Dear Sir or Madam,

Thank you very much for your feedback and suggestions, which in our opinion will result in significant improvements to this manuscript. My answers are as follows,

Comment 1: The paper is an interesting read with some relevant contributions. However, the experiments are based on assumptions about the data. These assumptions about the data are not evaluated. The method could also be tested on synthetic data that has been proved to have the assumed characteristics.

Answer: Much appreciate for this invaluable suggestion. We would like to forge a light synthetic data to vividly illustrate the challenges brought by the multimodal data. In addition, previous studies have showed that the stock data have the characteristics of interactive feature mode [1] and the heterogeneity of the data in terms of sampling times [2].

Comment 2: The article does not contain enough information on several very important parts of the model. The tensor stream X_t is not described well. For example,

- What goes into it? How is the fundamental information encoded by the tensor?
- The two operators that are used to compose the tensor X are not described.
- "The factor matrix R_k describes one distinct facet of the information space of the stock market." => What does this mean?

Answer: We appreciate this comment a lot, and the related parts in the manuscript are the tensor representation and tensor decomposition and reconstruction described in Section 3.1 and 3.2 respectively. In fact, as described in the manuscript, the tensor stream X_t consists of two modes, in which there are eight attributes in the fundamental's mode and three attributes in the news mode. Due to the limited pages, we did not cover all the description of the tensor processing sections. However, we would like to add more details in the revised version. In addition, we have released a full version of our work which has not been submitted to anywhere else for publication. This full version can be accessed at https://github.com/Anonymous440/Tensor-based-eLSTM/blob/master/Version1.pdf.

As described in Section 3.1, the two operators are the classic rules in the Tucker theory. That is, a tensor can be decomposed into a core tensor multiplied by a matrix along each mode. For better understanding, we will add the related reference paper [3] in the revised version. Based on the tensor theories, the factor matrix R_k is obtained by the tucker decomposition as described in Section 3.2. It can be described as one distinct facet of the market information (e.g. fundamentals or news mode). More detailed information is also available in the full version of the submitted manuscript.

Comment 3: It is not clear how the event information in E_t is represented. For example,

- What does every event e_t encode? I think that every e_t encodes the sentiment of one news article, but this is not clear.
- e_n is used in the figure, but the texts mentions e_t. This is confusing.
- What is the size of E_t. Is |E_t| = 92, the amount of companies? What if more news articles are published about the same company at the same time? What happens then? The representation is unclear.

Answer: Much appreciate for the comments. First, as described in Section 3.4, e_t is the total number of news articles at time t. The different representations with e_t and e_n in the figure are because that we evaluate the n-day-ahead influence caused by media information as described in Section 4. Therefore, we try to apply e_n for representing the previous n days influence in the day t as shown in Fig 3. Therefore, the size of E_t is n. We would like to make it clearer in the revised version. In addition, we have released a long full edition of the paper which only aims at publicity and has not yet been submitted to any conferences and journals. This version can be accessed at https://github.com/Anonymous440/ Tensor-based-eLSTM/blob/master/Version1.pdf. It provides more detailed information about the event-driven mechanism.

Comment 4: Thanks for sharing the code and the data. This is very valuable and much appreciated! Please version the code and data, so that readers can know that they are using the same code and data as was used in the experiments.

Answer: Much appreciate for this invaluable comment. We have released the data and source code with the manuscript. And we will follow this suggestion to reorganize the source code and experimental data for the final version.

Comment 5: The paper does not describe how the available data is divided into training, evaluation and test sets. Is nested cross validation used or is a part of the data taken out as a test set? How large part of the data is used as a test set? How does that influence the performance? A year of data seem little for such an experiment. I miss the persistence baseline. How well would that do in this case?

Answer: Much appreciate for this invaluable suggestion. Followed the previous work [2], the data were divided into two sets. The first 9 months of data were used to train the model, and the last 3 months were used for model evaluation and investment simulation. Note that, we adopted the sliding widows to incorporate the influence of previous k days. Therefore, the knowledge of the proposed model is kept learning while training. In addition, the experimental data is based on the daily level. According to the review on the media-aware stock movements [4], the size of one-year data is sufficient to evaluate the system performance. After submission, we have extended our experiments for three-years, and we would like to incorporate them in the revise version.

