arXiv:2301.06974v1 [cs.IR] 17 Jan 2023

利用知识图谱提升基于文本的信息检索可解释性

Boqi Chen*麦吉尔大学 加拿大魁北克省蒙特利尔 boqi.chen@mail.mcgill.ca

Kua Chen*麦吉尔大学 加拿大魁北克省蒙特利尔 kua.chen@mail.mcgill.ca

Yujing Yang*麦吉尔大学 加拿大魁北克省蒙特利尔 yujing.yang2@mail.mcgill.ca

Afshin Amini Aggregate Intellect 加拿大安大略省多伦多 afshin.emd@gmail.com

Bharat Saxena BMC Software Inc 印度马哈拉施特拉邦浦那 bharat_saxena@bmc.com

Cecilia Chávez-García Aggregate Intellect 伦敦，安大略省，加拿大 cecilia.uku@gmail.com

Majid Babaei 麦吉尔大学 蒙特利尔，魁北克省，加拿大 majid.babaei@mcgill.ca

Amir Feizpour Aggregate Intellect 蒙特利尔，魁北克省，加拿大 amir.fzpr@gmail.com

Dániel Varró 麦吉尔大学 蒙特利尔，魁北克省，加拿大 daniel.varro@mcgil.ca

摘要

得益于机器学习的最新进展，基于向量的方法已被应用于许多现代信息检索（IR）系统。这些方法虽然展现出良好的检索性能，却通常无法解释为何某个特定文档会作为查询结果被检索出来，这正是可解释信息检索（XIR）旨在解决的问题。知识图谱记录了有关实体的结构化信息以及天然可解释的关系。现有的大多数可解释信息检索方法都只关注检索模型，很少考虑如何利用现有知识图谱提供解释。本文提出了一种通用架构，可在检索过程的多个环节中引入知识图谱以实现可解释信息检索。此外，我们针对不同类型的解释，构建了该架构的两个实例。我们使用标准指标在知名信息检索基准上评估所提方法，并将其与作为基线方法的基于向量的方法进行比较。

CCS 概念

• **信息系统** → **信息检索**；• **计算方法** → **知识表示与推理**。

关键词

可解释的信息检索、知识图谱、实体链接、自然语言处理

*三位作者对本研究的贡献相同。

允许免费制作本作品的全部或部分内容的数字或纸质副本，用于个人或课堂用途，前提是这些副本不用于或分发以谋取利润或商业利益，并且每份副本首页均附有本声明和完整引文。对于本作品中由 ACM 以外的其他方拥有版权的组成部分，必须尊重其版权。注明出处后允许摘录。若要以其他方式复制、重新发表、发布到服务器或分发给他人，则须事先获得明确许可和/或支付费用。请通过 permissions@acm.org 申请许可。*TrustLOG-CIKM '22*，*2022*年*10*月*21*日，美国佐治亚州亚特兰大© 2022 计算机协会。ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00 https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM 参考格式：

Boqi Chen、Kua Chen、Yujing Yang、Afshin Amini、Bharat Saxena、Cecilia Chávez-García、Majid Babaei、Amir Feizpour 和 Dániel Varró。2022。《借助知识图谱提升基于文本的信息检索的可解释性》。载于 首届图上可信学习研讨会，与第*31*届 *ACM* 信息与知识管理国际会议联合举办（*TrustLOG-CIKM '22*），*2022*年*10*月*21*日，美国佐治亚州亚特兰大。ACM，美国纽约州纽约市，7页。https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 引言

动机。 给定自然语言查询作为输入，基于文本的信息检索（IR）旨在获取最相关的文档。现代信息检索系统通常利用文本嵌入，将文本映射为向量，使向量之间的距离能够表示文本的相似度。

在现实世界的信息检索应用中，如果检索结果附有额外解释，用户就能做出更好的决策。例如，当用户查询是什么因素导致心脏病时，解释原因有助于用户判断结果的相关性。这是一种局部 解释，说明文档 d 为何与用户查询 q 相关。一般而言，局部 解释应当满足以下条件：(1) 可解释 ，即人类能够理解；(2) 准确 ，即能够准确回答问题。

常见的解释可能会突出显示文档 [5] 中最相关的部分，或提供特征重要性分数 [25]。近期方法还会利用可解释特征，对基于向量的信息检索系统 [23]生成的候选结果进行重新排序。然而，现有方法很少明确纳入文本中的结构化信息，而这可能是解释的重要来源。

问题陈述。 本文聚焦于可解释信息检索 （XIR）任务，旨在提高检索算法的可解释性，即让检索结果更易于理解，并反映算法的内部行为。我们旨在探究如何利用具有显式语义关系的知识图谱作为外部信息源来提升 XIR。

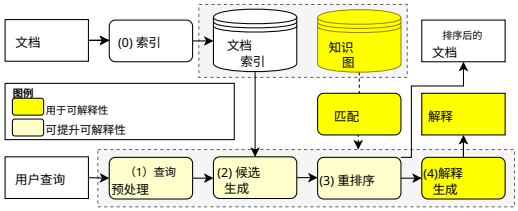


图1：使用知识图谱进行解释的总体架构

知识图谱（KG）是一种有向标记图，其中节点表示实体，边刻画节点之间的关系。已有研究表明，知识图谱有助于提升信息检索系统的性能 [27]，并为其他机器学习算法生成解释 [35]。然而，知识图谱在可解释信息检索中的应用仍局限于特定领域的使用场景 [1, 47]。

