Towards a Transformer-Based Reverse Dictionary Model for Quality Estimation of Definitions (Student Abstract)

Julien Guité-Vinet, Alexandre Blondin Massé, Fatiha Sadat

Université du Québec à Montréal, Canada H2X 3Y7

guite-vinet.julien@courrier.uqam.ca, blondin_masse.alexandre@uqam.ca, sadat.fatiha@uqam.ca

Abstract

In the last years, several variants of transformers have emerged. In this paper, we compare different transformer-based models for solving the *reverse dictionary task* and explore their use in the context of a serious game called *The Dictionary Game*.

Introduction

In its simplest form, a common language dictionary can be seen as an association between the meaning of a word and its definition. A task related to word sense disambiguation is called *the reverse dictionary task*. It focuses on the reverse association, *i.e.* guessing a word from its definition. A model capable of effectively solving this task should capture multiple semantic relationships such as synonymy, polysemy and homography. In order to solve the *reverse dictionary task*, some studies rely on information retrieval through lexical database (Shaw et al. 2013) or on graph-based approaches (Thorat and Choudhari 2016). Recurrent networks have also been shown to be expressive enough to take them into account (Morinaga and Yamaguchi 2018; Kartsaklis, Pilehvar, and Collier 2018; Zhang et al. 2020). More recent studies have proposed transformer-based models (Yan et al. 2020; Chen and Zhao 2022; Li et al. 2022; Mane et al. 2022).

In this paper, we investigate different processing of the task with XLNet (Yang et al. 2019), the distilled version of BERT (Devlin et al. 2019) and GPT-2 (Radford et al. 2018). We also examine how dictionaries can be semantically developed. To do so, we experiment with a serious game called *The Dictionary Game* which aims at better understanding the mental lexicon (Blondin Massé et al. 2008). It starts with a given root word that the player must define. The player must recursively define all new words in the definitions, until all words are defined. Then, we establish a qualitative measure of a dictionary using the fine-tuned models.

The Transformer-Based Reverse Dictionary

The objective of the *reverse dictionary task* is to build a model returning a set of candidates that maximize the probability that a word is paired with a dictionary definition. The model takes as input a definition in natural language for a word that a user can think of. It outputs a ranked set of word associated with the given definition. The Bidirectional Encoder Representations from Transformers (BERT), is a language model that learns from bidirectional representations of unlabeled text by jointly conditioning left and right contexts across all layers (Devlin et al. 2019). The Generative Pre-trained Transformer 2 (GPT-2) is another large language model pretrained on a very large corpus of English data. It is describes as an autoregressive transformer language model consisting of a stack of decoders (Radford et al. 2018). Decoders differ from encoders having a masked self-attention mechanism (Vaswani et al. 2017). XLNet is another autoregressive and bidirectional model (Yang et al. 2019). Instead of using a fixed token order, XLNet considers all possible permutations of the tokens.

Methodology

We retrieved pretrained models of the DistilBERT, DistilGPT-2 and XLNet transformers. The distilled versions retain most of the knowledge of the original versions while reducing the number of parameters (Sanh et al. 2019). We pooled their outputs, *i.e* a context representation of the input sentence and connected them to a dropout layer followed by a softmax layer. For DistilBERT, we used the output of the `[CLS]` special token, which embeds the input sentence for classification tasks. For DistilGPT-2, we used the output of the last token since that token learns about the previous tokens given the nature of the decoder's attention mechanism. As for XLNet, we used the average of the hidden states of all tokens in the output layer, excluding the tokens associated with the padding.

All models have been trained using the cross-entropy loss function and the AdamW optimizer. We added an auxiliary task following the same configuration as the previous task in order to forecast the POS of the target word. Besides, few definitions from the data exceed 50 tokens after being tokenized. Therefore, we limited the size of the input sentence, discarding word-definition pairs whose definition exceeds 50 tokens.

Evaluations

We evaluated our model on the data collected by Hill et al. according to three metrics: the median rank of the target words, the rate of occurrences of the target words in the top 1/10/100, and the rank variance of the target word. The data and metrics are used as a benchmark in several other studies (Hill et al. 2016; Morinaga and Yamaguchi 2018; Zhang et al. 2020; Mane et al. 2022).

According to our results, XLNet seems to better generalize the task, followed by DistilBERT, then DistilGPT-2. On the description test set, XLNet forecasts a median rank of 3, a top 1/10/100 of .38/.65/.85 with a rank variance of 364, slightly upgrading the top-1 forecast of previous state of the art models that we are aware of. Since XLNet and DistilBERT are bidirectional models, this could be a factor improving the accuracy for this task.

