

实体摘要系统的解释性评测

刘庆霞, 李俊宥, 程龚

南京大学计算机软件新技术国家重点实验室, 江苏 南京 210023

摘要

实体摘要任务是从知识图谱中描述实体的大量三元组中选取最优子集作为摘要。现有实体摘要系统通常以较复杂的方式集成多种摘要技术特征。已开展的评测工作对现有系统进行了总体效果的评测和对比,但未能解释系统所用各摘要特征对最终效果的作用。为此,提出对实体摘要系统开展解释性评测。提出两种新指标:特征效用率和特征显著率,两者分别度量各摘要特征在标准摘要和系统生成摘要中的显示度,两者对比分析的结果在一定程度上可为系统取得的最终效果提供解释。基于3个评测集实现了这种评测新方法,运用6种常见的摘要特征,对9个非监督实体摘要系统和两个有监督实体摘要系统进行了解释性评测,相关代码和数据已开源。

关键词

知识图谱;实体摘要;评测集

中图分类号:TP31

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2021023

An interpretive evaluation of entity summarization system

LIU Qingxia, LI Junyou, CHENG Gong

State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China

Abstract

The task of entity summarization (ES) is to select an optimum subset from a large set of triples describing an entity in a knowledge graph. ES systems often integrate many and various ES features in a complex way. While state-of-the-art ES systems have been evaluated and compared by recent benchmarking efforts, it was unclear whether and how much each constituent ES feature had contributed to the performance of an ES system. An interpretive evaluation of ES systems was proposed. Two novel evaluation metrics were proposed, feature effectiveness ratio and feature significance ratio, to characterize how much ground-truth summaries and machine-generated summaries exhibit each ES feature. Their comparison would help to interpret the performance of an ES system. Based on three benchmarks, metrics with six popular ES features were implemented, and an interpretive evaluation of nine unsupervised ES systems and two supervised ES systems were presented. The code and data are open source.

Key words

knowledge graph, entity summarization, benchmark

1 引言

知识图谱可描述实体属性及实体间关系,提供丰富的知识。这种结构化的知识表示有助于数据融合与推理,便于应用对信息进行智能化处理。知识图谱已成为大数据环境下一种常用的数据形式,有效支撑着各类互联网应用,在搜索、电商、社交网络^[1]等领域发挥着重要作用。在知识图谱中,实体、属性及其取值构成了描述该实体的一条三元组,DBpedia和Wikidata等知名知识图谱中包含的三元组总量达到十亿级,一些实体被许多三元组描述,例如,DBpedia中描述实体Barack Obama的三元组有一千多条。然而,在基于知识图谱为终端用户提供信息服务的应用中,为了避免用户信息过载,信息呈现空间通常有限,例如谷歌的知识卡片仅能呈现少量三元组,描述实体的三元组数量往往超过应用允许的限制。

研究者将解决上述问题的方法称作实体摘要(entity summarization, ES),目标是从知识图谱中描述实体的所有三元组中选取一个最优子集作为摘要呈现,在给定的容量限制内为用户提供实体的最关键信息。实体摘要已为多种下游应用提供了支持^[2-3]。现有研究已提出多种实体摘要方法,并实现了各类实体摘要系统。实体摘要系统通常是对多种实体摘要技术特征(以下简称摘要特征)的综合,不同摘要特征体现了实体摘要关注的不同方面,例如实体摘要系统FACES-E^[4]综合了频度、信息度和多样性等摘要特征。参考文献[5]对现有实体摘要系统及其覆盖的摘要特征进行了较详尽的介绍。

相比于对实体摘要方法和系统的研究,与实体摘要评测相关的工作较少。

评测对于实体摘要问题的研究具有长远意义,摘要系统效果的比较、摘要特征效用的评价等都亟须评测工作的支撑。近期开展的一项评测工作ESBM(entity summarization benchmark)^[6]提供了目前规模最大的实体摘要评测集,并基于该评测集对9个实体摘要系统进行了评测和比较。ESBM以黑盒的形式评测实体摘要系统的效果,但未能解释设计复杂的摘要系统表现出的具体效果的深层原因。为了推动摘要系统不断改进,研究者需要深入分析摘要系统的效果,理解系统各组件的具体效用。因此,有必要以白盒形式解释摘要系统的效果,从细粒度分析各项摘要特征的有效性。

为此,本文在ESBM的基础上,提出对实体摘要系统进行解释性评测,这项尝试被称为iESBM(interpretive ESBM)。相关代码和数据已发布在GitHub,并基于ODC-By协议开源。本文仍以通用型实体摘要系统为研究对象,但关注更细粒度层面的分析,从摘要特征的角度对摘要系统的效果进行解释。具体而言,对于每种摘要特征,使用特征效用率(feature effectiveness ratio, FER)度量该摘要特征在标准摘要中的显示度,使用特征显著率(feature significance ratio, FSR)度量该摘要特征在系统生成摘要中的显示度。FER和FSR分别量化了标准摘要和摘要系统生成摘要的特点,将两者进行对比,便可从摘要特征的维度对摘要系统的效果进行一定程度的解释。

