November 9, 2023

Dear Reviewer,

We would like to express our sincere gratitude to you for the time and effort you have spent reviewing our manuscript. We have carefully considered your comments and suggestions and have made the necessary revisions to the manuscript. Please find below a detailed response to each of the comments.

"Suggestion of title: Boosting Lightweight Sentence Embeddings with Knowledge Transfer from Advanced Models: A Model-Agnostic Approach"

We greatly appreciate your constructive suggestion regarding the paper's title. In light of your advice, we have updated the title to "Boosting Lightweight Sentence Embeddings with Knowledge Transfer from Advanced Models: A Model-Agnostic Approach," which we believe more accurately reflects the essence and the methodological breadth of our study.

This new title underscores both the enhancement aspect of our approach and its broad applicability, irrespective of the underlying models involved in the knowledge transfer process. We are confident that this change will offer readers a clearer understanding of the paper's contribution right from the outset.

"*In the abstract, it would be beneficial to have more insights or explanations about what "distorted, model-specific fashion" means in practical terms and the implications of such a distortion.*"

Upon reflection, we agree that the phrase "distorted, model-specific fashion" may have been too vague without the necessary context and could lead to confusion. As suggested, we have removed this phrase from the abstract to improve clarity. To provide more insight, we have added a brief explanation of the practical implications of the distortion caused by model-specific characteristics. This addition will help readers understand the significance of our findings and the potential impact on the field.

The revised abstract now reads as follows:

In this study, we investigate knowledge transfer between two distinct sentence embedding models: a computationally demanding, highly performant model and a lightweight model derived from word vector averaging. Our objective is to augment the representational power of the lightweight model by exploiting the sophisticated features from the robust model. Diverging from traditional knowledge distillation methods that align logits or hidden states of teacher and student models, our approach uses only the output sentence vectors of the teacher model for the alignment of with the student models's word vector representations. We implement two minimization techniques for this purpose: *distance minimization* and *distance & perplexity minimization*. Our methodology uses WMT datasets for training, and the enhanced embeddings are validated via Google's Analogy

tasks and Meta's SentEval datasets. We found out that our proposed models intriguingly retained and conveyed information in a model-specific fashion.

"However, there's some redundancy in the text. For instance, the concept of the student model trying to mimic or acquire representations from the teacher model is repeated multiple times.Some sentences are awkwardly phrased or have minor grammatical errors. E.g., "be lack of prediction capabilities" should be "lack prediction capabilities. The introduction mentions scenarios where distillation methods can be challenging due to limited access to model parameters or the need to redevelop models. However, the significance of these challenges and how they relate to the paper's primary focus isn't elaborated upon.The statement that "generating sentence vectors with FastText is very efficient and calculations rely on CPU" might need further explanation. Some might interpret that FastText exclusively uses CPU, which may not be accurate. The introduction states that sBert excels over FastText in terms of accuracy and similarity tests. However, this is introduced suddenly without a clear lead-in. The transition between discussing the models' technical specifications and their performance could be smoother."

- Redundancy in text: We reviewed the manuscript thoroughly and have eliminated redundant descriptions of the student model's mimicry of the teacher model. We have rephrased the text for brevity and clarity.
- Grammar and phrasing issues: We have carefully gone through the manuscript and corrected grammatical errors, including the one you highlighted ("be lack of prediction capabilities" has been changed to "lack prediction capabilities").
- Significance of challenges in distillation methods: We have expanded the discussion in the introduction to better articulate the relevance of the challenges associated with distillation methods to the focus of our paper. This should now clearly convey why these challenges are significant in the context of our research.
- FastText's computational efficiency: We have revised the statement regarding FastText to avoid any misinterpretation about its exclusive reliance on CPU. The revised text now explains that while FastText is designed for efficient computation on the CPU, it does not preclude the use of other processing units.

We revisioned all the text including the points that you have made.

