is not only comprised of closed ended questions, it also provides students with a space for textual comments, an open ended feedback to express their thoughts and experiences. It provides students with an opportunity to give suggestions and opinions about various teaching aspects not covered by quantitative metric. However, manual handling of qualitative opinions is a very tedious task, as a result, it is either being ignored altogether or full of biases. So there is a need to automate this process to analyze students' feedback and this task comes under the emergent area of opinion mining.

Although, there have been some research studies that attempted to solve this problem by using opinion mining [2], [25], [26], [29] but their work is confined to just finding the polarity of an overall sentence (i.e. positive or negative). On the contrast, students describe various aspects of a teacher like teaching methodology, communication skills, technical knowledge, etc. in their feedback. For instance consider the following statement,

"He is expert in this subject."

The associate editor coordinating the review of this manuscript and approving it for publication was Fatih Emre Boran.

Although, this statement is positive but should not be considered as a generalized opinion on teacher's performance because it covers only one aspect(i.e. Knowledge) and does not cover his/her overall teaching pedagogy or work ethics issues. So, we need a system that can identify the opinion orientation with respect to a particular aspect. Consequently, identifying relevant aspects of an entity is very significant as this will help the teacher to work on its specific areas of improvement.

The prior studies made in aspect based sentiment analysis (ABSA) considered customer's feedback on products or services such as restaurants or laptops etc. and are not suitable for mining students' opinion on teacher's performance. This research proposes an alternative approach that is equally effective of Employee Evaluation and may also be used for other domains with a slight modification of input parameters.

The proposed system performs ABSA on student's textual feedback to evaluate the teaching quality of a concerned faculty member. Though the contemporary ABSA systems [3], [4] have achieved satisfactory results in different domains (products or services) by using handcrafted features and through complex neural network models. The recent trend is more inclined to get competitive results by using automatic feature extraction and reducing model complexities. Considering Occam's razor principle [5] which says that simple models are always chosen over complex model especially in the case of real world applications. So considering these two points in view, we have designed a simplified model based on automatic feature extraction.

Our proposed framework is comprised of two layered LSTM model for aspect extraction and sentiment classification. The first layer classifies a review sentence in one of the six aspects including- Teaching Pedagogy, Behavior, Knowledge, Assessment, Experience, and General. Next, the second LSTM layer predicts the sentiment orientation ($+ve$, $-ve$ or *Neutral*) expressed towards that particular aspect. The advantage of using this two layered model is that if decoupled these two layers can still be used to perform the individual job of aspect extraction or sentiment polarity detection. We have also built and incorporated domain word embedding into these models. Both layers using domain embedding demonstrate reasonable performance at their respective tasks.

Our research contributions are:

1. Preparation of academic domain dataset, manually tagged by domain experts, comprising of more than 5000 instances. Each student feedback is tagged in one of six aspects including Teaching Pedagogy, Behavior, Knowledge, Assessment, Experience and General.

2. A two staged LSTM model for aspect extraction and sentiment classification for an opinionated text from student feedback in the academic domain.

3. Advancement in sentiment analysis to combine it with aspect extraction with deep learning techniques.

4. An LSTM model that is not domain specific and can be used in other domains with a slight variation of input and output parameters.

The rest of the paper is organized as follows: Section II presents the related work in the field of aspect based sentiment analysis. In Section III, we have discussed a detailed methodology describing the skip gram model for generating domain word embedding and LSTM network models for aspect extraction and sentiment prediction. Experiments and results are presented in section IV & V whereas section VI conclude the paper with future directions given in section VI.

II. LITERATURE REVIEW

In this section, we first describe the aspect based sentiment analysis and its two sub tasks namely aspect extraction and aspect sentiment classification. We also present research work related to each sub task and finally demonstrate the previous work conducted in academic domain specifically for automatic teaching performance evaluation.

A. ASPECT BASED SENTIMENT ANALYSIS (ABSA)

ABSA can be defined as a task of extracting people's opinions or sentiments from the available text with respect to the set of aspects. The term opinion in ABSA is well defined in [6] as a concept represented in a form of quintuple

$$(e_n; a_{nj}; oo_{njkl}; O_h; t_1) \tag{1}$$

where e_n is entity name, a_{nj} denotes aspect of that entity, oo_{njkl} is the Orientation of opinion on that aspect (i.e. +ve, -ve or neutral), O_h denotes the person who has given opinion and t_1 indicates time on which opinion was made. For discovering the quintuples from the text, these five sub tasks needed to be performed:

1. Entity Extraction
2. Aspect Extraction
3. Extracting Opinion Holder
4. Time when opinion was made
5. Aspect sentiment classification

In this study, we have explored two essential tasks of ABSA [7] i.e. Aspect Extraction (AE) and Aspect Sentiment Classification (ASC) for the reason that entity (teacher) and time of the opinion is already known, and opinion holder is needed to be kept anonymous in the academic domain.

1) ASPECT EXTRACTION

It can be defined as the task of identifying an entity's relevant features from the opinionated text [8]. Every entity has some sort of features or aspects associated, for which the opinions are being formed. Considering the academic domain this task can be defined as

Given a set of feedback about teacher T, the task is to identify x major aspects of the entity.

Most research studies were purely attempted to extract only aspects from the available text without classifying text orientation. They performed this task of aspect extraction through various supervised, semi-supervised, and deep learning approaches. In one of the earliest studies [9], the researcher proposed a method to perform ABSA on product reviews. Their study was based on the assumption that aspects are basically Nouns and Noun Phrases. Extraction of such noun phrases is being done through the association rule mining technique. This method was relatively simple and moderated by various other researchers. Another study [10] made improvements in the results by identifying valid aspects and filtering out the irrelevant aspects. They computed the PMI score between the aspect term and meronym discriminators.

$$PMI_{(x,y)} = \frac{hits(xNEARy)}{hits(x)hits(y)} \quad (2)$$

If the PMI score value exceeds the defined threshold value, it is extracted as a valid aspect. Semi-supervised techniques were also explored like in [11] researchers proposed a double propagation method to identify new aspects from the initially known seed words. This method considered the syntactic structure of aspects and sentiment terms. Some researchers addressed this task as a sequence labeling program [12]. Manual labeling of words is being performed and then they applied the Hidden Markov Model. By learning patterns from the tagged sentences, the system learns patterns and performs aspects extraction. Various supervised algorithms were also applied like Conditional Random Fields (CRF) [13]. This method proved to be effective in a single domain as well as in cross domain. In one of the study [14] SVM binary classifier was used to find target aspect and opinion terms where they used semantic and various lexical features (unigram, bigram, and POS) and achieved an average accuracy of 94%.

Deep learning methods were also explored for extracting aspects. In [15] seven layer deep convolution neural network was proposed for extracting aspects. They also used some linguistic patterns and fed them into the model for improving results. One of the novel and simplified model of CNN was proposed in [16] for performing supervised aspect extraction. In this study, they used the concept of double embedding. In [17] aspect extraction was performed through the IOB2 scheme which classifies the word in three tags [inside, beginning, outside] of the vector. LSTM neural network was used that takes all the feature vectors as an input and outputs the predicted tag sequence in IOB2 format.

2) ASPECT SENTIMENT CLASSIFICATION

The second task in ABSA is to predict the orientation of aspects mentioned in the sentence.In [18] after aspect extraction, opinion terms were extracted using using "10 syntactic dependency rules". They consider those adjectives as opinions which are from the 3- word distance to aspect. WordNet was used to calculate polarity of extracted opinion phrase. Another study performed aspect sentiment classification on movie reviews [19]. In addition, to calculate sentiment orientation, they also specified sentiment strength towards a particular aspect of the movie. SentiwordNet was used as a lexical resource for computing sentiment scores.

