Exercise Contextualized Embeddings

Natural Language Processing

Ulrich Heid, Christian Wartena, Johannes Schäfer

Summer Term 2022

- Install flair (see `https://github.com/flairNLP/flair`). For example, in colab you can use the following:

```
!pip install flair
```

- Use flair embeddings to represent tokens in sentences (see `https://github.com/flairNLP/flair/blob/master/resources/docs/TUTORIAL_3_WORD_EMBEDDING.md`).
- Represent the tokens by stacked embeddings of a forward and backwards RNN on top of the character embeddings (`flair_embedding_forward` and `flair_embedding_backward`). Note: you should not use the glove embeddings here, just the flair embeddings.

1 Tasks

- Load the dataset of movie lines which we used in the exercise last week.
- Select (at least) two words which occur at least 100 times in this dataset. One word should have only one meaning (or at least one meaning which you expect to be used very often while you expect the other meanings of this word to be very rare). The other word should have (very many) multiple meanings (for example, the most common verbs[1] often have very many meanings). You should check WordNet to get an idea of different meanings of the words, see `http://wordnetweb.princeton.edu/perl/webwn`.
- For each of the two words do the following[2]:
  - Count how often the word occurs in this dataset.
  - Compute the flair representations for all sentences where the word occurs in. From each of those sentence representations select the contextualized embedding vector of the word and save all these vectors in a list (if your word is extremely frequent, you may limit this list to a maximum size - so the following experiments are feasible).
  - Compute a centroid embedding vector for all contextualized embeddings of the word (compute the element-wise sum of all vectors and devide by the number of vectors - the resulting centroid vector is a vector in the same embedding space). Calculate the standard deviation of the cosine similarities of the contextualized embedding vectors to the centroid vector? (If you want, you can plot the distribution.)
- Compare the resulting values for both words. Did you expect this result?

[1] `https://www.educall.com.tr/blog/post/500-most-common-english-verbs`

[2] If you implement these steps in a loop and it does not take too long to compute, it might be nice to run this for a few more words.