A definition can be qualified as *perfect*, *good*, *mediocre* if it is in the top-1, top-10, top-100 forecasts respectively, or *wrong* if it lies outside. One way to get a more meaningful value is to consider the rank or the degree of certainty associated with the model's forecast for a given target word and definition. We noticed that DistilBERT tends to forecasts better average median ranks. Since the dictionaries from the data have fairly short definitions, this may have the effect of favoring bidirectional models.

Discussion

The degree of certainty seems to be inadequate to assess the quality of a dictionary since for *Merriam-Webster's*, DistilBERT forecasts an average degree of certainty of 0.55 while averaging the median rank at 0. This is mainly because the model considers synonyms in its probability analysis.

The average rank appears to be a more suitable metric. However, it can be subject to high variances when a good rank for a definition is assigned to a word of a small dictionary. It appears that the quality of a dictionary should depend on its size, as well as on the allowed vocabulary, according to the principle that by using several words, we can build deeper lexical semantics. However, this is more intricate since we have observed that a dictionary having less than 24 words can have a better average rank than a dictionary having more than 110 words.

Conclusion

We have introduced a framework that simulates natural language understanding through a transformer-based model to solve the *reverse dictionary task* and evaluate the quality of dictionaries. For those tasks, it suggests that performance gains are possible and offers insights into what works when involving transformer-based models and dictionary-based data sets. Also, dictionary definitions have a very specific grammatical structure, presumably more simpler than the general case of free text. Therefore, we consider adding more data sources such as urban dictionaries and crossword dictionaries.

References

Blondin Massé, A.; Chicoisne, G.; Gargouri, Y.; Harnad, S.; Picard, O.; and Marcotte, O. 2008. How is meaning grounded in dictionary definitions? In *Proceedings of the 3rd Textgraphs Workshop on Graph-Based Algorithms for Natural Language Processing, TextGraphs 2008*.

Chen, P.; and Zhao, Z. 2022. A Unified Model for Reverse Dictionary and Definition Modelling. In *Proceedings of AACL-IJCNLP 2022*.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the NAACL HLT 2019*, volume 1.

Hill, F.; Cho, K.; Korhonen, A.; and Bengio, Y. 2016. Learning to understand phrases by embedding the dictionary. *Transactions of the Association for Computational Linguistics*, 4.

Kartsaklis, D.; Pilehvar, M.; and Collier, N. 2018. Mapping text to knowledge graph entities using multi-sense LSTMs. In *Proceedings of the EMNLP 2018*.

Li, B.; Weng, Y.; Xia, F.; He, S.; Sun, B.; and Li, S. 2022. LingJing at SemEval-2022 task 1: Multi-task self-supervised pre-training for multilingual reverse dictionary. In *Proceedings of SemEval 2022*.

Mane, S. B.; Patil, H. N.; Madaswar, K. B.; and Sadavarte, P. N. 2022. WordAlchemy: a transformer-based reverse dictionary. In *2nd International CONIT*. IEEE.

Morinaga, Y.; and Yamaguchi, K. 2018. Improvement of Reverse Dictionary by Tuning Word Vectors and Category Inference. *Communications in Computer and Information Science*, 920.

Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I.; et al. 2018. Improving language understanding by generative pre-training. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf. Accessed: 2023-11-18.

Sanh, V.; Debut, L.; Chaumond, J.; and Wolf, T. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv:1910.01108.

Shaw, R.; Datta, A.; VanderMeer, D.; and Dutta, K. 2013. Building a Scalable Database-Driven Reverse Dictionary. *IEEE Transactions on Knowledge and Data Engineering*.

Thorat, S.; and Choudhari, V. 2016. Implementing a Reverse Dictionary, based on word definitions, using a Node-Graph Architecture. In *Proceedings of COLING 2016*.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Yan, H.; Li, X.; Qiu, X.; and Deng, B. 2020. BERT for Monolingual and Cross-Lingual Reverse Dictionary. In *Proceedings of the EMNLP 2020*.

Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R. R.; and Le, Q. V. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32.

Zhang, L.; Qi, F.; Liu, Z.; Wang, Y.; Liu, Q.; and Sun, M. 2020. Multi-Channel Reverse Dictionary Model. *Proceedings of the AAAI Conference on Artificial Intelligence*.