Recently, most research studies used different variants of a neural network for this task. Adaptive recursive neural network (AdaRNN) proposed in [20] which basically performs target dependent sentiment classification on tweets. It considered the root node as a feature and trained a softmax classifier with these features to predict the distribution of tweets over multiple classes. In [21] target dependent LSTM and its variant target connection LSTM was proposed. The target word is considered as a feature that is concatenated with more features of text to perform aspect sentiment classification. In order to capture the intra and inter sentence relations a hierarchical and Bidirectional LSTM was proposed. Sentence level word embedding was created and fed into biLSTM which finally produces distribution over sentiments. Another study [22] proposed recurrent attention network in order to predict the sentiment of sentences having a complicated context. Bitmask bidirectional LSTM was proposed in [23] for performing ABSA. In this study, they used a bitmask layer which keeps track on the position of aspect in the sentence. They also experimented with word embedding approaches like word2vec and Glove. They achieved the highest accuracy score in SE-ABSA 81.2%. Another study [24] reported 69.38% in aspect extraction and 89.3% using LSTM and embedding of commonsense knowledge by means of the lexicon. All the above mentioned research studies performed their study on mobile, laptops and restaurant reviews.

3) ASPECT BASED SENTIMENT ANALYSIS IN ACADEMIC DOMAIN

Considering the academic domain, several research studies have been conducted but their work is only confined to the task of sentiment analysis (classifying sentence as positive, negative or neutral) only. Supervised algorithms were used in one study to perform sentiment analysis on students feedback provided at the Facebook page [25]. The aim of this study was to evaluate a teacher's performance by identifying feedback orientation. In one of the study, researchers prepared Teaching senti−lexicon [26] consisted of teaching corpus, category and sentiment weight score. They experimented with SVM, ID3 and Naives Bayes algorithms. They concluded that satisfactory results were achieved by using teaching senti−lexicon rather than general lexicon (SentiWordNet). Sentiment analysis was performed on the self evaluated comments of students in [27] for early predicting the academic failure of a student. As this will further help the teacher in improving the teaching process. Another study developed a decision tree algorithm [28] to help out in the annual appraisal report generation. They not only considered students feedback but also used various performance measures for generating a report. A lexicon based sentiment analysis system proposed in [29] provides a sentiment analysis based metrics that seems highly correlated with the Likert scale. Moreover, they generated word cloud to provide more insights into the teacher's performance in which positive words were shown with green color and negative words with red color. Also, high frequency words were shown with a larger font size. Faculty Evaluation System [FES] proposed to evaluate student's textual feedback [30]. They divided the system into 3 subcomponents namely (Feature Extractor, sentiment analyzer, feature sentiment evaluator). Their results are quite promising in terms of accuracy reported. Another study [31] used SMS texts for performing teaching evaluation. They first identified the different categories discussed within the SMS text and then performed modeling of text by considering three models: the base model, the corrected model, and the sentiment model. They concluded that the sentiment model can be used as a model of choice for teaching evaluation. The teacher's performance evaluation tool was proposed in [32]. They classified the English and Filipino language comment in the positive and negative category based on the cumulative score of opinion terms. The scores were fetched from their own created polarity data set. They used the Naive Bayes classifier for performing the task of sentiment analysis. Various machine learning algorithms (SVM, Random Forest, Simple Logistics, Decision Tree) and one of deep learning method (multilayer perceptron) were used to perform the task of sentiment analysis on educational database [33]. The highest accuracy was achieved with SVM 78.7% and %78.3 with Multilayer Perceptron.

As most of the studies in academia are inclined towards only sentiment analysis. Though Few studies have turned their focus for aspect based sentiment analysis. In which they not only classified sentences as per their polarity but also identified aspects discussed within the student feedback. In [34], they proposed a system that first identifies the feature discussed within the feedback then, it classifies comment as per its sentiment orientation. For feature identification, they used conditional context tf−idf scores and reported that on average 81.06% features were identified correctly. However, in sentiment analysis, they reported an accuracy of 89.67% by using Naive Bayes classifier. Features used for training a classifier were unigram, bigram, and unigram in root forms. In [35] they performed aspect extraction by using the dependency relation between opinion term and noun/noun phrase present within the sentence. They also developed concept ontology for extracting only domain relevant aspects and applied pruning of those aspects having a value less than the specified threshold value. Naive Bayes classifier was used for performing this task and for sentiment analysis they used an online sentiment analyzer. They achieved F-score 0.80 in aspect extraction and 0.72 in sentiment detection. Deep learning techniques are not yet explored in the academic domain. To the best of our knowledge, our study is a first attempt that is using LSTM for performing aspect based sentiment analysis on students' feedback for faculty teaching performance evaluation.

III. METHODOLOGY

In this section, we discuss the methodology of our proposed model in detail: we first demonstrate the creation of academic domain data, followed by preprocessing step; afterward, we describe skip gram model for generating domain word embedding; lastly, we explain the working mechanism of our two layer LSTM neural network for aspect extraction and aspect sentiment classification. Figure 1 shows a complete flow of our methodology.

A. DOMAIN UNDERSTANDING

In order to identify the labeling categories of our academic domain data set, It was required to know the core aspects or parameters of a teacher, considered by management while evaluating faculty teaching performance. For this purpose, we needed the output of domain experts especially people designated at different managerial positions in academics. As these are the people who have a great deal of experience to share about different parameters on which the teaching performance is assessed. To collect their responses, we designed a structured interview based on 10 questions verified by the Quality Enhancement Cell(QEC) of Sukkur IBA University. Interview questions are mentioned in Appendix. The interview duration was set to 15 minutes. We chose a structured interview for data collection as these are easy to replicate and fairly quick to conduct that ultimately save the time of interviewer and Interviewee.

We used representative sampling strategy by choosing strata of 45 people (Director Quality Enhancement Cell (QEC), Head of Departments (HODs) and Academic coordinators) from different universities of Sukkur and Khairpur city. With the consent of competent authority, we took their appointment and conducted interviews at their respective locations in university. Finally, we got the appointment from 40 people. All the interviews were tape recorded for analysis. As the total number of participating audience was 40, data was analyzed manually by carefully jotting down the response of each Interview question. After reviewing all the interviews, we come up with 12 distinct aspect categories. Table 1 shows all the 12 aspect categories along with the total number of experts who recommended this aspect category as an important feature.

As most of the terms were synonyms of each other so we combined them in one broad category. We place teaching skill and teaching methodology in one general term namely teaching pedagogy. Similarly, all the terms relevant to teacher behavior like polite, unbiased, selfless, tolerant were collectively termed as behavior. We also discarded a few terms that were having the lowest scores and were not important from student perspective like research contribution. Also, we included another category with the name "General", for labeling comments that do not contain any aspect but rather discuss the teacher as a whole. After applying these filters, we finalize six aspect categories for labeling data set:

1. **Teaching Pedagogy**: refers to the instructional approaches implemented in the classroom and the interactions that take place during learning. If a student discusses the teaching style of a teacher including different teaching strategies, that comment will be labeled with Teaching Pedagogy as an aspect.

