

Using BiLSTM Structure with Cascaded Attention Fusion Model for Sentiment Analysis

J Sangeetha^{1*} & U Kumaran²

¹Department of Computer Applications, ²Department of Computer Science and Engineering, Noorul Islam Centre for Higher Education, Kumaracoil, Thuckalay 629 180, Tamil Nadu, India

Received 12 July 2022; revised 18 September 2022; accepted 06 October 2022

In the last decade, sentiment analysis has been a popular research area in the domains of natural language processing and data mining. Sentiment analysis has several commercial and social applications. The technique is essential to analyse the customer experience to develop customer loyalty and maintenance through better assistance. Deep Neural Network (DNN) models have recently been used to do sentiment analysis tasks with promising results. The disadvantage of such models is that they value all characteristics equally. We propose a Cascaded Attention Fusion Model-based BiLSTM to address these issues (CAFM-BiLSTM). Multiple heads with embedding and BiLSTM layers are concatenated in the proposed CAFM-BiLSTM. The information from both deep multi-layers is merged and provided as input to the BiLSTM layer later in this paper. The results of our fusion model are superior to those of the existing models. Our model outperforms the competition for lengthier sentence sequences and pays special attention to referral words. The accuracy of the proposed CAFM-BiLSTM is 5.1%, 5.25%, 6.1%, 12.2%, and 13.7% better than RNN-LSTM, SVM, NB, RF and DT respectively.

Keywords: CAFM, Deep learning, Deep neural network, Long short-term memory, Natural language processing

Introduction

Sentiment analysis is a set of tools, strategies, and procedures for detecting and extracting information. The data gathered comprises the users' attitudes and opinions. The primary goal is to determine whether a user has a negative, good, or neutral view of a product or something else. The quantity of articles on sentimental analysis has exploded in recent years. It is one of the most rapidly expanding research fields. Later, in the mid-2000s, firms mostly concentrated on product reviews that were available on the internet. This is quite beneficial in financial market forecasting.

It is impossible to do this task solely by human processing, given the constant expansion of data volume. This fact encourages sentiment analysis technologies to progress in a more sophisticated path.^{1,2} Machine Learning (ML) shines brightly in Natural Language Processing (NLP) problems because to its great modelling capacity.³ Although machine learning is quite advanced, its fundamental weaknesses have resulted in a slew of problems.⁴ To begin with, creating a sentiment dictionary is a labor-intensive and time-consuming procedure that necessitates some past knowledge. Second, the

implementation of features selection procedures and parameter tweaking is critical to their success. The use of these methods has been considerably curtailed as a result of the aforementioned flaws.

Deep Learning (DL) techniques are increasingly being included into NLP tasks.⁵ Deep learning's ability is the vast quantity of raw data without relying on prior predictive information makes these approaches particularly appealing for sentiment analysis. Deep learning could provide solutions to natural language ambiguity and complexity that aren't present in classic text mining methods. RNN with LSTM units, for example, uses hierarchical architectures. DL uses the attention mechanism of humans to give more weight to the essential pieces of information that are more important in making sentiment judgments.

In this paper, we propose a Cascaded Attention fusion model-based BiLSTM to address these issues (CAFM-BiLSTM). Multiple heads with embedding and LSTM layers are concatenated in the proposed CAFM-BiLSTM.

Related Works

The IMAN model, used an attention mechanism to represent the relationship between context and aspect word.⁶ To begin, they used pre-trained BERT word

*Author for Correspondence
E-mail: sangiprathap@gmail.com

vectors as the initial embedding, laying the groundwork for generating cutting-edge experimental results. However, in error analysis, the model occasionally failed to manage them properly. Clearly, present attentive procedures can be improved more. It is recommended to add disagreement regularization and applying sophisticated information aggregation to better leverage the diversity of multi-head attention.⁷ They proposed using the routing-by-agreement approach to better capture the various information dispersed in the extracted partial-representations. The estimation of rating score for a product depending on the target user is one of four crucial processes in the suggested recommendation system. The suggested system uses a filter based on demographic age group and then a product recommendation classifier based on fuzzy logic to consider known and unknown products using the similarity score. Depending on the reviews with the corresponding target user, the suggested technique determines the product rating score. Finally, in the decision-making process, fuzzy rules are used to anticipate related recommendation products using SA and ontology alignment. It was discovered that the output disagreement term and the EM routing algorithm produced the best results and were mutually beneficial.

