International Journal of Artificial Intelligence,
ISSN 0974-0635; Int. J. Artif. Intell.
2025, Volume 23, Issue Number 2

A Robust Approach to Use LSTM-RNN in Natural Language Processing

Nguyen T. Long[1,2], **Nguyen T. Huong**[1,3]

[1]Department of Informatics, Institute for Cybersecurity and Digital Technologies,
MIREA – Russian Technological University, Moscow, Russia,
nguen@mirea.ru, nguen_t@mirea.ru

[2]Faculty of Information Technology, Sao Do University
174261 Hai Duong, Viet Nam
long_nt@saodo.edu.vn

[3]University of Information and Communication Technology
Thai Nguyen University, 70000, Thai Nguyen, Viet Nam
nthuongcntt@ictu.edu.vn

ABSTRACT

The integration of LSTM-RNN networks into natural language processing (NLP) has shown promising potential for capturing spatial and temporal dependencies in sequential data. In contrast to conventional LSTM models, which focus primarily on temporal dynamics, LSTM-RNN incorporates convolutional operations to better capture local patterns and hierarchical structures in text. This hybrid architecture is particularly well suited for tasks where sequential and spatial relationships need to be modeled, such as machine translation, sentiment analysis, and text generation. Therefore, a robust approach using LSTM-RNN for natural language processing is proposed in this paper. This study proposes the application of LSTM-RNN for sentiment analysis and utilizes its ability to capture semantic and syntactic relations in texts. The model processes input sequences through an embedding layer to represent words as dense vectors, followed by one or more LSTM layers with dropout regularization to prevent overfitting. The final dense layer with a sigmoid or softmax activation function predicts sentiment polarity.

Keywords: LSTM-RNN, CNN, sentiment analysis, Natural Languages Processing, Deep Learning.

Mathematics Subject Classification: 68U10, 68T45.
Computing Classification System: 10010147.10010257, 10010147.10010178, 10010147.10010371.10010382.10010383, 10010147.10010178.10010224.10010245.10010247.

1 Introduction

Natural language processing (NLP) has made remarkable progress in recent years, driven by the rapid development of deep learning techniques. Among these, recurrent neural networks

(RNNs) and their variants, such as Long Short-Term Memory (LSTM) networks, have become popular for sequential data modeling due to their ability to capture long-term dependencies in texts. However, traditional LSTMs focus primarily on temporal relationships and often have problems with hierarchical feature extraction, which is crucial for understanding complex linguistic structures. To counteract this limitation, the convolutional LSTM (Conv-LSTM) has emerged as a powerful hybrid architecture that combines the strengths of convolutional neural networks (CNNs) and LSTMs and enables the simultaneous modeling of spatial and temporal dependencies in sequential data.

The application of Conv-LSTM in NLP involves various tasks, including text classification, named entity recognition, speech recognition and text generation. In text classification, the convolutional layers extract salient features from the word embeddings, while the LSTM layers model their contextual relationships, improving accuracy over conventional methods. In sequence-to-sequence tasks such as machine translation, the ability of Conv-LSTMs to hierarchically encode input sentences and decode them into the target language shows superior performance compared to standalone CNNs or RNNs. In speech processing, Conv-LSTM also effectively processes spectrogram input by capturing both frequency-localized features (via convolutions) and temporal dynamics (via LSTMs), making it suitable for end-to-end speech recognition systems.

In this paper, we propose the use of LSTM-RNN in NLP by evaluating its performance in several benchmarks and comparing it with state-of-the-art approaches. We have also presented architectural variations, such as stacked LSTM-RNN layers to address specific NLP challenges. By unifying spatial and temporal modeling, Conv-LSTM provides a versatile framework to improve language understanding, generation and translation. The paper is divided into the following sections: Introduction, Related Work, Proposed Methodology, Analysis of Results, Future Development Directions, and Conclusion.

