Austin Girouard and Aarya Patil

ATG180001 / AAP180014

Dr. Karen Mazidi

Assignment 8: ACL Paper Summary

Title: Cross-Lingual Phrase Retrieval

Authors: Heqi Zheng; Xiao Zhang; Zewen Chi; Heyan Huang; Tan Yan; Tian Lan; Wei Wei; Xian-Ling Mao

Affiliations: School of Computer Science and Technology, Beijing Institute of Technology; Beijing Engineering Research Center of High Volume Language Information Processing and Cloud Computing Applications; Huazhong University of Science and Technology.

Problem addressed: While there have been many effective studies on monolingual phrase retrieval and cross-lingual phrase retrieval, learning cross-lingual phrase representations is still challenging for 2 reasons. First, “a phrase is a conceptual unit containing multiple words, so it is necessary to model interaction between words, which is not considered in word-level methods” (4193). Second, “a phrase contains fewer words with less information compared to sentences which prevents sentence encoders from taking the advantage of the ability of understanding full-length sentences” (4193). So, to solve these two problems, a new method of cross-lingual phrase retrieval was proposed.

Prior work: Many recent studies have focused on phrase retrieval from monolingual phrase sets instead of cross-lingual phrase retrieval. These studies “learn dense representations of phrases, and achieve promising results in entity linking, slot filling, and open-domain question

answering tasks" (4193). There have been proposed methods to perform cross-lingual text retrieval. These methods typically take each language separately to train word embeddings, which are then used to create a mapping between the different languages. Then, "the bilingual word pairs can be retrieved between vocabularies using nearest neighbor search," also known as bilingual lexicon induction (4193).

Unique contributions: In order to solve some of the problems that other cross-lingual method propositions have, the author's developed XPR. What makes XPR different than other cross-lingual methods is that it "produces phrase representations using example sentences, which can be collected from unlabeled text corpora" instead of directly encoding the input text (4193). XPR also has the ability to "either directly serve as an unsupervised retriever or be further trained to produce better-aligned phrase representations" (4194).

How the authors evaluated their work: The authors evaluated their work by demonstrating the effectiveness of XPR on "eight language pairs under four evaluation settings" and by creating a "cross-lingual phrase retrieval dataset, which provides 65K bilingual phrase pairs with 4.2M example sentences in 8 language pairs" (4194). In comparison to other multilingual methods, such as Cross-Lingual Word Embeddings (CLWE) and Cross-Lingual Sentence Encoders (CLSE), XPR greatly outperformed in accuracy across unsupervised, supervised, zero-shot transfer, and multilingual supervised evaluation settings, improving over these methods by as much as 71% accuracy on average.

Citations Received (Google Scholar)

Heqi Zheng: N/A; Xiao Zhang: 99; Zewen Chi: 579; Heyan Huang: N/A; Tan Yan: N/A; Tian Lan: N/A; Wei Wei: 1622; Xian-Ling Mao: 1692.

The lead co-authors of the article are Heqi Zheng, Xiao Zhang, and Zewen Chi. The corresponding author is Dr. Xian-Ling Mao, associate professor and doctoral supervisor in Computer Science and Technology at the Beijing Institute of Technology. Dr. Mao is also the most cited author of the paper (according to Google Scholar). Dr. Wei Wei, a professor and doctoral supervisor at the Huazhong University of Science and Technology, follows closely in citations.

Reference:

https://aclanthology.org/2022.acl-long.288/