知识图谱在实体之间提供了明确的语义关系，这些关系可以作为解释，或帮助阐明查询的意图。通常，这类语义关系也会出现在文本中。例如，句子“肥胖会导致心脏病”体现了两个实体 肥胖 和 心脏病，它们通过 导致 关系相连。

实体和关系的链接，即识别并连接知识图谱与一段文本中的匹配部分，是一个众所周知的挑战 [8, 21]。虽然现有的实体链接工具可以作为起点，但利用知识图谱提高信息检索的可解释性还需要：（1）一个通用的集成架构，以及（2）一套详细方案，说明如何利用知识图谱信息来改进特定类型的解释。

目标与贡献。给定一组自然语言文档和一个查询，本文提出了一种通用架构，将知识图谱集成到基于向量的 信息检索系统中，为检索结果生成局部 解释。我们使用各种实体/关系链接方法，将知识图谱中最相关的部分与文本进行匹配，然后利用这些信息提高信息检索系统的可解释性。总而言之，具体贡献如下：

- 我们提出了一种通用架构，将信息检索划分为不同阶段，并着重说明如何在每个阶段利用知识图谱来改进解释。
- 我们通过整合开放知识图谱 Wikidata 和多种实体链接方法，以两种方式实例化该通用架构，从而实现可解释的信息检索。
- 我们在两个数据集 WIKIQA [48] 和 Robust04 [19]上，对两种特定的解释方法进行了初步实验评估。假设实体匹配器达到理想效果，我们的方法将识别最重要句子的性能提高了 6.96%，并将基础检索性能提高了 0.45%。

2 方法

2.1 通用架构

图 1 展示了由知识图谱增强的可解释信息检索系统的总体架构。在预处理阶段，（0）文档会被存储并整理到文档索引中（假设文档以向量化形式存储）。随后，（1）我们的框架接收自然语言形式的用户查询作为输入，并对查询进行预处理（例如通过查询扩展 [3]）。接下来，（2）候选生成会检索出与处理后的查询最相关的排序后的候选文档列表。（3）该候选文档列表可能会通过重排序重新排列。最后，（4）我们的框架输出排序后的文档以及解释生成的结果。

我们的架构将如何利用知识图谱中的额外知识，通过实体/关系**匹配**融入信息检索的步骤 (1)–(4)，从而逐步提升可解释性。

- **实体/关系匹配：**通过实体/关系匹配，可以将文本与知识库中的对应实体连接起来 [31]。此步骤首先使用实体识别来检测查询中的关键词和实体，例如专有名称 [18]。这些实体会与知识图谱中的节点进行匹配。随后，通过关系抽取识别语料数据的实例与概念之间的关系，从而将识别出的实体转换为结构化数据 [15]。
- **面向可解释性的查询预处理：**可以利用从知识图谱中检索到的信息扩展用户查询。知识图谱通常既包含实体信息（如描述、别名和属性），也包含关系信息（如相连的实体和关系类型）。可以将这些额外的知识添加到查询中，以改进查询理解。
- **可解释候选生成：**可解释候选生成可以使用直接从文档本身构建的知识图谱。该知识图谱由从文档中识别出的实体和关系构成，其中每个实体可能与一个或多个文档相关联。这样，候选生成就转化为一个知识图谱检索问题，可以通过精确或模糊实体匹配来完成。
- **可解释重排序：**知识图谱还可用于计算重排序所需的可解释特征。例如，可以先将查询和候选文档分别转换为一组实体。然后，根据这些实体在知识图谱中的距离和连通性，计算实体对之间的相关性。再将此类分数汇总为查询-文档相关性，用于重排序。
- **解释生成：**步骤 (1)-(3) 中的信息可用于生成更完善的解释，呈现给用户。步骤 (1) 中的扩展查询可用于找出检索到的文档中最相关的部分，并解释查询与文档之间的关联原因。步骤 (2) 中使用的知识图谱可提供查询与文档之间匹配实体的数量。最后，步骤 (3) 中计算的特征可为文档提供内在可解释的排序。

2.2 用于最重要句子的知识图谱

最重要句子（MIS）是文本片段中最相关的部分，说明了该文本片段与查询之间的关系。MIS 能吸引用户的注意力，并强调与查询相关的主要内容，使用户无需阅读整个文本片段，就能理解为何检索到该文本片段。形式上，设 q 为查询，S 为来自文档的一组句子，$sim(\cdot,\cdot)$ 为两个文本片段之间的某种相似度函数。

This tablespoon has a capacity of about 15 mL. Measuring Spoons. In the US and parts of Canada, a tablespoon is the largest type of spoon used for eating from a bowl. In the UK, Europe and most Commonwealth countries, a tablespoon is a type of large spoon usually used for serving. In countries where a tablespoon is a serving spoon, the nearest equivalent to the US tablespoon is either the dessert spoon or the soup spoon. A tablespoonful, nominally the capacity of one tablespoon, is commonly used as a measure of volume in cooking. It is abbreviated as T, tb, tbs, tbsp, tblsp, or tblspn. The capacity of ordinary tablespoons is not regulated by law and is subject to considerable variation. In the USA one tablespoon (measurement unit) is approximately 15 mL; the capacity of an actual tablespoon (dining utensil) ranges from 7 mL to 14 mL. In Australia one tablespoon (measurement unit) is 20 mL.