本文主要贡献包括以下3个方面。

- 提出实体摘要的解释性评测指标: 本文提出计算摘要特征的FER和FSR,并具体应用于现有实体摘要系统常用的4种三元组级特征和两种摘要级特征。
- 细粒度分析3个评测集的标准摘要:

本文分析标准摘要的FER, 从不同的摘要特征维度刻画标准摘要具有的性质。

- 对11个实体摘要系统进行解释性评测: 本文分析9个非监督实体摘要系统和两个有监督实体摘要系统生成摘要的FSR, 将结果与FER对比, 从摘要特征维度解释这些摘要系统的效果。

2 相关工作

2.1 实体摘要系统

对现有实体摘要系统的详细介绍可参见参考文献[5]。本文关注通用型实体摘要系统, 这类摘要系统可被广泛用于各类场景, 并非为特定领域、应用或用户定制。现有通用型实体摘要系统大多采用非监督方法, 利用各种模型集成多种摘要特征。例如, RELIN^[7]基于三元组间的随机游走综合信息度和相关度对三元组进行排序。DIVERSUM^[8]基于属性的多样度和流行度生成内容多样化的摘要。FACES^[9]及其扩展FACES-E^[4]基于词袋模型计算三元组间的相似度, 对三元组进行聚类, 并从不同聚类中选择信息度和流行度较高的三元组。CD^[10]通过构造并求解二次背包问题来选择高信息度和多样度的三元组。LinkSUM^[11]对PageRank值和反向链接值进行线性组合。BAFRE^[12]将三元组划分为两个层面, 元信息层面的三元组根据其内容在本体中的深度进行排序, 数据层面的三元组根据流行度进行排序。还有一些方法采用了更复杂的模型, 例如, ES-LDA^[13]、ES-LDA_{ext}^[14]和MPSUM^[15]等系统引入了主题模型隐含狄利克雷分布(latent Dirichlet allocation, LDA), 而KAFCA^[16]采用了形式概念分析(formal

concept analysis, FCA)。还有一些工作^[17-18]尝试了有监督的实体摘要方法, 采用深度神经网络进行建模, 从训练集中的标准摘要学习出摘要生成模型。

2.2 实体摘要评测集

实体摘要研究者已经构建了一些评测集^[6,9,19-21], 这些评测集为通用型实体摘要系统的评价提供了标准摘要。然而, 对实体摘要系统的全面评测仍较为少见, 近期的ESBM工作^[6]对9个非监督实体摘要系统进行了评测, 这些摘要系统由各类摘要特征综合而成, 包括ESBM在内的现有评测仅以黑盒的方式对摘要系统进行总体效果的评测和比较, 并未深入分析各摘要特征对总体效果所起到的作用。鉴于现有评测工作的这一局限性, 本文提出新的评测指标, 使实体摘要系统的评测不再局限于粗粒度的总体效果对比, 而是更细粒度地从摘要特征维度对摘要系统的效果进行解释。

3 评测指标与实现

3.1 基本概念

知识图谱的一种典型格式是资源描述框架(resource description framework, RDF)^[22], 这类知识图谱又被称作RDF图。一个RDF图 T 是由形式为<主语, 谓语, 宾语>的三元组构成的集合, 这些三元组描述的所有实体构成的集合记为 E 。对于一个实体 $e \in E$, 其描述 $\text{Desc}(e) \subseteq T$ 由所有以 e 为主语或宾语的三元组构成。一个三元组 $t \in \text{Desc}(e)$ 可写作<subj(t), pred(t), obj(t)>, 描述了实体 e 的属性prop(t)及其取值val(t):

$$e = \text{subj}(t), \text{prop}(t) = \text{pred}(t), \text{val}(t) = \text{obj}(t)$$

$$\text{或 } e = \text{obj}(t), \text{prop}(t) = \text{pred}(t)^-, \text{val}(t) = \text{subj}(t)$$

$$(1)$$

其中, $\text{pred}(t)^-$ 表示 $\text{pred}(t)$ 的反属性。由此, 描述实体 e 的三元组 $t \in \text{Desc}(e)$ 可简化表示为属性取值对 $\langle \text{prop}(t), \text{val}(t) \rangle$ 。

实体 e 的摘要 S 是 e 的描述的子集, 即 $S \subseteq \text{Desc}(e)$, 包含至多 k 个三元组。本文用 S_e 表示由摘要系统为实体 e 生成的待评测摘要, S_g 表示 e 的一个标准摘要。需要注意的是, 实体可能有多个标准摘要, 因为评测集通常会提供多位专家独立标注的标准摘要, 本文将 e 的所有标准摘要的集合记为 S_G 。

