Context Aware Contrastive Opinion Summarization

S. K. Lavanya[1(✉)] and B. Parvathavarthini[2]

[1] Jerusalem College of Engineering, Chennai, India
sklavanyasambath@gmail.com
[2] St. Joseph's College of Engineering, Chennai, India
parvathavarthini@gmail.com

Abstract. Model-based approaches for context-sensitive contrastive summarization depend on hand-crafted features for producing a summary. Deriving these hand-crafted features using machine learning algorithms is computationally expensive. This paper presents a deep learning approach to provide an end-to-end solution for context-sensitive contrastive summarization. A hierarchical attention model referred to as Contextual Sentiment LSTM (CSLSTM) is proposed to automatically learn the representations of context, feature and opinion words present in review documents of each entity. The resultant document context vector is a high-level representation of the document. It is used as a feature for context-sensitive classification and summarization. Given a set of summaries from positive class and a negative class of two entities, the summaries which have high contrastive score are identified and presented as context-sensitive contrastive summaries. Experimental results on restaurant dataset show that the proposed model achieves better performance than the baseline models.

Keywords: Opinion summarization · Attention model · Pointer-generator network · Context-aware sentiment classification

1 Introduction

Accumulating opinions from review documents is termed opinion summarization. There are two types of opinion summarization techniques: extraction- and abstraction-based summarization. A summary is developed by selecting certain prominent input sentences in extraction-based techniques. Abstractive summarization techniques produce novel summaries by rephrasing sentences which are rather abstractive in nature. A feature-based summarization is an abstraction-based method that presents the opinion distribution of each feature separately and helps users to make better decisions.

Though different summarization techniques have been proposed, there is still a need for summarizing different opinions about a particular feature of different products. Contrastive opinion summarization methods produce a summary consisting of a set of contrastive sentence pairs by choosing the most representative

and comparable sentences from two sets of positively and negatively opinionated sentences. Such a summary helps users easily interpret contrastive opinions about a product or service. Most of the present works on contrastive summarization select contrastive sentences only based on the explicit opinion words present. The context information is also necessary to understand the implicit opinion present in the sentences. For example, the sentence "city gas mileage is horrible", does not convey a negative sentiment always. It conveys that for city traffic, mileage is not good. It would be helpful for the users if the context (reason) is also considered while producing summaries.

Context refers to the circumstances in terms of which an event or entity can be fully understood. Text data is associated with rich context information. Meaning of an unknown word can be guessed by looking at the context words surrounding that word [7]. Deriving these context words using a machine learning algorithm is computationally expensive [3,6]. Model-based approaches rely on hand-crafted features produced by machine learning algorithm for producing a contrastive summary. In this work, an end-to-end deep learning framework is proposed for the problem of context-sensitive contrastive opinion summarization.

2   Related Work

The emergence of deep neural networks has motivated the researchers to model the abstractive summarization task using different network architectures. A sequence-to-sequence attention model [16] generates a short summary conditioned on the input sentence. A novel convolutional attention-based decoder [5] generates a summary by focusing on the appropriate input words while preserving the meaning of the sentence. A Feature-rich encoder [14] captures linguistic features in addition to the word embeddings of the input document. To address the problem of out-of-vocabulary words, pointer-generator model was used by the decoder.

A general single-document summarization framework [4] applies attention directly to select sentences or words similar to pointer networks for producing a summary. A read-again mechanism based model [24] computes the representation of each word by taking into account the entire sentence not just considering the history of words before the target word. A hierarchical document encoder and an attention-based extractor model [10] incorporated the latent structure information of summaries to improve the quality of summaries generated using abstractive summarization.

A selective encoding model [26] selects the encoded information using selective gate from encoder to improve encoding effectiveness and control the information flow to the decoder. A graph-based attention neural model [19] discovers the salient information of a document. In addition to the saliency problem, fluency, non-redundancy and information correctness are addressed using a unified framework. A hierarchical model [11] improves text summarization using significant supervision by the sentiment classifier. The summarizer captures the sentiment orientation of the text thus improving the coherence of the generated summary. An end-to-end framework [25] integrates the sentence selection

strategy with the scoring model to predict the importance of sentence based on the previously selected sentences. The keywords extracted from the text using an extractive model is encoded using a Key Information Guide Network (KIGN) [9]. These keywords guides the decoder in summary generation. Thus the method combines the advantage of extractive and abstractive summarization.

