

CHAPTER 12

CONCLUSIONS

In this closing chapter, we provide an overview of the whole thesis. In section 12.1 we will have another look at the research questions and the manner in which they were approached in each part of the thesis. We will summarise the achievements made through the course of this study in Section 12.2. Finally, we will provide a brief summary of the thesis.

12.1 Research Questions

In this section, we revisit the research questions and briefly discuss the ways they have been approached in our work.

RQ-A: *What are the available supervised and unsupervised STS methods, and how do they perform in multilingual and multi-domain environments?*

The first part of the thesis focussed on various supervised and unsupervised STS methods. We explored embedding aggregation based STS methods, sentence encoders, Siamese neural networks and transformers in STS. We evaluated all the methods in three popular English STS datasets. Furthermore, for each STS method, we analysed the ability of the method to perform in a multilingual and multi-domain setting by evaluating them in Arabic, Spanish and bio-medical STS datasets. Our results indicated that the best supervised STS method is

transformers and sentence encoders are the best unsupervised STS method.

RQ-B: *Can the neural STS methods be applied in TMs? How efficient and effective are they compared to the real-world TM tools?*

The second part of the thesis focussed on applying the developed neural STS methods in TMs. Considering the accuracy and efficiency, we picked three sentence encoders; Infersent, Universal Sentence Encoder and SBERT. We designed a TM matching algorithm based on them. We evaluated the proposed algorithm in a real-world TM; DGT-TM. We compared the results from each of the sentence encoders with the results from Okapi, which uses edit distance to acquire the best match from the translation memory. The results indicated that our approach returns better matches than Okapi in many cases. We confirmed this with an automatic evaluation as well as with a human evaluation. Furthermore, the proposed algorithm is very fast and can be used in real-world applications.

Our conclusion is that the STS methods have progressed to a point where they can improve TM matching and retrieval. Unlike the early days, the current neural models are optimised for real-world applications. Therefore these methods should be studied more in forgotten tasks such as TM matching and retrieval.

RQ-C: *Can the state-of-the-art STS methods be adapted in the QE task? Can these simple STS architectures outperform current complex QE methods?*

The third part of the thesis focussed on applying the developed neural STS methods in translation quality estimation. The current state-of-the-art in QE is neural models such as OpenKiwi and DeepQuest. However, these neural QE

architectures are complex and need a lot of computing resources to train a QE model. We proposed redefining the QE task as a cross-lingual STS task and applying the STS architectures we experimented with in Part I of the thesis to QE to address this limitation. We applied the state-of-the-art transformer based STS methods to sentence-level QE by changing the input embeddings to cross-lingual. This approach was very simple compared to the existing QE solutions. Our approach outperformed existing QE methods in fifteen language pairs providing state-of-the-art results. We extended the architecture to word-level QE and evaluated the proposed architecture in eight language pairs. Our simple word-level architecture outperformed other complex neural models in all the language pairs providing state-of-the-art results.

Our conclusion is that the state-of-the-art STS methods can easily be adapted in the QE task by changing the embeddings to cross-lingual. These architectures are simple compared to the existing neural architectures in OpenKiwi and DeepQuest, yet the proposed architectures in this study outperformed them in all the language pairs.

12.2 Achievements

This thesis was conceived in order to apply STS methods in the applications of translation technology. First, we studied and employed different STS methods and then applied them in translation memories and translation quality estimation.

For STS in Part I of the thesis, we started with unsupervised STS methods. In

Chapter 2, we employed vector aggregation based unsupervised STS methods. For the first time, we integrated contextual word embedding models such as BERT, ELMo and Flair in vector aggregation based unsupervised STS methods. We showed that vector aggregation based STS methods could be improved with contextual embeddings. This method we proposed is the current state-of-the-art in vector aggregation based STS methods in several datasets (Ranasinghe et al., 2019a). In Chapter 4, we employed Siamese neural networks in the STS task. In 2018, an LSTM based Siamese neural network was state-of-the-art in STS. We improved this architecture by including simple GRUs in the network. Our GRU based Siamese neural network outperformed the LSTM based Siamese neural network architecture in multiple datasets and provided state-of-the-art results before the transformer era. This GRU based Siamese neural network stands as one of the best STS methods based on traditional word embeddings for multiple STS datasets (Ranasinghe et al., 2019b).