2. **Behavior**: is the way in which the teacher acts, especially towards students. In this study, the focus is more

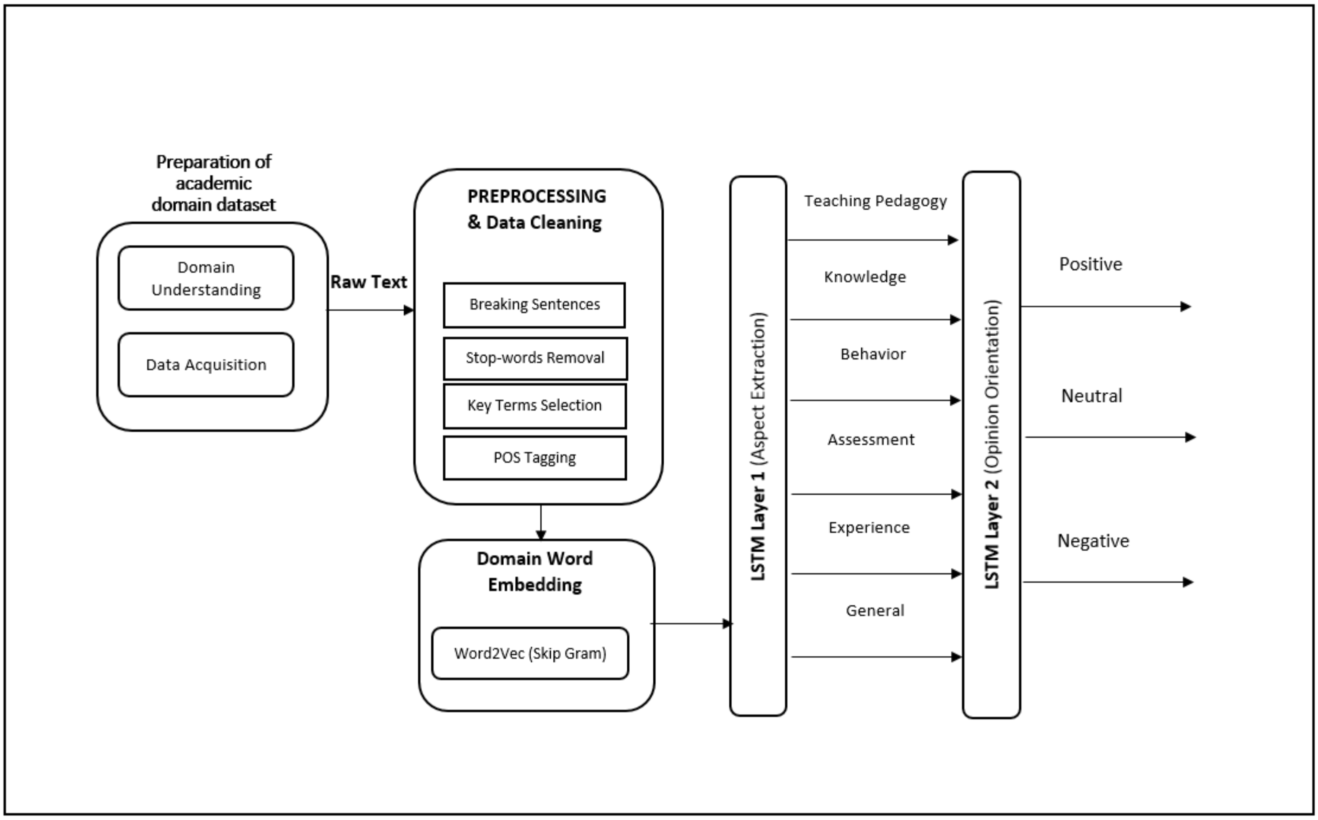

FIGURE 1. Flow diagram of methodology.

TABLE 1. Aspect categories with experts score.

S.No	Aspects	Aspect score as per No. of experts
1.	Behavior	18
2.	Punctual	4
3.	Knowledge	15
4.	Assessment	13
5.	Experience	15
6.	Polite	3
7.	Teaching Skill	19
8.	Selfless	1
9.	Teaching Methodology	1
10.	Unbiased	3
11.	Tolerant	6
12.	Research Contribution	2

towards relationship-oriented behavior as it includes terms like polite, kind, impartial, etc.

3. **Knowledge**: refers that how much the teacher has a theoretical or practical understanding of a course.

4. **Assessment**: indicates the variety of tools or methods used by the teacher to evaluate, measure, and document the learning progress and skill acquisition of students. So, if a student discusses how the teacher did their evaluation in their feedback than those comments will be tagged having an assessment as an aspect.

5. **Experience**: is the skill of a teacher which he has gained because of doing that similar activity for a long time. Any student feedback discussing such a feature of a teacher will be classified in this category.

6. **General**: When the review sentence is about the whole teacher and not covers any particular aspect.

B. DATA ACQUISITION

After identifying aspect categories, the next step was to acquire data of student's feedback. To the best of our knowledge, there is no any publically available data set of students' feedback for performing ABSA. For this study, we used last five years students feedback of Sukkur IBA University. Usually, the university is processing feedback manually by tagging each students comments in positive or negative category. The presence of negative comments within feedback is indicated through manual highlighting process done by domain experts. Despite doing this hectic work, this still does not specify the teacher's aspect discussed within that comment.

As the feedback's were already tagged with orientation, our next task was to assign a proper aspect label to each feedback. Initially, the data was in the raw form so we applied some pre-processing discussed in section III-C. Afterward, with the involvement of domain experts, aspect of each of the positive and negative feedback was identified and labelled. More than 5000 comments have now been tagged by domain experts to include aspect of each of the comment along with its sentiment orientation. Figure 2 shows a sample of 50 comments from our dataset after being tagged by domain experts for aspect and sentiment orientation.

1	Remarks	Aspect	Orientation
2	he is possessing high Teaching Skills	Teaching Pedagogy	Positive
3	he has skills to understand the capability of students to run the class accordingly	Teaching Pedagogy	Positive
4	Always degrade the students	Behavior	Negative
5	He is very fast!	Teaching Pedagogy	Negative
6	The teacher is well equipped with enough knowledge	Knowledge	Positive
7	He is Expert of the finance field	Knowledge	Positive
8	hH also was a bit speedy	Teaching Pedagogy	Negative
9	The pace of his teaching was fast so few points were missed	Teaching Pedagogy	Negative
10	He provides real time learning environment with his experience	Experience	Positive
11	Nice teacher	General	Positive
12	Well-Experienced	Experience	Positive
13	Always pay respect to students	Behavior	Positive
14	should be slow in pace while teaching	Teaching Pedagogy	Negative
15	we learned alot in the class	Experience	Positive
16	evaluation is unjustified	Assessment	Negative
17	Excellent Teaching Skills	Teaching Pedagogy	Positive
18	He has strong knowledge about course	Knowledge	Positive
19	he is very good teacher for case study especially	Knowledge	Positive
20	Brilliant Methodology	Teaching Pedagogy	Positive
21	Nice approach of teaching	Teaching Pedagogy	Positive
22	Keeps each student on same track	Behavior	Positive
23	He is good i think Sir a guru of finance at SIBA	Knowledge	Positive
24	sir taught us very effectively	Teaching Pedagogy	Positive
25	Sir still hasn't given result of 1st 2nd mid strongly disagree in fair assessment option	Assessment	Negative
26	His pace is very much high	Teaching Pedagogy	Negative
27	I have a very wonderful learning experience with respected sir	Teaching Pedagogy	Positive
28	one of the most talented teaching faculty	Knowledge	Positive
29	sir has alot practical knowledge we are very lucky beacuse we have learned many things	Knowledge	Positive
30	he is g great teacher in terms of teaching methods	Teaching Pedagogy	Positive
31	he explained less about the topics must work on his knowledge	Knowledge	Negative
32	he should teaches the students in proper way	Teaching Pedagogy	Negative
33	he only ask from students don't define himself	Teaching Pedagogy	Negative
34	The way of teaching is good	Teaching Pedagogy	Positive
35	He has good knowledge	Knowledge	Positive
36	a man with a grace	Behavior	Positive
37	He was very friendly to us	Behavior	Positive
38	Enhance effective learning	Teaching Pedagogy	Positive
39	well prepared in classs	Teaching Pedagogy	Positive
40	he has excellent teaching methodology	Teaching Pedagogy	Positive
41	exceptionally brilliant teacher i have ever learned from experienced teacher	Experience	Positive
42	his teaching method is quite learning for students	Teaching Pedagogy	Positive
43	The best teacher at IBA	General	Positive
44	please employ all teachers as he is	General	Positive
45	such a wounderful person in sukkur iba	General	Positive
46	very cooperative honest	Teaching Pedagogy	Positive
47	outstanding mentor	General	Positive
48	Sir gets personal with students on silly things too quickly	Behavior	Negative
49	sir seems to be suffering from objection	Behavior	Negative
50	shows too much anger	Behavior	Negative

FIGURE 2. Sample of 50 comments from our tagged dataset.