A multi-layered attentional peephole convolution LSTM network [15] automatically generates summary from a long text. Instead of traditional LSTM, convolution LSTM was used which allows access to the content of previous memory cell through a peephole connection. An LSTM-CNN based model [18] produces abstractive text summarization by taking semantic phrases rather than words as input. The model was able to produce natural sentences as output summary.

Most of the existing techniques for summary generation differ in network structure, inference mechanism, and decoding method. In this paper, a neural network model has been proposed for improving the contrastive summary generation process by selecting and encoding essential keywords.

3   Proposed Framework

The proposed context-sensitive contrastive summary model is a sequence-to-sequence model with attention and a pointer-generator mechanism. The input of the proposed model is documents of entities taken for comparison. These are a sequence of words passed to an attentive encoder, Contextual Sentiment LSTM Encoder. The encoder's outputs are then passed to an attentive decoder which generates a summary using pointer-generator mechanism. The document summaries are then passed to a multi-layer perceptron which computes a contrastive score. The summaries which have high contrastive score are then presented as contrastive summaries. The architecture of the proposed context-sensitive contrastive summary model is shown in Fig. 1. The components and variables in the figure are explained in detail in the forthcoming section.

3.1   Contextual Sentiment LSTM Encoder

An attention-based Long short-term memory (LSTM) model takes a sequence of words as input and produces a probability distribution quantifying the importance of the target word in each position of the input sequence. An attentive network, the Sentic LSTM was proposed in [12] which adds common sense knowledge concepts from the Senticnet through the recall gate to control information flow in the network based on the target. A Coupled Multi-layer Attention network for the automatic co-extraction of feature and opinion words was proposed in [20]. The authors furnished an end-to-end solution by providing a couple of attention with a tensor operator in each layer of the network.

The proposed Contextual Sentiment LSTM model is a hierarchical attention model which automatically learns the representation of context, feature and opinion words present in review sentences. The attention mechanism of the proposed Contextual Sentiment LSTM model is shown in Fig. 2.
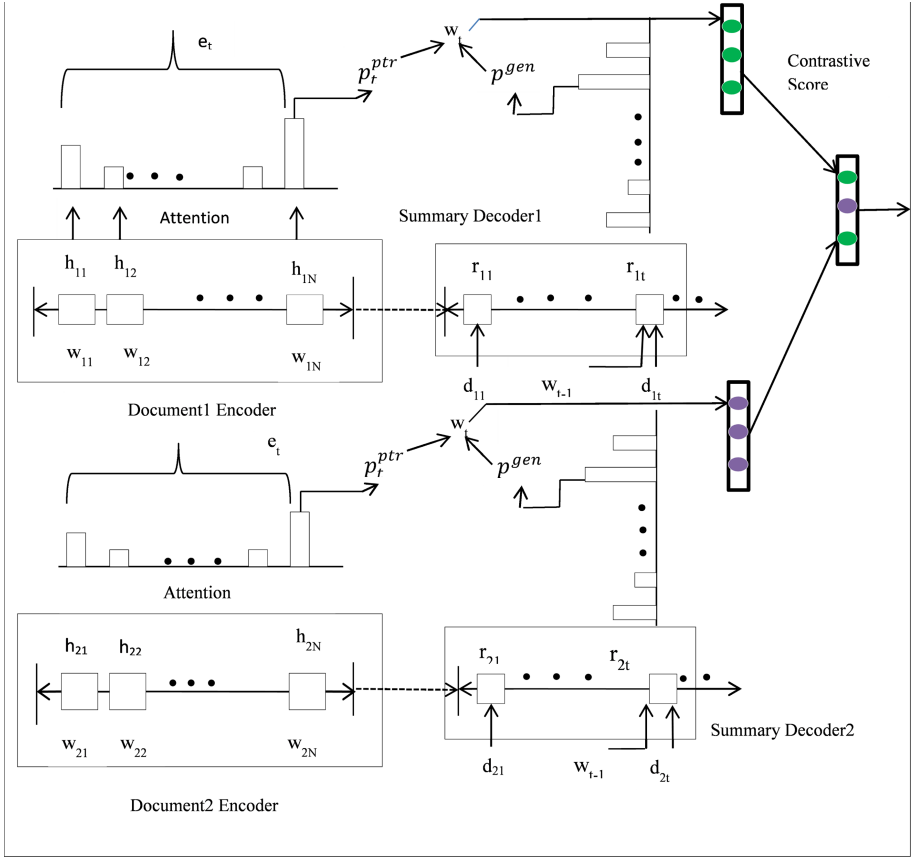
Fig. 1. Architecture of context-sensitive contrastive summary model.