"However, some transitions between topics or papers could be smoother. For example, "Apart from sentence generation model, knowledge distillation is a widely explored technique..." seems abrupt. Some methods or papers are explained in detail while others just get a passing mention. A bit more consistency would be beneficial. There are a few grammatical errors and awkward phrasings, such as "Models such as Bert[11] opened new opportunities for which can be exploited for…

Use smoother transitions between various topics to ensure the literature review flows seamlessly. The review ends by stating the unique focus of the current study, but it might be helpful to underscore this difference more prominently. For seminal works or models like BERT, consider providing a slightly more detailed explanation given their significance

in the field. Briefly explain why each cited work is relevant to the current study. This can help readers connect the dots more easily."

- Transitions between topics: We have revised the literature review to ensure smoother transitions between topics. For instance, the sentence you highlighted has been rephrased to "In addition to being a cornerstone in sentence generation models, knowledge distillation also enjoys widespread application as a technique...". This provides a more natural segue into the discussion of knowledge distillation.
- Consistency in explaining methods or papers: We have striven for a more consistent level of detail across the methods and papers discussed. While our focus is on a pioneering concept, we have ensured that each cited work is given sufficient explanation to establish its relevance clearly.
- Grammatical errors and phrasing: All identified grammatical errors and awkward phrasings have been corrected. We have paid particular attention to sentence construction to enhance the manuscript's readability.
- Highlighting the unique focus of the study: We have revised the conclusion of the literature review to more prominently underscore the unique aspects of our study.
- Explanation of seminal works: We have expanded the discussion around seminal models like BERT to provide the reader with a more comprehensive understanding of their significance in the field.

we have undertaken a comprehensive revision of the text with a focus on enhancing the flow and consistency of transitions between topics. We have ensured smoother transitions to facilitate a more coherent narrative throughout the manuscript.

We have also revisited the literature review section to achieve a balance in the level of detail provided for each reference. While our work introduces a pioneering concept, we recognize the importance of situating it within the existing body of research. To this end, we have expanded on the discussions of the three pivotal papers, references [14], [19], and [20], to clearly delineate their relevance to our study. Although reference [20] pertains to image processing, its methodology for knowledge transfer across disparate neural architectures is particularly pertinent to our work and has now been given a more detailed treatment.

"However, the information, especially around the reasons for evaluating only a subset and the underlying models, is dense. Some readers might find it difficult to quickly grasp the significance of specific results. The term 'FastTextupdated' appears twice in Table 3, which can be confusing. It's unclear if they represent different models or if this is a typographical error. There's a sudden mention of "Distance Min.," which seems to refer to "Distance Minimization." While abbreviating can save space, it can sometimes cause confusion if not introduced properly. While tables are effective, visual representations like bar graphs or line plots might help readers visualize the performance differences more effectively, especially when there are multiple models or methods to compare. Although the section does compare updated embeddings against original embeddings, it might benefit from a

comparison with standard or widely-accepted benchmarks in the field. There's a mention of sBert's word relations and sBert vectors, but without proper context, it's not immediately clear to unfamiliar readers what role these vectors play. Some of the explanations, like the description around "Distance Minimization" and "Distance & Perplexity Minimization," could be streamlined for clarity."

- Clarification of Tables: To address the confusion regarding 'FastTextupdated' mentioned in Table 3, we have carefully checked and corrected the typographical errors. We have now differentiated between models where necessary and provided clear definitions at the beginning of Section 6 to help readers understand the context of each model evaluated.
- Abbreviations and Terminology: We have introduced and explained abbreviations such as "Distance Min." (Distance Minimization) the first time they appear in the text to prevent any ambiguity.
- Visualization of Results: Instead of using plots, we added extra information to tables 3 and 4 for better interpretability.
- Benchmark Comparisons: We have elaborated on the SentEval benchmarks used for sentence evaluation to establish the context and significance of our results against standard industry benchmarks.
- Explanation of sBert Embeddings: We have revised the explanations of the word embeddings obtained from sBert, ensuring they are accessible to readers unfamiliar with these concepts.

"Resolve the repeated use of 'FastTextupdated' in Table 3 to ensure clarity. Consider including visual aids like graphs to represent performance differences more vividly. After presenting the tables, a brief summary of the main takeaways will help the reader consolidate their understanding. Introduce abbreviations and acronyms with care, ensuring that they don't add confusion. If possible, compare the results with well-established benchmarks in the field to provide context about the performance of the proposed models. Streamlining the language and breaking down complex sentences can enhance clarity and readability. Consider diving deeper into the reasons behind certain anomalies in results (e.g., why the Distance Minimization method loses all word context information)."

