

ACL 2018 | 清华大学：基于实体的神经信息检索模型-理解实体在神经信息检索模型中的作用

原创 热爱学习的 读芯术 2018-06-10

点击蓝色字关注读芯术 AI应用领域最专业的垂直类自媒体

你和“懂AI”之间，只差了一篇论文

号外！又一拨顶会论文干货来袭！

2018年6月9-10日，代表学术界和工业界的顶级交流盛会——由中国中文信息学会青年工作委员会和百度公司联合举办的【“AIS2018 (ACL、IJCAI、SIGIR) 论文预讲会”】在北京盛大举行。两天时间内，来自“情感分析”“推荐系统”“机器问答”“对话系统”等不同主题的顶级会议论文报告汇聚一堂，英雄相惜，华山论剑。

据芯君了解，本次预讲会在学术圈的火爆程度完全超出了主办方的想象，开放报名的短短几天就全面满额了，以至于主办方不得不设定条件筛选参会者。

读芯君作为本次预讲会的活动媒体，将全程跟随大会，为大家全程纪录活动中最前沿的观点，最有价值的成果，并特邀预讲会论文报告者联合为读者朋友们推出预讲会系列组文，向你展示顶会最新论文成果。

读芯术读者论文交流群，请加小编微信号：[zhizhizhuji](#)。等你。

这是读芯术解读的第53篇论文



作者：刘正皓，熊辰炎，孙茂松，刘知远

ACL 2018

基于实体的神经信息检索模型：理解实体在神经信息检索模型中的作用
Entity-Duet Neural Ranking: Understanding the Role of Knowledge Graph
Semantics in Neural Information Retrieval

清华大学

Tsinghua University

在本文中，我们提出了一种融合知识的神经信息检索方法（Entity-Duet Neural Ranking Model, EDRM），目的是使用人类的先验知识（Knowledge Graph）来增强神经检索模型的结果。



当前的神经信息检索模型（Neural Information Retrieval, Neural-IR）可以大致被分为两种：一种是基于表示的神经网络信息检索模型（Representation-based）；另一种是基于交互的神经信息检索模型（Interaction-based）。Xiong等人提出了基于kernel pooling的交互的神经信息检索模型（K-NRM）；Dai等人基于K-NRM提出了Conv-KNRM，利用CNN整合n-gram的信息，并达到了当前的最好水平。

基于实体的信息检索模型（Entity-oriented Information Retrieval）是基于人类的先验知识，将知识图谱的信息融合进信息检索模型中。当前的工作都是得到更好的问题、文档的表示，或者通过实体建立问题-文档的隐含关系来增强信息检索效果。但当前基于实体的信息检索模型都是基于人工特征应用到传统的信息检索模型中。因此在本文中我们使用信息抽取的方法，首次将知识结合到神经信息检索模型中，使得模型具有更好的效果以及泛化能力。

