

Understand Students Feedback Using Bi-Integrated CRF Model Based Target Extraction

K. Sangeetha^{1,*} and D. Prabha²

¹Department of Computer Science and Engineering, Panimalar Engineering College, Chennai, 600123, India

²Department of Computer Science and Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, 641008, India

*Corresponding Author: K. Sangeetha. Email: sangeethascholar1@gmail.com

Received: 09 April 2021; Accepted: 21 May 2021

Abstract: Educational institutions showing interest to find the opinion of the students about their course and the instructors to enhance the teaching-learning process. For this, most research uses sentiment analysis to track students' behavior. Traditional sentence-level sentiment analysis focuses on the whole sentence sentiment. Previous studies show that the sentiments alone are not enough to observe the feeling of the students because different words express different sentiments in a sentence. There is a need to extract the targets in a given sentence which helps to find the sentiment towards those targets. Target extraction is the subtask of targeted sentiment analysis. In this paper, we proposed the innovative model to find the targets of the given sentence using Bi-Integrated Conditional Random Fields (CRF). A Parallel fusion neural network model is designed to perform this task. We evaluate the model using the Michigan dataset and we build a dataset for target extraction from student reviews. The experimental results show that our proposed fusion model achieves better results compared to baseline models.

Keywords: Feedback; sentimental analysis; deeplearning; integrated CRF

1 Introduction

Student satisfaction is used to describe whether students are happy and contented and fulfilling their desires and need at studies. Many measures purport that students' satisfaction is a factor in students' motivation, students goal achievement in the institution. Factors contributing to student satisfaction include treating students with respect, empowering students, positive approach, teaching quality, student-staff relationship etc. When changes are introduced, there will be, in general, opposition from students. But if a student is very satisfied with their organization, they will whole heartedly accept, introduce, and implement the changes. Further, help keep the company on the right track. The facets of students' satisfaction measured vary from institution to institution, in which student feedback used to find the satisfaction level of a student. Student's feedback can highlight the different issues of the students about the lectures. They express their feelings or emotions about the lectures in the feedback. Student's feedback is the teaching-learning process equip the teachers as well as the students [1,2]. Using students feedback the faculty can try to improve the way of teaching style [3].



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Students can address different issues regarding teaching, courses, management, curriculum etc in the feedback. One of the best methods used to classify the emotions of the text is sentiment analysis. Sentiment analysis used in various applications to review the feedback. It is also powerful when it is applied to specific domain [4,5]. However, sentiment analysis at sentence level or document level cannot give a lot of careful information, therefore a fine-grained task, Target-based sentiment analysis. Target-based sentiment analysis consists of several subtask includes target extraction, target level sentiment classification, category detection of targets, etc. In which target extraction is a basic sub task. Targets are the opinion words used to express the sentiments explicitly. Most researchers work based on this subtask [6]. However, some fine-grained assessment is need in this technique.

We design a fusion model using deep learning methods to extract the targets. A core challenge in target extraction is to learn the context related to targets. To address this, we propose an architecture which combines Bidirectional LSTM and BIGRU with Glove embedding to encode the input sequence of sentences. This is used to find the relationship between the words. Bi-LSTM finds the future words depends on the current context. For example, in the below two sentences Bi-LSTM is used to find the next word after teddy is bears or Roosevelt.

- i) He said teddy bears on the sale.
- ii) He said teddy Roosevelt was a great president.

In addition, fusion architecture has included CRF. This allows better targets to be found with the help of probability scores.

Our main contributions are summarized as follows,

- We propose a Bi-Integrated CRF architecture for target extraction, which combines best approaches of deep learning includes Bi-directional LSTM, Bi-Directional GRU, CRF, Glove embeddings. It works like a parallel architecture.
- We evaluate our models using Michigan and real time data sets. We conduct extensive experiments on those datasets, and the result show that our integrated model performs better than the baseline models.
- We build the real-time data set using student reviews.