图2：MIS 示例

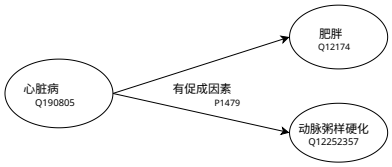


图3：实体和关系示例

最重要句子定义为

$$MIS = \arg\max_{s \in S} sim(q, s) \tag{1}$$

一种广泛用于衡量两段文本之间相似度的方法是嵌入。我们在本文的方法中采用这种方法，将文本嵌入为一个向量。随后，可以使用两个向量之间的标准相似度度量（例如余弦相似度或点积）来近似公式 1 中的相似度函数 sim。我们在本文的方法中选择了余弦相似度。

示例 1。图2展示了在一个段落中高亮 MIS 的示例。该段落介绍的是汤匙，MIS 包含汤匙的定义和容量。

一种简单的方法是使用句子检索器，通过计算查询嵌入与文本片段中每个句子的嵌入之间的余弦距离，来识别 MIS [33]。然而，这种方法通常会遗漏有关查询实体的一些语义信息。例如，“bank”一词既可能指金融机构，也可能指河岸。为避免查询中的语义歧义，我们使用参考知识图谱，通过实体链接和关系抽取来扩展查询。

我们扩展了查询预处理，以便更好地识别 MIS，同时保持架构的其他部分不变。具体而言，我们利用知识图谱中链接实体的额外信息，以三种不同的方式进行查询扩展。

(A) 如果查询包含具有直接关系的实体，则将链接实体视为潜在信息，并将其附加到查询中（见图3）。(B) 如果查询仅包含一个实体，则使用该实体的描述进行查询扩展。(C) 如果查询包含多个实体，但不包含任何关系，则仅使用实体进行查询扩展。

在情况 (A) 中，可以通过知识图谱导航，在知识图谱中查找与查询相关的信息。然而，许多查询不包含任何可用于知识图谱导航的关系。为了在这种情况下仍能利用知识图谱，我们在情况 (B) 和 (C) 中使用知识图谱中的实体信息。

扩展查询随后被（1）用于候选生成，以对候选项进行排序；（2）用于解释生成，以突出显示 MIS。

示例 2。图3展示了用于识别 MIS 的实体链接和关系抽取。查询为心脏病的病因，抽取出的实体和关系分别是心脏病 和具有促成因素。因此，链接得到的（唯一）实体是肥胖和动脉粥样硬化，它们由此用于查询扩展。

2.3 用于可解释重排序的知识图谱

用户研究表明，基于可解释特征的重排序可以提高排序结果的可解释性 [23]，因为用户只需观察特征值即可深入了解文本片段。这类可解释特征的例子包括词项统计、上下文词语和基于引用的流行度 [23]。

我们的框架融入了沿着可解释特征对文档进行重排序的理念，以便更好地解释文档的排序。我们提出了一种新的可解释特征，用于根据知识图谱中的实体匹配对候选文档进行重排序。

实体匹配通过以下方式确定两个文本片段之间的关联：(1) 从两个文本片段中提取实体 [37],，以及 (2) 计算两类实体之间实体对之间的相关性 [8]。[8] 的作者提出了一种相关性度量 R，其依据是维基百科知识图谱等图中两个实体之间入链的重叠情况。给定一对短语，该算法会在知识图谱中找到对应的实体 P_a 和 P_b，然后按照 [8] 中的建议，计算 P_a 与 P_b 之间的相关性，如下所示：

$$R(P_a, P_b) = \frac{log(max(|in(P_a), in(P_b)|)) - log(|in(P_a) \cap in(P_b)|)}{log(W) - log(min(|in(P_a), in(P_b)|))} \tag{2}$$

其中，$in(P_a)$表示实体 P_a 的入边集合，W 表示知识图谱中节点的总数。

为了计算查询与文档之间的实体相关性得分，我们先对文档中的实体相关性值取平均，再对查询中的实体相关性值求和。具体而言，查询-文档相关性得分定义如下：

$$QDR = \sum_{i=0}^{i=n} \frac{\sum_{j=0}^{j=m} R(E_{q_i}, E_{d_j})}{m} \tag{3}$$

其中，R 表示两个实体之间的相关性得分函数 [8]，E_q 表示查询中的一组 n 实体，E_d 表示文档中的一组 m 实体。此值可用作可解释特征，对候选生成提出的文档进行重新排序。