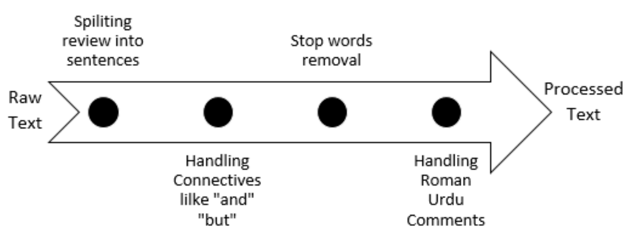

FIGURE 3. Preprocessing and data cleaning.

C. PREPROCESSING & DATA CLEANING

At first, when we reviewed our data set it was having comments about both teacher and course. So, we kept only those comments that were given for the teachers and discarded others. The filtered data set had near about 2000 comments given specifically for the teacher. Secondly, it consisted of paragraphs where students evaluated their teachers in various aspects. So in our first step, we used OpenNLP to break the paragraphs into multiple sentences. By default, it used period sign (.) for splitting a sentence but then again considering the data set we found sentences where connectives like "AND" & "BUT" were used. Their usage causes the inclusion of multiple aspects within a single sentence. For example, consider a sentence:

Sir was fair during the assessment and taught the concepts well.

Here, the student appreciated two aspects of a teacher that is Assessment and Teaching Skills by using connective "And". Similarly in another sentence "But" is used as a connective between behavior and teaching skill.

Good behavior but should be slow in pace while teaching.

So in order to capture a single aspect from a sentence, we further break a sentence on these connectives. In addition to this, data had a lot of irrelevant symbols like tags, colon, semicolon, and emoticons. So we performed the data cleaning to remove noise from the data. In this step, stop words like is, am, are, an etc. were further removed by using a list of stop words from Natural Language Toolkit(NLTK) corpus. Moreover, Roman Urdu comments were also discarded using lang. detector library of Python in which only English language comments were taken for further consideration. The whole preprocessing step is summarized in Figure 3.

D. DOMAIN WORD EMBEDDINGS

One of the main features that we used in our methodology is academic domain word embeddings to represent words semantically [36]. To the best of our knowledge no such pretrained word embedding exists for the academic domain so we utilized last 5 years students' comments from Sukkur IBA University to build domain specific word embeddings. By doing so, we expect that these trained domain embeddings capture semantic similarities between words in a better way. To generate these domain word embedding, we have used a skip gram model.

1) SKIP GRAM MODEL FOR DOMAIN EMBEDDINGS

The Skip-gram model exploits context words of the given word mentioned in the text for computing word embeddings. The prior step of using skip gram requires to define the context window: Number of words appearing left and right to the target word. For our model, we defined a context window size as 2 because we had a minimum sentence of length two in our data set. Suppose, we denote our target word as w_k and context window as z. As a result, the context window for w_k (including the target word w_k) is denoted as:

$$[w_{k-z}, \ldots, w_{k-1}, w_k, w_{k+1}, \ldots, w_{k+z}] \tag{3}$$

As an example, let's take a single sentence from our sample feedback.

''He has a lot of knowledge about subject.''

In our model we have used z = 2, so our data set becomes: *[has, a], He), ([He, a,lot], has), ([He, has, lot, of], a), ([has,a,of, knowledge],lot),...*

After specifying the context window, next step was to convert the data in the form of $< input, output >$ tuples, so we converted the whole data in the required form as shown in equation 4. Whereas input indicates the target word w_k and at the output, context words appearing left and right to the target word are shown.

$$[.., (w_k, w_{k-z}), .., (w_k, w_{k-1}), (w_k, w_{k+1}))..,(w_k, w_{k+z}), ..] \tag{4}$$

After forming these tuples, we build a neural network with one hidden layer for learning these domain word embeddings. For the input layer, the number of neurons is set to the number of words in the domain vocabulary V. All the input words are one hot encoded as a vector of size V. The size of the output layer is also kept same to input layer because in output we need vectors of all the words in the vocabulary. Next, we specified a *VXK* matrix to store word embeddings whereas $V = 12477$, which denotes the number of words in the vocabulary and we selected $K = 100$, representing the dimension of word embeddings as this embedding setting works better in our model. We referred this matrix as W_d with each row denoting a vocabulary word. This matrix is fed to the hidden layer as an input. Similarly, connections from the hidden layer to the output layer are represented through a matrix W_O of size *KXV* where each column denotes words from the vocabulary.

The Output layer of the network results the unnormalized predicted scores of words being a context word, denoted as $logit(w_k)$ as shown in equation 5. Whereas W and b are weights and biases matrices respectively.

$$logit(w_k) = h_k W + b \tag{5}$$

We applied the softmax activation function to the $logit(w_k)$ to get the normalized scores in the range of [0,1] denoted as y_k as shown in equation 7.

$$y_k = softmax(logit(w_k)) \tag{6}$$

whereas softmax function is defined in equation 7, which divides exponential power for each logit value with the sum of exponential parameters for each value in logit vector. In result, it generates the final values in the range of [0,1].

$$Softmax(logit(w_k)) = \frac{exp(logit(w_k))}{\sum_{k=1}^{n} exp(logit(w_k))} \tag{7}$$

In order to get the good word embeddings, we optimized the loss function $L(\theta)$ shown in equation 8. The objective of doing this loss function optimization was basically to minimize the probability of "all the non−contextual word''and maximize the probability of predicting a contextual words of a given target word. It force word embeddings to organize themselves well according to the meaning.

$$L(\Theta) = -(-1/V - 2z) \sum_{k=z+1}^{V-z} \sum_{j=k-z, j\neq k}^{k+z} \sum_{l=1}^{V} I_{x_k} logit(w_k) - log\left[\sum_{x_k \epsilon V} exp\left\{\sum_{l=1}^{v} I_{x_k} logit\ (w_k)\right\}\right] \tag{8}$$

Here, $logit(w_k)$ denotes the unnormalized probability scores, V denotes the total size of vocabulary, z indicates window size and $\prod_{x_k}$ denotes the one hot encoding of that particular target word w_k, and l is simply a loop variable.

We trained this model for 50 epochs and saved these domain embeddings to utilize them further in LSTM model described in section III-E. Fig 4 represents the whole skip gram model for learning domain embedding from data preparation to the final output vector of word embeddings.

E. LSTM MODEL FOR LAYER1 AND LAYER2

Now, we discuss the working mechanism of our proposed LSTM model that we have used in both steps of Aspect extraction and opinion orientation as shown in Figure 5. Unlike other studies, we have not modeled this problem as a sequence labeling task where the data is initially labeled through various tagging schemes like IOB2, rather we have simply used a tagged data set where we have manually tagged reviews as per their respective labels categories. So our proposed model classifies a review in the specified categories by using LSTM neural network.

LSTM is one of the special variants of RNN that overcomes the problem of vanishing gradient problem [37], [38]. It has the ability to store longer memory and better control on what to store and discard. Moreover, it has a gating mechanism to control information flow inside the LSTM cell [39]. These gates includes:input gate, forget gate and output gate denoted

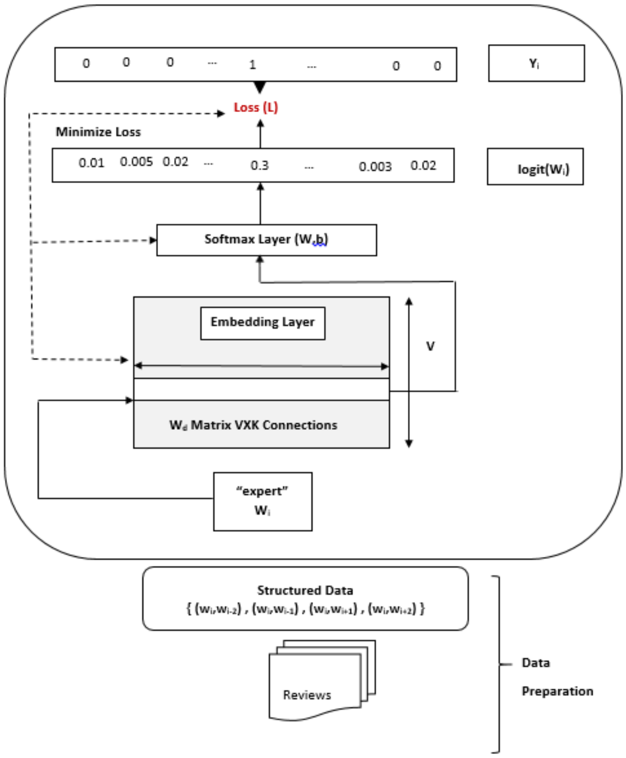

FIGURE 4. Learning of academic domain word embedding using skip gram model.