An intrinsic effect of user reviews on social recommendation was investigated in the literature which yielded prospective model for sentiment analysis.⁸⁻¹⁰ Consumer ratings and the No. of reviews are external influence aspects in determining the essential products for the consumer.^{8,11} The user internal effect component boosted the forecasting efficiency of the recommendation system, according to their findings. Following the collection of an adequate dataset, NLP-based approaches are used to reprocess the data (tweets), and then a feature extraction method is used to extract sentiment-relevant features. Finally, a model is trained and tested on test data using machine learning classifiers. The Apache Spark framework is utilized in this project. Customers' reviews were sorted into 8 emotions (fear, anger, trust, sadness, anticipation, surprise, joy and disgust) and two sentiments using the NRC emotion lexicon (+ve and -ve). Their findings suggest that SA can assist in identifying customer behaviors and overcoming risks in order to satisfy customers. To increase the resilience and accuracy of sentiment analysis approaches, However, considerable work on the Aspect Based Sentiment Analysis (ABSA) detailed in this research remains to

be done due to a lack of data. With regard to a self-constructed stacking system, the model includes fewer data points, and there is still space for development in tasks involving a solo domain.

The intelligent system that uses fuzzy logic and LSTM can classify customer review statements into four categories: highly positive, positive, negative, and highly negative.⁹ This model was tested using Amazon mobile phone reviews, Amazon video games reviews, and Amazon goods customer reviews benchmark datasets, yielding accuracy of 96.93%, 83.82%, and 90.92%, respectively. In addition, to improve sentiment identification, a feature focus attention approach was proposed. Lin *et al.* proposed a hybrid attention network for ABSA that can be employed for tasks with ease.¹⁰ To learn the new architecture for gathering, filtering, and evaluating YouTube comments for a specific product. Unproductive and comments related to videos were separated from comments about products using a classification approach. Following that, the classification findings were employed in the target product's sentiment analysis, where a novel approach was applied to identify both the sentiments of the mentioned subjects Arabic Sentiment Analysis (ASA) and overall sentiment Social media Sentiment Analysis (SSA) in the phrase would benefit ABSA's responsibilities.

Proposed CAFM-BiLSTM

Our model accepts user reviews as input and outputs the polarity of the reviews. Embedding, Multi-head attention, Dropout, BiLSTM and Dense layers are all used in this model.¹¹⁻¹³ The Overview of the proposed CAFM-BiLSTM for Sentiment analysis is shown in Fig. 1. The following are the main steps in the proposed model: i) Get the text's word embedding; ii) Enter the word vector into the Multi-head Attention mechanism; iii) get sentiment classification using BiLSTM.

Embedding Layer

The user text is converted to vectors by embedding layers. All of the data in the input should be encoded as integers. The EML accepts an integer as input and looks up the internal dictionary to return the dense vector associated with it. Its initial weight is determined at random. These word vectors are gradually adjusted using the Back Propagation (BP) technique. In our proposed model, we use the Glove embedding layer, which is one of the most prominent embedding approaches. It is a publicly accessible pre-

