Paper 1: "Better Word Embeddings by Disentangling Contextual n-Gram Information"

Abstract: Pre-trained word vectors are ubiquitous in Natural Language Processing applications. In this paper, we show how training word embeddings jointly with bigram and even trigram embeddings, results in improved unigram embeddings. We claim that training word embed- dings along with higher n-gram embeddings helps in the removal of the contextual information from the unigrams, resulting in better stand-alone word embeddings. We empirically show the validity of our hypothesis by outperforming other competing word representation models by a significant margin on a wide variety of tasks. We make our models publicly available.

Citation:
Gupta, P., Pagliardini, M., Jaggi, M. (2019,June). *Better Word Embeddings by Disentangling Contextual n-Gram Information*, Paper presented at the 2019 Conference of the North American Chapter of the Association for Computational Linguistics https://www.aclweb.org/anthology/N19-1098*.* Minneapolis, June 2019. Association for Computational Linguistics.

Why I like this paper: I think it is interesting how they propose that training word embeddings with the use of n-grams like bigrams and trigrams, produces embeddings that are more accurate. This is very relevant as it suggests that we can better model a language if we take into account different ways in which words are used, which to me makes sense. But they also go a step further and incorporate char n-grams into the calculation for the 1-gram embeddings which is very interesting.

Paper 2: "Character n-gram Embeddings to Improve RNN Language Model"

Abstract: This paper proposes a novel Recurrent Neural Network (RNN) language model that takes advantage of character information. We focus on character n-grams based on research in the field of word embedding construction (Wieting et al. 2016). Our pro- posed method constructs word embeddings from character n- gram embeddings and combines them with ordinary word em- beddings. We demonstrate that the proposed method achieves the best perplexities on the language modeling datasets: Penn Treebank, WikiText-2, and WikiText-103. Moreover, we con- duct experiments on application tasks: machine translation and headline generation. The experimental results indicate that our proposed method also positively affects these tasks.

Citation:
Takase, S., Suzuki, J., Nagata, M. (2019, July). *Character n-Gram Embeddings to Improve RNN Language Models,* Paper presented in the AAAI Conference on Artificial Intelligence https://www.aaai.org/ojs/index.php/AAAI/article/view/4440. Honolulu, Hawaii, July 2019. Association for the Advancement of Artificial Intelligence.

Why I like this paper: The objectives of research of this paper are very interesting, those are: " Can character n-gram embeddings improve the performance of state-of-the-art RNN language models? Do character n-gram embeddings have a positive effect

on infrequent words? Is multi-dimensional self-attention effective for word em- bedding construction as compared with several other similar conventional methods? How many n should we use?" Besides such questions, the paper focuses on the creation of word embeddings based on character n-grams constructed by a neural network.

Paper 3: "A Neural Conversational Model"

Abstract: Conversational modeling is an important task in natural language understanding and machine in- telligence. Although previous approaches ex- ist, they are often restricted to specific domains (e.g., booking an airline ticket) and require hand- crafted rules. In this paper, we present a sim- ple approach for this task which uses the recently proposed sequence to sequence framework. Our model converses by predicting the next sentence given the previous sentence or sentences in a conversation. The strength of our model is that it can be trained end-to-end and thus requires much fewer hand-crafted rules. We find that this straightforward model can generate simple con- versations given a large conversational training dataset. Our preliminary results suggest that, de- spite optimizing the wrong objective function, the model is able to converse well. It is able extract knowledge from both a domain specific dataset, and from a large, noisy, and general do- main dataset of movie subtitles. On a domain- specific IT helpdesk dataset, the model can find a solution to a technical problem via conversa- tions. On a noisy open-domain movie transcript dataset, the model can perform simple forms of common sense reasoning. As expected, we also find that the lack of consistency is a common fail- ure mode of our model.

Citation:
Vinyals, O., Quoc, J,L. (2015, July). *C*A Neural Conversational Model Paper presented in the ICML Deep Learning Workshop 2015 https://arxiv.org/abs/1506.05869 . July 2015.

Why I like this paper: I am interested in engines that allow humans to interact and produce conversations, seems like something I am very interested in knowing more about. Also, I think this will bring a lot of attention from the group because the greatest representation of speech processing tools we have nowadays are tools likes Siri, etc.