 --========  Review Reports  ========--  
  
The review report from reviewer #1:  
  
\*1: Is the paper relevant to ICDM?  
  [\_] No  
  [X] Yes  
    
\*2: How innovative is the paper?  
  [\_] 6 (Very innovative)  
  [\_] 3 (Innovative)  
  [X] -2 (Marginally)  
  [\_] -4 (Not very much)  
  [\_] -6 (Not at all)  
    
\*3: How would you rate the technical quality of the paper?  
  [\_] 6 (Very high)  
  [\_] 3 (High)  
  [X] -2 (Marginal)  
  [\_] -4 (Low)  
  [\_] -6 (Very low)  
    
\*4: How is the presentation?  
  [\_] 6 (Excellent)  
  [\_] 3 (Good)  
  [\_] -2 (Marginal)  
  [X] -4 (Below average)  
  [\_] -6 (Poor)  
    
\*5: Is the paper of interest to ICDM users and practitioners?  
  [\_] 3 (Yes)  
  [X] 2 (May be)  
  [\_] 1 (No)  
  [\_] 0 (Not applicable)  
    
\*6: What is your confidence in your review of this paper?  
  [\_] 2 (High)  
  [X] 1 (Medium)  
  [\_] 0 (Low)  
    
\*7: Overall recommendation  
  [\_] 6: must accept (in top 25% of ICDM accepted papers)  
  [\_] 3: should accept (in top 80% of ICDM accepted papers)  
  [\_] -2: marginal (in bottom 20% of ICDM accepted papers)  
  [X] -4: should reject (below acceptance bar)  
  [\_] -6: must reject (unacceptable: too weak, incomplete, or wrong)  
    
\*8: Summary of the paper's main contribution and impact  
  Summary: The authors use both the emotional and semantic aspects of article content and user comments to detect fake news. Using sentiments to detection fake news is not a novel approach but the authors' way of generating emotion vectors and leveraging the interaction (or fusing) between the sentiment vector and the semantics vector is interesting.  
  
\*9: Justification of your recommendation  
  Descriptions of data is vague. Some technical details need more elaborations (e.g. hand-crafted emotional features are not listed). Performance between datasets are significant, but no explanation was given.  
  
\*10: Three strong points of this paper (please number each point)  
  1: This paper's plots/graphs are simple but informative  
2: The authors incorporated numerous baseline models for performance comparisons  
3: the ablation section (Component Analysis) is detailed and well written  
  
\*11: Three weak points of this paper (please number each point)  
  1: Descriptions of data is vague.  
2: Some technical details need more elaborations  
3: The performance difference between Twitter data and Weibo data is high, but no explanation was given.  
  
\*12: Is this submission among the best 10% of submissions that you reviewed for ICDM'19?  
  [X] No  
  [\_] Yes  
    
\*13: Would you be able to replicate the results based on the information given in the paper?  
  [X] No  
  [\_] Yes  
    
\*14: Are the data and implementations publicly available for possible replication?  
  [\_] No  
  [X] Yes  
    
\*15: If the paper is accepted, which format would you suggest?  
  [\_] Regular Paper  
  [X] Short Paper  
    
\*16: Detailed comments for the authors  
  Summary: The authors use both the emotional and semantic aspects of article content and user comments to detect fake news.  
  
Data section: the description of the Weibo dataset is vague. What’s the time range of the dataset? How many users are included in the dataset? Rumor is a subcategory of fake news. Does the dataset contain different types of fake news or just rumors (the wording in the paper is a bit ambiguous)? Weibo is predominantly in Chinese, but is the dataset itself only in English?   
  
Model section: Is it possible for the authors to provide a complete list of the “Hand-crafted News Emotion Features”?  
  
Analysis section:   
authors choose to include top 5 comments for the content module. Have the authors done any robustness checks (e.g. using top 2 comments, or using top 10 comments?)  
The model performance difference between the Weibo and Twitter datasets are quite big (“87.2% on weibo dataset and 75.1% on Twitter dataset”), what are the possible reasons? Additionally, the “A PRELIMINARY ANALYSIS ON EMOTION SIGNALS section only included results from the Weibo dataset. Were observations for the Twitter dataset similar?  
Is the case study representative?  
  
Minor things:  
Grammar and incomplete sentences. E.g. “we give an overview of related.”, “the credibility of of tweets”,   
Some footnotes are unhelpful  
  
Finally, I would like to thank the authors for investing valuable time and resources to the study of fake news. It is definitely an important research endeavor.   
  