3.2 评测指标

对于实体摘要系统中的一个摘要特征, 本文计算两项指标: 特征效用率和特征显著率, 分别用于分析标准摘要和待评测摘要, 再通过两者的比较来解释摘要系统的效果。本文将现有摘要特征分为两类: 三元组级特征和摘要级特征。这两类特征的FER和FSR的计算方式略有不同。

(1) 三元组级特征的评测指标

三元组级特征为每个三元组 $t \in \text{Desc}(e)$ 计算一个打分, 记作 $\text{TScore}(t)$ 。例如, 属性频度就是一种常用的三元组级特征, 若该特征在摘要中确实有效, 即高频度(或低频度)的属性确实更常被选入摘要 S , 则应观察到出现在 S 中的三元组的 TScore (即属性频度)均值高于(或低于)实体描述 $\text{Desc}(e)$ 中的三元组的 TScore 均值。具体而言, 实体描述 $\text{Desc}(e)$ 与摘要 S 在上述均值上的差异可表示为:

$$R(S, \text{Desc}(e)) = \frac{\frac{1}{|S|} \sum_{t \in S} \text{TScore}(t)}{\frac{1}{|\text{Desc}(e)|} \sum_{t \in \text{Desc}(e)} \text{TScore}(t)}$$

$$(2)$$

对计算结果 R 的观察主要在于其偏离1的方向(即高于或低于1)以及偏离程度。基于 R , 三元组级特征的FER和FSR的定义如下。

给定实体 e 的所有标准摘要的集合 S_G , 三元组级摘要特征的FER定义为 S_G 中各标准摘要相应 R 值的均值:

$$\text{FER} = \frac{1}{|S_G|} \sum_{S_g \in S_G} R(S_g, \text{Desc}(e)) \quad (3)$$

给定摘要系统生成的待评测摘要 S_e , 三元组级特征的FSR定义为 S_e 相应的 R 值:

$$\text{FSR} = R(S_e, \text{Desc}(e)) \quad (4)$$

(2) 摘要级特征的评测指标

摘要级特征将摘要包含的三元组集合视作整体来计算一个打分, 记作 SScore 。例如, 摘要多样性就是一种常用的摘要级特征, 若该特征在摘要中确实有效, 即摘要 S 确实由较多样(或较相似)的三元组构成, 则应观察到 S 的 SScore 值(即多样性)高于(或低于)实体描述 $\text{Desc}(e)$ 的 SScore 值。具体而言, 实体描述 $\text{Desc}(e)$ 与摘要 S 在上述值上的差异可表示为:

$$R(S, \text{Desc}(e)) = \frac{\text{SScore}(S)}{\text{SScore}(\text{Desc}(e))} \quad (5)$$

摘要级特征的FER和FSR同样采用式(3)和式(4)计算, 仅将其中 R 的计算方法替换为式(5)。

(3) FER与FSR的意义

FER与1的偏离情况体现了摘要特征在标准摘要中的显示度, 可部分体现摘要特征的有效性。类似地, FSR体现了摘要特征在待评测摘要上的显示度。对于由多种摘要特征综合而成的摘要系统, 这些摘要特征的FSR与相应FER间的差异能够为摘要系统的最终效果提供部分解释。

例如, 某摘要系统以三元组属性频度为摘要特征, 若该摘要特征的FER较高

(远大于1), 则表明标准摘要包含较多具有高频度属性的三元组, 说明该摘要特征是有有效的, 使用该摘要特征有助于提高摘要质量。同时, 若该摘要特征的FSR也较高(远大于1), 则表明该摘要系统生成的摘要确实选取了较多具有高频度属性的三元组。若FER和FSR都较高且较为接近, 则属性频度这项摘要特征可作为该摘要系统取得较好效果的解释之一。

3.3 具体实现

本文选取若干常用摘要特征实现上述评测指标。近期的评测工作ESBM^[6]对9个非监督实体摘要系统进行了评测, 参考文献[5]也全面介绍了现有实体摘要系统。本文从这些工作提及的实体摘要系统中选取摘要特征, 包括4个三元组级特征和两个摘要级特征, 它们涵盖了参考文献[5]中归纳出的三大类摘要特征, 即频度/中心度特征、信息度特征、多样/覆盖度特征。

(1) 三元组级特征

本文具体实现了4个三元组级特征, 其TScore的计算方式互不相同, 分别记为属性局部频度(local frequency of property, LFoP)、谓语全局频度(global frequency of predicate, GFoP)、取值全局频度(global frequency of value, GFoV)和属性取值对的信息度(informativeness of property-value, IoPV)。

摘要系统DIVERSUM^[8]和LinkSUM^[11]采用三元组的LFoP作为摘要特征。对于实体描述中的三元组 $t \in \text{Desc}(e)$, 该摘要特征计算了三元组属性在实体描述中出现的次数:

$$\text{LFoP}(t) = |\{t' \in \text{Desc}(e): \text{prop}(t') = \text{prop}(t)\}| \quad (6)$$

摘要系统LinkSUM^[11]和BAFREC^[12]

