Natural Language Processing Article Reviews

Phoe Aung, Dylan Bortolon, Laurence Burden, and Amanda Burfield

Purdue Global University

IN400: Artificial Intelligence (AI) - Deep Learning and Machine Learning

Robert Kayl

20 September 2023

# 

Natural Language Processing Article Reviews

## Participants (Heading Level 2)

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan.

## Assessments and Measures

Ut blandit malesuada quam, ac varius tortor gravida eget. Vestibulum id ligula leo, ut accumsan mi. Sed tristique euismod convallis. Nulla facilisi. Etiam vestibulum est id orci interdum vitae porta enim blandit. Cras sit amet arcu dolor, at venenatis erat. Vestibulum accumsan placerat mauris. Morbi nec nibh nibh. Duis ultricies posuere nunc. Morbi at tellus quis magna vestibulum eleifend.

### Heading Level 3

Maecenas ullamcorper bibendum consequat. Pellentesque ultrices, eros eu tincidunt pretium, magna leo volutpat libero, non bibendum diam nunc eget urna. Vivamus eu tortor et dui aliquam vestibulum at vel augue. Vivamus elit dui, porttitor eget egestas at, rhoncus in justo. Curabitur tristique, elit ac venenatis volutpat, eros mauris iaculis diam, vitae rhoncus erat metus vitae eros.

**Heading level 4 is inline with the following paragraph, separated by a period.** Nulla congue egestas ante, id ultricies orci dignissim commodo. Fusce placerat, libero eu pharetra pulvinar, lorem dui pulvinar nisi, et semper orci orci vitae magna. Nullam sodales, felis id feugiat scelerisque, tortor nulla interdum mauris, ac porttitor odio dolor eget eros.

**Second level 4 heading.**Duis sit amet ipsum pretium erat accumsan iaculis vitae eget risus. Donec ut dui in lorem volutpat fermentum bibendum pulvinar libero. Nunc imperdiet eros et mi posuere pellentesque. Donec tincidunt ipsum eget nisl ullamcorper eu placerat libero ullamcorper. Maecenas id luctus ligula. Cras condimentum eleifend nibh sit amet iaculis. Suspendisse placerat sollicitudin mi, vel ornare augue hendrerit ac. Nulla sed suscipit sapien. Cras pellentesque orci lectus, eu consequat enim.

***Heading level 5 is also inline.*** Nulla congue egestas ante, id ultricies orci dignissim commodo. Fusce placerat, libero eu pharetra pulvinar, lorem dui pulvinar nisi, et semper orci orci vitae magna. Nullam sodales, felis id feugiat scelerisque, tortor nulla interdum mauris, ac porttitor odio dolor eget eros.

***Second level 5 heading.*** Duis sit amet ipsum pretium erat accumsan iaculis vitae eget risus. Donec ut dui in lorem volutpat fermentum bibendum pulvinar libero. Nunc imperdiet eros et mi posuere pellentesque. Donec tincidunt ipsum eget nisl ullamcorper eu placerat libero ullamcorper. Maecenas id luctus ligula. Cras condimentum eleifend nibh sit amet iaculis. Suspendisse placerat sollicitudin mi, vel ornare augue hendrerit ac. Nulla sed suscipit sapien. Cras pellentesque orci lectus, eu consequat enim.

# Peer Review

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan.

# Review of Sequence Classification with Human Attention

**By Laurence Burden**

## Selected Paper Summary

The provided paper was written by five researchers who had the goal of introducing natural human attention data into a natural language processing algorithm. The human attention process is described as the natural reading method by humans to omit certain function words and “focus on less predictable content words” (Barrett et al., 2018). This data is then used across NLP models to detect sentiment analysis, grammatical error detection, and detection of abusive language.

The researchers created a recurrent neural architecture using both the standard recurrent parameters and the attention function created from open-source eye-tracking data. They used a grouping of labeled sentences and sequences that associate a scalar value that represents human attention as their input set. The researchers point out that their protocol does not use, and does not need in-task eye-tracking recordings, and instead relies on already existing models of human reading attention (Barrett et al., 2018). Using these methods allows the researchers to compare human versus machine attention and to use the human attention data to influence the way a machine reads information. These methods were then used to create the various models for the three use cases they set out to compare against the usual model creation methods.