2 Related work

Deep Learning (Miguel, 2022) is applied in many fields, with numerous successful proposals that have brought significant benefits to life, particularly in road traffic (Huong and Long, 2022) and object surveillance (Jonathan, Carlos and Luis, 2018), (Colpas and Alberto, 2023)... All of these demonstrate the power of applying advanced technologies and optimized algorithms of DL to practical applications. Currently, NLP is one of the potential fields where DL algorithms can be applied and deployed to build control and interaction systems in a natural and context-sensitive manner. The application of hybrid deep learning models in natural language processing (NLP) has gained significant attention in recent years, with Conv-LSTM emerging as a powerful architecture that combines the strengths of convolutional networks and recurrent neural networks. This section reviews previous research on the use of Conv-LSTM and related hybrid models in NLP, covering foundational work, comparative studies, and recent advances. The Conv-LSTM architecture for precipitation prediction, which proved to outperform conventional LSTMs in modeling spatio-temporal sequences (Griffis, Patil, Bell and Dixon, 2023). The replacement of fully connected layers in LSTMs with convolution operations that allow the model to capture spatial hierarchies while preserving sequential dependencies. The research work began with the adaptation of Conv-LSTM for language tasks. The author in (Chen, 2024) investigated its use in text classification and showed that convolutional filters can extract meaningful word and sentence features before passing them to LSTM for sequence modeling. This hybrid approach outperformed standalone CNNs and LSTMs, especially on tasks requiring fine-grained feature detection, such as sentiment analysis and topic categorization. One of the main applications of Conv-LSTM in NLP is text classification. It has been shown that the model can effectively classify documents by first applying convolutional filters on word embeddings to recognize key phrases, followed by LSTM-based sequence aggregation. Their experiments on benchmark datasets showed consistent improvements over traditional methods, especially in the treatment of negations, where local context plays a crucial role.

Conv-LSTM has also been used in sequence-to-sequence tasks such as machine translation. A convolutional sequence-to-sequence model was presented that outperformed traditional RNN-based approaches in terms of both accuracy and training efficiency. While CNNs were primarily used in this work, later studies incorporated Conv-LSTM to better capture long-range dependencies (Lv, Wang, Gao and Zhao, 2021). They proposed a hybrid encoder-decoder framework in which the encoder uses Conv-LSTM for hierarchical encoding of output sentences, while the decoder uses an attention-based LSTM for fluent generation. Their model achieved competitive results in WMT English-to German translation, especially in preserving syntactic structure (Chen, Xu, He and Wang, 2017). The dynamic convolutional networks were introduced for machine translation. Their approach inspired later adaptations that combined convolutional layers with recurrent mechanisms to improve translation quality (Tang and Zhang, 2008). They optimized these architectures by incorporating multi-head attention, resulting in models that balanced computational efficiency and high performance (AlBadani, Shi and Dong, 2022).

Conv-LSTM is widely used in speech recognition and audio processing. In the study (Goh, Lau and Lee, 2019), it was shown that replacing the fully concatenated layers in LSTMs with convolutional operations improves the recognition accuracy of phonemes, as the model can better capture spectro-temporal patterns in speech signals. This approach was later extended to end-to-end speech recognition systems, where conv-LSTM layers were used to process raw audio waveforms before passing them to a connectionist decoder for temporal classification (CTC). More recently, researchers have explored Conv-LSTM for multimodal NLP tasks such as video subtitling and audiovisual speech recognition. The authors in (Liu and Guo, 2019) used Conv-LSTM to encode video images and generate descriptive captions, and showed that the model can effectively match visual and speech sequences. The Conv-LSTM model was used for lip reading (Kamyab, Liu and Adjeisah, 2021), where convolutional layers extracted mouth movement features, while LSTMs modeled temporal articulation patterns.

The study (Basiri, Nemati, Abdar, Cambria and Acharya, 2021) uses RNN in a system capable of answering paragraph questions. RNNs have been successfully applied to many problems, especially in the field of NLP (Natural Language Processing). Theoretically, it is true that RNNs have the ability to remember the computations (information) before them, but the traditional RNN model cannot remember the removed steps due to the loss of derivations, so the success of this model mainly comes from another improved model the LSTM. In (Saraswat, Srivastava and Shukla, 2018), a neural machine was proposed to generate short text dialogs. RNN models show good performance in classifying different opinions and answering questions. An adaptive RNN model (ADARNN) for classifying different opinions is used in Twitter promotions. These architectures have made it possible to solve a wide range of problems that were previously intractable, such as computer vision, machine translation, speech recognition, and natural language processing (NLP), with a solution quality that is now comparable to, and in some cases better than, human performance. Language models for real-world speech recognition or machine translation systems are based on huge amounts of data, and the conventional wisdom is that we just need more data. Models from research are usually complex and often only work for systems based on very limited amounts of training data. In fact, most of the proposed advanced language modeling techniques offer only tiny improvements over the simple basic models and are rarely used in practice.