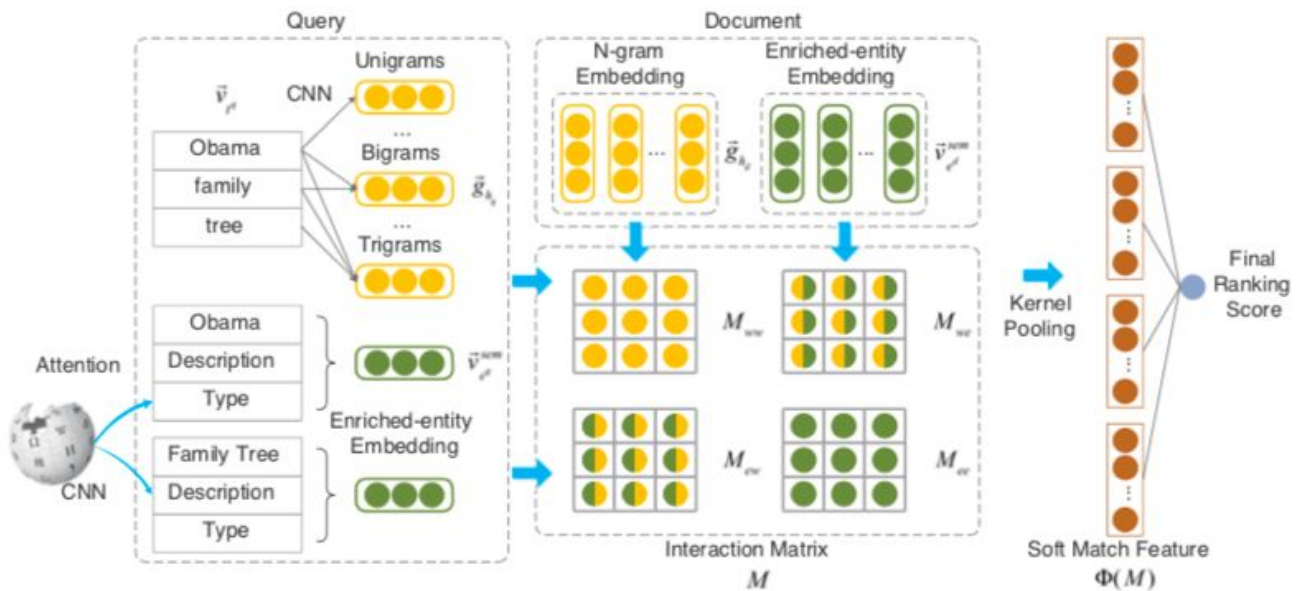


Figure 1: The architecture of EDRM.



基于以上，我们针对Interaction-based Ranking Model提出了Entity-Duet Neural Ranking Model (EDRM)。在EDRM模型中，每一个问题以及文档中的实体包含三种不同的语义表示：实体向量 (EntityEmbedding)，描述向量 (Description Embedding) 和类型向量 (Type Embedding)，并线性组合成增强实体表示 (Enriched-entityEmbedding)。为了更好的融合词级别以及实体级别的信息，我们提出了Neural Entity-Duet Framework，其中包含四种交互矩阵：问题词语-文档词语、问题词语-文档实体、问题实体-文档词语、问题实体-文档实体。通过神经网络的方法利用用户的点击信息端到端的训练优化。

我们在实验中通过将EDRM整合到两个state-of-the-art的基于交互的神经信息检索模型中 (K-NRM以及Conv-KNRM)，使用 and K-NRM 相同的实验设置，训练采用DCTR推断出的label进行训练，并设置三个测试场景：

Testing-SAME：测试采用DCTR推断出的label；

Testing-DIFF：测试时使用TACM推断出的label；

Testing-RAW：测试时采用MRR来测试最符合文档的排序结果。

通过表1我们可以看到相比较我们主要的基线模型 (K-NRM和Conv-KNRM) 来讲，模型在Testing-DIFF和Testing-RAW上面有更好的表现，证明模型具有更好的泛化能力。

Table 1: Ranking accuracy of EDRM-KNRM, EDRM-CKNRM and baseline methods. Relative performances compared with K-NRM are in percentages. †, ‡, §, ¶, * indicate statistically significant improvements over DRMM†, CDSSM‡, MP§, K-NRM¶ and Conv-KNRM* respectively.

Method	Testing-SAME				Testing-DIFF				Testing-RAW	
	NDCG@1		NDCG@10		NDCG@1		NDCG@10		MRR	
BM25	0.1422	-46.24%	0.2868	-31.67%	0.1631	-45.63%	0.3254	-23.04%	0.2280	-33.86%
RankSVM	0.1457	-44.91%	0.3087	-26.45%	0.1700	-43.33%	0.3519	-16.77%	0.2241	-34.99%
Coor-Ascent	0.1594	-39.74%	0.3547	-15.49%	0.2089	-30.37%	0.3775	-10.71%	0.2415	-29.94%
DRMM	0.1367	-48.34%	0.3134	-25.34%	0.2126†	-29.14%	0.3592§	-15.05%	0.2335	-32.26%
CDSSM	0.1441	-45.53%	0.3329	-20.69%	0.1834	-38.86%	0.3534	-16.41%	0.2310	-33.00%
MP	0.2184†‡	-17.44%	0.3792†‡	-9.67%	0.1969	-34.37%	0.3450	-18.40%	0.2404	-30.27%
K-NRM	0.2645	-	0.4197	-	0.3000	-	0.4228	-	0.3447	-
Conv-KNRM	0.3357†‡¶*	+26.90%	0.4810†‡¶*	+14.59%	0.3384†‡¶*	+12.81%	0.4318†‡¶*	+2.14%	0.3582†‡¶*	+3.91%
EDRM-KNRM	0.3096†‡¶*	+17.04%	0.4547†‡¶*	+8.32%	0.3327†‡¶*	+10.92%	0.4341†‡¶*	+2.68%	0.3616†‡¶*	+4.90%
EDRM-CKNRM	0.3397†‡¶*	+28.42%	0.4821†‡¶*	+14.86%	0.3708†‡¶*	+23.60%	0.4513†‡¶*	+6.74%	0.3892†‡¶*	+12.90%



接下来我们验证了实体在检索模型中的作用（如图3和图4所示）。首先从图3可以看出，SoftMatch所占比例远远高于Exact Match，这与K-NRM以及Conv-KNRM的结果一样，证明相比较传统的Exact Match来讲Soft Match更加重要；接下来我们可以看到cross-space以及entity-word之间的交互占了较大比重，从而证明了crossmatch 以及word-entity level match 的有效性。