It consists of a word encoder, a word attention, a sentence encoder, and a sentence attention layer. Word embeddings are given as input to the word encoder in order to avoid the assignment of random weights to the input words during training. The dependency-based word embeddings proposed in [8] is used to generate word embeddings corresponding to the input words. Dependency-based word embeddings are generated based on dependency relation between words and hence best suitable for the problem of opinion summarization.

The word encoder transforms the dependency-based word embeddings into a sequence of hidden outputs. Multi-factor attention is built on top of the hidden outputs: feature words attention, opinion words attention and context words attention. Each attention takes as input the hidden outputs found at the position of the feature, opinion or context words and computes an attention weight over these words by finding the similarity between the hidden vectors and vectors from knowledge bases using a bilinear operator. A sentence context vector is
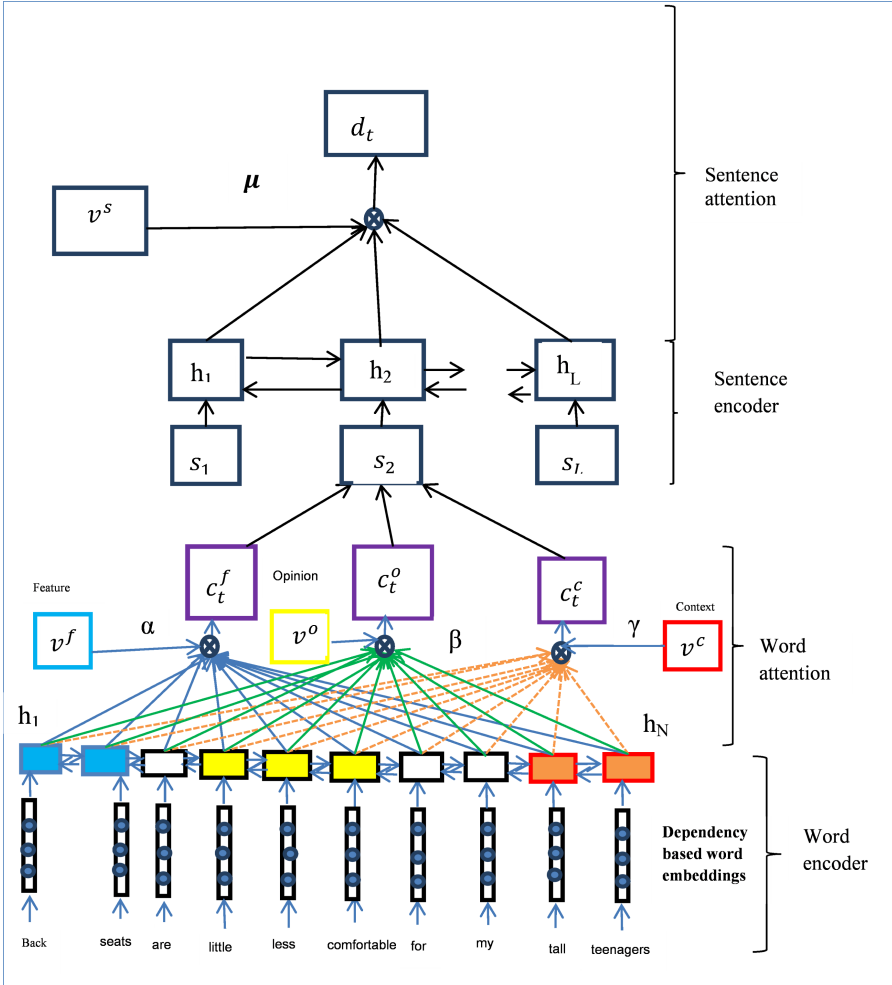
Fig. 2. Attention mechanism of the proposed Contextual Sentiment LSTM Model

generated by aggregating the context vectors of feature, opinion and context words produced using attention mechanism.

