Towards a Hybrid LSTM-Transformer Model for Financial Data: A Theoretical Approach

Basit Ali BASITALI71019@GMAIL.COM
NED University of Engineering & Technology,
Karachi, Sindh 75270 PK

Abstract

This paper proposes a hybrid LSTM-Transformer architecture to train a Named Entity Recognition (NER) model on financial data, such as receipts and invoices. These data types are unstructured and come in various formats, making them difficult to process. The proposed model combines the sequential pattern recognition capabilities of LSTM networks with the contextual sensitivity of Transformer self-attention layers, making it well-suited for financial data applications. This study establishes a modular, design-oriented framework, complete with pseudocode and architectural explanations, to serve as a foundation for future empirical testing. This conceptual work aims to set a benchmark in financial data modeling by addressing domain-specific challenges and providing a scalable structure for subsequent validation.

1. Introduction

Businesses and individuals face an overwhelming volume of financial documents, such as receipts and invoices, which contain important information regarding spending patterns. However, it is unfortunately difficult to extract meaningful insights from these data, as classical data processing models often do not take into account its variability and sequential dependencies (Hochreiter & Schmidhuber, 1997; Graves, 2013). Although conventional machine learning models, and even certain deep learning architectures, cannot adequately address the unique, multidimensional nature of financial documents, financial documents can be divided into sub-documents for learning purposes (Liu, Gegov, & Cocea, 2015).

Recently, research has been conducted to explore hybrid model architectures of different neural network models (Vaswani et al., 2017). In particular, integrating Long Short-Term Memory (LSTM) networks with Transformer models has been shown to be an effective approach to data that are both contextual and sequentially complex. While a word or sentence can be interpreted as "groups" of symbols, LSTM networks are useful for processing data sequences of any order (e.g., a time series), provided that order is important. However, unlike Transformers, self-attention mechanisms allow us to capture long-term dependencies in the data, providing a model with the ability to understand contexts at different levels (Xie et al., 2019).

In this study, we present a theoretical model of an LSTM-Transformer for financial data analysis. The architecture is divided into modules such that LSTM, which works best in sequential processing, and the long-range contextual power of the Transformer layers can be used to improve the accuracy of data processing (Ranjan & Daniel, 2023; Yang, Zhang, & Tao, 2022a; Bai & Tahmasebi, 2022). The pseudocode and detailed architectural design of this framework are also presented. Although empirical validation is outside the scope of this study, this theoretical model provides a new methodology for financial data study and sets a strong foundation for further research.

2. Literature Review

In recent years, people have started using advanced deep learning models to analyze financial data (Ozbayoglu, Gudelek, & Sezer, 2020). Financial data are often complex and continuously changing; therefore, special methods are needed that can handle both the order of events and context. This section looks at past research on how financial data is studied, different ways to combine models, and how LSTM and Transformer models are used in this field.

2.1 Financial Data Processing Challenges

Financial data such as invoices and receipts mostly consist of a combination of numeric, textual, and tabular texts. Older methods, like rule-based systems and standard machine learning, don't work well with the different formats and messy structures (Kuhn & Johnson, 2013).

2.2 LSTM Networks in Financial Applications

The widely used LSTM networks have been used for time series forecasting (Tang, Fan, Shi, Huang, & Ma, 2021) and sequential data analysis. Financial data with temporal patterns (e.g., expense trends or transaction histories) are ideally suited to data that can retain long-term dependencies, which makes them good. However, they do not perform well on tasks that require contextual understanding (Dai, Yang, Yang, Carbonell, Le, & Salakhutdinov, 2019).

2.3 Transformer Models and Self-Attention Mechanisms

Transformer models with self-attention mechanisms were first introduced for natural language processing (Vaswani et al., 2017). This mechanism allows the model to focus on relevant parts of the input, irrespective of their position, and makes Transformers powerful for unstructured data analysis. Recently, it has been shown how Transformers can be applied to financial documents such as invoices and contracts to extract useful information (Yang, Li, Dong, Zhang, & Smyth, 2022b). However, although they work well, Transformers can miss important details in tasks that depend on the order of the data.