3 Methodology

The LSTM (Fig. 1) is an extended version of the recurrent neural network (RNN), which is designed for solving problems with long-term dependencies. RNN is a neural network that contains loops. This network has the ability to store information that is transferred from one layer to another. The output of the hidden layer depends on the information from the other layers at all times. RNNs are often used in natural language processing or sequential data problems. However, due to the simple architecture of RNN, the ability to link layers with long distances is not good. It is basically unable to memorize information from long-distance data, and therefore the first elements in the input sequence often do not have much influence on the prediction results of the elements for the output sequence in the following steps (Fig. 2). The reason for this is that RNNs are affected by the gradual decrease of the derivative during the learning process - vanishing gradient. LSTM networks were developed to overcome this problem. The working mechanism of LSTM is to memorize only relevant information that is important for prediction while discarding other information.

LSTM networks can consist of several LSTM cells that are connected to each other. The idea of LSTM is to add an intra-cell state s_t and three gates to filter input and output information for the cell, including the forgetting gate, the input gate i_t and the output gate. At each time step, the gates receive the input value (which represents an element of the input sequence) and the value, which results from the output of the memory cells of the previous time. The gates all have the function of filtering information with different purposes. The gates are defined as follows: Forget gate: has the function of removing unnecessary information from the internal cell state. Input gate: helps to filter out the information that needs to be added to the internal cell state. Output gate: determines which information from the internal cell states is used as output. During execution, s_t and the output values h_t are calculated as follows:

Step 1: the LSTM cell decides which information from the internal cell states in the previous time should be discarded. The value of the forgetting gate in time step t is calculated on the basis of the current input value, the output value of the LSTM cell in the previous step and

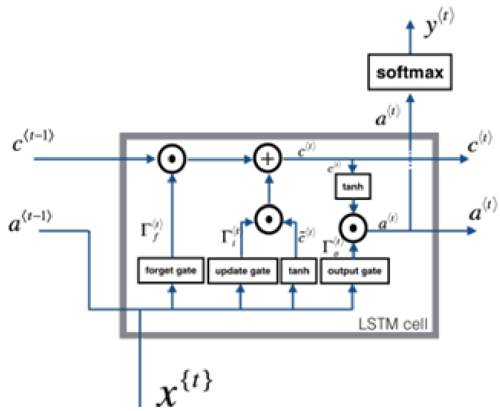

Figure 1: The diagram describes the working process between the gates of the LSTM network.

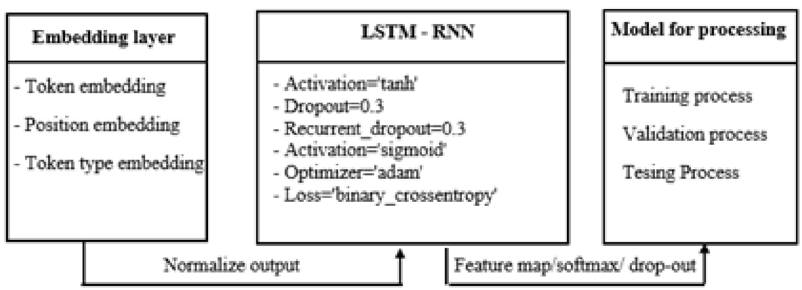

Figure 2: Diagram of LSTM-RNN for Sentiment analysis application.

the bias voltage of the forgetting gate. The sigmoid function transforms all activation values according to the formula into the value range between 0 and 1.
Step 2: the LSTM cell determines which information must be added to the internal cell states. This step includes two calculations and represents the information that can be added to the internal cell states.
Step 3: the new value of the internal cell state s_t is calculated based on the results obtained.