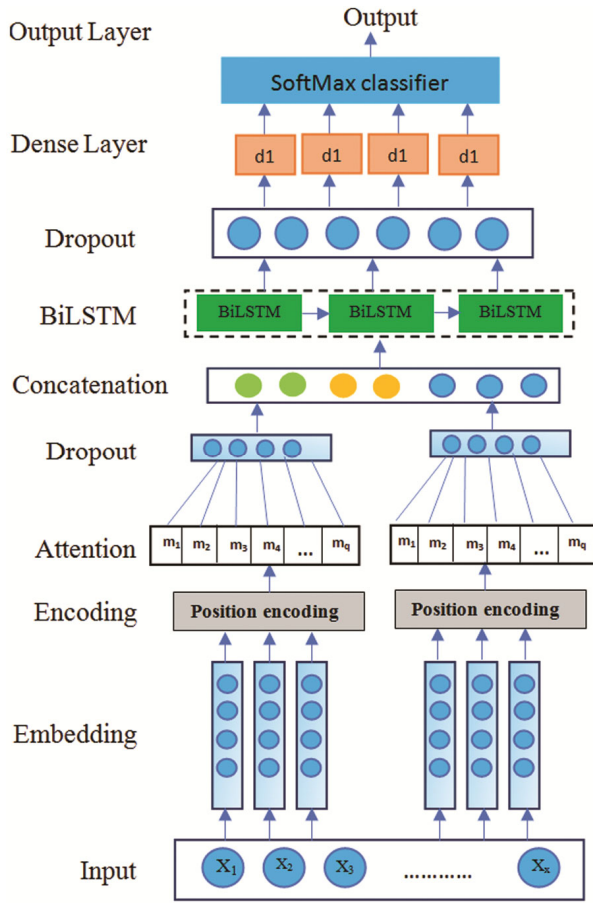


Fig. 1 — Overview of the proposed CAFM-BiLSTM for sentiment analysis

trained word embeddings approach. The words were trained using a data set. A variety of embedding vector sizes are available. The files will be available in text format after they have been downloaded. Glove creates a matrix of the provided word's co-occurrences.

$$N = \sum_{p,q=1}^m f(Y_{pq})(z_p \tilde{z}_q + a_p + \tilde{a}_q - \log(Y_{pq}))^2 \dots (1)$$

where, Y_{pq} is the co-occurrence matrix, $z_p \tilde{z}_q$ denotes the left right index and a_p, \tilde{a}_q are bias words.

Multi-head Attention Layer

Multi-head Attention is an improved version of the classic attention mechanism that also outperforms it. The structure of the MHAT mechanism is depicted in Fig. 2. To begin, K, M, and S are transformed. One head is calculated at a time in this method. There are many attention heads in use. The outputs are,

$$head_p = attention(KW_p^K, MW_p^M, SW_p^S) \dots (2)$$

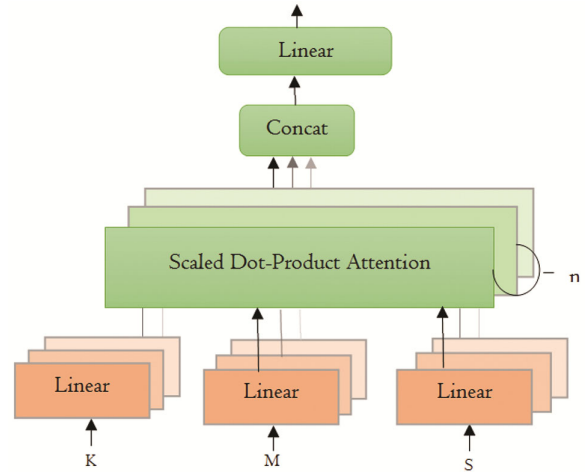


Fig. 2 — Overall structure of the MHA mechanism

$$Multihead(K, M, S) = Concat(head_1, \dots, head_n)W^o \dots (3)$$

BiLSTM Layer

The LSTM units serve as the building blocks for the BLSTM layers.^{14,15} The cell state, which comprises the information preserved by the LSTM unit, is the most important feature of an LSTM unit. The amount of information that the LSTM unit can include or erase from its cell state is controlled by gate structures. The LSTM unit additionally has a hidden state that records the observed sequence's history.

LSTM differs from traditional neural networks in that each LSTM unit has a unique feature. One memory cell M_t and 3 gates, comprising the forget gate o_t , the input gate g_t , and the output gate r_t , make up the unit. The three gates work together to manage the state of the memory cell M_t . The forget gate decides whether historical unit data is deleted or retained. The input gate is supplied to the unit from the inputs, while the output gate controls the unit's output. LSTM's forward calculation at time t can be expressed mathematically as follows.

$$o_t = \sigma(Y_o k_{t-1} + Q_o y_t + c_o) \dots (4)$$

$$g_t = \sigma(Y_g k_{t-1} + Q_g y_t + c_g) \dots (5)$$

$$M_t = o_t \odot M_{t-1} + g_t \odot \tanh(Y_h k_{t-1} + Q_h y_t + c_h) \dots (6)$$

$$r_t = \sigma(Y_r k_{t-1} + Q_r y_t + c_r) \dots (7)$$

$$k_t = r_t \odot \tanh(M_t) \dots (8)$$

where, k_t is the output of the LSTM unit t and σ is the sigmoid function, $\tanh()$ denotes activation function,

\odot is the Hadamard product, Y & Q are weight matrixes and c are the bias vector.