The document encoder transforms the sentence vectors into a sequence of hidden outputs. A document context vector is generated using sentence level attention. The document context vector captures the important information present in the input document.

4   Context-Sensitive Contrastive Summary Model

Given a review document, a Skip-Gram model generates dependency-based word embeddings for each word in the document. The dependency-based word embed-

dings are given as input to the bi-directional LSTM. The sequence of hidden outputs is produced by concatenation of both the forward and backward hidden outputs. The information flow at the current time step is controlled by input, output and forget gates of the LSTM cell as given by Eq. (1–6)

$$i_t = \sigma(W_i[h_{t-1}; w_t] + b_i) \tag{1}$$

$$f_t = \sigma(W_f[h_{t-1}; w_t] + b_f) \tag{2}$$

$$o_t = \sigma(W_o[h_{t-1}; w_t] + b_o) \tag{3}$$

$$\tilde{c}_t = \tanh(W_c[h_{t-1}; w_t] + b_c) \tag{4}$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \tag{5}$$

$$h_t = o_t * \tanh(c_t) \tag{6}$$

where i_t, f_t, o_t and c_t are the input gate, forget gate, output gate and memory cell respectively. W_i, W_f and W_o are the weight matrices, b_i, b_f, b_o and b_c are the bias for the input gate, forget gate, output gate and memory cell respectively, and w_t is the word vector for the current word. The sequence of hidden outputs produced is denoted by H $= \{h_1, h_2, ..., h_N\}$, where each hidden element is a concatenation of both the forward and backward hidden outputs.

4.1   Word Attention

It is essential to learn the embedding vector for each feature, opinion and context word present in a sentence for the problem of context-sensitive summarization. Hence three attentions: Feature attention, opinion attention and context attention are explored.

The ConceptNetNumberbatch embeddings proposed in [19] is used to guide the network to attend to the feature words present in a document for feature attention. The AffectiveSpace embeddings proposed in [2] is used as a source of attention for opinion words. The context hints gathered from different user surveys are utilized to build attention source vector for context words.

$$c_t^f = \sum \alpha_t h_t \tag{7}$$

$$c_t^o = \sum \beta_t h_t \tag{8}$$

$$c_t^c = \sum \gamma_t h_t \tag{9}$$

The sentence context vector is acquired by fusing together the vectors of feature, opinion and context. The context vector captures the most important features of a sentence with respect to feature, opinion and context words.

$$s_t = (c_t^f + c_t^o + c_t^c) \tag{10}$$

4.2   Sentence Attention

Sentence level attention guides the network to attend the most important sentences in a document that serve as clues for sentiment classification. A sentence attention weight is calculated by introducing a sentence attention source vector.

$$d_t = \sum \mu_t h_t \tag{11}$$

4.3   Context-Sensitive Sentiment Classification

The document context vector is fed into a unidirectional LSTM which converts it into a vector whose length is equivalent to the quantity of classes {positive, negative}. The softmax layer changes the vector to a conditional probability distribution.

$$p_d^r = softmax(W^r d_t + b_d^r) \tag{12}$$

4.4   Context-Sensitive Summary Generation

For constructing a context-sensitive summary of the input document a unidirectional LSTM is used as a decoder. The hidden state of the decoder is updated using the document context vector and the previous decoder state.

$$r_t = LSTM(r_{t-1}, d_t, w_{t-1}) \tag{13}$$

The pointer-generator network proposed in [17] is adopted on top of LSTM for generating summaries. The pointer-generator network generates summary by copying important words from the input document via pointing and generating words from a fixed vocabulary.