- Table Clarification: We have resolved the issue of the repeated 'FastTextupdated' label in Table 3. For clarity, we added information about the models in the introduction of section 6 (Results section). Each instance now distinctly identifies the relevant model configurations.
- Summary of Results: Following each table, we have added a concise summary highlighting the main findings to assist the reader in consolidating their understanding of the complex data presented.

- Abbreviations and Acronyms: We have taken care to introduce and define all abbreviations and acronyms upon their first appearance to prevent any potential confusion.
- Benchmark Comparisons: We gave 2 reference papers in the results section for comparing the results.
- Language and Readability: The manuscript has been carefully revised to streamline the language and simplify complex sentences, thereby enhancing overall clarity and readability.
- Anomalies in Results: We have delved deeper into the reasons behind certain anomalies, such as the relation between word domain coverage and classification results, and provided an explanation within the text.
- Execution Times and Additional Resources: A new sub-section detailing the execution times of the models has been added for comprehensive understanding. Furthermore, we have provided a direct link to the resources used for sBert sentence generation.

"However, the discussion, while detailing the approach and results, lacks an in-depth comparative analysis with other existing methods or similar studies. Statements such as "it can also be assumed that the updated word embeddings converge to sBert's word embedding layers" need more justification or evidence."

We decided to remove the statement "it can also be assumed that the updated word embeddings converge to sBert's word embedding layers" from our discussion.

"Sentences like "Our unique proposition entails utilizing the average of pre-logit hidden states which makes the distillation process only by using a single vector" could benefit from clarification. While differences between the conventional and the proposed methods are highlighted, more nuanced differences, pitfalls, or challenges faced could be discussed to provide a comprehensive view of the study's journey.

Moreover, investigate potential real-world applications or use-cases that could benefit from the improved embedding strategy. Compare and contrast the computational costs, time difficulties, and resource consumption of the proposed methods to those of traditional methods. Address the difficulties encountered during the study in greater depth, possibly expanding on how they were (or were not) solved. Consider how the procedures could be extended beyond the specific models mentioned. Can the method be applied to other models or frameworks? Discuss whether or not any hyperparameter tuning or optimization was undertaken, and if not, speculate on the potential consequences. Investigate whether the results have implications or applications in other fields or interdisciplinary areas."

In response, we have thoroughly revised the discussion and conclusion sections for clarity and comprehensive coverage of the study's scope and implications. Specifically, we have clarified complex sentences, such as the one you highlighted about our unique proposition, to ensure that our methodology is transparent and understandable.

We have expanded our discussion to include a nuanced examination of the challenges and pitfalls encountered during the study. This includes a new subsection detailing the

computational costs associated with our proposed methods compared to baseline models.

The Experiments section now contains additional information on hyperparameter optimization. Lastly, we have explored the broader implications of our results.

We entirely rewritten the discussion and conclusion sections. Also included information in Experiments section such as hyperparameter optimization.

"The paper's abbreviations are defined near the end. Please incorporate them into the main text."

We have ensured that all abbreviations are indeed introduced and defined within the main text for immediate reference, particularly in Table 1 where they are used extensively. This should facilitate the reader's understanding as they progress through the paper.

Additionally, we have retained the list of abbreviations at the end of the manuscript to comply with the journal's formatting requirements. This dual approach is intended to enhance readability while adhering to the prescribed structure of the publication.

We trust that these revisions will satisfactorily address your concerns and enhance the manuscript's contribution to the field.

The constructive critique provided has not only helped in addressing specific issues but also guided a more comprehensive refinement of the work presented.

We are particularly grateful for your recommendation on the title change, which has significantly improved the paper's precision in conveying its core premise and scope. Your expertise has been instrumental in this revision, ensuring that the paper communicates its contributions to the field more effectively.

Thank you once again for your constructive feedback and guidance. It is greatly appreciated.

Warm regards,

Kadir Günel

Davutpaşa Mah. Davutpaşa Caddesi 34220, Esenler / İSTANBUL TEL: +90-544-321-30-02
kadir.gunel@std.yildiz.edu.tr