Comment 6: There are very many hyperparameters that can be changed. How can we be sure that the chosen set of hyperparameters are the best set of hyperparameters in general and not only the set of hyperparameters that fits the data the best? Which data was used for setting the hyperparameters? K is chosen experimentally. Which data is used choose K? Training, validation or test data? Why would K=6 be the best K? I see the figure, but it seems random.

Answer: We appreciate this comment a lot, and the related parts in the manuscript is the model parameters described in the Section 4.1. In this study, k is selected experimentally. That is, we set k to the range of [1~30] to find the best performance. In figure 7, when k passed 6, the system performance goes down, even though there are some fluctuations in terms of both accuracy and MCC. Therefore, we chose the optimal k as 6. In fact, we conducted a series of experiments to find the optimal settings. Due to the limited pages, we did not cover all of our tuning work. However, we would like to add more details in the revised version. We have released a full version of our work, which has not been submitted to anywhere else for publication. This version can be accessed at https://github.com/Anonymous440/Tensor-based-eLSTM/blob/master/Version1.pdf. It provides extra descriptions of model tuning and model comparison.

Comment 7: You do not show any evidence that this actually is the case. As long as the data has not been shown to have this property, there is no evidence that the algorithm finds this pattern.

Answer: Much appreciate for this comment. Previous studies show that the stock data have the characteristics of interactive feature mode [1] and the heterogeneity of the data in terms of sampling times [2]. We would like to make it clearer in the revised version.

Comment 8: The accuracy of the sentiment algorithm is not evaluated. Hence, the effect of the sentiments is only evaluated though the fact that the performance of the algorithm is improved when using sentiment. Stock relevance is encoded by looking at how often stocks are mentioned in the same article. This is a hypothesis that is not tested in the paper.

Answer: Much appreciate for the comments. In this study, we adopted a simple algorithm proposed by Tetlock [5] published in journal of finance. Although there are lots of advanced sentiment algorithms to further improve the performance of sentiment analysis, this simple method has been proved its efficiency and adopted in a number of previous studies to measure the investors' sentiment for stock predictions [1,4]. In fact, the fluctuation of a stock is affected by its related stocks. This is a well-known problem called as stock comovement in Finance [6]. Li et al. pointed it out as one of the critical challenges for predicting stock movements [4]. With the development of social media, the stock relevance can be captured by the co-exposure in the news events [2]. Therefore, we explore the effectiveness of the stock relevance in Section 4.2 and have proved its power. More details can be seen in our released full edition. This version can be accessed at https://github.com/Anonymous440/Tensor-based-eLSTM/blob/master/Version1.pdf.

References

[1] Q. Li, Y. Chen, L. Jiang, P. Li, and H. C. Chen. A tensor-based information framework for predicting the stock market. ACM Transactions on Information Systems (TOIS), 34(2):11, 2016.
[2] X. Zhang, Y. Zhang, S. Wang, Y. Yao, B. Fang, and S. Y. Philip, "Improving stock market prediction via heterogeneous information fusion," Knowledge-Based Systems, vol. 143, pp. 236–247, 2018.
[3] T.G. Kolda and B.W. Bader. Tensor decompositions and applications. SIAM Review, 51(3):455–500, 2009.
[4] Q. Li, C. Yan, J. Wang, Y. Chen, and H. C. Chen, "Web media and stock markets: A survey and future directions from a big data perspective," IEEE Transactions on Knowledge & Data Engineering, vol. 30, no. 2, pp. 381–399, 2018.
[5] P.C. Tetlock, M. Saar-Tsechansky, and S. Macskassy, "More than words: quantifying language to measure firms' fundamentals", Journal of Finance., vol. 63, no. 3, pp. 1437–1467, 2018.
[6] Liu Ling, Wu Jing, Li Ping, and Qing Li. A social-media-based approach to predicting stock comovement. Expert Systems with Applications, 42(8):3893–3901, 2015.