采用三元组的GFoP作为摘要特征。对于实体描述中的三元组 $t \in \text{Desc}(e)$, 该摘要特征计算了三元组谓语在RDF图 T 中出现的次数:

$$\text{GFoP}(t) = \lg(|\{t' \in T: \text{pred}(t') = \text{pred}(t)\}| + 1) \quad (7)$$

摘要系统FACES^[9]、FACES-E^[4]和BAFREC^[12]采用三元组的GFoV作为摘要特征。对于实体描述中的三元组 $t \in \text{Desc}(e)$, 该摘要特征计算了取值 $\text{val}(t)$ 在RDF图 T 中出现的次数, 即RDF图中顶点 $\text{val}(t)$ 的度数, 对该值取对数, 以校正过于倾斜的度数分布, 于是得到:

$$\text{GFoV}(t) = \lg(|\{t' \in T: \text{obj}(t') = \text{val}(t) \text{ 或 } \text{subj}(t') = \text{val}(t)\}| + 1) \quad (8)$$

摘要系统RELIN^[7]、FACES^[9]、FACES-E^[4]和CD^[10]采用三元组的IoPV作为摘要特征。对于实体描述中的三元组 $t \in \text{Desc}(e)$, 该摘要特征考虑了RDF图描述的所有实体 E , 计算了“属性取值对 $\langle \text{prop}(t), \text{val}(t) \rangle$ 出现在实体 $e' \in E$ 的实体描述中”这一概率事件的自信息:

$$\text{IoPV}(t) = -\lg \frac{|\{e' \in E: \langle e', \text{prop}(t), \text{val}(t) \rangle \in \text{Desc}(e')\}|}{|E|} \quad (9)$$

(2) 摘要级特征

本文具体实现了两个摘要级特征, 其SScore的计算方式不同, 分别记为属性多样性(diversity of property, DoP)和取值多样性(diversity of value, DoV)。

摘要系统DIVERSUM^[8]和MPSUM^[15]通过避免选取属性相同的三元组来提高摘要多样性。本文将这一思路转化为摘要 S 的一种SScore值计算方式, 定义 S 的DoP为 S 中三元组包含的独特属性的占比:

$$\text{DoP}(S) = \frac{1}{|S|} |\{\text{prop}(t'): t' \in S\}| \quad (10)$$

摘要系统FACES^[9]、FACES-E^[4]和

CD^[10]通过避免选取取值相似的三元组来提高摘要多样性,相似性计算通常基于取值的文本形式。本文将这一思路转化为摘要 S 的一种SScore值计算方式,定义 S 的DoV为 S 中两两取值的文本差异度的均值:

$$\text{DoV}(S) = \frac{1}{\binom{|S|}{2}} \sum_{\substack{t_i, t_j \in S \\ i < j}} (1 - \text{ISub}(\text{val}(t_i), \text{val}(t_j))) \quad (11)$$

其中, $\text{ISub}^{[23]}$ 是常用的字符串相似度度量, $\text{ISub}(\text{val}(t_i), \text{val}(t_j))$ 返回取值 $\text{val}(t_i)$ 和 $\text{val}(t_j)$ 文本形式的相似度, 范围为0~1。具体到RDF图中取值的文本形式, 对于字面量, 取其字面形式 (lexical form); 对于非字面量, 首先检索其 rdfs:label 值, 若检索失败, 则取其本地名称 (local name)。

4 评测集设置

4.1 评测集

本文采用实体摘要领域常用且规模最大的两个评测集: ESBM和FED。

ESBM v1.2是目前最新的实体摘要评测集^[6], 分为两个评测子集: ESBM-D和ESBM-L。ESBM-D中的125个实体取自百科知识图谱DBpedia 2015-10, ESBM-L中的50个实体取自电影知识图谱LinkedMDB。该评测集为每个实体描述提供由不同专家标注的标准摘要, 包括6个以 $k=5$ 为容量限制的标准摘要和6个以 $k=10$ 为容量限制的标准摘要。

FED是为评价FACES而专门设计的评测集^[9]。该评测集中的50个实体来自百科知识图谱DBpedia 3.9。为了匹配FACES的处理能力, FED中的实体描述仅包含取值为实体的三元组, 不包含取值为类型或

字面量的三元组。FED为每个实体描述提供5~8个以 $k=5$ 为容量限制的标准摘要 (均值为7.32) 和5~8个以 $k=10$ 为容量限制的标准摘要 (均值为7.16)。

ESBM v1.2提供了对数据的五等份划分, 以支持统一的五折交叉验证, 每折中60%、20%、20%的实体分别用于训练、验证、测试。本文采用同样方式对FED数据进行划分。

4.2 参评系统

本节对如下11个实体摘要系统进行评测。

- 9个非监督的实体摘要系统, 这些系统也在ESBM^[6]的评测范围内, 即: RELIN^[7]、DIVERSUM^[8]、FACES^[9]、FACES-E^[4]、CD^[10]、LinkSUM^[11]、BAFRE^[12]、KAFCA^[16]和MPSUM^[15]。
- 两个有监督的实体摘要系统: ESA^[17]和DeepLENS^[18]。