For TMs in Part II of the thesis, we started with analysing existing TM Systems. The majority of these systems are based on edit distance. However, the edit distance based TM systems are weak and can not retrieve semantically similar segments from the TM. As a result, third-generation generation TM systems have been proposed. Even though these third-generation TM systems have addressed the limitations of the edit distance based TM systems, they are not popular in the community since they are largely inefficient, and there is not much performance gain in using them. Addressing this gap, in Chapter 7, we proposed to use deep learning based STS methods in TM retrieval task. Considering the

efficiency, we proposed a novel TM retrieval method based on sentence encoders and compared its performance against a popular TM system; Okapi. We showed that the proposed method could retrieve better matches than Okapi in many cases. Furthermore, the proposed method is fast and can be used in real-world applications. This is the first study done on utilising neural sentence encoders in TM retrieval task. We believe our contribution will bring a new direction to the third-generation TM systems (Ranasinghe et al., 2020a; Ranasinghe et al., 2021a).

For QE in Part III of the thesis, we started with analysing the existing QE systems. The majority of these QE systems are based on complex neural network architectures and need a lot of computing resources to train a QE model, which we have identified as a major limitation. To address this weakness, we propose to redefine the QE task as a cross-lingual STS task and apply the STS architectures we experimented in Part I of the thesis to QE, which are considerably simpler than the existing QE models. First, we proposed two STS architectures based on transformers for the sentence-level QE task. Then we extended the best sentence-level architecture for word-level QE by changing the output layer. We evaluated the proposed algorithms in 15 different language pairs. We showed that our simple method outperforms the current QE systems in all the languages in both word-level and sentence-level, providing state-of-the-art results. The approaches that we proposed for QE are rapidly increasing in popularity due to their superior performance, and we believe that our contributions have already revolutionised the QE field (Ranasinghe et al., 2020c; Ranasinghe et al., 2020b). Furthermore, the multilingual QE approaches we proposed in Chapter 11 would

be beneficial for many low resource languages where the training data is difficult to find (Ranasinghe et al., 2021b).

The key points of strength in this work can be summarised as follows:

The use of state-of-the-art methods - In this work, we employed state-of-the-art methodologies and kept improving the architecture of our models in newer experiments as the field of NLP was concurrently progressing. For instance, we adapted state-of-the-art transformers in our study. Even for the unsupervised STS approaches, we utilised contextual embeddings. Furthermore, we employed architectures and embeddings that have proven stable and reliable in a variety of tasks. For example, the cross-lingual embedding model we used in Part III of the thesis has been evaluated on various cross-lingual benchmarks, proving this approach's reliability.

Multilingualism - A common theme in our study was language-independent models. Almost all the experiments that were conducted in the course of this research are language independent. To demonstrate that, we constantly evaluated our models in a variety of languages. For example, the QE method we proposed in Part III of the thesis was evaluated in 15 language pairs, including low-resource languages such as Sinhala and Nepalese. We believe that the multilingual nature of our experiments would benefit many languages.

Replicability, and creation of new resources in the form of code - For almost all the experiments that were conducted in the course of this research,

the code and the machine learning models have been made available to let other researchers replicate or experiment further in the same direction. In most cases, we have used established and standard datasets to compare the proposed methods with the available methods. Out of the codebases released in the study, two projects have been extremely popular in the community. The transformer-based STS method in Chapter 5 was released as a Python library named *ststransformers* and has more than 3,000 downloads from the community¹. The QE framework; *TransQuest* which was released for the Part III of the thesis, has more than 9,000 downloads. Furthermore, the various pre-trained models released for *TransQuest* has more than 12,000 downloads². With these statistics, we believe that the software released in this study have been useful and have created a huge impact on the NLP community.

Competitiveness - In this study, we aimed to build the best approach possible for a given task. We always compared our method to the best systems available in each dataset. For instance, we evaluated the QE method we developed on the WMT 2020 QE shared task. Our approach won first place in all the language pairs outperforming all other participants across the world, proving our study's competitiveness.

¹The latest download statistics for *ststransformers* is available on <https://pepy.tech/project/ststransformers>

²The latest download statistics for *TransQuest* framework is available on <https://pepy.tech/project/ststransformers>. The number of downloads for each pre-trained *TransQuest* model is available on <https://huggingface.co/TransQuest>

12.3 Summary

To conclude, this thesis focused on employing deep learning based STS metrics in the applications of translation technology. We developed language-independent STS methods and utilised them in two applications in translation technology, translation memory systems and translation quality estimation. We made original contributions to each of the applications. We provided the best existing algorithm to perform translation memory matching and retrieval. Also, the QE method we proposed in this study using STS architectures provided state-of-the-art results outperforming other QE systems and winning the WMT 2020 QE shared task. Our models employed sound and reliable machine learning techniques and proved remarkably effective across multiple languages.