Capstone Project: Comparative exploration of co-reference resolution with transformer models

Aditya A. Landge

ABSTRACT

Co-reference resolution is the task of automatically determining the chain of expressions in language that refers to the same entity or antecedent. In this project, I am going to explore transformer models, including the BERT model and recent variants of its representation (e.g., ALBERT), adapting them to the co-reference resolution task and measuring performance comparatively. This will include both quantitative and qualitative model analysis, including visualization. Moreover, I will compare the performance of transformer models on languages demonstrating different pronominal complexity as well as English data which includes non-binary pronoun.

1. INTRODUCTION

Co-reference resolution is the computational linguistic / NLP task to determine the expressions that allude to the same entity (anaphora). Language users refer to the same entity using various nouns and deictic indexicals such as pronouns. Such co-reference phenomena can be thought of as chained hyperlinks in natural language. However, co-reference can make it difficult for natural language reasoning models that aim to process language more meaningfully. Co-reference resolution has been a task attracting broad interest for a while, yet it still remains a substantial challenge for NLP systems. It is useful for solving various other tasks such as named-entity recognition, question answering, etc. [9]

In recent years, pre-trained language models like BERT have been very effective for improving many NLP tasks. BERT (Bidirectional Encoder Representations from Transformers) is a natural language representation model that is trained on language modeling tasks, e.g. on predicting masked words given their context. The pre-trained BERT can be adapted, by adding an extra output layer, to transfer the model to perform a variety of NLP tasks.

In this capstone, I aim to answer these questions:

RQ1 Based on analysis of the technical framework on which BERT is based, what are the computational trade-offs between BERT when compared against at least one of its variant, such as ALBERT (A Lite BERT [14]).

RQ2 Based on quantitative performance results and qualitative analysis of outcomes (including visualization), what are the strengths of each model, and, more generally, their challenges for solving the co-reference resolution task?

RQ3 How do the models compare when learning to perform coreference resolution (i.e., identify coreference chains) in corpora of languages with distinct pronominal systems (and potentially quite different linguistic typologies), including an English dataset which also integrates various non-binary pronouns.

2. MILESTONES

The capstone goals are divided across 3 milestones that I aim to achieve for the capstone project.

1. Milestone 1
   - Read articles [10] [15] [8] [9] [13] [4] [1] [5] and understand the co-reference resolution problem and its background.
   - Set up a base end-to-end system with BERT Base.
   - Process and analyze the annotated co-reference datasets, split them into development and final held-out testing sets, and compute corpus statistics and fundamental linguistic characteristics on the development data only.
   - Understand the technical basis of BERT [8] works and begin to understand one of its variants.
   - Explore how to fine-tune BERT for co-reference resolution task and obtain base results.
2. Milestone 2
   - Read articles [12] [14] [2] [6] [16] [7] .
   - Add a BERT variant (e.g., ALBERT [14]) and adapt it to the co-reference resolution task.
   - Analyze results and compute comparative task & computational performance statistics on BERT.
   - Seek to visualize the model for interpretation.
3. Milestone 3
   - Read articles [3] [17] [19] [18] [11] .
   - Compare the models performance on at datasets of at least two languages with distinct pronominal systems and potentially different linguistic typologies.
   - Analyze models quantitatively and qualitatively to explore their strengths and challenges
   - Explore privacy preservation as well as gender bias (ethics) in the models as use applied cases.
4. Stretch Goal

- Generate synthetic data of additional instances to balance the dataset to compare issues with pronouns.
- Analyze the how the trained and fine-tuned BERT model works on the synthetically modified custom test data set.
- Alternatively, explore how BERT can be used for another sequence task, such as temporal event analysis.

3. REFERENCES

[1] F. Alfaro, M. R. Costa-jussà, and J. A. R. Fonollosa. BERT masked language modeling for co-reference resolution. In *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, pages 76–81, Florence, Italy, Aug. 2019. Association for Computational Linguistics.

[2] Y. Arase and J. Tsujii. Transfer fine-tuning: A BERT case study. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5393–5404, Hong Kong, China, Nov. 2019. Association for Computational Linguistics.

[3] Y. T. Cao and H. Daumé III. Toward gender-inclusive coreference resolution. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4568–4595, Online, July 2020. Association for Computational Linguistics.

[4] H. Chen, Z. Fan, H. Lu, A. L. Yuille, and S. Rong. Preco: A large-scale dataset in preschool vocabulary for coreference resolution, 2018.

[5] B. Cui, Y. Li, M. Chen, and Z. Zhang. Fine-tune BERT with sparse self-attention mechanism. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3548–3553, Hong Kong, China, Nov. 2019. Association for Computational Linguistics.

[6] T. I. Denk and A. Peleteiro Ramallo. Contextual BERT: Conditioning the language model using a global state. In *Proceedings of the Graph-based Methods for Natural Language Processing (TextGraphs)*, pages 46–50, Barcelona, Spain (Online), Dec. 2020. Association for Computational Linguistics.

[7] T. I. Denk and A. P. Ramallo. Contextual bert: Conditioning the language model using a global state, 2020.

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[9] P. Elango. Coreference resolution : A survey. 2006.

[10] A. Emami, P. Trichelair, A. Trischler, K. Suleman, H. Schulz, and J. C. K. Cheung. The knowref coreference corpus: Removing gender and number cues for difficult pronominal anaphora resolution, 2019.

[11] M. Ionita, Y. Kashnitsky, K. Krige, V. Larin, A. Atanasov, and D. Logvinenko. Resolving gendered ambiguous pronouns with BERT. In *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, pages 113–119, Florence, Italy, Aug. 2019. Association for Computational Linguistics.

[12] N. Jiang and M.-C. de Marneffe. Evaluating BERT for natural language inference: A case study on the CommitmentBank. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6086–6091, Hong Kong, China, Nov. 2019. Association for Computational Linguistics.

[13] T. Klein and M. Nabi. Attention is (not) all you need for commonsense reasoning, 2019.

[14] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut. Albert: A lite bert for self-supervised learning of language representations, 2020.

[15] E. Lapshinova-Koltunski, C. España-Bonet, and J. van Genabith. Analysing coreference in transformer outputs, 2019.

[16] H.-L. Trieu, A.-K. Duong Nguyen, N. Nguyen, M. Miwa, H. Takamura, and S. Ananiadou. Coreference resolution in full text articles with BERT and syntax-based mention filtering. In *Proceedings of The 5th Workshop on BioNLP Open Shared Tasks*, pages 196–205, Hong Kong, China, Nov. 2019. Association for Computational Linguistics.

[17] Y. Wang, W. Che, J. Guo, Y. Liu, and T. Liu. Cross-lingual bert transformation for zero-shot dependency parsing, 2019.

[18] S. Wu and M. Dredze. Are all languages created equal in multilingual BERT? In *Proceedings of the 5th Workshop on Representation Learning for NLP*, pages 120–130, Online, July 2020. Association for Computational Linguistics.

[19] S. Yoosuf and Y. Yang. Fine-grained propaganda detection with fine-tuned BERT. In *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pages 87–91, Hong Kong, China, Nov. 2019. Association for Computational Linguistics.