At each time step, the generator selects and outputs an important word from a distribution of words computed through a softmax layer. The pointer network directly copies a word based on attention mechanism from the input document to the output summary.

$$p(w_t) = p_t^{ptr} p^{gen}(w_t) + (1 - p_t^{ptr}) \sum_i d_{t,i} \tag{14}$$

4.5   Context-Sensitive Contrastive Summary Generation

Given a set of summaries from two entities, the summaries which have high contrastive score are identified and presented as contrastive summaries. To compute contrastive score, the element-wise product of summary vectors is given as input to a Multi-layer perceptron.

$$Score_{i,j} = \pi_{i=1}^N \pi_{j=1}^N sigmoid(MLP([s_i s_j]; \theta)) \tag{15}$$

5   Experimental Results

Experiments are conducted on SemEval 2014 task 4 restaurant reviews dataset in [13]. The restaurant dataset consists of 1,125,457 documents with a vocabulary size of 476,191. Since the restaurant dataset is large, a sample of 50,000 documents is taken for experimentation to reduce the training time of the model. The sentences are annotated with feature, opinion and context labels for word attention as given in Table 1. For context-sensitive sentiment classification, the feature words are annotated with their corresponding polarities as shown in Table 2. The statistics of the datasets used for context-sensitive summary generation is shown in Table 3. Initially, sentences are tokenized using sentiment-aware tokenizer and words that occur less than 100 times are filtered. All the word vectors are initialized by Dependency-based word embeddings in [8]. The dimension of word embeddings is set to 300. The dependency-based word embeddings capture arbitrary contexts based on dependency relation between words which helps the proposed CSLSTM model to learn better weights.

   The dependency-based word embeddings are converted into hidden representations through CSLSTM model implemented with Theano library stated in [1]. The number of hidden units is set to 50. For training, a batch size of 30 samples

Table 1. Dataset description for word attention.

Training	Testing	Total
3,041	800	3,841

Table 2. Feature distribution per sentiment class.

Features	Positive		Negative	
	Train	Test	Train	Test
Food	867	302	209	69
Price	179	51	115	28
Service	324	101	218	63
Ambience	263	76	98	21
Others	546	127	199	41
Total	2179	657	839	222

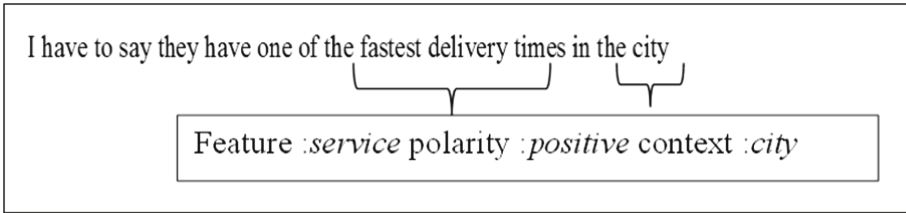**Table 3.** Statistics of the dataset for a context-sensitive summary generation.

Statistics	Training	Testing
Number of documents	45,000	5000
Number of document-summary pairs	283,326	11,360
Maximum document length	90	90
Maximum summary length	14	14

was used. AdaGrad optimizer is used for minimizing the training loss with a learning rate of 0.01.

5.1   Context-Sensitive Sentiment Classification Using Soft-Max Classifier

The sentiment classifier takes the feature, opinion and context words identified using CSLSTM as input and determines the polarity of each feature word. From the sample sentence shown in Fig. 3, it can be seen that the classifier detects a multi-word phrase "delivery times" as a feature with the help of feature attention source "service" and determines its polarity as "positive" based on context "city" detected using context attention source. The classifier predicts the correct polarity of a feature even if the sentence is long and has a complicated structure.

I have to say they have one of the fastest delivery times in the city

Feature :*service* polarity :*positive* context :*city*

Fig. 3. Context-sensitive classification of sample sentences from restaurant dataset.

5.2   Context-Sensitive Summary Generation Using Pointer-Generator Network

For the pointer-generator network, the generator vocabulary size is set to 20k based on most attended words using an attention mechanism. A sample context-sensitive summary generated for restaurant dataset is shown in Table 4. It can be seen that the model repeats some of the phrases in the reference summary. Rarely the model generates summary consisting of words that is not truthful with respect to the input document.