========================================================  
The review report from reviewer #2:  
  
\*1: Is the paper relevant to ICDM?  
  [\_] No  
  [X] Yes  
    
\*2: How innovative is the paper?  
  [\_] 6 (Very innovative)  
  [\_] 3 (Innovative)  
  [X] -2 (Marginally)  
  [\_] -4 (Not very much)  
  [\_] -6 (Not at all)  
    
\*3: How would you rate the technical quality of the paper?  
  [\_] 6 (Very high)  
  [\_] 3 (High)  
  [X] -2 (Marginal)  
  [\_] -4 (Low)  
  [\_] -6 (Very low)  
    
\*4: How is the presentation?  
  [\_] 6 (Excellent)  
  [\_] 3 (Good)  
  [X] -2 (Marginal)  
  [\_] -4 (Below average)  
  [\_] -6 (Poor)  
    
\*5: Is the paper of interest to ICDM users and practitioners?  
  [X] 3 (Yes)  
  [\_] 2 (May be)  
  [\_] 1 (No)  
  [\_] 0 (Not applicable)  
    
\*6: What is your confidence in your review of this paper?  
  [X] 2 (High)  
  [\_] 1 (Medium)  
  [\_] 0 (Low)  
    
\*7: Overall recommendation  
  [\_] 6: must accept (in top 25% of ICDM accepted papers)  
  [\_] 3: should accept (in top 80% of ICDM accepted papers)  
  [\_] -2: marginal (in bottom 20% of ICDM accepted papers)  
  [X] -4: should reject (below acceptance bar)  
  [\_] -6: must reject (unacceptable: too weak, incomplete, or wrong)  
    
\*8: Summary of the paper's main contribution and impact  
  1) A novel deep learning framework was proposed for fake news detection, which explored the possibility of using emotions expressed in the content of news and its social feedback. Rather than directly concatenating the representation of news content and user comments to classify microblogs from Weibo and Twitter (two languages involved), the proposed method used gate mechanism for fusion, which can learn the weights of input components. However, though this paper noticed that emotions haven’t been deeply mined to identify fake news, it didn’t clarify the boundary between  
fake news and, e.g., rumor and satire news, which might lead to the conclusions that are not very convincing. Additionally, though this paper comprehensively considered the available information from both news content and social context (user reposts and comments), there has been current work exploiting both news content and social context information to detect fake news (e.g., Shu et al., Beyond news contents: The role of social context for fake news detection, WSDM, 2019).   
  
2) The authors built their own Weibo dataset including user posts, reposts, comments and their corresponding emotions expressed. However, they neither provided the access to the dataset nor some important details in dataset construction.(details are in 16)  
  
3) This paper developed deep learning techniques for fake  
news detection, where both topics have attracted researchers’ attention particularly in recent years, which might lead this  
paper interest to a large proportion of ICDM users and practitioners.  
  
\*9: Justification of your recommendation  
  Providing in “detailed comments”.  
  
\*10: Three strong points of this paper (please number each point)  
  1) The motivation of this paper is persuasive, where it noticed  
that emotions haven’t been deeply mined to identify fake  
news and it comprehensively considered the available  
information from both news content and social context.  
  
2) Rather than directly concatenating the representation of  
news content and user comments to classify microblogs in  
Weibo and Twitter, the proposed method used gate  
mechanism for fusion, which can learn the weights of input  
components.  
  
3) This paper is easy to follow.  
  
\*11: Three weak points of this paper (please number each point)  
  1) It has a reproducibility issue with lacking key details of handcrafted  
features, in dataset construction and in the model  
training process.  
  
2) The contribution is marginal for the paper to be  
accepted in ICDM.  
  
3) It does not clarify the boundary between fake  
news and rumor, in work reviewing and writing (details are  
provided in 16).  
  
\*12: Is this submission among the best 10% of submissions that you reviewed for ICDM'19?  
  [X] No  
  [\_] Yes  
    
\*13: Would you be able to replicate the results based on the information given in the paper?  
  [X] No  
  [\_] Yes  
    
\*14: Are the data and implementations publicly available for possible replication?  
  [X] No  
  [\_] Yes  
    
\*15: If the paper is accepted, which format would you suggest?  
  [\_] Regular Paper  
  [X] Short Paper  
    
\*16: Detailed comments for the authors  
  The paper aims to detect fake news by mining the emotions revealed in the content of news and its social feedback. Rather than directly concatenating the representation of news content and user comments to classify microblogs in Weibo and Twitter, the proposed method EFND used gate mechanism for fusion, which can learn the weights of input components. The representation of either news content, or user comments was obtained by embedding the content (at word level) and emotions within the content,  
respectively, which would be further inputted into bi-GRU.  
  