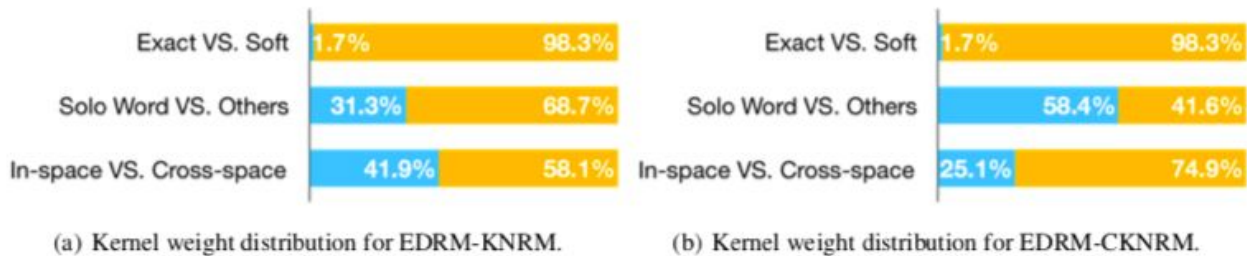


Figure 3: Ranking contribution for EDRM. Three scenarios are presented: Exact VS. Soft compares the weights of exact match kernel and others; Solo Word VS. Others shows the proportion of only text based matches; In-space VS. Cross-space compares in-space matches and cross-space matches.

为了进一步探究词语的n-gram和entity之间的交互比重，我们得到了图4所示的结果，从结果可以看到权重基本都是均匀分布，从而进一步证明了n-gram以及加入entity的必要性。

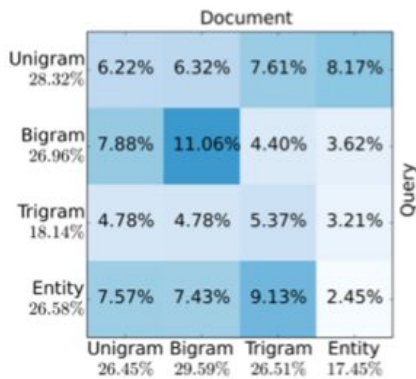


Figure 4: Individual kernel weight for EDMM-CKNRM. X-axis and y-axis denote document and query respectively.

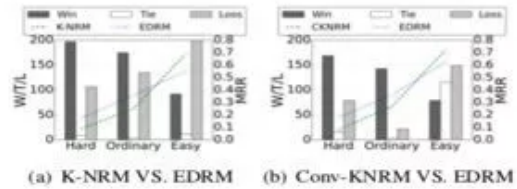


Figure 5: Performance VS. Query Difficulty. The x-axes mark three query difficulty levels. The y-axes are the Win/Tie/Loss (left) and MRR (right) in the corresponding group.

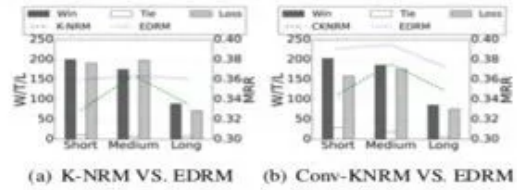


Figure 6: Performance VS. Query Length. The x-axes mark three query length levels, and y-axes are the Win/Tie/Loss (left) and MRR (right) in the corresponding group.

模型的提升主要是来源他的泛化能力，因此我们进一步分析其原因。首先我们做了消融实验，如表2所示。从表中可以看出，实体的描述对于增强模型的泛化能力有很大帮助，加入三种语义元素后对于模型的效果从不同方便有了一定提升。但是相对于Conv-KNRM来说，仅仅利用实体信息（EntityEmbedding）对于模型效果提升很不明显，我们认为，只考虑实体仅仅是一种更精确的 n-gram model，在这里Conv-KNRM已经学的很好。

Table 2: Ranking accuracy of adding diverse semantics based on K-NRM and Conv-KNRM. Relative performances compared are in percentages. †, ‡, §, ¶, *, ** indicate statistically significant improvements over K-NRM† (or Conv-KNRM†), +Embed‡, +Type§, +Description¶, +Embed+Type* and +Embed+Description** respectively.