Table 4. Sample context-sensitive summary generated using pointer-generator network for restaurant dataset.

Document All the appetizers and salads were fabulous, the steak was mouth watering and the pasta was delicious!!! The sweet lassi was excellent as was the lamb chettinad and the garlic naan but the rasamalai was forgettable. . . . . **Gold summary** Food was excellent with a wide range of choices and good services. It was a bit expensive though. **Context-sensitive summary** Excellent food with a wide range of choices and good services.

5.3 Summary Generation by Selecting Sentences Using Contrastive Score

The summaries that have high contrastive score are identified and chosen for a summary generation. Table 5 shows the sample contrastive summary pairs of Restaurant X and Restaurant Y.

Table 5. Sample contrastive summary pairs of Restaurant X and Restaurant Y.

Restaurant X	Restaurant Y
My fiance surprise 30th birthday dinner here could not be happier	Magnificent restaurant to celebrate my birthday the other night with my boyfriend
Great food at reasonable prices	Definitely not worth the price
Great place to dine for dinner	Place I would reconsider revisiting for dinner

6 Performance Analysis of Proposed Context-Sensitive Summary Method with Existing Methods

The performance of the proposed method is evaluated in terms of F_1 score, accuracy and ROUGE metrics.

6.1 Word Attention

The attention results of the proposed CSLSTM model is compared with two existing methods, RNCRF [21] and CMLA [20]. Attention in RNCRF model is based on pre-extracted syntactic relations. The CMLA model uses multi-layer attentions with tensors to learn the interaction between features and opinions.

F_1 Score. F_1 score is used to measure the accuracy of attention. It is based on the precision and recall score of the test dataset. It is represented by Eq. (16):

$$F_1 score = (1 + \beta^2) + (P * R)/(\beta^2.P) + R \qquad (16)$$

A comparison of F_1 scores of word attention by CSLSTM and existing methods for restaurant dataset is shown in Table 6. The RNCRF and CMLA models attend only to the feature and opinion words in a sentence. The CSLSTM model attends to the context words in addition to feature and opinion words. The proposed model outperforms existing models as the model attends to important words based on attention sources.

Table 6. Comparison of F_1 score of word attention by CSLSTM with existing models for restaurant dataset.

Model	Feature	Opinion	Context
RNCRF [21]	84.05	80.93	–
CMLA [20]	85.29	83.18	–
Proposed CSLSTM	85.70	82.16	78.12

6.2   Context-Sensitive Sentiment Classification

The sentiment classification results are compared with two existing methods, Sentic LSTM [12] and ATAE-LSTM [22]. The Sentic-LSTM model replaces the encoder with a knowledge embedded LSTM for filtering data that does not coordinate with ideas in the knowledge base. ATAE-LSTM model incorporates feature embeddings into the representation of a sentence.

Accuracy. Accuracy is the ratio of the correctly classified sentences to the total number of sentences in the dataset. It is represented by Eq. (17):

$$Accuracy \ = \ Number\ of\ correct\ predictions/Total\ number\ of\ predictions \tag{17}$$

Figure 4 illustrates the comparison of the accuracy of sentiment classification by proposed CSLSTM with existing methods for restaurant dataset with respect to sentence length.

The ATAE-LSTM performs better than the Sentic LSTM as it captures important information pertaining to a given feature for sentiment classification. The Sentic LSTM performs slightly lower level than the ATAE-LSTM, as knowledge base ideas may at times delude the network to manage words irrelevant to features. The CSLSTM network achieves the best performance of all, as it discriminates information in response to given feature, opinion and context inputs.

6.3   Context-Sensitive Summary Generation

The summary generated using the proposed CSLSTM model is compared with two existing models, MARS model [23] and pointer-generator model [17]. MARS is a sentiment-aware abstractive summarization system which leverages on text-categorization to improve the performance of summarization. Pointer-generator model introduces coverage based network to avoid repetition of phrases in the summary.

