7/9 Paper Intro

Pohan Chi

July 6, 2020

Outline

1. Dense Passage Retrival for Open-Domain Question Answering

2. Retrival-Augmented Generalization for Knowledge-Intensive NLP Task

3. Pretraining with Contrastive Sentence Objectives Improves Discourse Performance of Language Models

4. Transformers are RNNs: Fast Autoregressive Transformers with Linear Attention

Sentence Representation

Pretraining with Contrastive Sentence Objectives Improves Discourse Performance of Language Models

Dan Iter[**], **Kelvin Guu**[**], **Larry Lansing**[**], and **Dan Jurafsky**[*]

[*]Computer Science Department, Stanford University
[**]Google Research
{daniter, jurafsky}@stanford.edu
{kguu, llansing}@google.com

Accepted by ACL 2020

Highlight

- sentence-level CPC objective + MLM loss in pretraining stage
- Compare to SOP and NSP Pretraining Method

Overview

Overview

- Task Definitions
- Evaluation
- Methodology & Model
- Experiments
- Conclusion

Sentence Representation

Pretraining with Contrastive Sentence Objectives Improves Discourse Performance of Language Models

Dan Iter[**], **Kelvin Guu**[**], **Larry Lansing**[**], and **Dan Jurafsky**[*]

[*]Computer Science Department, Stanford University
[**]Google Research
{daniter,jurafsky}@stanford.edu
{kguu,llansing}@google.com

Accepted by ACL 2020

Task Definitions

Discourse

- Discourse is a coherent structured group of sentence
- ex: dialogue, document

ex:


Not l

Article Talk Read View source View history

Barack Obama

From Wikipedia, the free encyclopedia

(Redirected from Obama)

"Barack" and "Obama" redirect here. For other uses, see Barack (disambiguation), Obama (disambiguation), and Barack Obama (disambiguation).

Barack Hussein Obama II (/bəˈrɑːk huːˈseɪn oʊˈbɑːmə/ (listen);[1] born August 4, 1961) is an American politician and attorney who served as the 44th president of the United States from 2009 to 2017. A member of the Democratic Party, Barack Obama was the first African-American president of the United States. He previously served as a U.S. senator from Illinois from 2005 to 2008 and an Illinois state senator from 1997 to 2004.

Obama was born in Honolulu, Hawaii, making him the first president not born in the contiguous United States. After graduating from Columbia University in 1983, he worked as a community organizer in Chicago. In 1988, he enrolled in Harvard Law School, where he was the first black person to be president of the *Harvard Law Review*. After graduating, he became a civil rights attorney and an academic, teaching constitutional law at the University of Chicago Law School from 1992 to 2004. Turning to elective politics, he represented the 13th district from 1997 until 2004 in the Illinois Senate, when he ran for the U.S. Senate. Obama received national attention in 2004 with his March Senate primary win, his well-received July Democratic National Convention keynote address, and his landslide November election to the Senate. In 2008, he was nominated for president a year after his presidential campaign began, and after close primary campaigns against Hillary Clinton. Obama was elected over Republican John McCain and was inaugurated alongside Joe Biden on January 20, 2009. Nine months later, he was named the 2009 Nobel Peace Prize laureate.

Discourse Evaluation

- Sentence Position (SP)
- BSO (Binary Sentence Ordering) (same as SOP)
- DC (Discourse Coherence)
- SSP (Sentence Section Prediction)
- PDTB-E, PDTB-I, RST-DT

Model & Objectives

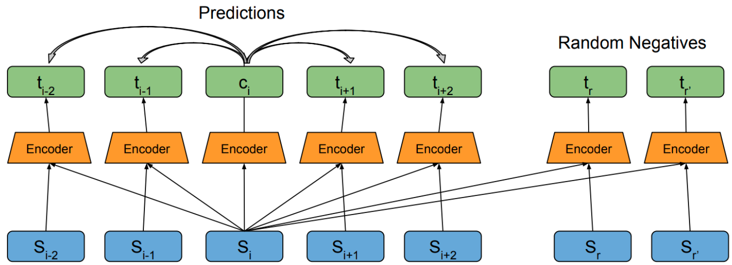


Figure 1: During training, a text segment is selected as the anchor (S_i). The anchor as well as all the targets, $S_{i-2}...S_{i+2}$ plus random samples S_r are encoded with the transformer masked language model. The encoded representation of the anchor is used to predict each target at its target distance. The S_i objects are raw text sentences, the *encoder* is the transformer model, and c_i and t_i are vectors.

