Cross-Lingual Learning with Distributed Representations

Matúš Pikuliak
Faculty of Informatics and Information Technologies
Slovak University of Technology, Bratislava, Slovak Republic
matus.pikuliak@stuba.sk

Abstract

Cross-lingual Learning can help to bring state-of-the-art deep learning solutions to smaller languages. These languages in general lack resource for training advanced neural networks. With transfer of knowledge across languages we can improve the results for various NLP tasks.

Introduction

Deep learning in recent years managed to improve the results for a plethora of *Natural Language Processing* (NLP) tasks. These new methods are however very data demanding and therefore can be effectively applied only to a handful of languages with sufficient resources. English language is in this regard the richest language while the others are behind by significant margin. Instead of laborious dataset assembly we can try to exploit these existing English resources and models and try to transfer the knowledge from them to other languages. If we can do this automatically we can significantly cut on the time required to build NLP solutions.

Cross-lingual Learning is one way of approaching this idea. From *Artificial Intelligence* point of view cross-lingual learning is a type of *Transfer Learning*. Transfer learning is a paradigm of machine learning dealing with cases when there is a difference between the data we are training on and the data we are using the trained model on. In cross-lingual setting we are trying to train a model using data from one language and then apply it to other languages. Of course the change in data distribution is quite severe in this case as we might transfer knowledge between languages that are completely different from linguistic point of view. In extreme cases they might even have a different alphabet, so there is not much connection to be found between them. To bridge this gap we can employ a variety of linguistic resources that can help us establish correspondence between languages. This could mean using bilingual dictionary, parallel corpora or machine translation systems *inter alia*.

Even though researchers already tried to perform such transfer with various deep NLP architectures, there is still a lot of open questions to be answered. First, the correspondence can be established on character-level, word-level or sentence-level. We can model all these with distributed representations. It is however not clear what are the advantages and disadvantages of these approaches, when should one choose one over the other nor how to combine them successfully. One could for example combine different transfer techniques for different language pairs to fully exploit their respective similarities. Currently usually only one language is used as source language but using multiple languages with different properties could be beneficial.

This is the basic open question I try to tackle, but there are other interesting questions at hand:

- What is the critical amount of various resources we need for successful transfer? What is the effect of imperfect machine translation on the resulting model? We are interested in the properties linguistic resources should have and how to predict their effect on learning.
- What exactly is being learnt by these neural networks when they are being trained in cross-lingual setting? We are interested in on what level is the correspondence between languages established and how does it compare to our existing knowledge about languages.
- How does the choice of neural network architecture affect the results? Are there qualitative or only quantitative differences between various approaches? A lot of architectures were proposed recently in NLP and they can be in usually in some way used for cross-lingual learning as well. Is any of them especially well-suited for transferring knowledge across languages?
- Can we combine knowledge from more languages to help us solve tasks? I am interested in scenario where each source language can help us in different way. E.g. we can use linguistically very similar language to learn building word representations from characters while we use other language for task supervision?

Related Work

One important technology that emerged in recent years are so called *multi-lingual word embeddings* (Upadhyay et al. 2016). It is an extension of widely used word embeddings in which the semantic vector space contains words from more than one language. Word representations in this space preserve semantic similarity of words in terms of their geometric similarity. Words with similar meaning, independently on the language they come from, are put close together based on distributional semantics. To construct connection between languages while building such space researchers use parallel corpora or bilingual dictionaries. Multi-lingual embeddings are then often used for cross-lingual learning as we have language independent representation of words at hand. Model trained on word representations from one language can then be used on data from other language as well.

Similarly we can try to train our models to create shared vector space for sentence representations (Hermann and Blunsom 2014). This can be done even in supervised setting when we want to make the representations useful for certain NLP tasks, such as sentiment analysis (Zhou, Wan and Xiao 2016). In such cases we have one shared deep model that can process and evaluate sentences multi-lingually with positive transfer of knowledge between languages. In such models we can have various parameters shared. In deep learning this practically always mean sharing weight matrices across languages. Based on what we need we can e.g. share only parameters responsible for creating word representations from characters across languages with similar vocabulary or parameters responsible for combining word embeddings into sentence representations (Yang, Salakhutdinov and Cohen 2017). Although multiple other solutions were proposed, I feel like proper evaluation techniques are still lacking. Right now there is no single solution preferred by community to create higher level representations for cross-lingual learning. What is even more surprising is the lack of understanding we have about these models. No proper techniques were proposed to help us understand how does the transfer of knowledge work in these works.

Research Plan

So far I have done an experiment when I tried project the information about word sentiment across languages using multi-lingual word embeddings. I trained multi-lingual word vector space and used existing English sentiment lexicons to teach simple feed-forward neural network to predict the sentiment of words based on their vector representation. Then I was able to use trained model to predict sentiment of words from other languages. This was preliminary experiment which results are still to be published.

Up until now I was experimenting only with transfer of sentiment, but it can be performed for other word properties as well, such as part-of-speech or whether it is named entity for named entity recognition. I also performed experiment only for transfer from English to German, as German has manually created sentiment lexicons ready which I can use for evaluation. During this experiment I was able to achieve results similar to state-of-the-art German sentiment lexicons with fully automatic method. Sentiment lexicons are basic resource used for sentiment analysis and which are quite laborious to assemble. My method can significantly reduce the amount of human work that need to be put in them.

I plan to extend this experiment with other languages, tasks and multilingual word embeddings models. Currently I am also writing a survey paper on cross-lingual learning in general. This is as far as I know the first attempt to systemize cross-lingual learning methods in survey format.

Aforementioned experiments are planned to be completed by February 2018, when AAAI takes place. Then I plan to work with sentence-level representations that are able to encode sentences from different languages into one shared semantic vector space. These methods were not yet scrutinized on how well they preserve various information about sentences, such as sentiment. Up until now most of the work in field was ad hoc and not a lot is known about properties of various approaches to encode semantics multi-lingually. This line of experimenting will also be in progress by February. I believe that by answering these questions I can help other NLP practitioners build their solutions for resource-poor languages.

Acknowledgments

This work was partially supported by the Scientific Grant Agency of the Slovak Republic, grants No. VG 1/0667/18, VG 1/0646/15. The authors would like to thank for financial contribution from the STU Grant scheme for Support of Young Researchers.

References

Hermann, K. M.; Blunsom, P. 2014. Multilingual Models for Compositional Distributed Semantics. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 58-68. Baltimore, Maryland: ACL.

Upadhyay, S.; Faruqui, M.; Dyer, Ch.; and Roth, D. 2016. Cross-lingual Models of Word Embeddings: An Empirical Comparison. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 1661-1670. Berlin, Germany: ACL.

Yang, Z.; Salakhutdinov, R.; Cohen, W. W. 2017. Transfer Learning for Sequence Tagging with Hierarchical Recurrent Networks. ArXiv cs.CL:1703.06345.

Zhou, X.; Wan, X.; Xiao, J. 2016. Cross-Lingual Sentiment Classification with Bilingual Document Representation Learning. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 1661-1670. Berlin, Germany: ACL.