Dialogue System based on Semantic Textual Similarity

No Author Given

No Institute Given

1 Introduction

Semantic Textual Similarity (STS) aims at computing the proximity of meaning transmitted by two fragments of text. We see a clear application of STS for mapping pairs of sentences that, though possibly having different surface forms, do have the same or similar meaning (e.g., *What is the procedure for participating?* and *How can I enter?*). This capability is extremely useful for a dialog system that answers questions in natural language. Following this idea, we trained a model for STS in Portuguese, in the ASSIN collection [2], and used it as the engine of a dialog system. The primary goal of this system is to answer questions on a domain, according to a list of Frequently Asked Questions (FAQs).

2 Feature Engineering for Portuguese STS

We extracted a set of lexical, syntactic and semantic features from Portuguese texts and then follow a supervised learning approach for learning a function of Portuguese STS from the ASSIN collection [2], which contains 10,000 sentence pairs (6,000 for training and 4,000 for testing) and their similarity value, based on the opinion of several human judges. Similarity ranges between 1 (completely different) and 5 (equivalent). Text was pre-processed with the NLPPy-Port toolkit [1] and features were extracted with this and other libraries in Python. Based on the features, various regression algorithms, with implementation available in scikit-learn, were explored for learning the STS model. The performance of these models was assessed with metrics commonly used in STS, namely Pearson correlation ($-1 < \rho < 1$) and Mean Squared Error (MSE) between the values computed and those in the collection.

The best results were obtained with a Support Vector Regressor (SVR), with the default scikit-learn parameters. Having in mind the scability of the model, we further tried to reduce the dimensionality of the feature set and achieved $\rho = 0.72$ and $MSE = 0.63$ with the following 12 features: the Jaccard, Dice and Overlap coefficients for token 1-grams and character 3-grams; the Jaccard coefficient for character 2-grams; the cosine similarity between the sentence vectors computed using the TF-IDF matrix; the fastText.cc word embeddings; and the NILC Word2vec, fastText.cc and PTLKB word embeddings weighted with the TF-IDF value of each token.

3 STS in a Dialogue System

The best STS model was trained in the full ASSIN collection and then integrated in a dialogue system that answers questions related to the exercise of economic activity in Portugal. For experimentation purposes, a list with 379 FAQs on this domain was used as the system's knowledge base.

In order to test how well the STS model worked in this domain, and bearing in mind that, in most cases, users will not search for the exact question, variations of the original questions were created: (i) Using the Google Translate API[1] for generating paraphrases of the original questions, with the following sequence of actions: translation of Portuguese to English and back to Portuguese (VG1); the previous result back to English and back to Portuguese (VG2). (ii) Manually produced by a group of native Portuguese speaking students, including paraphrases of the original questions and closely-related questions, possibly with minor spelling mistakes (VUC).

Testing consisted of feeding the system with each variation generated and checking if it was able to match it correctly with the original question of the FAQs. As expected, due to the nature of each variation group, the best results were obtained with VG1 and VG2 which achieved an accuracy of 91% and 90% respectively. The accuracy in the VUC group was around 58%, which resulted in a system overall performance of about 85%.

4 Conclusions

We described the integration of a STS model in a Portuguese Dialogue agent. When testing this agent, we noticed that, when questions are made in a similar way to how they are stored in the agent's knowledge base, the model is quite successful. However, when differences are more noticeable, performance decreases.

In the future, we will work both on the improvement of the STS model and its evaluation in a conversational scenario.

References

1. Ferreira, J., Gonçalo Oliveira, H., Rodrigues, R.: Improving NLTK for processing Portuguese. In: Symposium on Languages, Applications and Technologies (SLATE 2019). OASIcs, vol. 74, pp. 18:1–18:9. Schloss Dagstuhl (June 2019). https://doi.org/10.4230/OASIcs.SLATE.2019.18, http://drops.dagstuhl.de/opus/volltexte/2019/10885
2. Fonseca, E., Santos, L., Criscuolo, M., Aluísio, S.: Visão geral da avaliação de similaridade semântica e inferência textual. Linguamática **8**(2), 3–13 (2016)

[1] https://cloud.google.com/translate/docs/