2.4 Hybrid Architecture: Combining LSTM and Transformer Strengths

In order to overcome the limitations of individual models, we have investigated hybrid architectures composed of LSTM and Transformer models (Sakatani, 2021; Fu et al., 2024). A hybrid model for customer behavior prediction by fusing LSTM for sequence analysis and a Transformer for context-aware prediction has been proposed. Similar approaches have yielded promising results in the healthcare and e-commerce domains but have received very little attention in applying these hybrids to financial data.

2.5 Research Gap and Contribution

Although many different hybrid models have been developed in different domains, no study has yet explicitly examined the use of hybrid LSTM-Transformer model in financial data analysis. Financial data have unique characteristics that call for a specialized architecture that supports these requirements, as financial data combine sequential patterns with contextual dependencies. This paper addresses this gap by proposing a modular LSTM-Transformer hybrid model for financial data processing, providing a theoretical framework and pseudocode for its implementation. This framework provides a basis for future validation and practical applications.

3. Methodology

In this study, we propose a hybrid architecture between LSTM networks and Transformer models to predict financial health and spending from unstructured financial data, including receipts and invoices. We begin by explaining the individual architectures of LSTM and Transformer, followed by a detailed explanation of how they are combined into a unique hybrid model suited specifically for financial data.

3.1 LSTM Architecture Overview

Recurrent Neural Networks (RNNs) can use time-series data and predict their future; however, they struggle with problematic cases such as long-term dependencies in data, where past events are highly predictive of future events (Lin, 2022). Unlike RNNs, LSTMs are different in the sense that information loss in long sequences takes place with a vanishing gradient, and the input, forget, and output gates are used to pass information with long sequences. Therefore, LSTMs are very good for time-series prediction tasks, such as predicting stock prices and financial level trends (Yadav, Jha, & Sharan, 2020).

In financial data, the sequences of transactions, receipts, and invoices have many temporal dependencies, that LSTM can model. The model uses historical spending patterns to learn what past behaviors mean to future financial health. For instance, if an individual is likely to spend certain categories in a certain sequence, LSTM can learn about such trends and adopt for future prediction (Lin, 2022; Niu et al., 2022).

3.2 Transformer Architecture Overview

In the paper Attention is All You Need (Vaswani et al., 2017), the Transformer model introduced has revolutionized sequence processing tasks. Unlike RNNs, Transformers do not process data one after the other and have the ability to process data in parallel (Zhang & Zhao, 2023; Zhang, Liu, Zhang, & Wang, 2023). Instead, they use self-attention mechanisms to weigh each element of the sequence against all the others. Owing to its global attention mechanism, the Transformer is capable of understanding long-range dependencies for contexts, which makes it extremely powerful for tasks that require contextual understanding over long-range dependencies.

Transformers are particularly effective in financial data, where the context between different financial events (e.g., a large purchase and a small payment) may vary and require an understanding of both local and global contexts. The ability of Transformers to pay attention to important parts of a sequence, based on attention scores, makes them ideal for applications such as invoice categorization, fraud detection, and financial forecasting (Yu et al., 2024).

3.3 Proposed Hybrid Architecture for Financial Data

The proposed hybrid model takes advantage of both the sequential modeling ability of LSTM and the global contextual understanding of Transformer networks. To solve the complexity of financial data, the architecture is built using both long-term dependencies (LSTM) and relationships between multiple financial events (Transformer). In this order, our approach following steps.

3.3.1 Data Preprocessing

This stage involves:

1. **Data Normalization:** Standardizing the amounts and dates of the transactions.
2. **Text extraction:** OCR (optical character recognition) to extract text from scanned receipts and invoices, and turn it into structured data (e.g. date, item, and price).
3. **Data Annotations:** Extract the text data and create annotations such that we can feed these annotations into the neural networks.

In this step, we obtain a clean, structured dataset that is suitable for the hybrid model.

3.3.2 Model Architecture

1. **LSTM Encoding for Sequential Data**

   The financial data are processed by the LSTM network once the temporal aspects are learned by the network. The LSTMs will receive receipts for data analysis; for example, they may understand from past data how regular categories of spending often occur and predict future spending behavior.