The first use case the team tackled was sentiment analysis. Sentence-level analysis was selected, rather than word or phrase analysis. They also reduced the sentiment outcome to a binary-level outcome. Either the sentiment was negative or non-negative. Next, the team looked at grammatical error detection.

Grammatical error detection software and plugins have become widely popular for both students and professionals. The team used a dataset that contains essays penned by students who were in the process of learning English. The dataset had professional annotations for where the grammar was not written correctly. The model for this task also provided a binary result. Either a sentence was considered grammatically correct or it was not. The researchers then moved on to their final use case.

Hate speech detection was the final model created by this team. They used two datasets containing 20,940 manually annotated tweets (Barrett et al., 2018). The team reduced any tweets marked as either racist or sexist into one category for their final model. The team then set about conducting their comparative experiments, after creating their three models using the human attention parameters.

The results of the test are measured using the *F*1 score. This is calculated using both precision and recall. The team’s finding is that their “human attention model, based on regularizations from mean fixation durations in publicly available eye-tracking corpora, consistently outperforms the recurrent architecture with learned attention functions” (Barrett et al., 2018). The mean error reduction over the baseline claimed by the paper is 4.5%.

## Paper Analysis

### Research Use and Contributions

The chosen paper lends its research to improving the accuracy and usefulness of natural language processing. Sentiment analysis is widely used on large amounts of text from hundreds and thousands of posts across social media to determine the public’s general feelings on a topic. Increasing the accuracy of this analysis will lead companies to have more actionable data on how to navigate a market, politicians will be able to better craft their messages, and researchers will have a better understanding of public consensus.

### Advantages and Disadvantages of the Discussed Research

The idea of having a computer read text the same way that a human does comes with multiple advantages. This allows the computer to better approximate the same understanding of a piece of text as a human would. This can lead to a greater level of agreement between what the computer says the text means and a plurality of humans.

Some disadvantages may occur when using the ideas outlined in this paper. Namely, overreliance on the accuracy of this model may be used by some governments to generate a sentiment analysis model that is used to flag those considered dissidents. The paper also states that the new model was not tested against many models that came before that also artificial human attention in their analysis. In fact, there are studies out there that suggest attention is not a reliable means to assist in neural network development

### Additional Research on the Topic

This paper is not the only one on this topic. Others have attempted similar uses of artificial human attention both before and since the creation of the reviewed paper. Sood et al. found a similar increase in *F*1 score, 5%, with their own model using artificial human attention parameters (2020). The field of NLP can clearly benefit from the use of these new parameters to help computers analyze text in a way that more closely approximates how humans do. We must ensure that we are acting ethically, regardless of how useful this research looks on the surface.

### Reflections on Ethical Considerations

Dr. Ian Malcolm once stated, “Your scientists were so preoccupied with whether they could that they didn’t stop to think if they should” (Spielberg, 1993). No matter how useful a new technology looks on the surface, we must ensure its use is ethical in nature. The continued use of NLP models has the capacity to raise concerns about bias, privacy, and transparency (Bhattacharyya, 2023).

Bias concerns arise when a trained model provides results that a beneficial or detrimental to one group of individuals over another. Privacy concerns arise when normally private information is shared with people who should not have access to it through the use of NLP models. Finally, transparency needs to be an over-arching goal of all researchers so that anyone affected by an algorithm can know how the decision was reached.

### Future Research

More research will need to be conducted on this topic to fully know how useful artificial attention is to creating NLP models. Continuing to develop models that are harder to distinguish from real human beings may not be the best outcome for society. This process will surely help in general research of publicly available text, but there is a high likelihood of it being used for harmful outcomes, too.

***Conclusion***

Researchers will continue to make machine learning algorithms more closely model that of real human behavior. Adding parameters that mimic that of a human’s attention during reading has been shown by multiple papers to generate more accurate models. This has not been shown to be definitively true, though. The ethics of this process is an area that still requires more research, as well.

# Peer Review

# Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan.