尽管这种方法忽略了句子中的结构信息（例如实体的顺序），但可以针对每个实体单独并行高效地进行计算。我们将探索句子的结构信息作为未来工作。

3 评估

本节将探讨以下研究问题：

- RQ1：知识图谱如何影响突出显示最重要句子的效果？
- RQ2：通过基于可解释特征进行重排序，知识图谱如何影响检索性能？

3.1 实验设置

参考知识图谱。维基数据[40] 是一个开放的知识库，包含来自多个领域的 99,959,412 个数据项。我们选择维基数据作为参考知识图谱，因为它常用于链接预测、知识图谱检索和推荐等任务 [10], [41] [2]。在实验中，我们首先使用实体链接器从查询中识别实体和关系，然后从匹配到的实体和关系出发遍历知识图谱（使用 *qwikidata* 软件包 [34]）。与此同时，我们还使用维基百科页面图 [42] 来确定两个实体之间的相关性。

评估数据集。为了评估识别 *MIS*的效果，我们选择了 WIKIQA [48], ，这是信息检索领域广泛使用的热门多领域问答数据集和基准。与大多数其他信息检索数据集不同，该数据集中的句子被标注为相关查询的答案。我们选择该数据集主要有两个原因：（1）高质量答案取自维基百科页面摘要；（2）回答查询的句子均有标注，可用于评估句子级检索的效果。

为了评估使用可解释指标对文档进行重排序的效果，我们使用 TREC 2004 Robust（Robust04） [19], 测试集，该数据集是评估基于文本的信息检索效果的热门基准。Robust04 包含 249 个查询和 528,000 篇新闻语料文档。每个查询都对应多个相关文档，用于评估文档排序。此外，该数据集为查询分配了不同的文档相关性等级，因此适合用于评估排序效果。

比较的方法。 为了展示来自知识图谱的额外知识所带来的益处，我们在所有比较方法中均采用搭载预训练句子转换器模型 [9, 26] 的预训练嵌入检索器。我们使用 Haystack [22] 为所有比较方法构建信息检索流程，并使用 FAISS [14] 基于余弦相似度快速进行向量排序。

在识别 *MIS*的情况下，我们首先将检索到的文档拆分为单独的句子。搭载预训练嵌入模型的检索器会根据句子与查询之间的余弦相似度对句子进行排序。最后，将相似度得分最高的句子视为 MIS（见公式 1），并将其与真实标注句子进行比较，以评估性能。与基线方法（不进行查询扩展）不同，我们的方法利用知识图谱中的信息扩展查询。

在可解释重排序的情况下，基线方法也使用嵌入检索器，根据余弦相似度对文档进行排序。基线方法直接采用嵌入检索器给出的排序，不会对文档重新排序。我们的方法则使用同一检索器生成候选文档，再根据第2.3节中的查询-文档相关性 对其重新排序。

评估指标。为了评估这些方法的性能，我们采用了信息检索社区广泛使用的若干指标。具体而言，**准确率** 用于评估识别 MIS 的性能。为了评估重排序的性能，我们使用精确率（**P**）和**召回率** 来衡量检索文档的质量。然而，由于这些指标未考虑排序，我们使用前k项平均精度均值（**MAP@k**）和前k项归一化折损累计增益（**NDCG@k**）来评估前k 个文档的排序质量。

	段落检索器	句子检索器
基线	97.53	66.67
我们的方法 1	97.12	62.71
我们的方法 2	**97.94**	**71.31**

表 1：句子检索的准确率（以 % 计）

3.2 RQ1：MIS 作为解释

在此 RQ 中，我们旨在评估将知识图谱中的额外知识纳入查询后，对高亮显示 MIS 的效果所产生的影响。

理由

设置。 我们使用 WIKIQA 数据集中的查询和维基数据知识库进行查询扩展。基线方法使用原始查询检索最相关的文本片段，然后在该片段中检索 MIS。我们的方法使用扩展后的查询执行相同的检索过程。评估时，我们计算两种方法检索排名最高的 MIS 的准确率。我们使用包含 243 个查询和 1081 篇文档的测试集，衡量识别 MIS 的效果。

我们衡量查询扩展中两种方法变体的效果：（1）使用现有的实体匹配工具（**我们的 1**）；（2）使用理想的实体匹配方法（**我们的 2**）。两种情况下，我们使用相同的参考知识图谱。在 **我们的 1** 中，我们使用 WAT [21] 识别查询中的实体，并将其链接到维基数据条目。然而，这一自动化过程可能会出错，也未必能体现知识图谱额外信息带来的益处。为了展示知识图谱的潜在益处，在 **我们的 2** 中，我们人工检查 WAT 的结果，识别具有关系的实体，并将它们链接到正确的实体。在这种方法中，我们假设查询扩展所用的实体匹配器具有理想的性能，并测试句子检索的效果。

讨论。 在表 1 中，文本片段检索器 一栏展示了原始信息检索任务的准确率，即找出与查询最相关文档的准确率。句子检索器 则展示了识别 MIS 的准确率。

对于三种方法，文本片段检索器的准确率都很高（超过 98。00%）。**我们的方法 1** 的准确率比基线方法低 0。77%，而**我们的方法 2** 的准确率比基线方法高 0。24%。句子检索器的表现也呈现出类似趋势。**我们的方法 1** 的准确率比基线方法低 1。79%，而**我们的方法 1**的结果比基线方法高 7。29%。