In general, this paper is easy to follow. The motivation of this paper is persuasive, where it noticed that emotions haven’t been deeply mined to identify fake news and it comprehensively considered the available information from both news content and  
social context. However, in my opinion the contributions are marginal, which have been specified in 8. In addition, reproducibility is one of my major concerns, where  
the current version of this paper lacks some key details in the proposed model and in experimental setup. Specifically, below are my concerns, suggestions and questions for this paper:  
  
1) Reproducibility issue. “For a given post pj , we extract the  
emotion features included in work [2] and also add some  
extra emoticon features. There are 19 features regarding  
emotion aspects of news, including numbers of  
positive/negative words, sentiment score, etc.” I have no  
access to clearly understand what exactly these 19 features are and  
how they can be computed.  
  
2) I looked at the source provided to collect real news  
(“newsverify.com”) on Weibo and had a question. I saw the  
credibility of news is evaluated by 0 to 5 stars (as well as a  
score from 0% to 100%) instead of directly labeling the  
news as real or not (binary). What’s the star/score threshold  
for authors to take the news as real? The authors didn’t  
provide such details. Thought real news should be assigned  
5 stars or 100% intuitively while I didn’t saw any news  
satisfying it (most are at most 4.5 stars or approach 100%).  
  
3) Again, many experimental details were not provided in the  
paper. For example, how did the loss change during the  
training and if it finally converged? How did the authors  
divide each dataset as training and testing part? Did the  
authors use cross-validation to train the model?  
  
4) The authors aligned each word with a 32d-vector and with  
a 16d-vector for emotional embedding. Why did the  
authors set these two different sizes? I have this question as  
two dimensions are both comparatively small compared to  
that other researchers usually set when using word2vec.  
Did the authors take a look at the change of the  
performance when two dimensions vary, and/or their  
difference vary?  
  
5) This paper worked on detecting fake news, which was  
defined as intentionally false information, while reviewed  
many papers working on detecting rumor, as the  
information whose correctness hasn’t been verified. Rumor  
detection does relate to fake news detection; rumor-related  
studies are also valuable. However, it leads to  
missing many state-of-the-art fake news detection papers.  
  
6) Again, the review and summary for current fake news  
studies is limited. For example, in introduction, I agree with  
“existing work on fake news detection mainly focuses on  
news content and social context.” However, for “featurebased  
classification models extract basic semantic and  
emotion features from content, and statistical features from  
users [2]. Propagation-based models construct the  
relationship network inside the event, and incorporate  
social conflicting viewpoints towards event in the network  
[3] [4]”, propagation-based models are not just limited in  
this way of detecting. Some reviews and surveys are  
suggested for having a bigger view:  
o Shu, K., Bernard, H. R., & Liu, H. (2019). Studying fake news  
via network analysis: detection and mitigation. In Emerging  
Research Challenges and Opportunities in Computational Social  
Network Analysis and Mining (pp. 43-65). Springer, Cham.  
o Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., &  
Liu, Y. (2019). Combating fake news: A survey on identification  
and mitigation techniques. ACM Transactions on Intelligent  
Systems and Technology (TIST), 10(3), 21.  
o Zhou, X., & Zafarani, R. (2018). Fake news: A survey of  
research, detection methods, and opportunities. arXiv preprint  
arXiv:1812.00315.  
o Oshikawa, R., Qian, J., & Wang, W. Y. (2018). A survey on  
natural language processing for fake news detection. arXiv  
preprint arXiv:1811.00770.  
5  
  
7) Speaking of emotions, rather than intentionally false  
information (as the definition of fake news in this paper),  
the first thing coming to my mind was to detect satire news  
(as well as hate speech). Whether fake news includes satire news  
hasn't been clarified, while in the work of Rubin et al.  
(“Deception detection for news: three types of fakes”), they defined  
satire news as a sub-category of fake news. Did the authors  
regarded satire news as fake news? If not, is there any  
emotional difference between satire and fake news? The  
authors can discuss it a little bit in the paper.  
  
8) It is not recommended to describe datasets after Section III,  
where the authors have analyzed the data.  
  
9) It can be observed from Fig. 3 that fake news exhibits emotional difference from real news, e.g., the intensities of emotion anger in fake news are stronger than in real news. But is such difference statistically significant? I suggest the authors to calculate statistical significance tests, e.g., provide p-values.  
  