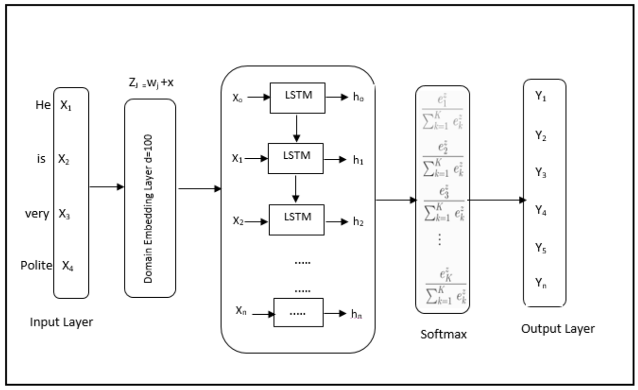

FIGURE 5. LSTM model configuration.

as f_{gt},i_{gt} and o_{gt} respectively. All these gates are made up of the sigmoid neural net layer and pointwise multiplication operation. The output of this sigmoid layer is between 0 and 1. The present cell state of LSTM at timestamp t is denoted as c_t.

Our two layered LSTM model first processes the input sentences where each word is denoted as $x_1, x_2, x_3, ..x_n$. Next, we have used domain embedding layer in which input words are concatenated with the 100-dimensional word vectors w_d and passed to the LSTM network layer as shown in equation 9.

$$z_d = w_d + x_t \tag{9}$$

whereas w_d is the 100 dimensional word embedding vector, x_t is the vector of input words and their concatenation is represented by z_d.

When the LSTM model receives the z_d, it first decides what information to remove from previous output state $h_{(t-1)}$ so for this purpose LSTM uses it's forget get as shown in equation 10 which ultimately results a number between 0 and 1. Whereas 1 indicates "Information Needed" and 0 indicates "Information not needed", but initially the previous output state remains empty as no output is displayed so far.

$$f_{gt} = \sigma(W_{fg}.[h_{t-1}, z_d] + b_{fg}) \tag{10}$$

where f_{gt} indicates the forget gate, W_{fg},b_{fg} are the weights and bias values of forget gate, h_{t-1} is the previous output state at time stamp $t-1$

Next, the LSTM layer use it's input gate denoted as i_{gt} to decide what to add from the current input z_d into present cell state c_t. This is done through sigmoid and tanh layer whereas sigmoid layer decides what to update in the current cell state as shown in equation 11 and tanh layer create a new vector comprised of updated values denoted as c'_t in equation 12.whereas W_{ig} and b_{ig} are weights and bias of input gate.

$$i_{gt} = \sigma(W_{ig}.[h_{t-1}, z_d] + b_{ig})] \tag{11}$$

$$c'_t = tanh(W_C.[h_{t-1}, z_d] + b_c) \tag{12}$$

After this step, old cell state c_{t-1} is being updated with new a cell state c_t by multiplying the values of forget gate f_{gt} with the old cell state c_{t-1} and adding new candidate values scaled by factor i_{gt} as shown in equation 13. In this way LSTM only remembers the key information from the current input.

$$c_t = f_{gt} * c_{t-1} + c'_t * i_{gt} \tag{13}$$

Finally, to generate what to show as an output, LSTM uses it's output gate denoted as o_{gt} which selects what to output from the updated cell state c_t then we pass the cell state through tanh (to get the values between 1 and -1) and multiply it with the output gate to display the desired output.

$$o_{gt} = \sigma(Wo_i.[h_{t-1}, z_d] + b_{og}) \tag{14}$$

$$h_t = o_t * tanh(c_t) \tag{15}$$

All the values from final hidden state h_t is forwarded to densely connected layer producing another hidden state $h'_t \epsilon \mathbb{R}^{32}$. At last, a dense layer with a softmax activation function takes h'_t and generate a probability distribution over K possible outcomes. Finally, the highest probability value is selected as a predicted label for the corresponding input.

1) INPUT AND OUTPUT PARAMETERS OF LSTM LAYER1 (ASPECT EXTRACTION)

In our first layer of LSTM, we have provided review sentences along with domain embeddings as input, in result it provides 6-dimensional vector $v \epsilon \mathbb{R}^6$. The output vector from layer 1 depicts a probability distribution over 6 aspect labels L = (General, Assessment, Behavior, Knowledge, Teaching

TABLE 2. Summary of layer 1 parameters.

Layer(Type)	Output Shape	Param#
Input_6 (Input Layer)	(None, 12477)	0
Embedding_6 (Embedding)	(None,2000,100)	129800
LSTM_6(LSTM)	(None,64)	42240
dense_11(Dense)	(None,32)	0
Dropout_6 (Dropout)	(None,32)	0.2 Low
dense_12(Dense)	(None,6)	198

TABLE 3. Summary of layer 2 parameters.

Layer(Type)	Output Shape	Param#
Embedding_6 (Embedding)	(None,2000,100)	600000
spatial_dropout1d_2 (Spatial)	(None,2000,100)	389648
LSTM_5(LSTM)	(None,196)	389648
dense_8(Dense)	(None,3)	788

Pedagogy and Experience). We choose the highest probability value as a predicted aspect label. In order to avoid overfitting during the training process, the dropout value of 0.2 is incorporated. Table 2 shows the values of the parameters used for layer 1.

2) INPUT AND OUTPUT PARAMETERS OF LSTM LAYER 2(OPINION ORIENTATION)

The second layer of our model is responsible for the prediction of a sentiment label of an extracted aspect term. We again use the LSTM neural network for performing this task but with little difference in a configuration as shown in table 3. As an input, this layer takes review sentences, domain embeddings and aspects predicted by layer 1. The Output of this LSTM layer is a 3 dimensional vector $v\epsilon\mathbb{R}^3$. This vector depicts a probability distribution over 3 sentiment orientations denoted as O = (Negative, Positive and Neutral). Finally, the highest probability value is considered a predicted orientation.

IV. EXPERIMENTS AND EVALUATION

In order to see the impact of the individual feature on the overall system, we conduct different experiments regarding the model parameters on training data. Based on that we selected a final model configuration that we used on test data.

A. DATASET AND RESOURCES

We evaluated our model performance on two datasets: the subset of Semeval-14 and on our own created academic domain dataset. unfortunately, there is no benchmark dataset containing students' reviews with ground truth values recorded for aspect categories and their relative polarities. Consequently, we used our constructed dataset having 5k tagged student's comments as per their aspect category (Teaching Pedagogy, Knowledge, Experience, Assessment, Behavior, and General) and orientation (Negative, Positive, Neutral). Table 4 depicts some statistics of our academic domain dataset.

Most research studies in ABSA evaluated their models on the standard data set of SemEval-14,15,16 which contain laptop and restaurant reviews. To show the generalizability of our model, we selected a subset of the SemEval-14 data set comprised of restaurant reviews, as the domain of this data set is quite different from academia. Table 5 shows some statistics of SemEval Data set.

TABLE 4. Academic domain dataset statistics.

No of Reviews	2180
No of Aspects	6
No of Sentences	5015
No of Orientation categories	3

TABLE 5. SemEval-14 dataset statistics.