   (a) **LSTM Layer Design:** To overcome these further dependencies, we designed an LSTM encoder with a large number of hidden units. We determine the number of layers and hidden units, and argue that the resulting network is optimized for accuracy and computational efficiency.

2. **Transformer Decoder for Contextual Analysis**

   The sequential data processed are passed into the Transformer model, which determines the contextual understanding of the data. With this, the Transformer with the self-attention mechanism will be able to point into the sequence (e.g., a large transaction or change in spending behavior) to locations that are more important in predicting future outcomes.

   (a) **Self-Attention Mechanism:** The attention mechanism gives more weight to financial events (e.g., big transactions and unusual spending) relevant to predicting financial health.

(b) **Multi-Head Attention:** To enable the model to attend to various parts of the sequence (i.e., to learn short-term and long-term spending patterns simultaneously), using multiple attention heads is imperative.

3. **Fusion Layer for Model Integration**

   The outcomes of the LSTM and Transformer models were combined in the final layer. The final prediction output of financial health or spending behavior is produced by combining the strengths of the two models into this layer.

   (a) **Concatenation:** A unified representation of the financial data is constructed by concatenating the outputs from the LSTM and Transformer modules.

   (b) **Dense Layer:** These outputs are combined using a dense fully connected layer to predict financial outcomes (e.g., overall spending or future financial health).

4. **Output Layer**

   For the final prediction of financial health, spending trends, and anomaly detection results, the output of the model corresponds to the respective predictions for each task.

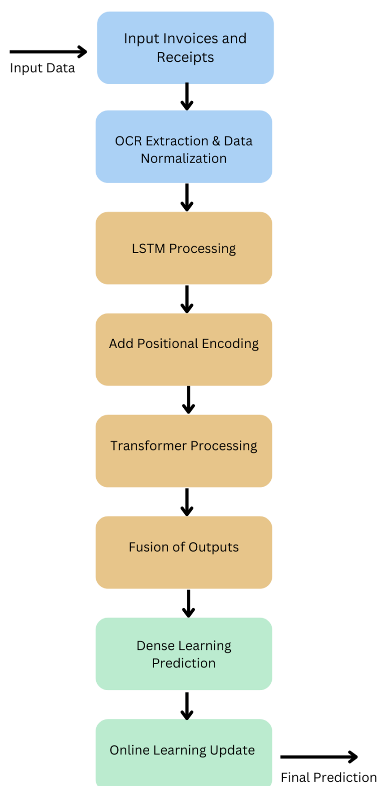

Figure 1: Flow Diagram of LSTM-Transformer Hybrid Architecture

5. **Pseudocode**

The pseudocode used to train the hybrid LSTM-Transformer model is as follows:

Algorithm 1: Hybrid LSTM-Transformer Model for Financial Data

```
procedure HybridModel
/* Initialization                                                   */
Initialize LSTM_parameters: hidden_units, num_layers;
Initialize Transformer_parameters: attention_heads, num_layers;
Initialize Positional_Encoding: max_sequence_length;
/* Financial Data Processing                                        */
Extract OCR_text ← OCR_extraction(financial_data);
Transform normalized_data ← normalize(OCR_text);
/* LSTM Layer Processing                                            */
foreach sequence in normalized_data do
    LSTM_output ← Process_with_LSTM(sequence, LSTM_hidden_state,
     LSTM_cell_state);
/* Add Positional Encoding for Transformer Processing               */
Transformer_input ← Add_Positional_Encoding(LSTM_output);
/* Transformer Layer Processing                                     */
foreach Transformer_layer do
    Attention_output ← MultiHead_Attention(Transformer_input);
    Transformer_output ← Apply_Feedforward_NN(Attention_output);
/* Fusion and Prediction                                            */
Fused_representation ← Concatenate(LSTM_output, Transformer_output);
Prediction ← Dense_layer(Fused_representation);
/* Unique Online Learning Update                                    */
Update_Model(Prediction, new_data);
return Prediction;
```

This pseudocode has been inspired from "Advanced hybrid LSTM-transformer architecture for real-time multi-task prediction in engineering systems" (Cao, Zhang, & Huang, 2024).