4 Analyze the result of applying LSTM-RNN for sentiment analysis in NLP

All experiments evaluating the proposed LSTM-RNN method compared to other methods are performed on the Social Media Sentiments Analysis Dataset (Parmar, 2023). This dataset consists of social media posts from various platforms. It includes both positive and negative sentiment labels, allowing for training sentiment analysis models on real-world social media data.
LSTM-RNN demonstrate strong performance in sentiment analysis with accuracy typically greater than 98% on the propose database (Fig. 3). Their effectiveness stems from the ability to process sequential text data while maintaining memory of long-range dependencies, which is crucial for interpreting sentiment-bearing phrases and contextual nuances. LSTM-RNN excel

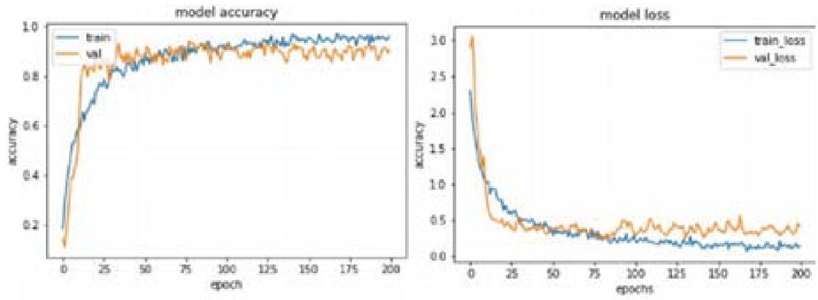

Figure 3: The result of accuracy and loss of training/validation process

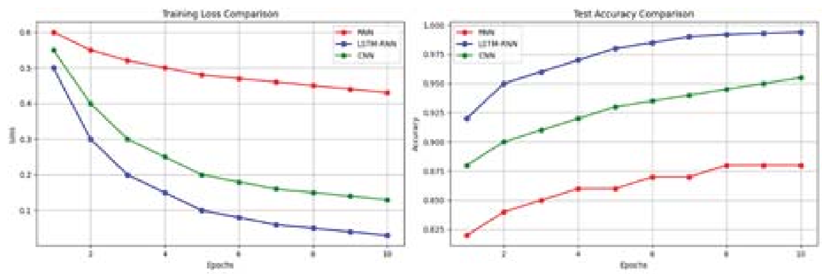

Figure 4: The result of comparing between LSTM-RNN with CNN and RNN for sentiment analysis.

at capturing such linguistic subtleties through their gated architecture.

In this study, we also applied two methods RNN and CNN in parallel in sentiment analysis. The results showed that the proposed method LSTM-RNN is completely suitable for the research objectives. LSTM-RNNs demonstrate superior performance over traditional RNNs and CNNs for sentiment analysis due to their ability to effectively capture long-range dependencies in text (Fig. 4). LSTM's gated architecture selectively retains important contextual information. CNN (Szegedy, Liu, Jia, Sermanet, Reed, Anguelov and Rabinovich, 2015), though efficient at extracting local features through filters, struggle with sequential relationships critical for sentiment understanding.

In Table 1 shown typical Mean Absolute Error (MAE) (Hanin, 2019) values and Mean Squared Error (MSE) (Allen-Zhu, Li and Liang, 2019) values on proposed datasets. The low MAE reflects the model's ability to minimize deviations in sentiment polarity predictions, while the smaller MSE indicates effective handling of outlier errors.

LSTM-RNN has to faces challenges, such as high computational costs and the need for large-

Table 1: Evaluation results comparing LSTM-RNN, CNN, and RNN in sentiment analysis.

Model	Accuracy (%)	Training time (min)	Loss	MAE	MSE
LSTM-RNN	98.7	2	0.03	0.47	0.36
CNN	94.03	5	0.04	0.2	0.09
RNN	91.5	5	0.12	0.35	0.15

scale training data to optimize its deep architecture. Additionally, designing optimal filter sizes and strides for convolutional operations in text - where word and sentence boundaries are less rigid than in image pixels - requires careful tuning. Recent work has explored integrating attention mechanisms with Conv-LSTM to enhance focus on relevant features or combining it with transformer-based architectures for scalability. These innovations aim to balance the model's representational power with efficiency, particularly for real-time applications.

5 Conclusion

LSTM-RNN represents a significant advance in NLP by bridging the gap between convolutional paradigms and recurrent paradigms. Its ability to utilize both hierarchical and sequential structures in texts makes it a promising tool to cope with the complexity of human language and paves the way for more robust and interpretable language models. The result of this study shows that LSTM-RNN is a versatile and powerful architecture for NLP that can handle a wide range of tasks by effectively combining spatial and temporal modeling. While there are still challenges in terms of efficiency and scalability, ongoing innovations in attention mechanisms, model compression and hybrid architectures continue to improve performance. This review highlights the evolution of LSTM-RNN in NLP and lays the foundation for further research to optimize and extend its applications.