对于9个非监督的实体摘要系统, 本文使用其在ESBM^[6]中的实现及配置, 具体而言: RELIN、CD和LinkSUM系统采用超参数对多种摘要特征进行线性组合, 本文在0和1之间以0.01为步长调整这些超参数。对于两个有监督的实体摘要系统, 采用其开源实现及配置, 在模型训练时采用早停机制, 根据验证集上的性能在1~50的范围内选择训练迭代次数。

5 评测结果

本节首先展现传统的非解释性评测结果, 然后基于FER分析各摘要特征的有效性, 最后对比FSR和FER, 并解释各实体摘要系统的效果。限于篇幅, 本节仅展现容量限制 $k=5$ 时的评测结果, $k=10$ 时的结果请参见GitHub。

5.1 非解释性评测

本文沿用ESBM^[6]采用的评测指标F值(F-score)对实体摘要系统的总体效果进行评价。具体而言,针对实体 e ,将待评测摘要 S_c 与各标准摘要 $S_g \in S_c$ 分别进行比较并计算F值,再对这些F值求均值,作为该待评测摘要的最终F值。这里采用以0.01为统计显著水平的双边双样本t检验,其零假设为:两个摘要系统在评测集上的平均F值相同。表1给出了各摘要系统在各评测集上F值的均值和标准差。表1还给出了两两系统间F值对比的t检验结果,若拒绝零假设,则用箭头标出,向上的箭头表示当前系统显著高于对比系统,向下的箭头表示当前系统显著低于对比系统,方块表示两者差异不显著。由于FACES和LinkSUM无法处理取值为类型或字面量的三元组,因此表1未给出它们在ESBM-D和ESBM-L上的结果。

可以看出,作为采用深度神经网络的系统,DeepLENS在3个评测集上显著优于其他系统。此外,BAFREC和ESA在

ESBM-D和ESBM-L上取得了较优结果,而ESA和LinkSUM在FED上取得了较有竞争力的结果。

通过表1,可以对不同摘要系统的效果进行粗粒度的对比。然而,这些结果无法对每个摘要系统的效果进行具体解释。接下来,本文通过FER和FSR进行细粒度的解释。

5.2 摘要特征的有效性

对于第3.3节介绍的6个摘要特征,首先分别计算其在3个评测集上的FER。这里采用以0.01为统计显著水平的双边单样本t检验,其零假设为:摘要特征的FER均值等于1。表2给出了各摘要特征在各评测集上FER的均值和标准差。若t检验结果显示拒绝零假设,则用箭头标出,向上的箭头表示FER的均值显著高于1,向下的箭头表示FER的均值显著低于1。因此,箭头直接标记出了有效的摘要特征。

(1) LFoP与DoP

这两个摘要特征在3个评测集上的FER均值都与1存在显著差异,且差异的方向在各评测集上一致。

表1 各评测集上的F值(均值 ± 标准差), 均值显著高于或低于其他摘要系统的t检验结果(即 $p < 0.01$) 分别用 ↑ 或 ↓ 标出; 不具备显著差异的结果用 · 标出; 不适于比较的用 - 标出

评测集	ESBM-D	ESBM-L	FED
RELIN	0.242±0.120 --↓↓-↓↓↓↓	0.203±0.125 --↓-↓·↓↓↓	0.118±0.080 -····↓↓·↓↓↓
DIVERSUM	0.249±0.136 ---·↓-↓↓↓↓	0.207±0.127 ---↓-↓·↓↓↓	0.113±0.079 -····↓↓·↓↓↓
FACES	不适用	不适用	0.146±0.090 ·-·-↓···↓
FACES-E	0.280±0.142 ↑·-·-↓↓↓↓	0.313±0.116 ↑↑-↑-↓↑·↓	0.146±0.090 -··-↓···↓
CD	0.283±0.134 ↑↑-·-↓↓·↓	0.217±0.101 ·-↓-↓·↓	0.138±0.076 -··-↓↓·↓
LinkSUM	不适用	不适用	0.237±0.120 ↑↑↑↑↑-↑↑·
BAFREC	0.335±0.140 ↑↑-↑↑-···↓	0.360±0.078 ↑↑-↑↑-↑↑·↓	0.195±0.098 ↑↑·↑·-·-↓
KAFCA	0.314±0.128 ↑↑-↑↑-···↓	0.244±0.117 ·-↓-↓-↓	0.143±0.086 -····↓-↓
MPSUM	0.314±0.123 ↑↑-↑-···-↓	0.272±0.112 ↑↑-·-↓-↓	0.169±0.103 ↑↑···↓-↓
ESA	0.331±0.159 ↑↑-↑↑-···↓	0.350±0.143 ↑↑-↑-↑↑-↓	0.247±0.132 ↑↑↑↑↑-↑↑-·
DeepLENS	0.404±0.126 ↑↑-↑↑-↑↑↑↑-	0.469±0.085 ↑↑-↑↑-↑↑↑↑-	0.279±0.101 ↑↑↑↑↑-↑↑↑-·