It would be very helpful for the sequence of tasks if you could access future information in the same way you could access prior information. The following is the operating principle: the forward layer captures the sequence's previous information, while the backward layer captures the sequence's future information.

$$\vec{k}_t = L(y_t, \vec{k}_{t-1}, M_{t-1}) \quad \dots (9)$$

$$\overleftarrow{k}_t = L(y_t, \overleftarrow{k}_{t+1}, M_{t+1}) \quad \dots (10)$$

where, $L(\cdot)$ signifies the LSTM hidden layer's operation. The text feature is obtained by combining the forward and backward output vectors, which are \vec{k}_t and \overleftarrow{k}_t respectively. It's important to know that H stands for the number of hidden layer cells:

$$H_t = \vec{k}_t \parallel \overleftarrow{k}_t \quad \dots (11)$$

Dropout Layer

Instead of looking into irrelevant parameters, this layer is utilized to focus on the target future parameters. When the network is utilized in training, full utilization of all data which is called true state and this time no data points are drooped. The model.fit, training python inbuilt function is suitably set to True inevitably.

Dense Layer

The network layer's neurons are contained in a dense layer. The neurons in the dense layer receive information from all other layers.

$$d = b(x.w) + a \quad \dots (12)$$

where, b is the element-wise argument, w is the weights matrix, and a is the layer's bias vector. Value of a is called as bias value which allows you to shift the activation function by adding a constant to the input.

SoftMax Classifier

We send the generated vector directly to the SoftMax layer. The following is the outcome of the prediction:

$$x = \text{softMax}(V) = \frac{\exp(V)}{\sum_{k=1}^c \exp(V)} \quad \dots (13)$$

Working of CAFM-BiLSTM for Sentiment Analysis

Given a set of n input sequences each containing m words, Let $X = X_1, X_2, X_3, \dots, X_n$ be the set of reviews and $X_n = c_1, c_2, \dots, c_m$ be the set of words.

The features are extracted from the raw data after it has been pre-processed. Each input sentence in our study has 70 features, which will be supplied into the parallel embedding layers Glove. The embedding layer generates 300 features. The features are again sent to the position encoding layer, which uses the distance between the words in the sequence to determine the location of the word. This layer produces 300 features as a result of its work. The embedding layer's output is sent into the many heads, with a 512 output feature. The concatenation layer, which consists of 512 features, will combine the results from multiple heads. The total number of features will be 1024. This will be fed into the BiLSTM model, which will return 100 features. The BiLSTM algorithm determines the association between the target words and decreases the number of features from 1024 to 100. At the end of the BiLSTM, a dropout layer is introduced. The three features will be dense layer output, dense layer input, and dense layer output.

Results and Discussion

The proposed model CAFM-BiLSTM was analyzed using Amazon Product reviews dataset. The sample of amazon product review is shown in Fig. 3.

We utilize recall, precision-score and accuracy values to evaluate the classification models' performance that are computed as in following equations respectively.

$$PR = \frac{TP}{TP+FP} \quad \dots (14)$$

$$R = \frac{TP}{TP+FN} \quad \dots (15)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad \dots (16)$$