2 Related Work

Thinking of others has always been an important information. Student feedback is the information used to equip the teacher as well as the students. There are different mediums are available to collect student's feedback like oral feedback, survey based on questions, clickers, social media etc. Previous research narrated there are several mediums are used to collect students' feedback, but they are not useful without analysis. Traditional way of student's feedback based on specific question called as reflective prompt [7]. The opinion of the students collected based on this prompt. Different aspect of the students cannot be observed through this method. Another method to obtain the feedback is asking question directly to the students. In this method all students will not answer the questions [8,9] Clickers is a handheld device through which students gives the feedback for the questions asked by the lectures. Clickers has some demerits they are i) Students may cause damage to the device, ii) They forget to bring the device to the class, iii) Device distracts the students. Social media is very popular among students. Participation is more in providing feedback. But it has some disadvantages i) students get distracted, ii) time wasting and even students spend more time in social media because it is multipurpose tool. Sentiment Analysis is an application of Natural Language Processing. It is used to identify the polarity from a given text that is positive, negative or neutral. Sentiment analysis used in various applications like, movies, advertisements, products, cars, smart phones, tourism, and education [10,11] argues that students do not have proper equipment to express their emotions. They also argue that emotion of the students plays an important role

in learning because positive emotions influence the students to perform better, and negative emotions distract the students from learning. Previous methods have some challenges in sentiment analysis including i) similar lexical features shows different sentiments ii) Different style of writing but same sentiment iii) subject name and professor can appear in different forms, iv) sentiment lexicons are not sufficient for sentiment analysis v) feedback may not be genuine [12–14]. Socher argues that sentiment detection requires more supervised training and powerful evaluation features. In this paper we proposed the model which identifies the student's sentiment towards the courses and instructors as stated in their remarks. Targets are words through which opinion are expressed. It is very important because without knowing the target, opinion of the sentence has limited use.

3 Methodology

3.1 Bi-Integrated CRF Fusion Model

3.1.1 Task Formulation

Given sentence s is converted into tokens or words $(w_1, w_2 \dots w_n)$ using tokenization techniques. Stop words are removed and stemming is done to reduce the words to root by dropping unnecessary suffixes. Sentences are labelled using tags to simplify the target extraction process. (Ramshaw and Mitch Marcus. 1995) BIO Tagging scheme is used as a tagging technique. Each word w_i in sentence s is tagged as $t_i \in \{i, o\}$ (i-inside, o-outside). Our work concentrates only on courses and instructors. Hence, we tagged the courses and instructors tagged as i and others as o. For example, sentence TOC was not too hard is tagged in w_i/t_i style as follows:

Toc/i was/o not/o too/o hard/o (Target: TOC)

Further the length of words (n_words) and length of tags (n_tags) to equalize the length of the sentence to maximum length. Maximum length of sentence in 70.

3.1.2 BI-Integrated CRF Fusion Architecture

Fig. 1 shows the fusion architecture of our methods. We proposed an integrated fusion model to extract the targets from students' feedback. The fusion model is built using a parallel architecture that integrates BILSTM, BIGRU, and dropout layers with a global vector for word representation (Glove). The result of the parallel architecture is processed using CRF which finds out the eminent target. Dense layer is added to reduce the dimensions of the fully connected network. Detail description of each part explains in the following sections.

3.1.3 Word to Vectors

First step is to generate the input vectors for each word using glove word embedding. For this dictionary is used, which is represented as $\in R^d$, where d is embedding dimension. The sequence of words $(w_1, w_2 \dots w_n)$ converted to word vectors $((w_{v1}, w_{v2} \dots w_{vi} \dots w_{vm}))$ by dictionary mapping where $w_{vi} \in R^d$.

3.1.4 Parallel Architecture

Learning of context of the word is biggest challenge in the target extraction. As the places of the words differ, so does their context. It is evident that more processing is needed to find the targets. Therefore, integrating the connections between target word and context words is required when developing a learning system. So, we separately design two models: model_1 & model_2 to extract the target in a better way. Both models are built using BILSTM, BIGRU and dropout.

3.1.5 BILSTM

Bi directional LSTM (BILSTM) [15] used to find past and future context information. The input of the BILSTM is the sequence of word vectors from the embedding layer and it is processed along with the hidden

states. BiLSTM works in both forward $\overrightarrow{F_{at}}$ and backward direction $\overleftarrow{B_{bt}}$. Forward LSTM starts from first word to the target and backward LSTM starts from last word to target. Fig. 2 shows forward and backward LSTM. The output of the BiLSTM is the concatenation of the both the LSTM represented as y_t^1 for model_1, y_t^2 for model_2:

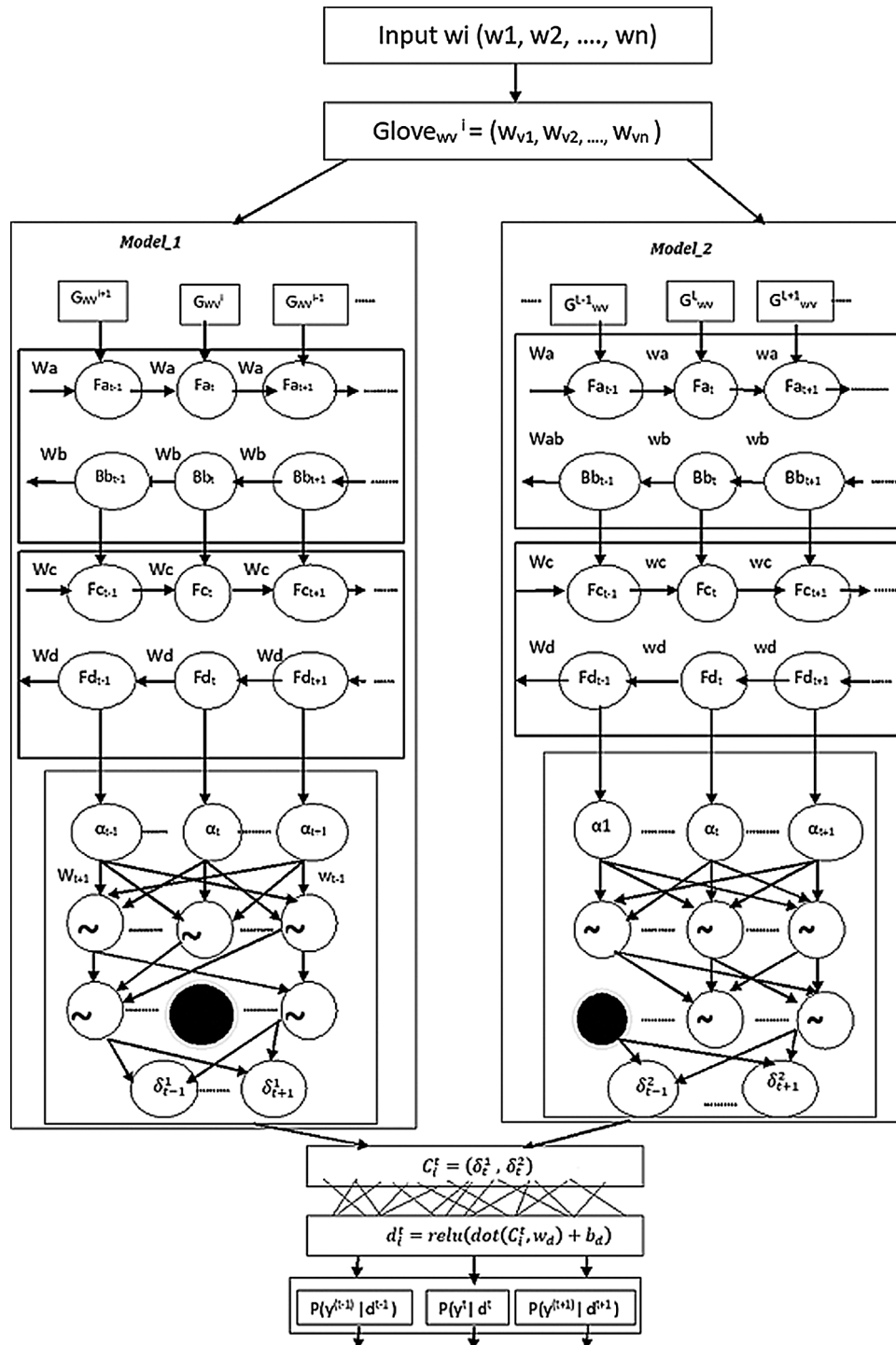


Figure 1: Bi-integrated CRF fusion model

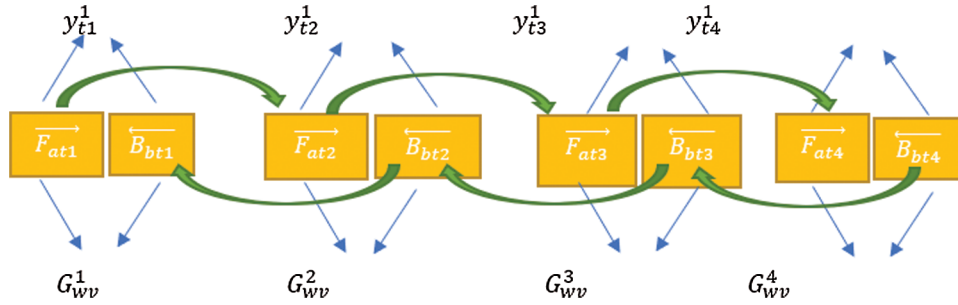


Figure 2: Forward and backward LSTM

We obtain forward and backward representation of model_1 Bi-LSTM represented as follows:

$$\overrightarrow{F}_{at}^1 = LSTM \left(w_F^1, \overrightarrow{F}_{at-1}^1 \right) \quad \forall at \in [1, \dots, i-1] \tag{1}$$

$$\overleftarrow{B}_{bt}^1 = LSTM \left(w_B^1, \overleftarrow{B}_{bt-1}^1 \right) \quad \forall bt \in [l+1, \dots, n] \tag{2}$$

$$y_t^1 = \sigma \left(W_y^1 \left[\overrightarrow{F}_{at}^1, \overleftarrow{B}_{bt}^1 \right] + b_y^1 \right) \tag{3}$$

Forward and backward representation of model_2 Bi-LSTM represented as follows:

$$\overrightarrow{F}_{at}^2 = LSTM \left(w_F^2, \overrightarrow{F}_{at-1}^2 \right) \quad \forall at \in [1, \dots, i-1] \tag{4}$$

$$\overleftarrow{B}_{bt}^2 = LSTM \left(w_B^2, \overleftarrow{B}_{bt-1}^2 \right) \quad \forall bt \in [l+1, \dots, n] \tag{5}$$

$$y_t^2 = \sigma \left(W_y^2 \left[\overrightarrow{F}_{at}^2, \overleftarrow{B}_{bt}^2 \right] + b_y^2 \right) \tag{6}$$

The outputs of BiLSTM layers are scores of each target represented as:

$$y_t^1 = (y_1^1, y_2^1 \dots y_n^1) \tag{7}$$

$$y_t^2 = (y_1^2, y_2^2 \dots y_n^2) \tag{8}$$

3.1.6 BiGRU

The scores of the targets are given to the BiGRU layer. BiGRU [16] is a combination of two GRU (forward and backward) which is used to find the key features using two gates i) reset gate ignores the irrelevant words and ii) update gate reserves the important information. The gates shall get the value between 0 and 1 where 0 means the resultant data is unimportant and 1 means resultant data is important. At time step t, GRU includes three parameters: reset gate r_t , update gate z_t and hidden state h_t . All three gates are represented using the following equations:

$$r_t = \sigma(W^r x_t + U^r h_{t-1}) \tag{9}$$

$$z_t = \sigma(W^z x_t + U^z h_{t-1}) \tag{10}$$

$$\tilde{h}_t = \tanh(W^h x_t + U^h \cdot (h_{t-1} \circ r_t)) \quad (11)$$

$$h_t = Z_t \cdot h_{t-1} + (1 - Z_t) \circ \tilde{h}_t \quad (12)$$

BiGRU is the concatenation of forward GRU $\overrightarrow{F_{ct}}$ and back ward GRU $\overleftarrow{B_{dt}}$ represented using the following equation,

$$\overrightarrow{F_{ct}} = GRU \left(w_F^1, \overrightarrow{F_{ct-1}} \right) \quad \forall at \in [1, \dots, i-1] \quad (13)$$

$$\overleftarrow{B_{dt}} = GRU \left(w_B^1, \overleftarrow{B_{dt-1}} \right) \quad \forall bt \in [l+1, \dots, n] \quad (14)$$

$$\alpha_t^1 = \sigma \left(w_\alpha^1 \left[\overrightarrow{F_{ct}}, \overleftarrow{B_{dt}} \right] + b_y^1 \right) \quad (15)$$

Forward and backward representation of model_2 Bi-LSTM represented as follows:

$$\overrightarrow{F_{ct}^2} = GRU \left(w_F^2, \overrightarrow{F_{ct-1}^2} \right) \quad \forall at \in [1, \dots, i-1] \quad (16)$$

$$\overleftarrow{B_{dt}^2} = GRU \left(w_B^2, \overleftarrow{B_{dt-1}^2} \right) \quad \forall bt \in [l+1, \dots, n] \quad (17)$$

$$\alpha_t^2 = \sigma \left(w_\alpha^2 \left[\overrightarrow{F_{ct}^2}, \overleftarrow{B_{dt}^2} \right] + b_y^2 \right) \quad (18)$$