ROUGE. The results of summarization are evaluated using three different metrics provided by ROUGE. In addition to the two metrics ROUGE-1 and ROUGE-2, ROUGE-L is taken as the third metric for evaluation.
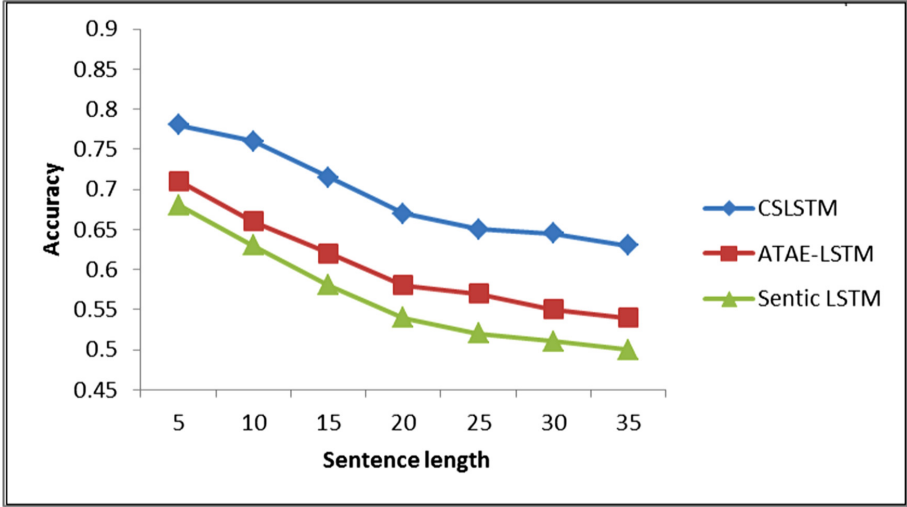
Fig. 4. Comparison of accuracy of sentiment classification by proposed CSLSTM with existing methods for restaurant dataset with respect to sentence length.

ROUGE-1: The number of unigrams in common between human summary and model summary.
ROUGE-2: Captures bigram overlap, thus measuring the readability of summaries.
ROUGE-L: The longest common sequence between the reference summary and model summary.

Table 7 shows the comparison of $ROUGE_{F1-score}$ of the summary produced by CSLSTM and existing methods. The attention of important phrases in the document by the proposed CSLSTM model results in great $ROUGE-1_{F1-Score}$ of around 36 in restaurant dataset. The proposed model also achieves a high $ROUGE-L_{F1-Score}$ compared to existing models.

Table 7. Comparison of $ROUGE_{F1-score}$ of summarization by CSLSTM with existing models for restaurant dataset.

Model	$ROUGE-1$	$ROUGE-2$	$ROUGE-L$
Pointer-generator model [17]	31.83	11.70	28.3
MARS model [23]	35.66	15.2	32.42
Proposed CSLSTM	36.34	17.6	33.57

7   Summary

The proposed attentive neural network architecture improves the performance of context-sensitive contrastive summarization using attention mechanism. Different from existing works, the proposed architecture includes multi-factor attention mechanism into the encoder-decoder used for summary generation. The three kinds of attention (i.e., feature attention, opinion attention, and context attention) selectively attend to the significant information when decoding summaries. The context vector obtained by encoding feature, opinion and context information is classified using a softmax classifier. Experimental results show that the classifiers incorporating context-sensitive information outperform the state-of-the-art models in terms of classification accuracy. The summary generated using the attention of a few important phrases in the document achieved a good rouge score than existing methods.

