

A Light-weight Strategy for Restraining Gender Biases in Neural Rankers

Amin Bigdeli Negar Arabzadeh Shirin Seyedsalehi Morteza Zihayat Ebrahim Bagheri
 {abigdeli, shirin.seyedsalehi, mzihayat, bagheri}@ryerson.ca, Narabzad@uwaterloo.ca

Problem

Abstract

In light of recent studies that show neural retrieval methods may intensify gender biases during retrieval, the objective of this paper is to propose a simple yet effective sampling strategy for training neural rankers that would allow the rankers to maintain their retrieval effectiveness while reducing gender biases. Our work proposes to consider the degrees of gender bias when sampling documents to be used for training neural rankers. We report our findings on the MS MARCO collection and based on different query datasets released for this purpose in the literature. Our results show that the proposed light-weight strategy can show competitive (or even better) performance compared to the state-of-the-art neural architectures specifically designed to reduce gender biases.

Top 3 re-ranked documents by the original BERT model.

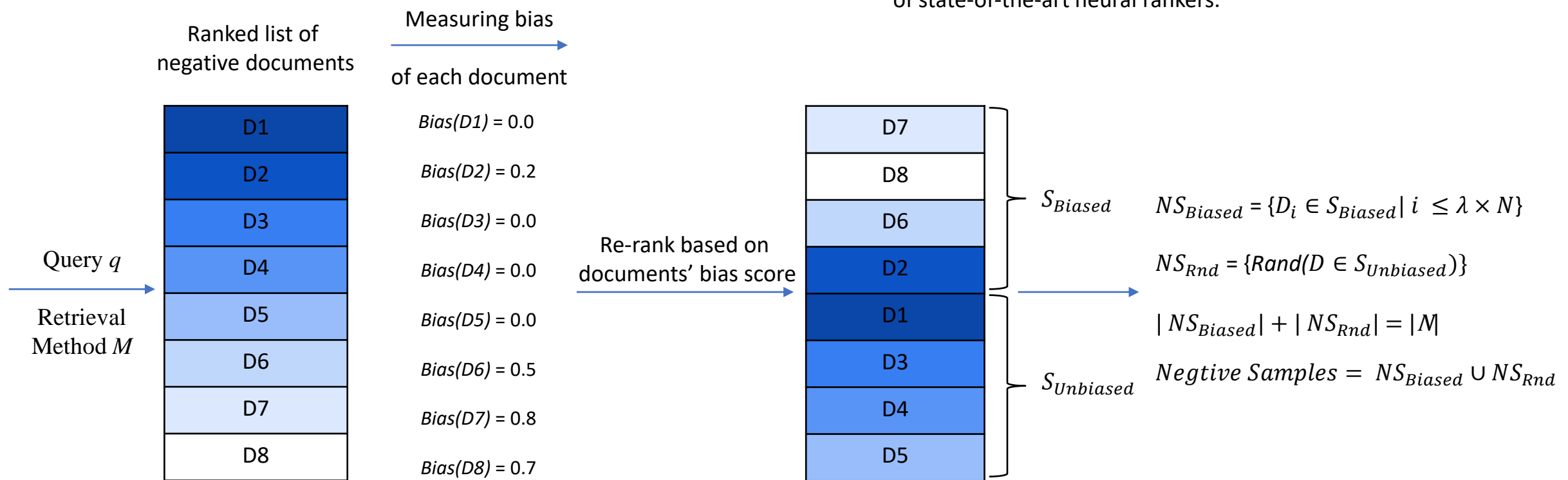
Query: is a supervisor considered a manager?
So, a manager or supervisor who has control or [...], is considered an employer and subject to the rights and obligations of an employer under the AEPA. Ranking Position in Ours: 1
A manager or supervisor of agricultural employees may also be considered an employer for the purposes of the AEPA. The AEPA defines employer [...] Ranking Position in Ours: 2
It becomes clear that the core of the role and responsibility of a supervisor [...]. A supervisor in any department has more or less the same work experience as the other members in his team, but he is considered to be the leader of the group. The word manager comes from the word management, and a manager is a person who manages men . To manage is to control and to organize things, men , and events. Managers do just that. They ensure smooth running of the day to day functioning of a workplace [...]. Ranking Position in Ours: 22

Solution

Proposed Approach: Instead of randomly sampling N negative samples from $top-k$ retrieved documents by $BM25$, we systematically select N negative samples such that the neural ranker is exposed to stereotypical gender biases that need to be avoided when ranking documents.

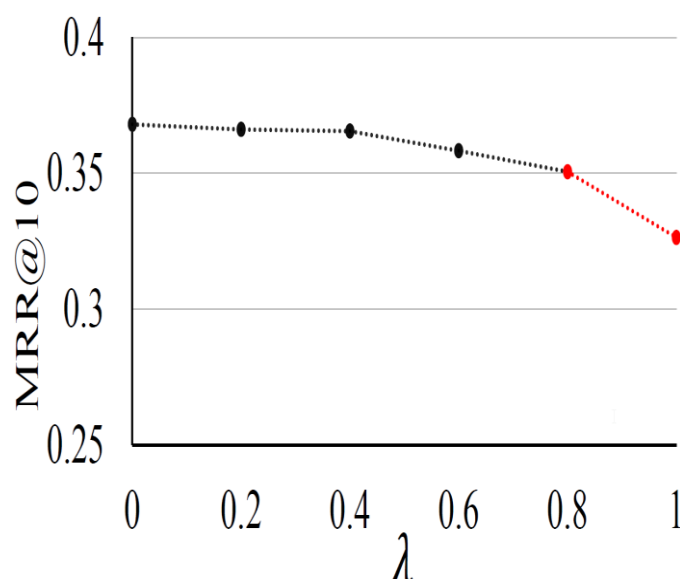
Advantages of the Proposed Approach:

1. Decreases the level of gender biases in neural ranking models, while maintaining a comparable level of retrieval effectiveness.
2. Does not require any changes to the architecture of state-of-the-art neural rankers.



Evaluation

Impact of λ on BERT model's performance on MS MARCO Dev Set. The red points indicate statistically significant drop in performance.



Comparison between the Performance of the base ranker and the ranker trained based on our proposed negative sampling strategy when $\lambda = 0.6$ on MS MARCO Dev Set.

Neural Ranker	MRR@10		Change
	Original	Ours	
BERT (base)	0.3688	0.3583	-2.64%
DistilRoBERTa (base)	0.3598	0.3475	-3.42%
Electra (base)	0.3332	0.3351	+0.57%

Comparing the base rankers vs our proposed approach and ADVBERT in terms of performance and Fairness (NFaiRR) when using different pre-trained language models on a fairness-sensitive query set.

