A Study of Lexical Matching in Neural Information Retrieval - Abstract*

Thibault Formal[1,2], Benjamin Piwowarski[2,3] and Stéphane Clinchant[1]

[1]*Naver Labs Europe, Meylan, France*
[2]*Sorbonne Université, Institute for Intelligent Systems and Robotics, Paris, France*
[3]*CNRS*

Abstract

Neural Information Retrieval models hold the promise to replace lexical matching models, e.g. BM25, in modern search engines. While their capabilities have fully shone on in-domain datasets like MS MARCO, they have recently been challenged on out-of-domain zero-shot settings (BEIR benchmark), questioning their actual generalization capabilities compared to bag-of-words approaches. Particularly, we wonder if these shortcomings could (partly) be the consequence of the inability of neural IR models to perform lexical matching off-the-shelf. In this work, we propose a measure of discrepancy between the lexical matching performed by any (neural) model and an "ideal" one. Based on this, we study the behavior of different state-of-the-art neural IR models, focusing on whether they are able to perform lexical matching *when it's actually useful*, i.e. for important terms. Overall, we show that neural IR models fail to properly generalize term importance on out-of-domain collections or terms almost unseen during training.
This paper is an extended abstract of a short paper accepted at ECIR22.

Keywords

Neural Information Retrieval, BERT, Lexical Matching

CIRCLE (Joint Conference of the Information Retrieval Communities in Europe) 2022, July 04–07, 2022, Samatan, France

*Corresponding author.
† These authors contributed equally.
thibault.formal@naverlabs.com (T. Formal); benjamin@piwowarski.fr (B. Piwowarski); stephane.clinchant@naverlabs.com (S. Clinchant)