**Topic # 5 by Amanda Burfield**

**Universal Language Model Fine-tuning for Text Classification**

**By Jeremy Howard and Sebastian Ruder**

**Article Summary**

The article I chose to review is the Universal Language Model File-tuning for Text Classification (ULMFiT). The purpose of ULMFiT is to apply it to a task within Natural Language Processing (NLP) and allow it to fine-tune the model. The article states it can reduce error between 18 and 24 percent on datasets. The model and coding were proposed to be open source. (Howard et al, 2018)

Computer vision (CV) models use fine-tuning on different datasets. It is rare or unlikely that the data within CV models are trained to perform a certain way. Text classification usages in real-world applications can be the spam filter in your email, fraud, and ect. The fine tuning within text classification and inductive transfers targets the first layer within a model and has been shown to be unsuccessful for NLP. (Howard et al, 2018)

The article continues to explain that the two authors don’t appear to be very confident in their ability to train a language model (LM). The two lack the knowledge on training the model themselves resulting in overfitting and catastrophic forgetting with classifiers. In turn they recommended a new method. ULMFiT will address the issues and allow for the fine-tuning for certain models and allow for six text classification tasks. (Howard et al, 2018)

They proposed the following contributions of ULMFiT which included CV-like transfer learning for NLP tasks, fine-tuning with triangular learning and gradual unfreezing to retain information and avoid catastrophic forgetting. They also state that ULMFiT will outperform six representative text classification datasets with the previously stated error reduction percentage, ablation analysis, and enable wider adoption. (Howard et al, 2018)

ULMFiT has three stages that consists of the LM being trained to capture the different layers of the language, fine-tuning the data and slanted triangular learning rates (STLR) to learn specifics, and fine-tuning the classifier with gradual unfreezing and ect. The fine tuning to each layer is emphasized in importance as well. It was also suggested that fine tuning the last layer’s learning rate would assist in fine-tuning of the other layers within the model. (Howard et al, 2018)

They later used ULMFiT in experiments to test out the text-classification in relation to real-world applications. They evaluated datasets with different numbers and documents in different lengths. Doing the evaluation allowed for them to use text classification and learning approaches to allow for sentiment analysis and question classification. The sentiment analysis used IMDB movie reviews and Yelp reviewed datasets that were compiled by Zhang et al (2015). The use of an open domain with fact questions was also applied. With elevation of the datasets, it allowed them to place in special text-classifications for uppercase words, repetitions, and ect. They then established baselines and comparison models between other datasets that were similar. In conclusion they reported the results with three different charts depicting validation error rate by percentage as the y-axis and # of training examples as the x-axis. They used from scratch, supervised learning, and unsupervised learning to depict the experiment. The from scratch showed a very distinct line near the top of the y-axis and decreasing whereas the supervised and unsupervised learning showed gradual decreases in reference to the validation error rate. (Howard et al, 2018)

They later concluded the article they proposed ULMFiT as a sample-efficient transfer learning method to apply to any NLP task. They also raised the suggestions to prevent catastrophic forgetting for ULMFiT and stating once again it outperforms other transfer learning techniques with six representative text classification tasks. (Howard et al, 2018)

NLP presents possibility of ethical concerns. Within the article, the authors discussed that the use of ULMFiT with any NLP outperforms other applications in real-world applications. The best way to ensure that it remains ethical is to ensure there is no bias within the datasets themselves. The use of an IMDB dataset is great but if you are comparing the ratings of different movies, it may require the use of other datasets that have moving ratings besides IMDB. It’s also good to ensure that there is transparency within the datasets and allows for privacy. When ensuring these things, it’s best to locate various laws as well that pertain the privacy and security and adhere to local and federal laws around them.