Method	Testing-SAME				Testing-DIFF				Testing-RAW	
	NDCG@1		NDCG@10		NDCG@1		NDCG@10		MRR	
K-NRM	0.2615	-	0.4197	-	0.3000	-	0.4228	-	0.3447	-
+Embed	0.2743	+3.68%	0.4296	+2.35%	0.3134	+4.48%	0.4306	+1.86%	0.3641†	+5.62%
+Type	0.2709	+2.41%	0.4395†	+4.71%	0.3126	+4.20%	0.4373†	+3.43%	0.3531	+2.43%
+Description	0.2827	+6.86%	0.4364†	+3.97%	0.3181	+6.04%	0.4306	+1.86%	0.3691 †§*	+7.06%
+Embed+Type	0.2924†	+10.52%	0.4533†§¶	+8.00%	0.3034	+1.13%	0.4297	+1.65%	0.3544	+2.79%
+Embed+Description	0.2891	+9.29%	0.4443††	+5.85%	0.3197	+6.57%	0.4304	+1.80%	0.3564	+3.38%
Full Model	0.3096 †§§	+17.04%	0.4547 †§§¶	+8.32%	0.3327 †*	+10.92%	0.4341 †	+2.68%	0.3616†	+4.90%
Conv-KNRM	0.3357	-	0.4810	-	0.3384	-	0.4318	-	0.3582	-
+Embed	0.3382	+0.74%	0.4831	+0.44%	0.3450	+1.94%	0.4413	+2.20%	0.3758†	+4.91%
+Type	0.3370	+0.38%	0.4762	-0.99%	0.3422	+1.12%	0.4423†	+2.42%	0.3798†	+6.02%
+Description	0.3396	+1.15%	0.4807	-0.05%	0.3533	+4.41%	0.4468†	+3.47%	0.3819†	+6.61%
+Embed+Type	0.3420	+1.88%	0.4828	+0.39%	0.3546	+4.79%	0.4491†	+4.00%	0.3805†	+6.22%
+Embed+Description	0.3382	+0.73%	0.4805	-0.09%	0.3608	+6.60%	0.4494†	+4.08%	0.3868†	+7.99%
Full Model	0.3397	+1.19%	0.4821	+0.24%	0.3708 †§§	+9.57%	0.4513 ††	+4.51%	0.3892 ††	+8.65%

为了进一步探究实体对于检索模型的影响，我们测试了不同场景下的模型的表现（如图6），并发现模型在基线模型比较难做对的问题以及短的问题上，效果更加明显，这进一步说明了引入实体的重要性。



最后我们做了样例分析如表3所示。首先，实体描述能够解释实体的含义。例如“美图秀秀网络版”和“美丽说”是两个分别提供图像处理和购物服务的网站。他们的描述提供了额外的检索信息，从而使得文本的表述得以增强。其次，实体类型可以在问题和文档之间建立深层联系。例如，不同问题中实体“蜡笔小新”和“银魂”是共享相同的实体类型；在问题和文档中，也是有这种隐藏的关联，比如：“鲁大师”和“系统优化”。

Table 3: Examples of entity semantics connecting query and title. All the examples are correctly ranked by EDRM-CKNRM. Table 3a shows query-document pairs. Table 3b lists the related entity semantics that include useful information to match the query-document pair. The examples and related semantics are picked by manually examining the ranking changes between different variances of EDRM-CKNRM.

(a) Query and document examples. <i>Entities</i> are emphasized.	
Query	Document
<i>Meituxiuxiu web version</i>	<i>Meituxiuxiu web version</i> : An online picture processing tools
Home page of <i>Meilishuo</i>	Home page of <i>Meilishuo</i> - Only the correct popular fashion
<i>Master Lu</i>	Master Lu official website: <i>System optimization</i> , hardware test, phone evaluation
<i>Crayon Shin-chan</i> : The movie	<i>Crayon Shin-chan</i> : The movie online - Anime
<i>GINTAMA</i>	<i>GINTAMA</i> : The movie online - Anime - Full HD online watch

(b) Semantics of related entities. The first two rows and last two rows show entity descriptions and entity types respectively.

Entity	Content
<i>Meituxiuxiu web version</i>	Description: Meituxiuxiu is the most popular Chinese image processing software, launched by the Meitu company
<i>Meilishuo</i>	Description: Meilishuo, the largest women's fashion e-commerce platform, dedicates to provide the most popular fashion shopping experience
<i>Crayon Shin-chan</i> , <i>GINTAMA</i>	Type: Anime; Cartoon characters; Comic
<i>Master Lu</i> , <i>System Optimization</i>	Type: Hardware test; Software; System tool

本文首次将知识融合到信息检索模型，证明了模型的效果以及泛化能力，进一步的验证了实体的重要性以及模型取得泛化能力的原因。



留言 点赞 发个朋友圈
我们一起探讨AI落地的最后一公里

推荐文章阅读

ACL2017 论文集：34篇解读干货全在这里