Domain	Restaurant
No of Reviews	3041
No of Aspects	5
No of Orientation categories	2

TABLE 6. Word embedding statistics.

Word Embedding	Vocabulary	Dimension	Dataset
Academic Domain	12477	100	Students' feedback
OpinRank	18k	100	OpinRank Reviews
Glove.6B.100D	400K	100	Wikipedia 2014

We also evaluated our model performance by using three different word embeddings: Domain word embedding, standard pre-trained embeddings i.e.glove.6B.100D [40] and the word embedding generated from the OpinRank Review Dataset. The purpose behind using these two other embeddings was to evaluate which embedding works better for our model. Table 6 demonstrates some details about these word embeddings including their dimension, vocabulary size and from which dataset they were created.

B. MODEL TRAINING

As discussed earlier that the proposed system is comprised of two subtasks namely aspect extraction and orientation detection. To perform an experiment, we used 70% of data for model training, 10% for validation and 20% for testing. We used stochastic gradient descent to train our two layered LSTM model. We start our model training with a learning rate of 0.01 for 100 epochs (100 iterations over all samples in mini-batches of 34 samples). Simultaneously, we also monitored the loss and accuracy on the validation data. As it can be seen in Figure 6, the training loss was decreasing and the training accuracy was increasing with every epoch. At the nine epoch, both the loss and accuracy seem to reach peak value and after the 10th epoch, the model started over optimizing on the training data. In this case, to prevent such overfitting we trained a new network from scratch for nine epochs and then evaluated it on test data. We used categorical cross-entropy as a loss function as shown in equation 16.

$$E(y, y') = -\sum y(l)\, log\, y'(l) \tag{16}$$

Whereas y and y' denotes the expected and predicted probabilities for label l.

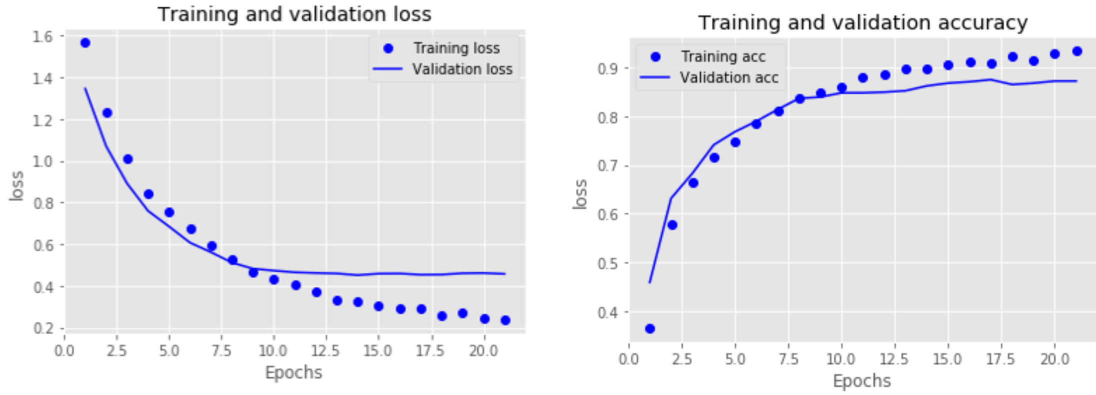

FIGURE 6. Optimal epoch.

The softmax function is used as an activation function in both layers. The reason behind using softmax function is that it generates a range of probability values as an output from 0 to 1, and the sum of those probabilities is equal to one. especially in the case of multi-classification models, it generates the probability for each specified class in which the target class will have the highest probability. As our model prior to softmax function generates output in different ranges, so to align those output values in the range of 0 and 1 and predicting the target class as in our case (Aspect and Orientation) softmax function is applied at the output layer as shown in equation 17.

$$Softmax(h'_t) = \frac{exp(h'_t)}{\sum_{k=1}^{k} exp(h_t)} \tag{17}$$

where, h'_t is the final hidden state output of LSTM layer ranges from -infinity to + infinity.

V. RESULTS AND FINDINGS

We evaluated our two layered LSTM model separately to better see the individual performance of each layer for its respective task. To evaluate the model performance, we selected four metrics used extensively in the domain of ABSA: Precision, Recall, F1 Score and Accuracy. The calculation of these metrics is done by using equation 18, 19, 20 and 21.

$$Precision = \frac{TruePositive}{(TruePositive + FalsePositive)} \tag{18}$$

$$Recall = \frac{TruePositive}{(TruePositive + FalseNegative)} \tag{19}$$

$$FScore = \frac{2 * Precision * Recall}{(Precision + Recall)} \tag{20}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{21}$$

First, We performed experiments by using different word embeddings in our model including glove.6B.100D, OpinRank and word embedding generated from the academic dataset. Result of using different word embedding in a model is demonstrated in Table 7 that clearly shows that model gives a reasonable performance (91% accuracy in aspect extraction and 93% in sentiment polarity detection) when domain word embedding is used as compared to pre-trained word embedding. The result of the first layer of aspect extraction is demonstrated in Table 11. The estimated F1 score for aspect extraction was found to be 0.85 with the precision and recall of 89% and 83% respectively. We have also demonstrated the results of each category. As it can be seen that we got the low F1 score in the assessment category because of the reason that training instances for assessment were less in number and due to that LSTM could not learn well about this aspect label category. So if we provide more training instances for this category, the score can be further improved.

The performance of our second layer of sentiment orientation detection demonstrated variant precision and recall values for the targeted category. The details are mentioned in Table. 10. The F1 values ranged from 0.79 to 0.96, with an average value of 0.86.

We have compared our model performance with the baseline studies conducted in the academic domain as shown

TABLE 7. Comparison of different word embeddings.

Word Embedding	Aspect Extraction	Sentiment Detection
glove.6B.100D	82%	85%
Domain Embedding	91%	93%
OpinRank embeddings	74%	79%

TABLE 8. Comparison of proposed model with other state-of-the-art.

	Sentiment Orientation Detection			
Approach	Precision	Recall	F1 Score	Accuracy
Online Sentiment Analyzer [35]	0.69	0.76	0.72	-
Supervised SVM [33]	-	-	-	78.7%
Naive Bayes [34]	-	-	-	89.67%
SVM Linear [25]	-	-	-	80%
Lexicon + Naive Bayes [29]	-	-	-	80%
Proposed Model	**0.88**	**0.85**	**0.86**	**93%**
	Aspect Extraction			
Approach	Precision	Recall	F1 Score	Accuracy
Naive Bayes [35]	0.76	0.84	0.80	-
Lexicon + Naive Bayes [34]	-	-	-	83.07%
Proposed Model	**0.89**	**0.83**	**0.85**	**91%**

TABLE 9. Comparison of proposed model on SemEval-14 with other state-of-the-art approaches.

Approach	Sentiment Orientation at SemEval-14 Accuracy
LSTM +Aspect Embedding [21]	77.2%
Sentic LSTM + Target attention [41]	76.47%
Aspect Fusion LSTM [42]	75.44%
BiGRU + Aspect Attention [43]	84.1%
Proposed Model	**85%**

Approach	Aspect Extraction at SemEval-14 F1 Score
Hierarchical LSTM [44]	77%
LSTM + Target attention [41]	73.82%
Bitmask Bidirectional LSTM [23]	80.1%
Proposed Model	**82%**

TABLE 10. Results of sentiment orientation detection.

	Precision	Recall	F1 Score
Positive	0.96	0.95	0.96
Negative	0.90	0.81	0.85
Neutral	0.79	0.80	0.79
Overall	**0.88**	**0.85**	**0.86**

TABLE 11. Results of aspect extraction.

