Digital Media Laboratory

Weekly Report August 30, 2021

Ali Jandaghi Alaee
alialaee98@gmail.com

Report No. 5
00/05/26- 00/05/1

Abstract

On this week, another experiment on learning embedding of queries and categories has been executed. In addition, a great solution for turning two matrices into one has been found and this problem seems to be solved. Furthermore, some time was spent on getting familiar with the evaluation methods described in the search2vec's experiments section.

Description

- Another experiment was executed this week with more data entries. Specific details of this execution are illustrated on the below table. Fortunately, the simple averaging on two learning matrices (target and context matrices) could effectively merge these two matrices into one, with keeping their embedding related to each other. Now, the single output matrix contains the expected embedding properly.

Vocabulary size	500 queries + 1056 categories
Dimension of representative vectors	300
Number of Training samples	2'631'805
Number of epochs	1

  The code is accessible on *this google collaboration*.

- Evaluation methods that had been discussed on the main article, were read and studied. Among them, the implementation of nDCG method seems to be convenient (other methods mostly were related to online examinations). However, the question is why we need to implement this self-assessment, considering that the competition is going on and it's judgement engine might be available. In fact, still there are a few vague issues remaining about the use of nDCG that should be fixed as soon as possible (For instance, how can we find the best ideal rank? with the help of number of clicks associated with every query or the previous assigned ranks?)

Next Week

1. According to the answer of the question stated above about nDCG, maybe an implementation of this algorithm will be needed.
2. Developing a piece of code to find K nearest neighbours to a point.
3. Probably, putting some effort to establish an algorithm for tale queries.

References