DrugProt

Orpailleur Team

September 2021

1 Orpailleur members

- Yannick Toussaint (Yannick.Toussaint@loria.fr)

- Laura Zanella (laura-alejandra.zanella-calzada@loria.fr)

- Fatima Habib (fatima.habib2@etu.univ-lorraine.fr)

2 Systems description

We have four proccesses to run. The first three are the following:

- Convolutional neural network (CNN) based model: this model consists of two convolutional layers, where each layer is followed by a max pooling layer. The input to this models are the word embeddings, the position embedding and the target entities embeddings.

- SciBERT + CNN based model: we add a SciBERT layer before the previous model. The SciBERT layer gives the sentence embedding and the entities embeddings, those embeddings are concatinated with the position embedding and used as inputs to the model.

- SciBERT + LSTM + CNN model: we insert an LSTM layer between the SciBERT layer and the CNNs layers. The inputs to this model are the same of the SciBERT + CNN model.

The fourth model is a Random Forest (RF) classifier combining the three models.