	Precision	Recall	F1 Score
Assessment	0.90	0.62	0.73
Behavior	0.82	0.85	0.83
Teaching Pedagogy	0.89	0.83	0.85
Experience	0.93	0.90	0.91
Knowledge	0.87	0.88	0.87
General	0.95	0.90	0.92
Overall	**0.89**	**0.83**	**0.85**

in Table 8. As shown, our model outperformed the baseline approaches in both tasks of aspect extraction and sentiment orientation detection. We got 93% accuracy in the task of sentiment orientation detection and 91% accuracy in aspect extraction. Most research studies in ABSA evaluated their models on the standard data set of SemEval-14,15,16 which contain laptop and restaurant reviews. In order to check whether the proposed model can work in the different domain by slightly changing the input and output parameters, we also selected SemEval-14 data set, consisted of restaurant reviews, as the domain of this data set is quite different from academia. Table 9 shows the parameter set for this standard set. As it can be seen that SemEval-14 dataset has five labelled aspect categories and two sentiment orientations, so we changed the output parameters of aspect extraction and sentiment classification layer from 6 to 5 and from 3 to 2. Rest of the parameters were kept similar as shown in table 2 and 3. Moreover, we also applied our preprocessing step to this standard dataset before fedding to the LSTM model because without this step, model performance was degraded to some extent. Table 9 shows the performance of our proposed model on the SemEval-14 dataset along with the state of the art approaches conducted on this dataset. As depicted that the proposed model achieved reasonable performance on the standard dataset of SemEval-14 with 85% accuracy in sentiment polarity detection and F1 score of 82% in aspect extraction. From this, we can conclude that our model can also work in different domains for aspect extraction and sentiment polarity detection.

VI. CONCLUSION

Students while performing teacher's evaluation discuss several aspects of a teacher in their comments. Manual processing of these comments and reviewing every mentioned aspect polarity is a laborious job. Though, their thorough understanding acts as a metric for evaluating faculty teaching performance.

In this research, we proposed a system for automatically extracting aspects from the available text and their corresponding orientation. As no predefined aspects categories were available for teacher evaluation, therefore with domain expert's guidance we finalized six aspect categories including Teaching Pedagogy, Knowledge, Assessment, Experience, Behavior and General. Our model used two LSTMs models for the task of aspect extraction and polarity detection respectively. We also evaluated our model performance by using different word embedding and got promising results when we used domain embedding as an embedding layer. These domain embeddings were created through the skip gram model. We got satisfactory results by achieving 91% accuracy in aspect extraction and 93% accuracy in sentiment detection. Apart from using a manually tagged domain dataset, we also evaluated our model on a standard dataset of SemEval-14 and got satisfactory results.

LIMITATIONS AND FUTURE WORK

Although, our system demonstrated a reasonable performance in both tasks of aspect extraction and sentiment orientation detection shown in table 8 but there were few comments that were not classified correctly. One possible reason for such misclassification could be the presence of multiple aspects within the review sentence without any usage of connectives like "AND" & "BUT". For instance, consider these reviews from our academic domain data set:

"Outstanding in teaching as well as behavior towards the students."

"Sir was fair during the sessions he taught well the concepts"

More than one aspect could be addressed without using any connector, as shown in the above example. As per the demand of any supervised classification model, each sentence should be labeled with only one aspect. Therefore, such types of statements were misclassified and have been a major setback while calculating the accuracy of the system. So this is one of the limitation of our work that could be handled in future work. Moreover, this work of ABSA could be further extended by incorporating word dependencies, sentiment lexicon, and various NLP tools as an additional feature for creating domain embedding. Moreover, various neural network mechanisms like Gated RNN or CNN can also be used instead of LSTM to check system performance. As most of the domains now work on separately classifying the implicit and explicit aspects mentioned within the aspects. The task of aspect extraction could be further divided into these two subtasks. As this work relies on a supervised approach for ABSA, several unsupervised techniques like LDA can also be explored by specifying the number of clusters. Furthermore, the current system only handles comments given in the English language. As most of the students write comments in Roman Urdu so in future this system is further expanded by processing Roman Urdu and other natural language comments. Another aspect is to check the association of emoticons like a happy face, sad face, etc. with sentiment orientation. Usually, when students are writing feedback about a teacher, they take help from many symbols, apparently, the usage of different symbols and emoticons is orientations specific so its exploitation might increase orientation accuracy.

APPENDIX
INTERVIEW QUESTIONS

1. Do you consider conducting student feedback at the end of semester is a good way to measure teachers' performance?
2. Do you think the students evaluate the teacher fairly?
3. Have you ever gone through the teacher evaluation form of Sukkur IBA?
4. If yes? Does it cover all the parameters to evaluate faculty performance? Or would you like to suggest some improvements?
5. As an HOD/Registrar what do you think are the best traits of a teachers?
6. What core aspects would you consider while evaluating a faculty performance?
7. What aspects do you think should be only evaluated by students?
8. What type of question should never be asked from Students in evaluation?
9. From the management perspective, what additional aspects would you suggest to assess teachers performance?
10. Any further suggestion or comments.

REFERENCES

[1] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107, Mar./Apr. 2016.
[2] N. Altrabsheh, M. Cocea, S. Fallahkhair, and K. Dhou, "Evaluation of the SA-E system for analysis of students," in *Proc. IEEE 17th Int. Conf. Adv. Learn. Technol. (ICALT)*, Jul. 2017, pp. 60–61.
[3] S. Poria, E. Cambria, and A. Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network," *Knowl.-Based Syst.*, vol. 108, pp. 42–49, Sep. 2016.
[4] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao, "Coupled multi-layer attentions for co-extraction of aspect and opinion terms," in *Proc. 31st AAAI Conf. Artif. Intell.*, Feb. 2017, pp. 3316–3322.
[5] P. Schorer, "Occam's razor and computer-program testing," *IEEE Trans. Rel.*, vol. R-31, no. 3, p. 316, Aug. 1982.
[6] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*. Springer, 2012, pp. 415–463.
[7] A. Mukherjee and B. Liu, "Aspect extraction through semi-supervised modeling," in *Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, Long Papers*, vol. 1, Jul. 2012, pp. 339–348.
[8] E. Brunova and Y. Bidulya, "Aspect extraction and sentiment analysis in user reviews in Russian about bank service quality," in *Proc. IEEE 11th Int. Conf. Appl. Inf. Commun. Technol. (AICT)*, Sep. 2017, pp. 1–4.
[9] J. Khan and B. S. Jeong, "Summarizing customer review based on product feature and opinion," in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, Jun. 2016, pp. 158–165.
[10] S. Kalamdhad, S. Dubey, and M. Dixit, "Feature based sentiment analysis of product reviews using modified PMI-IR method," *Int. J. Comput. Trends Technol.*, vol. 34, no. 2, pp. 115–121, 2016.
[11] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," *Comput. Linguistics*, vol. 37, no. 1, pp. 9–27, 2011.
[12] W. Jin and H. H. Ho, "A novel lexicalized HMM-based learning framework for Web opinion mining," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, Jun. 2009, pp. 465–472.
[13] C.-Y. Zhang, X.-G. Hong, Z.-H. Peng, and Q.-Z. Li, "Extracting Web entity activities based on SVM and extended conditional random fields," *J. Softw.*, vol. 23, no. 10, pp. 2612–2627, 2012.
[14] A. S. Manek, P. D. Shenoy, M. C. Mohan, and K. R. Venugopal, "Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier," *World Wide Web*, vol. 20, no. 2, pp. 135–154, 2016.
[15] B. Wang and M. Liu, "Deep learning for aspect-based sentiment analysis," Stanford Univ., Stanford, CA, USA, Tech. Rep., 2015, pp. 1–9.
[16] H. Xu, B. Liu, L. Shu, and P. S. Yu, "Double embeddings and CNN-based sequence labeling for aspect extraction," May 2018, *arXiv:1805.04601*. [Online]. Available: https://arxiv.org/abs/1805.04601
[17] S. Jebbara and P. Cimiano, "Aspect-based sentiment analysis using a two-step neural network architecture," in *Semantic Web Challenges* (Communications in Computer and Information Science). Springer, 2016, pp. 153–167.
[18] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in *Natural Language Processing and Text Mining*. Springer, 2005, pp. 9–28.
[19] T. T. Thet, J.-C. Na, and C. S. G. Khoo, "Aspect-based sentiment analysis of movie reviews on discussion boards," *J. Inf. Sci.*, vol. 36, no. 6, pp. 823–848, 2010.
[20] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, "Adaptive recursive neural network for target-dependent twitter sentiment classification," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, 2014, pp. 49–54.
[21] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 606–615.
[22] M. Al-Smadi, O. Qawasmeh, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews," *J. Comput. Sci.*, vol. 27, pp. 386–393, Jul. 2018.
[23] A. Cahyadi and M. L. Khodra, "Aspect-based sentiment analysis using convolutional neural network and bidirectional long short-term memory," in *Proc. 5th Int. Conf. Adv. Inform., Concept Theory Appl. (ICAICTA)*, Aug. 2018, pp. 124–129.
[24] M. Tubishat, N. Idris, and M. A. M. Abushariah, "Implicit aspect extraction in sentiment analysis: Review, taxonomy, oppportunities, and open challenges," *Inf. Process. Manage.*, vol. 54, no. 4, pp. 545–563, 2018.
[25] A. Ortigosa, J. M. Martín, and R. M. Carro, "Sentiment analysis in Facebook and its application to e-learning," *Comput. Hum. Behav.*, vol. 31, pp. 527–541, Feb. 2014.
[26] C. Pong-Inwong and W. S. Rungworawut, "Teaching senti-lexicon for automated sentiment polarity definition in teaching evaluation," in *Proc. 10th Int. Conf. Semantics, Knowl. Grids*, Aug. 2014, pp. 84–91.
[27] L. C. Yu, C. W. Lee, H. I. Pan, C. Y. Chou, P. Y. Chao, Z. H. Chen, S. F. Tseng, C. L. Chan, and K. R. Lai, "Improving early prediction of academic failure using sentiment analysis on self-evaluated comments," *J. Comput. Assist. Learn.*, vol. 34, no. 4, pp. 358–365, 2018.
[28] M. A. Ullah, "Sentiment analysis of students feedback: A study towards optimal tools," in *Proc. Int. Workshop Comput. Intell. (IWCI)*, Dec. 2016, pp. 175–180.
[29] Q. Rajput, S. Haider, and S. Ghani, "Lexicon-based sentiment analysis of teachers' evaluation," *Appl. Comput. Intell. Soft Comput.*, vol. 2016, p. 1, Oct. 2016.
[30] A. Kumar and R. Jain, "Faculty evaluation system," *Procedia Comput. Sci.*, vol. 125, pp. 533–541, Feb. 2018.
[31] C. K. Leong, Y. H. Lee, and W. K. Mak, "Mining sentiments in SMS texts for teaching evaluation," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 2584–2589, 2012.