DoP的FER均值显著高于1,表明标准摘要通常内容较为多样,包含不同属性。

LFoP的FER均值显著低于1,表明标准摘要包含较多低频度属性。然而,该结论可能是受上述属性多样性影响的结果。具体而言,局部频度高的属性通常在实体描述中涉及较多三元组,而标准摘要对多样度的倾向将导致这些具有相同属性的三元组中仅有一个被选入标准摘要,从而降低了LFoP的FER。为此,本文追加了一个实验,尝试消除多样性带来的影响。该实验在FER的计算中,对实体描述中具有相同属性的不同三元组仅计一次属性频度值,这样重新计算的FER在各评测集上都显著高于1,表明由于追求多样性而选择不同属性时,标准摘要实际上仍偏向于选择局部频度较高的属性,如rdf:type和dct:subject。

(2) GFoV与IoPV

这两个摘要特征在3个评测集上的FER均值都与1存在显著差异,但在不同评测集上的差异方向有所不同。

GFoV在ESBM-L和FED上的FER均值显著高于1,表明标准摘要倾向于选择取值全局频度较高的三元组。在ESBM-L的实体描述中,大部分取值是实体(83%);而FED中所有取值都是实体。因此,这些评测集上的高FER表明标准摘要更偏好流行度高的实体。然而,在ESBM-D上该摘要特征的FER均值显著低于1,这是由于ESBM-D的组成与另外两个评测集不同,其实体描述中大部分取值(63%)为类型或字面量,标准摘要倾向于选择全局频度较低的

类型和字面量,以提供更具体的信息。

IoPV与GFoV截然相反,其FER均值在ESBM-D上显著高于1,而在ESBM-L和FED上显著低于1。这一结果是符合预期的,因为IoPV和GFoV在原理上通常起到相反作用:包含全局频度较低取值的三元组的信息量通常较大。

(3) GFoP和DoV

这两个摘要特征在一些评测集上的FER均值与1的差异不显著。

GFoP在ESBM-D上的FER均值显著低于1,表明标准摘要选入了较多全局频度不高的属性。类似于之前对LFoP的分析,GFoP呈现这一现象的原因同样是受到标准摘要倾向于选择多样属性的影响。通过追加实验消除多样性带来的影响之后,ESBM-D和ESBM-L上GFoP的FER均值显著高于1,表明由于追求多样性而选择不同属性时,标准摘要实际上仍倾向于选择全局频度较高的属性,如rdf:type、dct:subject、movie:director和movie:actor。然而,在FED上,重新计算的FER均值与1并无显著差异。

DoV在ESBM-D和ESBM-L上的FER均值显著高于1,在FED上略高于1,体现了标准摘要对取值多样度的偏好。但上述差异的绝对值并不大,这是由于实体描述中取值相似的情况本就不多见。

5.3 摘要系统效果的解释

本节尝试基于6个摘要特征初步解释11个实体摘要系统在各评测集上的摘要效

表2 各评测集上的FER(均值±标准差),均值显著高于或低于1的结果(即 $p < 0.01$)分别用↑或↓标出

评测集	LFoP	DoP	GFoV	IoPV	GFoP	DoV
ESBM-D	0.561±0.165 ↓	2.478±0.747 ↑	0.759±0.125 ↓	1.275±0.175 ↑	0.913±0.052 ↓	1.016±0.054 ↑
ESBM-L	0.581±0.125 ↓	3.093±2.394 ↑	1.349±0.188 ↑	0.864±0.057 ↓	0.998±0.029	1.061±0.068 ↑
FED	0.821±0.205 ↓	1.699±0.480 ↑	1.148±0.153 ↑	0.958±0.044 ↓	1.012±0.066	1.016±0.046

果。对于各摘要系统在各评测集上生成的摘要，计算各摘要特征的FSR。对同一摘要特征的FSR与FER进行比较，采用以0.01为统计显著水平的双边单样本 t 检验，其零假设为：摘要特征的FSR均值和FER均值相等。表3、表4和表5分别给出了评测集ESBM-D、ESBM-L和FED上各系统各摘要特征FSR的均值和标准差。若接受零假设，则标记为方块，称为“相符”，即待评测摘要（对应于FSR值）与标准摘要（对应于FER值）在该摘要特征上的显示度一致。

