Does BERT Make Any Sense?
Interpretable Word Sense Disambiguation with Contextualized Embeddings

Rohan Rathore

Gregor Wiedemann, Steffen Remus, Avi Chawla, Chris Biemann

Scientific workshop "Big Data Analytics"
Novosibirsk State University, Russia

December 1, 2020

Table of Content

- Introduction
- Prerequisites Revision
- Task Details
- Model
- Results
- My Implementation
- References

Introduction

- Contextualized word embeddings (CWE) are a major recent innovation in NLP (resolves polysemy of words)
- CWE implicitly provide a model for word sense disambiguation (WSD)
- The paper presents an approach to WSD task using nearest neighbor classification (k-NN) based model on CWE

Prerequisites Revision

- Contextualized word embeddings (CWE)
  - CWE provide semantic vector representations of words depending on their respective context
  - Advantage over static word embeddings for a number of tasks, such as text classification, sequence tagging, or machine translation
- Word sense disambiguation (WSD)
  - WSD is the task of associating a word in context with one of its sensible meanings
  - For the English language, WordNet is the most commonly used sense inventory providing more than 200K word-sense pairs
- Nearest neighbor classification (k-NN)
  - Algorithm used for classification and regression problems
  - For classification, case is classified by being assigned to the class most common among its k nearest neighbors

Task Details

- Train and evaluate k-NN based WSD system with Bert CWE
- Train model on SemCor dataset and on Princeton WordNet gloss corpus (WNGT)
- Evaluate on four standard WSD dataset - SensEval-2 (SE-2), SensEval-3 (SE-3), SemEval 2007 Task 7 (S7-T7), SemEval 2007 Task 7 (S7-T17)

k-NN Model for WSD

- Cosine distance of the CWE vectors as distance metric
- Since BERT uses subword units of words instead of lemmas in their embedding, therefore in the preprocessing, re-tokenization of lemmatized sentence was done, followed by averaging all subword units CWEs that belong to the target word
- k range $\{1,..., 10, 50, 100, 500, 1000\}$
- To counter class imbalance in training dataset, majority voting of kNN was set to $\min(k, \mod(Vs))$ where Vs is the set of CWE with the least frequent training examples for a given word sense 's'

Results - I

Bert based based F1 scores achieving new state of the art

k	SE-2	SE-3	S7-T7	S7-T17
1	76.10	78.62	81.11	59.82
2	76.10	78.62	81.11	59.82
3	**76.52**	79.66	80.94	59.38
4	**76.52**	79.55	80.94	59.82
5	76.43	79.79	81.07	60.27
6	76.43	79.81	81.07	60.27
7	76.50	80.02	81.03	60.49
8	76.50	79.86	81.03	60.49
9	76.40	79.97	81.03	60.49
10	76.40	**80.12**	81.03	60.49
50	76.43	79.66	81.11	60.94
100	76.43	79.63	81.20	60.71
500	76.38	79.63	81.11	60.71
1000	76.38	79.63	81.11	60.71
MFS	54.79	58.95	70.94	48.44
Kågebäck (2016)	*66.90*	*73.40*	-	-
Yuan et al. (2016)	-	-	84.30	63.70
Vial et al. (2018a)	-	-	86.02	66.81
Vial et al. (2019)	-	-	***90.60***	***71.40***

	SE-3 (train)	SE-3 (test)
(1)	President Aquino, admitting that the death of Ferdinand Marcos had sparked a wave of sympathy for the late dictator, urged Filipinos to stop weeping for the man who had laughed all the way to the **bank**[A Bank Building].	They must have been filled in at the **bank**[A Bank Building] either by Mr Hatton himself or else by the cashier who was attending to him.
(2)	Soon after setting off we came to a forested valley along the **banks**[Sloping Land] of the Gwaun.	In my own garden the twisted hazel, corylus avellana contorta, is underplanted with primroses, bluebells and wood anemones, for that is how I remember them growing, as they still do, along the **banks**[Sloping Land] of the rive Greta
(3)	In one direction only a little earthy **bank**[A Long Ridge] separates me from the edge of the ocean, while in the other the valley goes back for miles and miles.	The lake has been swept clean of snow by the wind, the sweepings making a huge **bank**[A Long Ridge] on our side that we have to negotiate.
(4)	However, it can be possible for the documents to be signed after you have sent a payment by cheque provided that you arrange for us to hold the cheque and not pay it into the **bank**[A Financial Institution] until we have received the signed deed of covenant.	The purpose of these stubs in a paying – in book is for the holder to have a record of the amount of money he had deposited in his **bank**[A Bank Building].
(5)	He continued: assuming current market conditions do not deteriorate further, the group, with conservative borrowings, a prime land **bank**[A Financial Institution] and a good forward sales position can look forward to another year of growth.	Crest Nicholson be the exception, not have much of a land **bank**[Supply or Stock] and rely on its skill in land buying.
(6)	The marine said, get down behind that grass **bank**[A Long Ridge], sir, and he immediately lobbed a mills grenade into the river.	The guns were all along the river **bank**[Sloping Land] as far as I could see.

My Implementation

- Trained on SensEval-2 train dataset
- Evaluated on SensEval-2 test dataset
- k=3
- F1 score - 74.18 %
- Code - https://colab.research.google.com
  /drive/137FZ9bO7I00gmq6Q7VYN8L9pUCp39NI3?usp=sharing

References

- Alan A., Duncan B., Roland V., Contextual string embeddings for sequence labeling, Proceedings of the 27th International Conference on Computational Linguistics, pages 1638–1649, Santa Fe, NM, USA, 2018

- Sergey B., Dmitry K., Anton O., Dmitry V., Breaking sticks and ambiguities with adaptive skipgram, Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS), pages 130–138, Cadiz, Spain, 2016

- Jose C., Mohammad T.P., From word to sense embeddings: A survey on vector representations of meaning, Journal of Artificial Intelligence Research, 63(1):743–788, 2018