

Abstract

First successful implementation of Search2Vec on the most repeated entries of data set

Description

On the unofficial mid week report a few problems explained with the embedding space results. Fortunately, after investigation for the problem, it was found that only one of two matrices of method's output were considered. In fact, according to the fact that training samples (X, Y) are only constructed with X as query and Y as category (or a specific product), the target matrix (first one) shows the embedding in the query environment and the context matrix illustrates the relations among the categories (or products). The latest experiment's properties are demonstrated in the below table.

Vocabulary size	100 queries + 756 categories
Dimension of representative vectors	300
Number of Training samples	1'327'707
Number of epochs	1

Until now, the quality of the results were evaluated by arbitrary checking on the Embedding projector.

Next Week

1. The main question right now is which is the best way to convert these two matrices into a single one.
2. Another task might be implementation of an assessment method. This method will be subsequently used to evaluate the embedding space acquired from the neural network.
3. Expansion of data could be possible for future implementations.

References