Paper Review

Author: Yunfei Luo

May 12, 2020

Character-level Convolutional Networks for Text Classification, by Zhang, Zhao, and LeCun
Reference Paper URL: https://arxiv.org/abs/1509.01626

Introduction/Main Goal
The paper was published in 2016. The main goal of the paper is to exemplify the performance of ConvNets (Convolutional Networks) on text classification, based on Character-level. That is, to show the ability of ConvNets understand the text (for the task of classification) without knowing the knowledge from words and semantic structure of the language.

What's New/Improvement
The authors are trying to improve upon the traditional text classification. As pointed by the authors, most of the machine learning classifiers for text classification nowadays are based on words. However, the authors offers a new approach to the task, i.e. treating the text as raw signal at character level, rather than word level. More specifically, instead of present word at each entry of the vector representation for the text, the signal present single character at each entry. Such approach is supposed to be more effective and perform better with large scale dataset.

Observations
This is also a paper on Neural Networks. The input is a text, and the output is a class that the text belongs to. More specifically, the input text will be numerically transformed into a matrix with dimension $m \times l$, such that m defines the domain of unit character depends on the input language, and l denotes the total length of the sequence of characters that we care about in the text. In this paper, the authors believe that $l = 1014$ is an appropriate length, since it is enough for catching the key information in a text. The output class is simply in string format, for example, "sport", "finance", "entertainment", etc. Although in the paper, the authors use supervised learning to train the Networks, all the class of the given texts are known, in practice, there do exist some unknown state. The hidden node would be the actual class that the texts belong to. The topology of the hidden nodes could be several independent trees, each with a broad class as root. For example, the class "sport", there would be some sub-class under this class, such as "basketball", "swimming", "mixed martial arts", etc.

Result
The result that the authors obtained is shown as followed (page 6 in the paper):

Table 4: Testing errors of all the models. Numbers are in percentage. "Lg" stands for "large" and "Sm" stands for "small". "w2v" is an abbreviation for "word2vec", and "Lk" for "lookup table". "Th" stands for thesaurus. ConvNets labeled "Full" are those that distinguish between lower and upper letters

Model	AG	Sogou	DBP.	Yelp P.	Yelp F.	Yah. A.	Amz. F.	Amz. P.
BoW	11.19	7.15	3.39	7.76	42.01	31.11	45.36	9.60
BoW TFIDF	10.36	6.55	2.63	6.34	40.14	28.96	44.74	9.00
ngrams	7.96	2.92	1.37	**4.36**	43.74	31.53	45.73	7.98
ngrams TFIDF	**7.64**	**2.81**	**1.31**	4.56	45.20	31.49	47.56	8.46
Bag-of-means	**16.91**	**10.79**	**9.55**	**12.67**	**47.46**	**39.45**	**55.87**	**18.39**
LSTM	13.94	4.82	1.45	5.26	41.83	29.16	40.57	6.10
Lg. w2v Conv.	9.92	4.39	1.42	4.60	40.16	31.97	44.40	5.88
Sm. w2v Conv.	11.35	4.54	1.71	5.56	42.13	31.50	42.59	6.00
Lg. w2v Conv. Th.	9.91	-	1.37	4.63	39.58	31.23	43.75	5.80
Sm. w2v Conv. Th.	10.88	-	1.53	5.36	41.09	29.86	42.50	5.63
Lg. Lk. Conv.	8.55	4.95	1.72	4.89	40.52	29.06	45.95	5.84
Sm. Lk. Conv.	10.87	4.93	1.85	5.54	41.41	30.02	43.66	5.85
Lg. Lk. Conv. Th.	8.93	-	1.58	5.03	40.52	28.84	42.39	5.52
Sm. Lk. Conv. Th.	9.12	-	1.77	5.37	41.17	28.92	43.19	5.51
Lg. Full Conv.	9.85	8.80	1.66	5.25	38.40	29.90	40.89	5.78
Sm. Full Conv.	11.59	8.95	1.89	5.67	38.82	30.01	40.88	5.78
Lg. Full Conv. Th.	9.51	-	1.55	4.88	38.04	29.58	40.54	5.51
Sm. Full Conv. Th.	10.89	-	1.69	5.42	**37.95**	29.90	40.53	5.66
Lg. Conv.	12.82	4.88	1.73	5.89	39.62	29.55	41.31	5.51
Sm. Conv.	15.65	8.65	1.98	6.53	40.84	29.84	40.53	5.50
Lg. Conv. Th.	13.39	-	1.60	5.82	39.30	**28.80**	40.45	**4.93**
Sm. Conv. Th.	14.80	-	1.85	6.49	40.16	29.84	**40.43**	5.67