(1) 非监督摘要系统

对于这些摘要系统，直接将其FSR结果与系统设计用到的摘要特征进行对照

分析。

作为较早的实体摘要系统之一，RELIN在3个评测集上的 F 值都低于其他系统。从FSR可分析出该结果的两个原因。其一，RELIN的IoPV特征的FSR在各摘要系统中最高（同时GFoV的FSR最低），事实上RELIN在设计时便强调偏好属性取值对信息度高（取值全局频度低）的三元组。对于IoPV特征，RELIN在ESBM-L和FED上的FSR与该摘要特征的FER在1的两侧，即RELIN生成摘要的IoPV与标准摘要的倾向相反；而在ESBM-D上，其FSR过高，远超标准摘要对该摘要特征的倾向程度。例如，在ESBM-L上，RELIN常选择属性movie:filmid和movie:actor_actorid

表3 评测集ESBM-D上的FSR（均值 ± 标准差），与FER均值不存在显著差异的项用·标记

摘要系统	LFoP	GFoP	GFoV	IoPV	DoP	DoV
RELIN	0.280±0.228	0.869±0.075	0.277±0.098	1.801±0.297	2.351±0.791 ·	0.749±0.253
DIVERSUM	0.649±0.192	0.910±0.048 ·	0.854±0.167	1.175±0.198	2.753±0.925	1.037±0.086
FACES-E	0.623±0.281 ·	0.914±0.079 ·	0.911±0.208	1.142±0.241	2.494±0.881 ·	0.972±0.118
CD	0.334±0.181	0.863±0.062	0.414±0.136	1.620±0.250	2.742±0.886	1.061±0.051
BAFREC	0.585±0.169 ·	0.954±0.056	0.908±0.177	1.117±0.198	2.586±0.755	0.980±0.107
KAFCA	0.361±0.248	0.850±0.083	0.646±0.244	1.377±0.294	2.505±0.829 ·	0.993±0.116 ·
MPSUM	0.434±0.201	0.876±0.072	0.730±0.250 ·	1.304±0.291 ·	2.742±0.886	0.891±0.187
ESA	0.266±0.212	0.848±0.084	0.529±0.179	1.535±0.282	2.303±0.827	0.930±0.160
DeepLENS	0.302±0.219	0.854±0.076	0.656±0.190	1.407±0.267	2.415±0.801 ·	0.957±0.115
FER	0.561±0.165	0.913±0.052	0.759±0.125	1.275±0.175	2.478±0.747	1.016±0.054

表4 评测集ESBM-L上的FSR（均值 ± 标准差），与FER均值不存在显著差异的项用·标记

摘要系统	LFoP	GFoP	GFoV	IoPV	DoP	DoV
RELIN	0.688±0.432 ·	0.991±0.047 ·	0.600±0.137	1.154±0.050	2.775±2.012 ·	0.967±0.189
DIVERSUM	0.870±0.381	0.993±0.038 ·	1.006±0.247	0.991±0.077	3.869±3.357	1.091±0.075
FACES-E	0.536±0.163 ·	0.962±0.079	1.341±0.296 ·	0.872±0.092 ·	3.848±3.352	1.059±0.103 ·
CD	0.470±0.199	0.996±0.073 ·	1.009±0.212	0.959±0.079	3.869±3.357	1.102±0.071
BAFREC	0.562±0.201 ·	1.020±0.053	1.598±0.491	0.781±0.144	3.485±3.228	1.007±0.088
KAFCA	0.234±0.200	0.954±0.056	1.309±0.395 ·	0.884±0.108 ·	3.869±3.357	1.104±0.102
MPSUM	0.568±0.201 ·	0.979±0.046	1.249±0.428 ·	0.908±0.131 ·	3.869±3.357	1.083±0.104 ·
ESA	0.514±0.235 ·	1.029±0.037	1.241±0.352 ·	0.892±0.116 ·	3.125±2.613 ·	1.013±0.154 ·
DeepLENS	0.361±0.163	1.004±0.037 ·	1.412±0.409 ·	0.840±0.129 ·	3.496±2.343	1.000±0.081
FER	0.581±0.125	0.998±0.029	1.349±0.188	0.864±0.057	3.093±2.394	1.061±0.068

表5 评测集 FED 上的 FSR (均值 ± 标准差), 与 FER 均值不存在显著差异的项用 · 标记

摘要系统	LFoP	GFoP	GFoV	IoPV	DoP	DoV
RELIN	0.911±0.481 ·	1.028±0.157 ·	0.652±0.329	1.123±0.097	1.473±0.579	0.761±0.209
DIVERSUM	1.339±0.220	0.962±0.056	1.043±0.191	0.989±0.069 ·	1.783±0.517	0.981±0.097 ·
FACES	0.860±0.314 ·	0.936±0.081	1.489±0.245	0.886±0.084	1.714±0.514 ·	1.019±0.124 ·
FACES-E	0.860±0.314 ·	0.936±0.081	1.489±0.245	0.886±0.084	1.714±0.514 ·	1.019±0.124 ·
CD	0.799±0.206 ·	1.042±0.075	0.699±0.226	1.118±0.076	1.783±0.517	1.060±0.066
LinkSUM	0.976±0.353	0.987±0.080 ·	1.656±0.250	0.797±0.089	1.460±0.474	1.062±0.074
BAFREC	0.928±0.273	0.949±0.078	1.658±0.304	0.811±0.089	1.768±0.516	0.975±0.119 ·
KAFCA	0.636±0.248	0.999±0.116 ·	0.864±0.363	1.024±0.092	1.699±0.518 ·	0.909±0.134
MPSUM	0.878±0.245 ·	0.918±0.067	1.344±0.289	0.949±0.095 ·	1.783±0.517	0.821±0.225
ESA	0.842±0.323 ·	1.090±0.113	0.813±0.232	1.039±0.075	1.378±0.408	0.875±0.136
DeepLENS	0.823±0.476 ·	1.056±0.095	1.166±0.375 ·	0.926±0.124 ·	1.481±0.486	0.863±0.131
FER	0.821±0.205	1.012±0.066	1.148±0.153	0.958±0.044	1.699±0.480	1.016±0.046