深入研究查询扩展后，我们发现，**ours 1**中使用的外部实体匹配工具对某些查询的匹配并不准确。在这种情况下，查询会被误导性信息扩展，因而无法选出最佳的 MIS，导致性能下降。然而，使用理想的实体匹配器（**ours 2**）后，识别 MIS 的性能显著提升。

	P	召回率	MAP@20	NDCG@20
基线	31.09	15.71	10.38	20.87
ours	31.09	15.71	8.13	17.89

表2：重排序评估结果（%）

研究问题1：我们发现，高亮 *MIS* 所带来的性能提升取决于实体匹配的性能以及知识图谱的相关性。使用理想的实体匹配器时，知识图谱提供的附加信息能够显著提高句子检索的准确率，而无需修改检索模型。

3.3 RQ2：可解释重排序

我们将实体匹配和查询-文档相关性作为可解释特征来重新排序文档，然后将重排序结果与基线方法进行比较。虽然这些特征可以提高排序的可解释性，但我们旨在评估它们对排序性能的影响。

理由

实验设置。本部分评估使用了 Robust04 数据集。在基线方法中，使用嵌入检索器根据原始查询与数据集中文档之间的余弦相似度检索并排序文档。我们的方法（**我们的方法**）使用相同的嵌入检索器检索文档，但根据查询-文档相关性得分对文档进行排序（见公式 3）。具体而言，该方法使用 REL [37] 进行实体链接，并使用 TagMe [8] 计算实体相关性分数，再结合这两种技术计算查询-文档实体相关性得分，以便对文档排序。排序后的文档的质量由 TERC 评估工具 [36]衡量。

讨论。如表2所示，两种方法的**P** 和**召回率** 相同，因为重排序不会从候选生成中排除现有候选项。此外，这两个指标并不能衡量排序性能。虽然我们的方法提高了结果的可解释性（这得益于从知识图谱中得出的指标），但其**MAP@20** 为8.13%，**NDCG@20**为17.89%，略低于基线方法（分别低2.25%和2.98%）。

为了进一步探究性能下降的原因，我们比较了两种方法的结果排序，发现采用新方法后，只有17.95%的查询获得了更好的排序结果，而大多数查询的排序保持不变。在考察查询的实体链接情况后，我们发现实体链接的性能同*RQ1*中一样有限。这一发现表明，更好的实体链接也有望提升重排序性能。

研究问题2：我们发现，知识图谱中的可解释特征虽然提高了排序的可解释性，却可能降低检索性能。我们怀疑，这种下降源于实体链接器的不完善。此外，我们的发现与其他研究 *[23]* 一致，即可解释重排序会降低排序性能。

3.4 效度威胁

内部效度。嵌入模型的训练通常具有非确定性。为了尽量减少随机性的影响，我们使用了预训练嵌入模型 [9, 26]，并在整个实验过程中保持不变。

外部效度。在所有实验中，我们都使用维基数据作为参考知识图谱。显然，使用不同领域的知识图谱可能会得到不同的改进。在*RQ1*实验中，我们使用现有的实体匹配器 [21, 37] 测量性能，并假设实体匹配器是理想的。实际应用中的性能可能介于两者之间。同样，在*RQ2*实验中，我们利用知识图谱中的实体关联性对文档重新排序；若使用知识图谱中的不同可解释特征，也可能得到不同的结果。

构念效度。我们使用经典信息检索指标，在广泛使用的信息检索基准上评估所比较方法的性能，这符合信息检索评估的最佳实践。

4 相关工作

4.1 可解释人工智能

一般而言，可解释人工智能方法可分为两类：事前 方法和事后 方法（即在模型训练完成后使用外部解释器） [11, 39]。后者又可分为两类：模型特定 方法和模型无关方法。

事前方法会构建透明模型，因此可以直接从模型中生成解释，例如基于树的模型 [13, 29] ，或扩展神经网络的架构以便进行解释 [6, 28, 30]。此外，也有许多关于用于解释的事后方法的综述 [11, 38, 50]。

特定于模型的方法通过利用模型内部表示和学习过程获取解释，因此仅适用于特定模型，并且可以直接根据模型输出生成解释。然而，大多数方法只适用于简单且可解释的模型，例如线性模型、最近邻和基于树的模型 [11]。相比之下，模型无关方法只分析输入输出对，不考察模型内部，从而提高可解释性。一种流行的模型无关方法是*LIME* [49] ，它旨在以可解释且忠实的方式解释任意分类器的预测。另一种流行的方法是*SHAP* [17] ，它基于Shapley回归值，这是一种用于计算特征重要性的方法，并能在需要对所有特征子集训练模型的情况因计算不可行时进行快速近似。然而，这些方法存在全局解释效果不佳以及相关特征等问题 [12]。