The outputs of BiGRU layers are scores of each target represented as:

$$\alpha_t^1 = (\alpha_1^1, \alpha_2^1 \dots \alpha_n^1) \quad (19)$$

$$\alpha_t^2 = (\alpha_1^2, \alpha_2^2 \dots \alpha_n^2) \quad (20)$$

3.1.7 Dropout

Dropout layers are added at the end of both models to remove the overfitting value. Dropout is a regularization method used to approximate an input vectors while training a large number of neural networks with different architectures in parallel. During training some of the layer outputs are randomly ignored or dropped out. To achieve this a new hyperparameter is used that specifies the probability at which output of the layers are dropped out and when to retain. 0.5 probability is used for retaining output of each node and value close between 0.8 to 1.0 for retaining input. The dropout range for this model is given 0.01. The output from the dropout at time t is represented using the following equations:

$$\delta_t^1 = (\delta_1^1, \delta_2^1 \dots \delta_n^1) \quad (21)$$

$$\delta_t^2 = (\delta_1^2, \delta_2^2 \dots \delta_n^2) \quad (22)$$

3.1.8 Concatenation

The output from model_1 and model_2 is concatenated using the concatenation layer represented using the following equation:

$$C_i^t = (\delta_i^1, \delta_i^2) \quad (23)$$

3.1.9 Deep Layer

Each neuron in a layer is fully connected to all the neurons in the previous layer. All the combinations of features are learned from the previous layer. The values in the matrix are trained and updated using backpropagation. The output of the dense layer is 'm' dimensional vector used to change the dimensions of the vector.

$$d_i^t = \text{relu}(\text{dot}(C_i^t, w_d) + b_d) \quad (24)$$

Activation function is used for performing element wise activation and weight matrix is created by layer, bias vector is created by layer.

3.1.10 Conditional Random Field (CRF)

Input of CRF is the sequence of target scores d from dense layers and y a corresponding sequence of tags. CRF find the score of whole sequence using the correlations between tags in neighborhoods. We use a linear chain CRF using conditional probability:

$$P(y/d) = \frac{\exp(w.F(d, y))}{\sum_{y' \in Y} \exp(\text{score}(d, y'))} \quad (25)$$

$$F(d, y) = \sum_{i=0}^n f(y_{i-1}, y_i, d, i) \quad (26)$$

F is a feature function in a CRF defined using set of feature functions $f_1 \dots f_d$ where each feature function f_i analyses the whole input sequence d, the current word position y_i , the previous word position y_{i-1} , and current position i of word in the sentence. Expanded linear-chain CRF equation is:

$$P(y/d) = \frac{\exp\left(\sum_i \sum_j w_j f_j(y_{i-1}, y_i, d, i)\right)}{\sum_{y' \in Y} \exp\left(\sum_i \sum_j w_j f_j(y'_{i-1}, y'_i, d, i)\right)} \quad (27)$$

Length of input sequence d is represented as i, sum over all feature function is represented as j and weight for given feature function is w_j , f_j are feature functions. Y is sum over all possible label sequence. d feature function can access all of observation. Finally, CRF converts scores in to probabilities between 0 and 1. If the result is 0, CRF omits the word, 1 means takes the word. Results of the CRF is a sequence of eminent targets $T_i = T_{v1}, T_{v2} \dots T_{vi} \dots T_{vn}$.

4 Experimental Analysis

4.1 Datasets

We train our model on two datasets: i) Michigan Datasets, ii) Real time datasets represent in table. Both data sets comprise student reviews and their sentiment.

4.1.1 Michigan Datasets

Sentences are extracted from a Facebook student group and also from a survey with the students where students define their experience with classes in the Computer Science department at the University of Michigan. The data set consists of 1,042 sentences written by both graduates and undergraduates, describing about both classes and instructors. Sentences are annotated to identify the course and instructors. Sentiments are labelled as positive, negative and neutral. If there is no explicit sentiment and

no evident in the given sentence, then it is considered as neutral. [Tab. 2](#) shows the students feedback with the annotation contains the class id, instructors and sentiment towards that statement are expressed. In total, the 1,042 sentences include 976 class mentions and 256 instructor mentions, for a total of 1,232 entities.