Explanation of Pseudocode

(a) **Hybrid Model Initialization**

i. Initialize LSTM and Transformer for hybrid model development.

ii. Positional Encoding is added to the LSTM output to make the sequence order explicit to the Transformer.

(b) **Financial Data Processing**

i. Extracts textual data from receipts and invoices and transforms raw text into numerical data suitable for processing by the LSTM and Transformer layers.

(c) **LSTM Layer Processing**

i. LSTM layers process sequentially normalized data to capture time-series patterns of financial transactions (Lin, 2022).

ii. The final LSTM layer is an output that contains localized patterns in the data, for example, short-term spending trends.

(d) **Transformer Layer Processing**

i. Positional Encoding was then added to the LSTM output to provide sequence-order awareness.

ii. Finally, Transformer layers compute the relationships between all elements in the sequence at once and identify complex dependencies.

(e) **Fusion of Outputs**

i. Both the LSTM and Transformer outputs are concatenated to combine local (LSTM) and global (Transformer) information on financial data.

ii. By combining the strengths of both architectures, this fusion proved to be well suited for financial data with a variety of dependencies.

(f) **Final Prediction**

i. The fused representation then proceeds to a dense layer to produce predictions, such as predicting financial health or grouping spending patterns.

(g) **Online Learning Updates**

i. The model is designed to learn continuously using online learning, allowing real-time integration of new financial data to improve its applicability to dynamic financial environments (Cao et al., 2024).

4. Discussion

The proposed hybrid LSTM-Transformer model is expected to be exceptionally good for predicting spending behavior and financial health from unstructured data. This section discusses the limitations of this approach and future related research.

4.1 Limitations

Although the hybrid model provides substantial performance, there are some limitations to consider.

1. **Theoretical Validation:** The proposed architecture has not been trained or validated on real-world datasets because of the unavailability of preprocessed financial data. Consequently, performance metrics, such as accuracy, efficiency, and scalability, are yet to be quantified.

2. **Computational Requirements:** Although effective, the Transformer component's self-attention mechanism is computationally expensive. Limited scalability of the model in large or real time scale applications without enough computational resources may be a consequence of this.

3. **Data Privacy and Security:** Financial data that usually contain sensitive information causes privacy and compliance issue. Finally, the model needs to be implemented further in order to work with data protection laws.

4. **Domain-Specific Tuning:** Predicting financial health from receipts and invoices is one where the model's effectiveness in such cases may not generalize to other types of financial data without more training and tuning.

4.2 Future Research Directions

Future work could address the limitations identified by exploring the following avenues:

1. **Implementation and Training:** A preprocessed dataset and training the proposed architecture with the dataset is a key next step to evaluate the performance of the architecture against benchmarks.

2. **Transfer Learning for Domain Generalization:** The performance is enhanced using pre trained models such as, for financial task, fine-tuned version to reduced data.

3. **Broader Applicability:** Extend architecture to support other domains like healthcare billing, supply chain finance in same modularity fashion that design (Li et al., 2022; Zhang & Zhao, 2023)

4. **Validation Through Implementation:** There will be a follow up paper where we will train this architecture using real world data, validate its performance and compare it with the current state-of-art models. The theoretical foundation laid here will serve as the starting of this progression.

5. Conclusion

This paper proposed a hybrid architecture that combines LSTM networks and Transformer models for training an NER model for financial data analysis. This architecture is designed to overcome the limitations of traditional models in handling financial documents such as invoices and receipts. By integrating the sequential learning capabilities of LSTM with the attention-based strengths of the Transformer (Bai & Tahmasebi, 2022; Zhang et al., 2023; Zhang & Zhao, 2023), this work introduces a modular approach for improving accuracy and predictions in financial data analysis.

The paper provides a detailed theoretical framework; however, the model has not yet been empirically validated due to the unavailability of preprocessed financial data. Nevertheless, the proposed design represents a major step forward, offering solutions to challenges such as contextual understanding and long-term dependency modeling, which are critical in financial document analysis. This work opens up opportunities for future research in testing, model optimization, and real-world deployment. It further prepares for the follow-up work, which subsequently validates the theoretical foundations described in this paper through practical implementations, thereby contributing to both academic literature and industry practices.