The table of error shows that overall, the character-level ConvNets performed well. A more detailed comparison of the errors with other text classification model is offered at page 7:

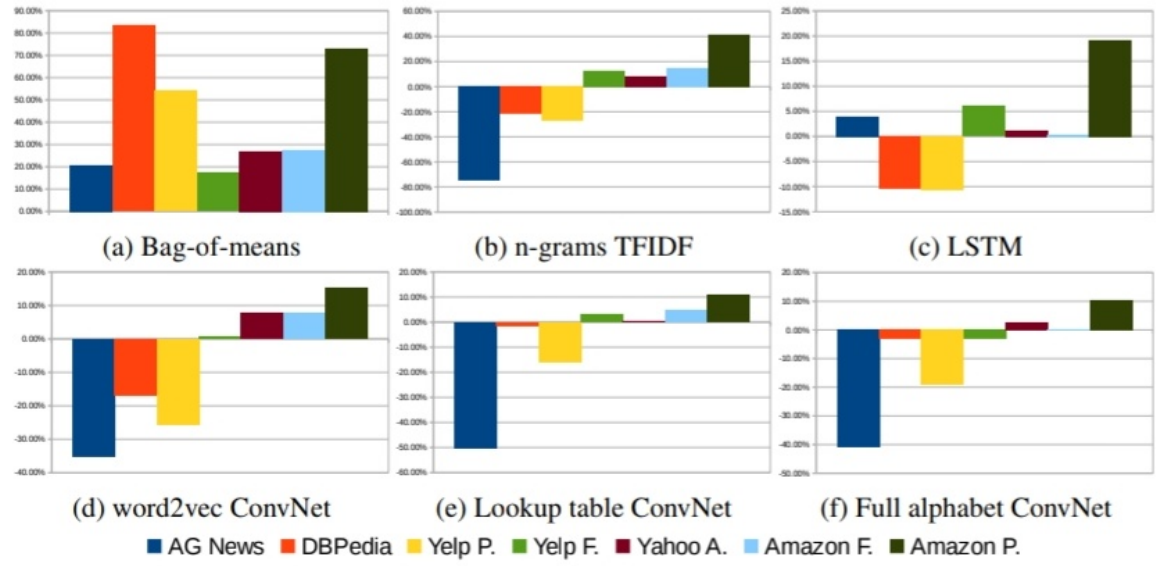

Figure 3: Relative errors with comparison models

The figure above shows that although the character-level ConvNets does not perform that well with the dataset "AG News" compare with other models (except Bag-pf-means, because this model has more errors in all the dataset cases), the overall values of errors made by the character-level ConvNets is considerable. The most important point is that the character-level based ConvNets does not need dictionary for training, so that the entire training process is much more effective than the other models.

Extension/Follow up

Some deficiency would be obvious that the character-level ConvNets might not perform well with non-symbol-based language, such as Chinese. In this paper, the authors deal with this case (the dataset of Sogou News) by transform the Chinese characters into Pinyin. However, Pinyin and Chinese characters are not an one-to-one mapping, but rather an onto mapping, from characters to Pinyin. Placing different characters at the same place could mean extremely different things. So would it be better if we split each character into strokes (such as point and across, which is also a finite set), then use this sequence of strokes as raw input signals? Such follow-up questions could form the extension for the work presented in the paper.