Table 1: Dataset task details

Dataset	Task	Details
Michigan Datasets	Sentiment Classification	1062 sentences
Real time datasets	Sentiment Classification	300 sentences

Table 2: Students feedback with annotations

S. No	Student's feedback	Annotation
1	The labs in EECS 588 are very engaging	{class id: 588, department: EECS, Sentiment: positive}
2	When I took 561, Rosita Williford was the professor	{class id: 561, instructor: Rosita Williford, Sentiment: neutral}
3	I didn't like EECS 417	{class id = 417, department = EECS, Sentiment = negative}

4.1.2 Realtime Dataset

Real time data creation includes collecting data, pre-processing and sentiment annotation. [Tab. 3](#) shows the samples of real time dataset.

Table 3: Realtimestudents' feedback

Class	ANT	CN	THT	OOAD	TOC	MPMC
III CSE A	Good	Excellent. Need Notes	Good	Can Understand easily	Nice Teaching	Excellent. Need Notes
III CSE A	good in explanation	excellent	more real time examples needed	hands on training needed	need to solve more problems	Good
III CSE C	Good understandable	Good	Good	Good	Good	Good
III CSE B	Screen is visible late	Nil	Voice is breaking	Nil	Nil	Nil

4.1.3 Data Collection

We collect feedback from Panimalar engineering college in Chennai, which was collected through a google student survey form by third-year compute science students at the beginning of semester August 2020. Students give their views and issues of their courses, teachers, and facilities of the current semester through google forms. Datasets contain 300 sentences.

4.1.4 Pre-Processing

We conduct the pre-processing to normalize sentences. Feedback are short messages written freely by students about every course and instructors of current semester. It contains a lot of acronyms, spelling mistakes, emotions, etc.

4.1.5 Annotation

We annotated the sentences with sentiments useful for the classification process. [Tab. 3](#) shows the distribution of comments of student feedback of real time data set.

4.2 Parameters Used for Evaluations

70% is for training and 30% is for testing. For training the maximum length of sentences is configured as 70 using padding to equalize the length. In our model binary cross entropy-based loss function is used to learn the model parameters of neural network [17,18]. For training we use ReLu as an activation function, the size of initial word vectors of 300, the dropout rates are range between 0.2 to 0.5. A back propagation algorithm with rmsprop optimization method is used to train the network with learning rate of 0.0001. After each training epoch the network is tested on valid data [19]. The parameters used in the model are represented in [Tab. 4](#).

Table 4: Parameters in fusion model

Parameters	Value
Word vector dimension	300
Activation	Relu
Loss	Binary cross entropy
Optimizer	Rmsprop

5 Result and Discussion

5.1 Comparison Methods

We compare the proposed method with other baseline models in terms of prediction accuracy. For this purpose, we use the following baseline models [15]

- LSTM:** This method consists of three gates input gate, forget gate and output gate. Hidden vectors of sentences are feed into a linear layer which produces a output as a class number and add a softmax layer to output the probability of classifying the sentences as positive, negative and neutral.
- TD-LSTM:** Target dependent LSTM (modified LSTM) is used to process the left and right context surrounding the text string. This process is used to find the best features. This model improves the accuracy.
- TC-LSTM:** Target connection LSTM (modified TD-LSTM) which captures the interaction between target word and each context.
- Bi-Directional-CRF:** classifies the sentences according to the number of targets present in the sentences. Classified sentences are feed into a CNN separately for sentiment classification.

Among all models, the basic LSTM performs worst. This is not surprising because this work requires understanding of target dependent text semantics, while the basic LSTM model does not capture any target information. So that it predicts the same result for different targets in a sentence. TD-LSTM and TC-LSTM performs better than LSTM, the results reveal the important of target extraction. Comparing

with baseline models we find that Bi-Integrated CRF fusion model improves the accuracy. We believe that integrating BILSTM, BIGRU, CRF with glove improves the performance of proposed model.

The experiment results of different models on Michigan dataset are shown in Tab. 5 and their plots are visualized using the Fig. 3. There are two evaluation metrics accuracy and F1 scores are used to evaluate the Bi-Integrated CRF fusion model. Among 5 approaches, our approach outperforms than all the models. Our model gives relative improvement of 0.98% and 0.82% compared to other models on Michigan dataset. 32% and 17% improvement than LSTM, 28% and 12% improvement than TD-LSTM. 27% and 12% improvement than TC-LSTM, 0.36% and 5.59% improvement than Bi-directional CRF.