Future research directions include exploring LSTM-RNN for low-resource languages, integration with pre-trained language models and application to novel NLP tasks such as code generation. Challenges remain, including increased computational complexity and the need for large datasets to effectively train deep LSTM-RNN models. Future research could investigate optimized architectures, attention mechanisms, and transfer learning techniques to increase efficiency. Overall, LSTM-RNN represents a versatile approach in NLP that bridges the gap between convolutional paradigms and recurrent paradigms for improved language modeling and understanding.

References

AlBadani, B., Shi, R. and Dong, J. 2022. A novel machine learning approach for sentiment analysis on twitter incorporating the universal language model fine-tuning and svm, *Appl. Syst. Innov.* **5**: 13.

Allen-Zhu, Z., Li, Y. and Liang, Y. 2019. Learning and generalization in over-parameterized neural networks, going beyond two layers, *Proc. Advances Neural Inf. Process. Syst. (NeurIPS)*, pp. 6158–6169.

Basiri, M., Nemati, S., Abdar, M., Cambria, E. and Acharya, U. 2021. Abcdm: An attention-based bidirectional cnn-rnn deep model for sentiment analysis, *Future Gener. Comput. Syst.* **115**: 279–294.

Chen, N. 2024. Exploring the development and application of lstm variants, *Appl. Comput. Eng.* **53**(1): 103–107.

Chen, T., Xu, R., He, Y. and Wang, X. 2017. Improving sentiment analysis via sentence type classification using bilstm-crf and cnn, *Expert Syst. Appl.* **72**: 221–230.

Colpas, A.C.and Isabel, A. and Alberto, M. 2023. Rtlahar: A model proposal based on reinforcement and transfer learning for the adaptation of learning in human activity recognition, *International Journal of Artificial Intelligence* **21**: 154–175.

Goh, Y., Lau, K. and Lee, Y. 2019. Audio-visual speech recognition system using recurrent neural network, *2019 4th International Conference on information technology (InCIT)*, pp. 38–43.

Griffis, E. J., Patil, O. S., Bell, Z. I. and Dixon, W. E. 2023. Lyapunov-based long short-term memory (lb-lstm) neural network-based control, *IEEE Control Syst. Lett.* **7**: 2976–2981.

Hanin, B. 2019. Universal function approximation by deep neural nets with bounded width and relu activations, *Mathematics* **7**(10): 992.

Huong, N. and Long, N. 2022. Deep learning methods for classification of road defects, *International Journal of Artificial Intelligence* **19**(2): 178–192.

Jonathan, R., Carlos, J. H. and Luis, M. G. 2018. Ernead: Training of artificial neural networks based on a genetic algorithm and finite automata theory, *International Journal of Artificial Intelligence* **16**: 214–253.

Kamyab, M., Liu, G. and Adjeisah, M. 2021. Attention-based cnn and bi-lstm model based on tf-idf and glove word embedding for sentiment analysis, *Appl. Sci.* **11**: 11255.

Liu, G. and Guo, J. 2019. Bidirectional lstm with attention mechanism and convolutional layer for text classification, *Neurocomputing* **337**: 325–338.

Lv, J., Wang, C., Gao, W. and Zhao, Q. 2021. An economic forecasting method based on the lightgbm-optimized lstm and time-series model, *Comput. Intell. Neurosci.* **2021**(1): 8128879.

Miguel, A. M. 2022. A component free derivation of backpropagation in deep learning, *International Journal of Artificial Intelligence* **20**: 93–104.

Parmar, K. 2023. Social media sentiments analysis dataset, https://www.kaggle.com/datasets/kashishparmar02/social-media-sentiments-analysis-dataset.

Saraswat, S., Srivastava, G. and Shukla, S. 2018. Classification of ecg signals using cross-recurrence quantification analysis and probabilistic neural network classifier for ventricular tachycardia patients, *Int J Biomed Eng Technol* **26**(2): 141–156.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D. and Rabinovich, A. 2015. Going deeper with convolutions, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9.

Tang, D. and Zhang, M. 2008. Deep learning in sentiment analysis, *Deep Learning in Natural Language Processing* pp. 219–253.