References

Bai, T., & Tahmasebi, P. (2022). Attention-based lstm-fcn for earthquake detection and location. *Geophys. J. Int.*, *228*(3), 1568–1576.

Cao, K., Zhang, T., & Huang, J. (2024). Advanced hybrid lstm-transformer architecture for real-time multi-task prediction in engineering systems. *Scientific Reports*, *14*(1).

Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q. V., & Salakhutdinov, R. (2019). Transformer-xl: Attentive language models beyond a fixed-length context. In *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, pp. 2978–2988.

Fu, Y., Song, J., Guo, J., Fu, Y., & Cai, Y. (2024). Prediction and analysis of sea surface temperature based on lstm-transformer model. *Regional Studies in Marine Science*, *78*, 103726.

Graves, A. (2013). *Supervised Sequence Labelling with Recurrent Neural Networks*, Vol. 385. Studies in Computational Intelligence.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Kuhn, M., & Johnson, K. (2013). *Classification trees and rule-based models* (1st ed. edition). Applied Predictive Modeling.

Li, Y., Tong, Z., Tong, S., Westerdahl, D., & Pastoriza, T. (2022). A data-driven interval forecasting model for building energy prediction using attention-based lstm and fuzzy information granulation. *Sustain. Cities Soc.*, *76*.

Lin, Y. e. (2022). A new attention-based lstm model for closing stock price prediction. *Int. J. Financ. Eng.*, *9*(3).

Liu, H., Gegov, A., & Cocea, M. (2015). Network-based rule representation for knowledge discovery and predictive modelling., 1–8.

Niu, D., Sun, L., Yu, M., & Wang, K. (2022). Point and interval forecasting of ultra-short-term wind power based on a data-driven method and hybrid deep learning model. *Energy*, *254*.

Ozbayoglu, A. M., Gudelek, M. U., & Sezer, O. B. (2020). Deep learning for financial applications: A survey. *Applied soft computing*, *93*.

Ranjan, R., & Daniel, A. K. (2023). Cobico: A model using multi-stage convnet with attention-based bi-lstm for efficient sentiment classification. *International Journal of Knowledge-Based Intelligent Engineering Systems*, *27*(1), 1–24.

Sakatani, Y. (2021). Combining rnn with transformer for modeling multi-leg trips. In *Proc. ACM WSDM WebTour*, pp. 50–52.

Tang, Q., Fan, T., Shi, R., Huang, J., & Ma, Y. (2021). Prediction of financial time series using lstm and data denoising methods. *arXiv preprint*.

Vaswani, A., et al. (2017). Attention is all you need. In *Proceedings of NeurIPS*, pp. 5998–6008.

Xie, Y., Liang, R., Liang, Z., Huang, C., Zou, C., & Schuller, B. (2019). Speech emotion classification using attention-based lstm. *IEEE/ACM Trans. Audio Speech Language Process*, *27*(11), 1675–1685.

Yadav, A., Jha, C. K., & Sharan, A. (2020). Optimizing lstm for time series prediction in indian stock market. *Procedia Computer Science*, *167*, 2091–2100.

Yang, F., Zhang, H., & Tao, S. (2022a). Travel order quantity prediction via attention-based bidirectional lstm networks. *J. Supercomput.*, *78*(3), 4398–4420.

Yang, L., Li, J., Dong, R., Zhang, Y., & Smyth, B. (2022b). Numeric-oriented hierarchical transformer model for multi-task financial forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36, pp. 11604–11612.

Yu, C., Xu, Y., Cao, J., Zhang, Y., Jin, Y., & Zhu, M. (2024). Credit card fraud detection using advanced transformer model. In *2024 IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom)*, pp. 343–350.

Zhang, N., & Zhao, L. S. (2023). Method for real-time prediction of cutter wear during shield tunnelling: A new wear rate index and mcnn-gru. *Methodsx*, *10*.

Zhang, W., Liu, X., Zhang, L., & Wang, Y. (2023). Intelligent real-time prediction of multi-region thrust of epb shield machine based on ssa-lstm. *Eng. Res. Express*, *5*(3).