Table 5: Experimental results of different deep learning models on michigandataset

Models	Accuracy	F1Score
LSTM	66.50%	64.70%
TD-LSTM	70.80%	69.00%
TC-LSTM	71.50%	69.50%
Bi-directional-CRF	98.40%	75.98%
Bi-Integrated-CRF	98.76%	81.57%

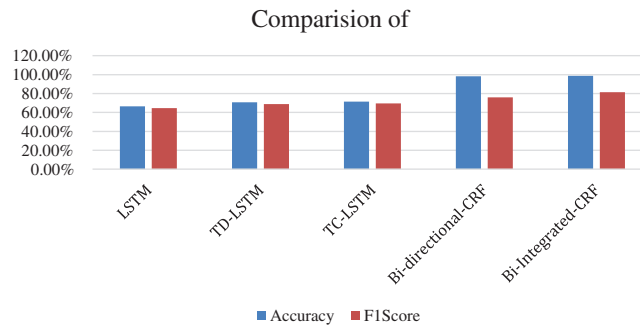


Figure 3: Experimental results of different deep learning models on Michigandataset

5.2 Evaluation Metrics

To measure the performance of the proposed model we used 4 evaluation metrics such as accuracy, F1 Score, Precision and Recall.

i) Accuracy

Sum of true positive and true negative divided by total number of value

$$TP = \frac{TP + TN}{TP + TN + FP + FN} \quad (28)$$

ii) Precision

It is the number of correct positive results divided by the number of positive results predicted by the classifier [19]

$$TP = \frac{TP}{TP + FP} \tag{29}$$

iii) Recall

It is the number of correct positive results divided by the number of all relevant samples [19]

$$TP = \frac{TP}{TP + FN} \tag{30}$$

iv) F1 Measure

F measure is used to evaluate the accuracy and is a harmonic mean of the precision and recall

$$F_1 = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{31}$$

The experiment results of different models on Real time dataset are shown in Tab. 6 and their plots are visualized using the Fig. 4. There are four evaluation metrics accuracy, precision, recall, F1 scores are used to evaluate the Bi-Integrated CRF fusion model. There is a slight improvement in the accuracy results of the proposed model i.e., 0.7% and 0.3% than LSTM and Bidirectional CRF Improvement of F1 score of the proposed model is 17% and 0.3% than LSTM and Bidirectional CRF. Improvement of precision of proposed model is 18% and 17% than LSTM and Bidirectional CRF. Improvement of precision of proposed model is 17% and 2% than LSTM and Bidirectional CRF.

Table 6: Experimental results of different deep learning models on real time dataset

Models	Accuracy	F1Score	Precision	Recall
LSTM	98.76%	64.99%	64.67%	65.85%
Bi-directional-CRF	99.13%	79.37%	65.85%	80.16%
Bi-Integrated-CRF	99.48%	82.38%	82.53%	82.38%

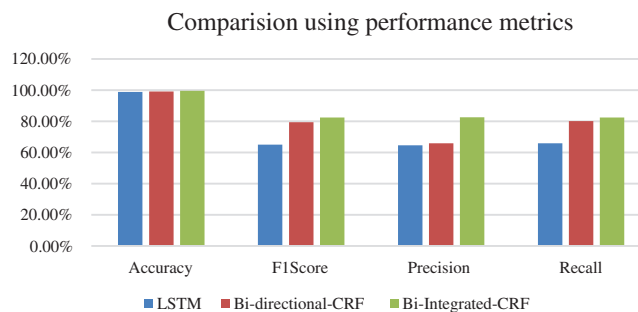


Figure 4: Experimental results of different deep learning models on real time dataset

The experimental results on the Michigan dataset, real time data sets. It can be observed that BI-Integrated CRF fusion model outperforms all the other approaches on the two data sets. Because, BI-Integrated CRF has more complicated networks and offers structure compatibility than other approaches. CRF integrated with parallel architecture gives better performance. BILSTM and BIGRU combination improves performance because they capture features on both past and future. Comparing

LSTM, TD-LSTM, TC-LSTM, Bi-Directional CRF combining parallel architecture with CRF can improve the performance of the model.