本文提出了一个通用框架，用于为基于模型的事后 嵌入的信息检索模型生成解释。

4.2 可解释信息检索

信息检索中的可解释性可根据结果形式分为：基于高亮、基于特征重要性、基于规则以及混合方法 [11]。

基于高亮的解释的一个著名例子是谷歌搜索，它会在搜索结果旁显示内容片段和关键词。类似地，Chios 和 Verberne [5]提出为深度信息检索模型高亮显示最重要的片段。基于特征重要性的解释方法可以根据查询和文档计算分数 [5, 23, 25]。一些方法

会根据查询中的词项重要性进行操作，另一些则侧重于使用可解释特征对文档进行重新排序 [23]。基于规则的解释旨在通过提供一个简化模型来解释决策，从而回答有关信息检索模型内部机制的特定问题 [32] 。最后，混合式解释结合所有方法，以提供全面的可解释性 [5, 23]。

虽然可解释信息检索大多采用事后方法，但也存在一些使用可解释嵌入对文档进行排序的方法 [20, 24]，使得向量空间中的距离能够解释查询与文档之间的相关性。

与这些方法相比，我们的通用框架能够为具有外部知识图谱的信息检索系统生成多种形式的解释。此外，我们提出了该框架的一种具体实现，用于突出最重要句子并进行可解释重排序。

4.3 用于信息检索的知识图谱

知识图谱中的实体可用于信息检索系统，以帮助理解用户的意图、查询和文档，超越仅依靠词元所能达到的效果 [27]。Reinanda 等人 [27] 总结了如何利用知识图谱中的面向实体的信息的相关方法：基于扩展的方法、潜在因子建模、语言建模和深度学习方法。与此同时，知识图谱也被用于一些特定领域应用，以解释信息检索系统。

基于扩展的方法 通过扩展查询和/或文档，在检索过程中丰富面向实体的信息 [4], 例如，使用同义词扩展查询。

潜在因子建模 试图在查询和文档中找到概念。 [44] 的作者提出了一种利用外部数据和知识库改进排序的新技术。该技术将外部对象视为查询与文档之间的潜在层，以学习判断文档相关性。

语言建模 方法在构建查询和文档的语言模型时考虑语义信息。例如，Ensan 等人 [7] 提出了一种文档检索器，利用语义链接系统构建文档和查询的图表示，其中节点表示文档中的概念，边表示概念之间的语义相关性。

用于可解释信息检索的知识图谱 深度学习方法在神经排序系统中使用知识图谱嵌入。示例包括将查询和文档嵌入实体空间 [46] ，以及构建查询与实体表示之间的交互矩阵。 [16, 45]

用于可解释信息检索的知识图谱 Hasan 等人 [1] 提出了一个框架，通过融入领域知识图谱，为特定领域的信息检索应用生成解释。 [47] 的作者借助注意力分数构建知识感知路径，应用知识图谱来解释信息检索模型。同样， [43] 的作者设计了一个可解释推荐系统，其中包含利用知识图谱进行决策的显式推理，从而使推荐具有可解释性。

在这项工作中，我们提出了一个通用框架，利用知识图谱提升信息检索系统的可解释性 。

5 结论与未来工作

本文探讨可解释信息检索（XIR）任务，旨在解释为何某个文档与某个查询相关。现有方法采用高亮和特征重要性，但较少关注利用外部知识库生成解释。

我们提出了一种通用架构，利用来自知识图谱的语义信息来提升信息检索系统的可解释性。我们利用现有的实体和关系匹配方法，识别知识图谱中与文本片段最相关的部分，并通过导航知识图谱，帮助解释信息检索各个处理环节产生的检索结果。我们通过两个信息检索示例展示了知识图谱的有效性：高亮最重要句子，以及利用可解释特征进行重排序。

我们使用多个指标对所提出的方法进行了初步实验评估，主要发现是，该方法的性能在很大程度上取决于实体匹配的质量。当实体链接器出现错误时，其性能可能低于基线方法。然而，在理想实体匹配器的情况下，我们的方法使最重要句子识别性能提升了 6.96%，并使基础检索性能提升了 0.45%。

我们认为，这种通用架构为利用知识图谱提升信息检索的可解释性开辟了许多方向。未来，我们计划将该通用架构应用于以下方面：(1) 将该架构与更先进的实体/关系匹配方法相结合；(2) 采用各种知识图谱检索方法来提升候选生成的可解释性；(3) 将知识图谱与嵌入模型相结合，创建可解释的文本嵌入；(4) 捕捉文档和查询中的结构信息；以及 (5) 通过用户研究，从用户视角评估我们的方法如何影响信息检索系统的可解释性。实验所用的代码和相关资源可在 [1]中找到

6 致谢

本论文部分由 NSERC RGPIN-2022-04357 项目和 FRQNT-B2X 项目（档案编号：319955）资助。本研究是麦吉尔大学 2022 年工程本科生暑期研究项目（ECSE-025）的一部分。

