

Master project for SIN/SSC/DS students

Exploiting Transfer Learning to predict the economic potential of companies

Martin Rajman
Martin.Rajman@epfl.ch

Project description:

The goal of this project is to design a Deep Neural Network (DNN) able to predict the economic potential of a company from a (short) textual description of this company.

For this purpose, a collection of 165'000 company descriptions has been collected, and 5'000 of them have been associated with an economic potential by a set of human experts.

To cope with the small size of the supervised training set, the idea is to exploit the multi-layered nature of the DNN to perform a specific type Transfer Learning, where:

1. the DNN is first trained to predict the n-th word in a description from the same description where the n-th word has been anonymized; for this task, the whole corpus of collected company descriptions can be used as training set, independently from the (small) size of the set of descriptions annotated by human experts; and then
2. the higher layers of the DNN are adapted and the DNN re-trained on the 5'000 available annotated descriptions to predict the targeted economic potentials.

The underlying assumption (to be validated) is that the set of parameters to be re-trained gets sufficiently manageable to make the training on the available 5'000 annotated descriptions feasible, while the preserved lower layers of the DNN remain relevant enough to continue to efficiently contribute to a satisfying prediction performance. To transform the textual descriptions into dense vector representations better suited for the training of a DNN, pre-trained word embeddings shall be used to feed the first (convolution and max-pooling) layers of the DNN.

Proposed for:

1 Master student in Computer science, Communication systems or Data Science