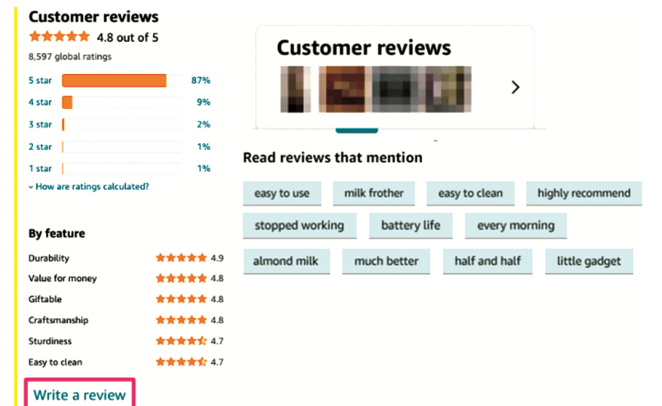


Fig. 3 — Sample for Amazon product review

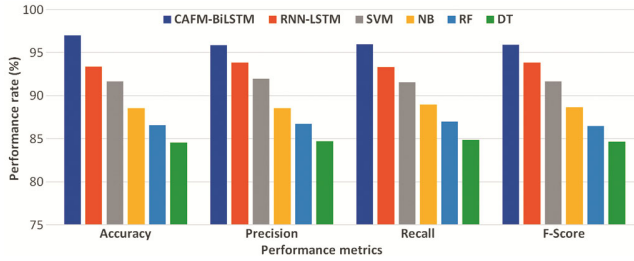


Fig. 4 — Comparison of performance metrics

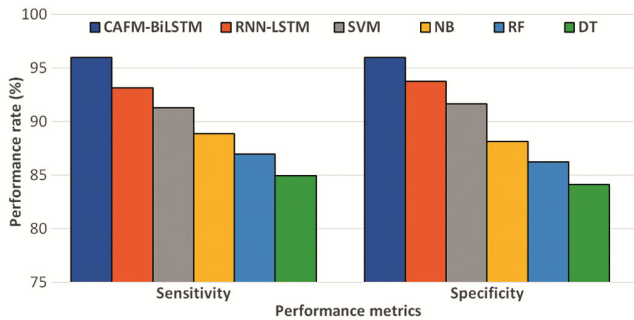


Fig. 5 — Comparison of sensitivity and specificity

$$F - Score = \frac{2(PR.R)}{PR+R} \dots (17)$$

where, If a positive review is classified as positive, TP is used. TN is used when a negative review is classified as negative. Negative reviews are labelled as bad, positive reviews are labelled as positive, and negative tweets are labelled as negative. FP denotes false positive sometimes in appropriate results induces FP.

We can see from Fig. 4, that the proposed model CAFM-BiLSTM has a higher accuracy of 96.988%. The proposed model has an F-score of 95.90%, precision of 95.89%, and recall of 95.98%. The proposed model has 2.12% more precision and 2% higher F-score than the existing RNN-LSTM model.

The proposed model CAFM-BiLSTM has a Sensitivity of 95.95%, and Specificity of 95.98% as shown in Fig. 5. The proposed model had more sensitivity and greater specificity than the existing RNN model when compared to the existing models.

The true positive rate vs the false positive rate for a multi classifier at various thresholds is calculated and plotted to form the ROC curve. The proposed model CAFM-BiLSTM is closer to the top-left corner, indicating greater efficiency, as seen in the Fig. 6.

The percentage of your test data that is correctly categorised is referred to as accuracy. We use the training data to train the model and test its performance on both the training and testing sets Figs. 7 (a & b). As the number of epochs grows, both the testing and training accuracy curves move upward.