参考文献

[1] Hasan Abu-Rasheed、Christian Weber、Johannes Zenkert、Mareike Dornhöfer 和 Madjid Fathi。2022。基于知识图谱的可迁移框架：用于在特定领域的智能信息检索中生成可解释结果。载于 *Informatics*，第 9 卷。MDPI，6。[2] Kholoud AlGhamdi、Miaojing Shi 和 Elena Simperl。2021。学习向维基数据编辑者推荐项目。载于 *International Semantic Web Conference*。Springer，163–181。[3] Hiteshwar Kumar Azad 和 Akshay Deepak。2019。信息检索的查询扩展技术：综述。*Information Processing & Management* 56，5（2019），1698–1735。[4] Saeid Balaneshinkordan 和 Alexander Kotov。2016。用于查询扩展的术语关联与知识图谱的实证比较。载于 *European conference on information retrieval*。Springer，761–767。[5] Ioannis Chios 和 Suzan Verberne。2021。通过向深度相关性匹配模型添加可解释元素来帮助评估结果。*arXiv preprint arXiv:2106.05147*（2021）。[6] Orsolya Csiszár、Gábor Csiszár 和 József Dombi。2020。基于连续值逻辑和多准则决策算子的可解释神经网络。*Knowledge-Based Systems* 199（2020），105972。[7] Faezeh Ensan 和 Ebrahim Bagheri。2017。通过语义链接实现的文档检索模型。载于*Proceedings ofthe tenthACM internationalconference onweb search and data mining*。181–190。

