NOTE: Comments with explanations underneath (which are underlined) are those which we did not follow. We took heed of and made changes accordingly for all other feedback.

==========
REVIEWER #1
==========

Reviewer's Scores

Appropriateness: Appropriate (most submissions)
Clarity (1-5): 3
Soundness (1-5): 4
Replicability (1-5): 3
Overall recommendation: Accept

Detailed Comments

Paper follows the guidelines for semeval system papers, the structure is good, the main sections are present.
The sections don't follow the recommendations by heart but that is not a problem from my point of view.

1.Introduction

I believe that since the semeval task was presented this would have been a great place to discuss rankings and results of the model presented in the competition.

=> Cagri said to save rankings/results for the results section. Our results can also be found in the abstract.

2. Background

I believe this sentence "The data consisted of pairs of English sentences, one of which made sense and the other against commonsense." would sound better like this: "The data consisted of pairs of English sentences, one of which made sense and the other which did not."

2.1 Pre-Trained Models

Here I have a small remark, the first sentence I understand (by the way there is a missing dot at the end of the phrase) however the second sentence
"When the fine-tuned ELMo was applied to a SemEval 2018 common sense dataset, it outperformed previous state-of-the-art performance by scoring 74.1 percent accuracy (Wang et al., 2019)." I believe it's missing the main idea, maybe you should why this result is important or how it influenced your decisions in the design of the model, something such as "thus it can be seen language models are a powerful tool and it is this reason we chose to use it as a main driver for our model".

3. System Overview

I don't really understand this sentence: "Between the then computed word vectors the cosine distance is measured and averaged.", more specifically this "Between the then computed word".

5. Results

I can't really find you on the results file. Maybe you should make it more clear what team name do you have.

Also I can't seem to find a team with 75% accuracy, there is no submission.

There are also more than 28 teams.

You should make sure to specify in what phase did you get the result, practice, evaluation or post-evaluation?

5.1 Error Analysis

"We performed the analysis only on a portion of the data." please be more specific about the portion.

6 Discussion

I believe a link or citation for FrameNet must be added.

I noticed you have very long paragraphs, maybe it would be a good idea to try and split them up, sometimes they are hard to follow.

=> We felt the lengths of the paragraphs were fine and well-organized based on the topic of discussion.

Also related to this hardship of following maybe some visual aid would help better understand the system, a model figure with the pipeline or even examples of how data get transformed between each step.

=> We declined to add a figure, as we thought this would be unnecessary considering the simplicity of the model. It seemed that some reviewers were confusing our different approaches. We adjusted multiple sentences in “System Overview” to make it apparent that we had several separate approaches, and we further specified the exact calculations we made to solve this task.

Maybe in the system model you can add some code snippets or hyperparameters because the description is very general, I don't think it will be easily reproducible.

=> We did not alter any hyperparameters or change the model in any way. We feel this was clearly stated in System Overview.

==========

REVIEWER #2

==========

Reviewer's Scores

Appropriateness: Appropriate (most submissions)
Clarity (1-5): 4
Soundness (1-5): 4
Replicability (1-5): 4
Overall recommendation: Accept

Detailed Comments

For well-prepared reader this papar is well-writen and well-structured which is understandable by most readers. It will be beter if some inportant equations attached. Although the accury if the proposed system is not high, I think the system is well-chosen because the choice of the system is based on two conclusions after analyzing the data. This paper takes perplexity of each sentence and distance between word vector of words as the focus of the task. Although these two points cannot contain all the situations we should consider, this is still the key breakthrough of the subtask A. In addition, at the end of the paper, full results analysis is listed. In addition to the above, I hope to provide more formulas and model diagrams to help readers understand and reproduce the paper more easily.

=> We addressed these same concerns from Reviewer No. 1.

==========

REVIEWER #3

==========

Reviewer's Scores

Appropriateness: Appropriate (most submissions)
Clarity (1-5): 2
Soundness (1-5): 2
Replicability (1-5): 2
Overall recommendation: Accept

Detailed Comments

This paper tried word vector distances, siamese network and GPT for ComVE. The authors perform some interesting discussions for the model and data. The writing is good and easy to follow. However, the result of the system is somewhat low. Why not use GPT-2 or RoBERTa?

=> We feel we motivated our intention with our model. We considered many during the project and felt the language model would perform the best.