6 Conclusion

This paper has presented a novel subtask of target-based sentiment analysis. Bi-Integrated Fusion model use to extract the targets form a given students feedback sentence. Glove embedding layer is used to convert the input words in to corresponding vectors. Parallel architecture is built using BiGRU, BiLSTM and dropout layer can effectively extract the target information using future and past context. Then we combine the parallel architecture models using dense layer. The output of dense layer is feed into CRF which finds the probability scores of targets. The experimental results demonstrate that our models achieve the best performance than baseline models in both Michigan and Real time data sets. We observed that LSTM considers the polarity of whole sentence and ignores the targets. Integration of Bi-LSTM and Bi-GRU with CRF address this issue and improves the performance of the fusion model. In future work Bi-integrated can be utilized to improve the performance of SA task.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interests: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] B. Sortkar, "Feedback for everybody? exploring the relationship between students' perceptions of feedback and student's socioeconomic status," *British Educational Research Journal*, vol. 45, no. 4, pp. 717–735, 2019.
- [2] L. Murtagh, "The motivational paradox of feedback: Teacher and student perceptions," *Curriculum Journal*, vol. 25, no. 4, pp. 516–541, 2014.
- [3] N. Zarmeen, R. Quratulain and H. Sajjad, "Sentiment analysis of student feedback using machine learning and lexicon-based approaches," in *Proc. ICRIIS*, Langkawi, Malaysia, pp. 1–6, 2017.
- [4] T. Songbo, X. Cheng, Y. Wang and H. Xu, "Adapting naive bayes to domain adaptation for sentiment analysis," in *Proc. European Conf. on Information Retrieval*, Berlin, Heidelberg: Springer, pp. 337–349, 2009.
- [5] C. Tao, R. Xu, Y. He and X. Wang, "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," *Expert Systems with Applications*, vol. 72, pp. 221–230, 2017.
- [6] J. Liu, Y. Yang, L. V. S. Wang and H. Chen, "Attention-based BiGRU-CNN for chinese question classification," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–12, 2019.
- [7] W. Luo, F. Liu and D. Litman, "An improved phrase-based approach to annotating and summarizing student course responses," in *Proc. ICCL*, Japan, pp. 53–63, 2016.
- [8] R. Stowell and M. Nelson, "Benefits of electronic audience response systems on student participation learning, and emotion," *Teaching of Psychology*, vol. 34, no. 4, pp. 253–258, 2017.
- [9] T. Mantoro, M. A. Ayu, E. Habul and A. U. Khasanah, "Survey vote: A free web based audience response system to support interactivity in the classroom," in *Proc. ICOS*, Kuala Lumpur, Malaysia, pp. 34–39, 2010.
- [10] Z. Kechaou, B. A. Mohamed and A. M. Alimi, "Improving e-learning with sentiment analysis of users opinions," in *Proc. EDUCON*, Vienna, Austria, pp. 1032–1038, 2011.
- [11] C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno and J. Caro, "Sentiment analysis of facebook statuses using naive bayes classifier for language learning," *Information, Intelligence, Systems and Applications*, vol. 4, pp. 1–6, 2013.
- [12] D. M. E. M. Hussein, "A survey on sentiment analysis challenges," *Journal of King Saud University Engineering Sciences*, vol. 30, no. 4, pp. 330–338, 2018.
- [13] N. Ambreen, Y. Rao, L. Wu and L. Sun, "Issues and challenges of aspect-based sentiment analysis: A comprehensive survey," *IEEE Transactions on Affective Computing*, 2020.

- [14] R. Socher, A. Perelygin, J. Wu, J. Chuang, D. M. Christopher *et al.*, “Recursive deep models for semantic compositionality over a sentiment treebank,” in *Proc. EMMNLP*, pp. 1631–1642, 2013.
- [15] D. Tang, B. Qin, X. Feng and T. Liu, “Effective LSTMs for target-dependent sentiment classification,” in *Proc. ICLTP*, Dublin, Ireland, pp. 3298–3307, 2016.
- [16] L. Yuan, Z. Zeng, Y. Lu, X. Ou and T. Feng, “A character-level BiGRU-attention for phishing classification,” in *Proc. ICICS*, Beijing, China, pp. 746–762, 2019.
- [17] G. E. Hinton, N. Srivastava and A. Krizhevsky, “Improving neural networks by preventing coadaptation of feature detectors,” *Computing Research Repository*, 2012.
- [18] C. Welch and R. Mihalcea, “Targeted sentiment to understand student comments,” in *Proc. ICCLL*, Osaka, Japan, pp. 2471–2481, 2016.
- [19] A. Mishra, “Metrics to evaluate your machine learning algorithm,” *published in towards data science*, Feb 24, 2018. [Online]. Available: <https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234>.