

BAI

智源学者成果展示——智源青年科学家

作者 徐君、兰艳艳（中国人民大学、中国科学院计算技术研究所）

2020年6月

项目负责人：徐君、兰艳艳

单位 / 团队：中国人民大学、中国科学院计算技术研究所

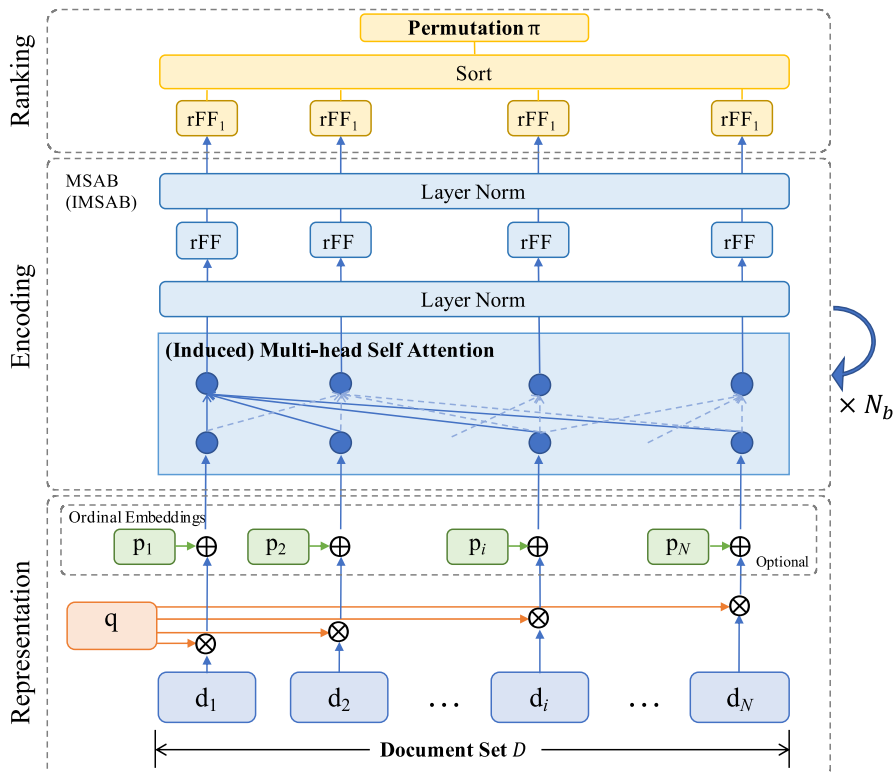
成果名称：

SetRank: Learning a Permutation-Invariant Ranking Model for Information Retrieval

成果简介：

信息检索中的排序学习旨在对给定文档集合的每个文档进行相关性排序。因此，理想的排序学习模型应该是从文档的集合到排列的映射，基于此需要满足两个关键要素：

- (1) 应具有对跨文档交互进行建模，以捕获查询中上下文信息的能力；
- (2) 应满足排列不变性，这意味着输入文档的任何排列都不会改变输出的排列。已有的排序学习方法，要么基于单变量评分函数，即分别对每个文档评分，未能建模跨文档交互；要么考虑顺序构建跨文档的多元评分函数，但不可避免地会牺牲排列不变性。为了解决上述问题，我们提出了 SetRank 排序学习模型，可在任意大小的文档集合上，定义的排列不变的排序学习函数。SetRank 使用多头自我关注结构（及其 Induced 变种）作为其关键组件，以便学习集合中所有的文档的表达。自我关注机制不仅可以帮助 SetRank 从跨文档交互中捕获本地上下文信息，而且还可以学习输入文档的排列不变表示，从而实现排列不变的排序学习模型。在三个大型公开数据集的实验上，SetRank 的性能明显优于包括传统的学习排名模型和最新的神经 IR 模型在内的基线模型。



Beijing Academy of Artificial Intelligence



微信关注
北京智源人工智能研究院