References

1. Bastien, F., et al.: Theano: new features and speed improvements. arXiv preprint arXiv:1211.5590 (2012)
2. Cambria, E., Poria, S., Bajpai, R., Schuller, B.: SenticNet 4: a semantic resource for sentiment analysis based on conceptual primitives. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 2666–2677 (2016)
3. Chen, G., Chen, L.: Augmenting service recommender systems by incorporating contextual opinions from user reviews. User Model. User-Adapt. Interact. **25**(3), 295–329 (2015). https://doi.org/10.1007/s11257-015-9157-3
4. Cheng, J., Lapata, M.: Neural summarization by extracting sentences and words. arXiv preprint arXiv:1603.07252 (2016)
5. Chopra, S., Auli, M. Rush, A.M.: Abstractive sentence summarization with attentive recurrent neural networks. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 93–98 (2016)
6. Hariri, N., Mobasher, B., Burke, R., Zheng, Y.: Context-aware recommendation based on review mining. In: Proceedings of the 9th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems, p. 30 (2011)
7. Lahlou, F.Z., Benbrahim, H., Mountassir, A., Kassou, I.: Inferring context from users' reviews for context aware recommendation. In: Bramer, M., Petridis, M. (eds.) Research and Development in Intelligent Systems XXX, pp. 227–239. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-02621-3_16
8. Levy, O., Goldberg, Y.: Dependency-based word embeddings. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pp. 302–308 (2014)
9. Li, C., Xu, W., Li, S., Gao, S.: Guiding generation for abstractive text summarization based on key information guide network. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2, pp. 55–60 (2018)
10. Li, P., Lam, W., Bing, L., Wang, Z.: Deep recurrent generative decoder for abstractive text summarization. arXiv preprint arXiv:1708.00625 (2017)

11. Ma, S., Sun, X., Lin, J., Ren, X.: A hierarchical end-to-end model for jointly improving text summarization and sentiment classification. arXiv preprint arXiv:1805.01089 (2018)
12. Ma, Y., Peng, H., Cambria, E.: Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In: Proceedings of 32nd Association for the Advancement of Artificial Intelligence, pp. 5876–5883 (2018)
13. Manandhar, S.: SemEval-2014 task 4: aspect based sentiment analysis. In: Proceedings of the 8th International Workshop on Semantic Evaluation, pp. 27–35 (2014)
14. Nallapati, R., Zhou, B., Gulcehre, C., Xiang, B.: Abstractive text summarization using sequence-to-sequence RNNs and beyond. arXiv preprint arXiv:1602.06023 (2016)
15. Rahman, M., Siddiqui, F.H.: An optimized abstractive text summarization model using peephole convolutional LSTM. Symmetry **11**(10), 1290 (2019)
16. Rush, A.M., Chopra, S., Weston, J.: A neural attention model for abstractive sentence summarization. arXiv preprint arXiv:1509.00685 (2015)
17. See, A., Liu, P. J., Manning, C. D.: Get to the point: summarization with pointer-generator networks. arXiv preprint arXiv:1704.04368 (2017)
18. Song, S., Huang, H., Ruan, T.: Abstractive text summarization using LSTM-CNN based deep learning. Multimed. Tools Appl. **78**(1), 857–875 (2018). https://doi.org/10.1007/s11042-018-5749-3
19. Speer, R., Havasi, C.: ConceptNet 5: a large semantic network for relational knowledge. In: Gurevych, I., Kim, J. (eds.) The People's Web Meets NLP. Theory and Applications of Natural Language Processing, pp. 161–176. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-35085-6_6
20. Wang, W., Pan, S.J., Dahlmeier, D., Xiao, X.: Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In: Proceedings of Association for Advancement of Artificial Intelligence, pp. 3316–3322 (2017)
21. Wang W., Pan, S.J., Dahlmeier D., Xiao, X.: Recursive neural conditional random fields for aspect-based sentiment analysis. arXiv preprint arXiv:1603.06679 (2016)
22. Wang, Y., Huang, M., Zhu, X., Zhao, L.: Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 606–615 (2016)
23. Yang, M., Qu, Q., Shen, Y., Liu, Q., Zhao, W., Zhu, J.: Aspect and sentiment aware abstractive review summarization. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 1110–1120 (2018)
24. Zeng, W., Luo, W., Fidler, S., Urtasun, R.: Efficient summarization with read-again and copy mechanism. arXiv preprint arXiv:1611.03382 (2016)
25. Zhou, Q., Yang, N., Wei, F., Huang, S., Zhou, M., Zhao, T.: Neural document summarization by jointly learning to score and select sentences. arXiv preprint arXiv:1807.02305 (2018)
26. Zhou, Q., Yang, N., Wei, F., Zhou, M.: Selective encoding for abstractive sentence summarization. arXiv preprint arXiv:1704.07073 (2017)