10) Some expressions are subjective without persuasive  
arguments or any references. For example,  
a. In intro, “these fake news do serious harm to not only the public credibility, but also social stability and economic market.”  
b. Also in intro, “since a large part of fake news doesn’t contain emotional signals in contents, difference of emotion between fake news and real news content may  
not be obvious.” How could the authors get the conclusion that “a large part of fake news doesn’t contain emotional signals in content”? Is there any analysis supporting it or any reference having indicated it?  
  
========================================================  
The review report from reviewer #3:  
  
\*1: Is the paper relevant to ICDM?  
  [\_] No  
  [X] Yes  
    
\*2: How innovative is the paper?  
  [\_] 6 (Very innovative)  
  [\_] 3 (Innovative)  
  [X] -2 (Marginally)  
  [\_] -4 (Not very much)  
  [\_] -6 (Not at all)  
    
\*3: How would you rate the technical quality of the paper?  
  [\_] 6 (Very high)  
  [X] 3 (High)  
  [\_] -2 (Marginal)  
  [\_] -4 (Low)  
  [\_] -6 (Very low)  
    
\*4: How is the presentation?  
  [\_] 6 (Excellent)  
  [\_] 3 (Good)  
  [X] -2 (Marginal)  
  [\_] -4 (Below average)  
  [\_] -6 (Poor)  
    
\*5: Is the paper of interest to ICDM users and practitioners?  
  [X] 3 (Yes)  
  [\_] 2 (May be)  
  [\_] 1 (No)  
  [\_] 0 (Not applicable)  
    
\*6: What is your confidence in your review of this paper?  
  [X] 2 (High)  
  [\_] 1 (Medium)  
  [\_] 0 (Low)  
    
\*7: Overall recommendation  
  [\_] 6: must accept (in top 25% of ICDM accepted papers)  
  [\_] 3: should accept (in top 80% of ICDM accepted papers)  
  [X] -2: marginal (in bottom 20% of ICDM accepted papers)  
  [\_] -4: should reject (below acceptance bar)  
  [\_] -6: must reject (unacceptable: too weak, incomplete, or wrong)  
    
\*8: Summary of the paper's main contribution and impact  
  This paper analyzes the rationality of using emotion for fake news detection not only from experimental results but also from the analysis of datasets. The analysis of the model and the rationality is also very complete. The paper also suffers from several problems, which are listed as follows.  
  
\*9: Justification of your recommendation  
  This paper provides a complete implementation of the comprehensive use of text information and emotional information in the fake news detection problem and combines the information of the fake news itself and related comments to improve the performance of the model. The analysis of the model and the rationality is also very complete. Although the specific implementation of the model mainly uses some mature methods, it is also a meaningful contribution to solve specific problems well.  
  
\*10: Three strong points of this paper (please number each point)  
  1, This work is timely and focus on an interesting and important topic, and the paper is generally easy to follow and understand.   
 2, The proposed framework EFND exploits a deep neural network to learn representations from publisher emotion, social emotion and content simultaneously, for fake news detection.   
 3, The experiment section is well designed, especially the design of the component analysis.  
  
\*11: Three weak points of this paper (please number each point)  
  1, Why should Gate\_N, Gate\_M, and Gate\_C be designed in different structures and why are they so designed? I think it is worth giving some more detailed explanation.  
2, Some typos exist in the paper, and some polish work are needed.   
3，The experimental results in Table3 may make readers feel confused, because there exist more than one bold-font results with one column or one row. I recommend that the table can be split into multiple tables.  
  
\*12: Is this submission among the best 10% of submissions that you reviewed for ICDM'19?  
  [X] No  
  [\_] Yes  
    
\*13: Would you be able to replicate the results based on the information given in the paper?  
  [X] No  
  [\_] Yes  
    
\*14: Are the data and implementations publicly available for possible replication?  
  [X] No  
  [\_] Yes  
    
\*15: If the paper is accepted, which format would you suggest?  
  [\_] Regular Paper  
  [X] Short Paper  
    
\*16: Detailed comments for the authors  
  1, This work is timely and focus on an interesting and important topic Fake News Detection.  
2, The experiment section is well designed and the authors list the key questions and use experimental results to answer them.  
3, This paper analyzes the rationality of using emotion for fake news detection not only from experimental results but also from the analysis of datasets in section 3. In response to such practical problems, I feel that this method of analyzing the data itself is solid.