[1]https://github.com/Aggregate-Intellect/xir

[8] Paolo Ferragina 和 Ugo Scaiella。2012。使用维基百科页面对短文本进行快速且准确的标注。IEEE Software 29，1（2012），70–75。[9] flax-sentence embeddings。2021。all_datasets_v3_mpnet-base。https://huggingface.co/flax-sentence-embeddings/all_datasets_v3_mpnet-base [10] Mikhail Galkin、Max Berrendorf 和 Charles Tapley Hoyt。2022。知识图谱归纳式链接预测的开放挑战。ArXiv abs/2203.01520（2022）。[11] Mir Riyanul Islam、Mobyen Uddin Ahmed、Shaibal Barua 和 Shahina Begum。2022。针对不同应用领域和任务的可解释人工智能：系统综述。Applied Sciences 12，3（2022），1353。[12] Sheikh Rabiul Islam、William Eberle、Sheikh Khaled Ghafoor 和 Mohiuddin Ahmed。2021。可解释人工智能方法：综述。arXiv preprint arXiv:2101.09429（2021）。[13] Sarah Itani、Fabian Lecron 和 Philippe Fortemps。2020。基于核密度估计的单类分类决策树。Applied Soft Computing 91（2020），106250。[14] Jeff Johnson、Matthijs Douze 和 Hervé Jégou。2019。使用图形处理器进行十亿规模的相似性搜索。IEEE Transactions on Big Data 7，3（2019），535–547。[15] Yankai Lin、Zhiyuan Liu、Maosong Sun、Yang Liu 和 Xuan Zhu。2015。学习用于知识图谱补全的实体和关系嵌入。载于第二十九届AAAI人工智能大会。[16] Zhenghao Liu、Chenyan Xiong、Maosong Sun 和 Zhiyuan Liu。2018。实体双重神经排序：理解知识图谱语义在神经信息检索中的作用。arXiv preprint arXiv:1805.07591（2018）。[17] Scott M Lundberg 和 Su-In Lee。2017。解释模型预测的统一方法。神经信息处理系统进展 30（2017）。[18] Alireza Mansouri、Lilly Suriani Affendey 和 Ali Mamat。2008。命名实体识别方法。International Journal of Computer Science and Network Security 8，2（2008），339–344。[19] NIST信息技术实验室（ITL）信息访问部门（IAD）的检索组。2022。TREC 2004 Robust赛道指南。https://trec.nist.gov/data/robust/04.guidelines.html [20] Abhishek Panigrahi、Harsha Vardhan Simhadri 和 Chiranjib Bhattacharyya。2019。Word2Sense：稀疏且可解释的词嵌入。载于计算语言学协会第57届年会论文集。5692–5705。[21] Francesco Piccinno 和 Paolo Ferragina。2014。从 TagME 到 WAT：一种新的实体标注器。载于第一届实体识别与消歧国际研讨会论文集（黄金海岸，昆士兰州，澳大利亚）（ERD '14）。计算机协会，美国纽约州纽约市，55–62。[22] Tanay；等 Pietsch、Malte；Soni。2022。Haystack。https://github.com/deepset- ai/haystack/ [23] Sayantan Polley、Atin Janki、Marcus Thiel、Juliane Hoebel-Mueller 和 Andreas Nuernberger。2021。Exdocs：基于证据的可解释文档搜索。载于 ACM SIGIR搜索与推荐因果关系研讨会。[24] M Atif Qureshi 和 Derek Greene。2019。Eve：使用维基百科的可解释向量嵌入技术。Journal of Intelligent Information Systems 53，1（2019），137–165。[25] Jerome Ramos 和 Carsten Eickhoff。2019。透明信息检索系统中的可解释性。[26] Nils Reimers 和 Iryna Gurevych。2019。Sentence-BERT：使用孪生BERT网络的句子嵌入。载于2019年自然语言处理经验方法会议论文集。计算语言学协会。[27] Ridho Reinanda、Edgar Meij、Maarten de Rijke 等。2020。知识图谱：信息检索视角。Foundations and Trends® in Information Retrieval 14，4（2020），289–444。[28] Isabel Rio-Torto、Kelwin Fernandes 和 Luís F Teixeira。2020。理解卷积神经网络的决策：模型内方法。Pattern Recognition Letters 133（2020），373–380。[29] Omer Sagi 和 Lior Rokach。2020。可解释决策森林：将决策森林转换为可解释树。Information Fusion 61（2020），124–138。[30] Anirban Sarkar、Deepak Vijaykeerthy、Anindya Sarkar 和 Vineeth N Balasubramanian。2022。通过概念学习事前可解释模型的框架。载于IEEE/CVF计算机视觉与模式识别会议论文集。10286–10295。[31] Wei Shen、Jianyong Wang 和 Jiawei Han。2014。基于知识库的实体链接：问题、技术与解决方案。IEEE Transactions on Knowledge and Data Engineering 27，2（2014），443–460。[32] Jaspreet Singh 和 Avishek Anand。2019。Exs：使用局部模型无关可解释性进行可解释搜索。载于第十二届ACM国际网络搜索与数据挖掘会议论文集。770–773。[33] Amit Singhal 等。2001。现代信息检索：简要概述。IEEE Data Eng. Bull. 24，4（2001），35–43。[34] Kensho Technologies。2019。qwikidata。https://github.com/kensho-technologies/ qwikidata。[35] Ilaria Tiddi 和 Stefan Schlobach。2022。作为可解释机器学习工具的知识图谱：综述。Artificial Intelligence 302（2022），103627。[36] Christophe Van Gysel 和 Maarten de Rijke。2018。Pytrec_eval：用于 trec_eval 的超快速 Python 接口。发表于 *SIGIR*。ACM。[37] Johannes M. van Hulst、Faegheh Hasibi、Koen Dercksen、Krisztian Balog 和 Arjen P. de Vries。2020。REL：站在巨人肩膀上的实体链接器。发表于 第*43*届 *ACM SIGIR* 信息检索研究与发展国际会议论文集（*SIGIR '20*）。ACM。[38] Sahil Verma、John Dickerson 和 Keegan Hines。2020。机器学习的反事实解释：综述。*arXiv* 预印本 *arXiv:2010.10596*（2020）。[39] Giulia Vilone 和 Luca Longo。2020。可解释人工智能：系统性综述。*https://arxiv.org/abs/2006.00093*（2020）。[40] Denny Vrandečić 和 Markus Krötzsch。2022。维基数据：主页。https: // www.wikidata.org/wiki/Wikidata:Main_Page [41] Andra Waagmeester、Gregory Stupp、Sebastian Burgstaller-Muehlbacher、Benjamin M Good、Malachi Griffith、Obi L Griffith、Kristina Hanspers、Henning Hermjakob、Toby S Hudson、Kevin Hybiske 等。2020。科学论坛：作为生命科学知识图谱的维基数据。*eLife* 9（2020），e52614。[42] 维基百科贡献者。2004。剽窃——维基百科，自由的百科全书。https://en.wikipedia.org/w/index.php?title= Plagiarism&oldid=5139350 [Online；访问于7月22日-2004].[43] Yikun Xian、Zuohui Fu、Shan Muthukrishnan、Gerard De Melo 和 Yongfeng Zhang。2019。用于可解释推荐的强化知识图谱推理。发表于 第*42*届 *ACM SIGIR* 信息检索研究与发展国际会议论文集。285–294。[44] Chenyan Xiong 和 Jamie Callan。2015。Esdrank：通过外部半结构化数据连接查询与文档。发表于 第*24*届 *ACM* 信息与知识管理国际会议论文集。951–960。[45] Chenyan Xiong、Jamie Callan 和 Tie-Yan Liu。2017。用于文档排序的词语—实体二重表示。发表于 第*40*届 *ACM SIGIR* 信息检索研究与发展国际会议论文集。763–772。[46] Chenyan Xiong、Russell Power 和 Jamie Callan。2017。通过知识图谱嵌入实现学术搜索的显式语义排序。发表于 第*26*届万维网国际会议论文集。1271–1279。[47] Zuoxi Yang。2020。结合知识图谱、用于可解释精准医学的生物医学信息检索。发表于 第*43*届 *ACM SIGIR* 信息检索研究与发展国际会议论文集。2486–2486。[48] Yang Yi、Yih Wen-tau 和 Christopher Meek。2015。WikiQA：开放域问答挑战数据集。计算语言学协会，2013–2018。[49] Yujia Zhang、Kuangyan Song、Yiming Sun、Sarah Tan 和 Madeleine Udell。2019。"为什么你应该相信我的解释？"理解 LIME 解释中的不确定性。*arXiv* 预印本 *arXiv:1904.12991*（2019）。[50] Jianlong Zhou、Amir H Gandomi、Fang Chen 和 Andreas Holzinger。2021。评估机器学习解释的质量：方法与指标综述。*Electronics* 10，5（2021），593。