**ULMFiT Usage with NLP**

ULMFiT involves a transfer learning technique applied within NLP tasks. Therefore, to perform, it needs to be paired with an NLP to complete its transfer and perform its purposes. It requires various datasets and can be used in supervised or unsupervised learning processes. (Kriplani, 2019) Advantages of using ULMFiT in real world applications can revolve around social media for examples. The collection of data between various post on Facebook, Instagram, and ect and training it to perform with ULMFiT. Once trained and identifying text-classifiers, the set can then be trained to identify information for specified tasks. (Faltl, et al, 2019)

**ULMFiT Contributions to NLP**

ULMFiT’s contributions included CV-like transfer learning for NLP tasks, fine-tuning with triangular learning and gradual unfreezing to retain information and avoid catastrophic forgetting. (Howard et al, 2018) It was found later that ULMFiT was “dethroned by BERT and BERT was then dethroned by XLNet for text classification. But overall, the use of ULMFiT was a steppingstone moving forward with Universal Language Model Fine-Tuning for transfer learning. (Kriplani, 2019)

**Advantages and Disadvantages of ULMFiT to NLP**

Advantages described in the article allowed for lower percentage rates for errors as well as being a top tier use of Universal Language Model Fine-Tuning. Another advantage is that it requires a small amount of data to train for a deep learning model. Disadvantages were that it needs to be trained before use and starting from the bottom layer of data and ultimately working in reverse. Lastly, it can be timely to train before being able to utilize ULMFiT. (Karbowy, 2021)

**Other ULMFiT Artciles**

The additional articles I found relating to ULMFiT gave various mixed reviews as two of them state that ULMFiT was “dethroned” by other NLP software. However, all three articles do come to conclusion of the important impact ULMFiT had on the development of text-classification and its development in data analytics. Overall, they all seem to support the argument of the original article and allow for a better understanding of how technology is always evolving.

**Future research with ULMFiT**

# Technology and various software is constantly evolving to suit the needs of day-to-day or real life applications. The future of ULMFiT was stated in the article by Howard Kriplani stating that it was dethrone by another application and later that another application had dethroned the previous. Overall, because of the ever-evolving world of technology and various environments the future of ULMFiT may produce more fine-tuning software down the line that will allow for more specific task identification.

# Peer Review

# Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan.

# References

Barrett, M., Bingel, J., Hollenstein, N., Rei, M., & Søgaard, A. (2018). *Sequence classification with human attention*. ACL Anthology. https://aclanthology.org/K18-1030/

Galassi, A., Lippi, M., & Torroni, P. (2021, October 10). *Attention in Natural Language Processing*. IEEE Explore. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7169508

Sood, E., Tannert, S., Muller, P., & Bulling, A. (2020). *Improving natural language processing tasks with human gaze ... - neurips*. Proceedings Neural Information Processing Systems. https://proceedings.neurips.cc/paper/2020/file/460191c72f67e90150a093b4585e7eb4-Paper.pdf

Spielberg, S. (1993). *Jurassic Park.* Universal Pictures.

Bhattacharyya, S. (2023, June 13). *The ethical considerations of Natural Language Processing (NLP)*. Analytics Steps. https://www.analyticssteps.com/blogs/ethical-considerations-natural-language-processing-nlp

Faltl, S., Schimpke, M., & Hackober, C. (7 February 2019). Universal Language Model Fine-Tuning. Humboldt Univesrity in Berlin. [ULMFiT: State-of-the-Art in Text Analysis (humboldt-wi.github.io)](https://humboldt-wi.github.io/blog/research/information_systems_1819/group4_ulmfit/)

Howard, J., & Ruder, S. (23 May 2018). Universal Language Model Fine-tuning for Text Classification. ARXIV. [1801.06146.pdf (arxiv.org)](https://arxiv.org/pdf/1801.06146.pdf)

Karbowy, H. (12 July 2021). Why ULMFiT can still matter: training and inference times. Edrone. [Why ULMFiT can still matter: training and inference times | edrone | CRM for e-commerce & marketing automation](https://edrone.me/blog/why-ulmfit-can-still-matter-training-and-inference-times)

Kriplani, H. (4 September 2019). Understanding language modelling (NLP) and Using ULMFIT. Towards Data Science. [Understanding language modelling(NLP) and Using ULMFIT | by Himanshu Kriplani | Towards Data Science](https://towardsdatascience.com/understanding-language-modelling-nlp-part-1-ulmfit-b557a63a672b#:~:text=Universal%20Language%20Model%20FIne%2DTuning,by%20XLNet%20in%20text%20classification%5D.)