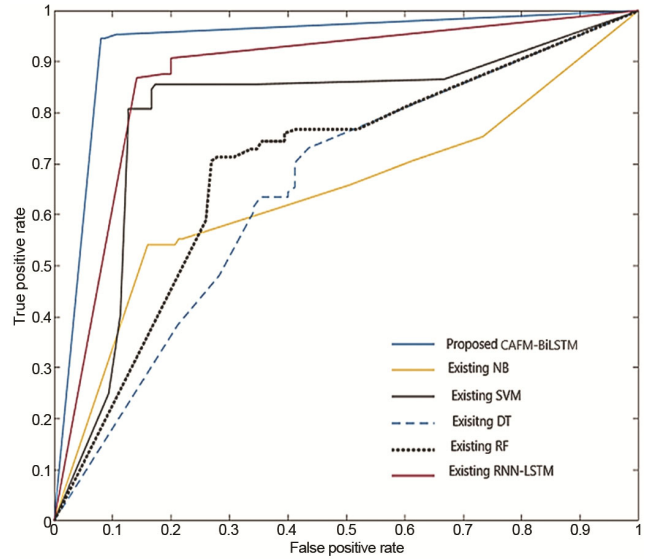


Fig. 6 — ROC curve

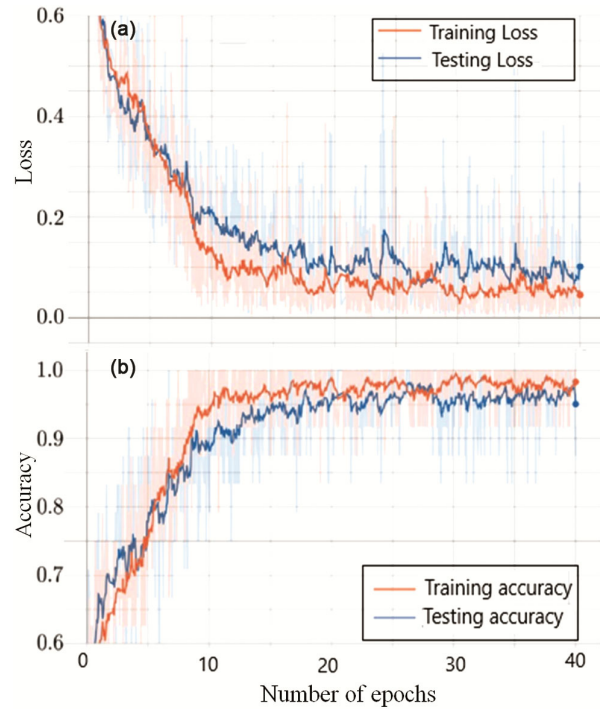


Fig. 7 — Performance curves: (a) Loss, (b) Accuracy

Conclusions

The BiLSTM with CAFM mechanism is used in this paper to introduce a sentiment analysis approach. As a result, the attention mechanism is discussed in this paper. Multiple heads with embedding and BiLSTM layers are concatenated in the proposed method. To improve the accuracy of the proposed method, our proposed model outperforms for longer sentence sequences and gives special

attention to referral words. The proposed model has a 95.90% F-score, 95.89% precision, and 95.98% recall. When compared to the current RNN-LSTM model, the suggested model has 2.12% more precision and a 2% higher F-score. Compared to RNN-LSTM, SVM, NB, RF, and DT, the proposed CAFM-BiLSTM is 5.1%, 5.25%, 6.1%, 12.2%, and 13.7% more accurate.

References

- 1 Mohbey K K, Sentiment analysis for product rating using a deep learning approach, *Int Conf Artif Intell Smart Syst (IEEE)* 2021, 121–126.
- 2 Balakrishnan V, Shi Z, Law C L, Lim R, Teh L L & Fan Y, A deep learning approach in predicting products' sentiment ratings: a comparative analysis, *J Supercomput*, **78(5)** (2022) 7206–7226.
- 3 Naresh A & Venkata Krishna P, An efficient approach for sentiment analysis using machine learning algorithm, *Evol Intell*, **14(2)** (2021) 725–731.
- 4 Dubey T & Jain A, Sentiment analysis of keenly intellectual smart phone product review utilizing SVM classification technique, In *2019 10th Int Conf Comput Commun Netw Technol (IEEE)*, 2019, 1–8.
- 5 Hasan MR, Maliha M & Arifuzzaman M, Sentiment analysis with NLP on Twitter Data, *Int Conf Comput Commun Chem Mater Electron Eng (IEEE)* 2019, 1–4.
- 6 Zhang Q, Lu R, Wang Q, Zhu Z & Liu P, Interactive multi-head attention networks for aspect-level sentiment classification, *IEEE Access*, **7** (2019) 160017–160028.
- 7 Li J, Wang X, Tu Z & Lyu M R, On the diversity of multi-head attention, *Neurocomput*, **454** (2021) 14–24.
- 8 Cheng L C, Chen Y L & Liao Y Y, Aspect-based sentiment analysis with component focusing multi-head co-attention networks, *Neurocomput*, **489** (2022) 9–17.
- 9 Wang X, Tang M, Yang T and Wang Z, A novel network with multiple attention mechanisms for aspect-level sentiment analysis, *Knowl Based Syst*, **227** (2021) 107196.
- 10 Lin Y, Fu Y, LiY, Cai G & Zhou A, Aspect-based sentiment analysis for online reviews with hybrid attention networks, *World Wide Web*, **24(4)** (2021) 1215–1233.
- 11 Yang J, Zou X, Zhang W & Han H, Microblog sentiment analysis via embedding social contexts into an attentive LSTM, *Eng Appli Artif Intell*, **97** (2021) 104048.
- 12 Xu Q, Zhu L, Dai T & Yan C, Aspect-based sentiment classification with multi-attention network, *Neurocomput*, **388** (2020) 135–143.
- 13 Xu G, Meng Y, Qiu X, Yu Z & Wu X, Sentiment analysis of comment texts based on BiLSTM, *IEEE Access*, **7** (2019) 51522–51532.
- 14 Rehman A U, Malik A K, Raza B & Ali W, A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis, *Multimed Tools Appl*, **78(18)** (2019) 26597–26613.
- 15 Muhammad P F, Kusumaningrum R & Wibowo A, Sentiment analysis using Word2vec and long short-term memory (LSTM) for Indonesian hotel reviews, *Proc Comput Sci*, **179** (2021) 728–735.