Experiments - 1

Model	SP	BSO	DC	SSP	PDTB-E	PDTB-I	RST-DT	avg.
BERT-Base	53.1	68.5	58.9	80.3	41.9	42.4	58.8	57.7
BERT-Large	53.8	69.3	59.6	**80.4**	**44.3**	43.6	59.1	58.6
RoBERTa-Base	38.7	58.7	58.4	79.7	39.4	40.6	44.1	51.4
BERT-Base BSO	53.7	72.0	71.9	80.0	42.7	40.5	**63.8**	60.6
CONPONO *isolated*	50.2	57.9	63.2	79.9	35.8	39.6	48.7	53.6
CONPONO *uni-encoder*	59.9	74.6	72.0	79.6	40.0	43.9	61.9	61.7
CONPONO (k=2)	**60.7**	**76.8**	**72.9**	**80.4**	42.9	**44.9**	63.1	**63.0**
CONPONO std.	±.3	±.1	±.3	±.1	±.7	±.6	±.2	-

Table 1: CONPONO improves the previous state-of-the-art on four DiscoEval tasks. The average accuracy across all tasks is also a new state-of-the-art, despite a small drop in accuracy for PDTB-E. BERT-Base and BERT-Large numbers are reported from Chen et al. (2019), while the rest were collected for this paper. We report standard deviations by running the evaluations 10 times with different seeds for the same CONPONO model weights.

Experiments - 2

Model	RTE	COPA
BERT-Base	66.4	62.0
BERT-Base BSO	71.1	67.0
CONPONO	70.0	69.0
BERT-Large	70.4	69.0
ALBERT	86.6	-

Table 3: Our model improves accuracy over BERT-Base for RTE and COPA benchmarks. Improvements are comparable to BERT-Large but still lag behind much larger models trained on more data, such as ALBERT. All scores are on the validation set.

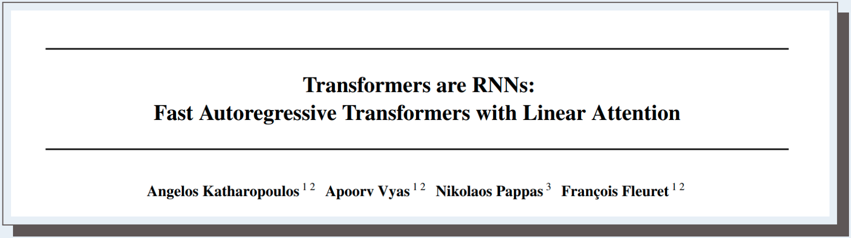
Model	Accuracy
BERT-Base	61.2
CONPONO	63.2
BERT-Large	69.8 [EM]

Table 4: CONPONO is more effective at classifying the most plausible sentence from the extended context than BERT-Base. We report the BERT-Large exact match score, where the model selects only the target entity from the context, for reference. All scores are on the validation set.

Conclusion & A lot

- benchmark in DiscoEval (sentence representation)
- Also improve performance in RTE, COPA, ReCoRD (Reading Comprehension).
- uni-encoder verse isolated coding comparison

Linear Attention


Transformers are RNNs:
Fast Autoregressive Transformers with Linear Attention

Angelos Katharopoulos [1 2] Apoorv Vyas [1 2] Nikolaos Pappas [3] François Fleuret [1 2]

Accepted by ICML 2020

Highlight

- Reduce $O(N^2)$ to $O(N)$
- Connect Transformer to RNN form
- 4000X faster than origin Transformer
- Compare to Reformer (LSH attention)

Overview

Overview

- Refomulate Equation
- Experiments
- Conclusion