[32] F. F. Balahadia, M. C. G. Fernando, and I. C. Juanatas, "Teacher's performance evaluation tool using opinion mining with sentiment analysis," in *Proc. IEEE Region Symp. (TENSYMP)*, May 2016, pp. 95–98.
[33] J. Sultana, N. Sultana, K. Yadav, and F. Alfayez, "Prediction of sentiment analysis on educational data based on deep learning approach," in *Proc. 21st Saudi Comput. Soc. Nat. Comput. Conf. (NCC)*, Apr. 2018, p. 1–5.
[34] A. Kumar and R. Jain, "Sentiment analysis and feedback evaluation," in *Proc. IEEE 3rd Int. Conf. MOOCs, Innov. Technol. Educ. (MITE)*, Oct. 2015, pp. 433–436.
[35] G. S. Chauhan, P. Agrawal, and Y. K. Meena, "Aspect-based sentiment analysis of students' feedback to improve teaching–learning process," in *Proc. Inf. Commun. Technol. Intell. Syst.*, 2018, pp. 259–266.
[36] L. M. Rojas-Barahona, "Deep learning for sentiment analysis," *Lang. Linguistics Compass*, vol. 10, no. 12, pp. 701–719, 2016.
[37] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[38] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing [review article]," *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, Aug. 2018.
[39] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
[40] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Oct. 2014, pp. 1532–1543.
[41] Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, "Sentic LSTM: A hybrid network for targeted aspect-based sentiment analysis," *Cogn. Comput.*, vol. 10, no. 4, pp. 639–650, 2018.
[42] Z. Toh and W. Wang, "Dlirec: Aspect term extraction and term polarity classification system," in *Proc. 8th Int. Workshop Semantic Eval.*, Aug. 2014, pp. 235–240.
[43] J. Cheng, S. Zhao, J. Zhang, I. King, X. Zhang, and H. Wang, "Aspect-level sentiment classification with HEAT (hierarchical attention) network," in *Proc. Conf. Inf. Knowl. Manage.*, Nov. 2017, pp. 97–106.
[44] Y. Ding, J. Yu, and J. Jiang, "Recurrent neural networks with auxiliary labels for cross-domain opinion target extraction," in *Proc. 31st AAAI Conf. Artif. Intell.*, Feb. 2017, pp. 3436–3442.

IRUM SINDHU was born in Sukkur, Pakistan. She received the B.S. degree in computer science and the M.S. degree in data and knowledge engineering from Sukkur IBA University, Khairpur, Pakistan, in 2014 and 2019, respectively. From 2015 to 2017, she was an Instructor of IBA-IET Khairpur. She has been a Lecturer with the Computer Science Department, Sukkur IBA University, since 2018. Her research interests include opinion mining, social media mining, and deep learning.

SHER MUHAMMAD DAUDPOTA received the M.Sc. degree from Sindh University, Jamshoro, in 2002, and the M.E. and Ph.D. degrees from the Asian Institute of Technology, Thailand, in 2008 and 2012, respectively. He is a computer scientist, with a major interest in big data analytics, data mining, data sciences, multimedia data mining, and deep learning. Most of his research publications are in these areas. In addition, alongside his computer science career, he has also been involved in quality assurance work with Sukkur IBA University and nationally. He has also worked on a five years partnership between Sukkur IBA, the Johnson County Community College, and the University of Missouri Kansas City (UMKC). He has served as the project director on this partnership project. Under this partnership, IBA Sukkur with the support from the U.S. State Department, learned the best practices in assessment, distance education, early childhood education, mathematics, and English.

KAMAL BADAR is a Postdoctoral Fellow of the College of Business Administration, Prince Sultan University, Riyadh, Saudi Arabia. He studies the antecedents and consequences of strategic networks (inter- and intra-organizational networks). Adopting a social capital perspective, he empirically tests causal models comprising network and non-network variables. Most of his research has been carried out in a developing country or transition economy context. Previously, he has published in *Long Range Planning*, the *Journal of Manufacturing Technology Management*, *Scientometrics*, and the *Aslib Journal of Information Management*.

MAHEEN BAKHTYAR received the master's and Ph.D. degrees from AIT, Thailand, with one year research experience from the National Institute of Informatics, Tokyo, Japan. She is currently an Assistant Professor with the Department of Computer Science and Information Technology, University of Balochistan, Pakistan. Her research interests mainly include information/knowledge management and retrieval, natural language processing, sentiment analysis, text processing, language understanding, question answering systems, and ontology processing.

JUNAID BABER received the M.S. and Ph.D. degrees in computer science from the Asian Institute of Technology, Thailand. He spent one year as a Research Scientist with the National Institute of Informatics, Tokyo. He is currently a Faculty Member of the University of Balochistan, Quetta. His research interests include machine learning, high performance computing, and data analytics.

MOHAMMAD NURUNNABI received the Ph.D. degree from CMA, SFHEA FRSA FAIA (Acad). He is the Aide to the Rector on Research and Internationalization and the Chair of the Accounting Department, Prince Sultan University, Saudi Arabia. He is an Academic Visitor with Antony's College, University of Oxford, U.K. He serves as the Editor-in-Chief of *PSU Research Review*, an international journal.