等来描述实体唯一标识的属性, 这些属性信息度极高, 但一般用户很少希望在摘要中看到, 因此极少被选入标准摘要。其二, RELIN的DoV特征的FSR在各摘要系统中最低, 且与FER分布在1值的两侧。这是由于RELIN的设计倾向于取值相似的三元组, 造成摘要内容的冗余并降低DoV。CD系统作为对RELIN的改进, 其F值略高于RELIN。从FSR结果可知, 原因之一是: CD相对于RELIN在多样性上的效果更好, CD的DoV特征的FSR非常高, 远高于RELIN的FSR, 这得益于CD最大化取值多样性的设计。然而, 由于同样倾向于选取属性取值对信息度高的三元组, CD的IoPV特征的FSR也显得过高。

FACES和FACES-E通常具有比RELIN和CD更高的F值。FACES-E在ESBM-L上的F值显著优于RELIN和CD, FACES-E的GFoV和IoPV特征的FSR与FER相符, 这是由该系统在信息度和取值流行度之间相互平衡导致的。然而在ESBM-D和FED评测集上, FACES-E的GFoV特征的FSR过高, 对取值频度的偏好过强。例如, FACES-E常选择以owl:Thing为类型取值

的三元组, 这种三元组意义不大。在FED评测集上, FACES和FACES-E的DoP和DoV特征的FSR与FER相符, 这是由于这两个系统都通过三元组聚类有效提升了摘要多样性。三元组聚类技术在FACES中用于处理取值为实体的三元组, 而FACES-E通过自动识别字面量类型将该技术扩展到取值为字面量的三元组。然而, 对于ESBM-D和ESBM-L这两个包含取值为字面量的评测集, FACES-E的DoP和DoV特征的FSR有时与FER不完全相符, 这说明FACES-E对FACES的扩展尚不够完善。

LinkSUM系统仅处理取值为实体的三元组。在FED评测集上, LinkSUM在非监督实体摘要系统中的F值最高, 其IoPV的FSR最低而GFoV的FSR高于大部分系统, 并且这两个摘要特征的FSR与1的偏离方向均与FER一致。LinkSUM主要设计原则为选择PageRank值较高的实体取值, 这通常对应取值较高的全局频度和较低的自信息, 其GFoV和IoPV的FSR印证了该设计原则在最终生成的摘要中起到了正面效果。然而, LinkSUM的DoP特征的FSR相对于FER偏低, 这是因为其设计未考虑属性多样性。

KAFCA系统是ESBM-D上表现较好的非监督实体摘要系统之一。在ESBM-D评测集上, KAFCA的DoP和DoV特征的FSR都与FER相符, 并且在其他摘要特征上的FSR与1的偏离方向均与FER相同。这是因为KAFCA是基于FCA的设计原理实现的。KAFCA优先选择局部频度低的属性, 这些属性通常全局频度也较低, 并对应局部频度较低的取值。KAFCA偶尔也选择包含局部频度较高词汇的取值, 这有助于选入一些尽管携带高频度属性但仍描述具体信息的三元组, 同时避免过于强调属性和取值多样性。然而, 这种机制在ESBM-L上效果不明显, KAFCA的DoP和DoV特征的FSR相比于FER显得过高。例如, ESBM-L中局部频度较高的属性movie:actor及其反向属性movie:actor⁻被KAFCA完全摒弃, 而标准摘要常选入这些属性。可见, 由于缺乏对全局信息的感知, KAFCA无法对具体实体进行进一步区分。

BAFREC是非监督实体摘要系统中在ESBM-D和ESBM-L上F值最高的系统, 在FED上的F值也较高。BAFREC将三元组分为两个层面, 采用不同的排序选取策略, 但都有多样性要求。对于元信息层面的三元组, BAFREC倾向于全局频度高的属性和全局频度低的取值(类型或字面量); 对于数据层面的三元组, 则倾向于全局频度高的取值(实体)。这些倾向与第5.2节分析的标准摘要的GFoP、GFoV、IoPV和DoP特征的FER较一致, 因而印证了BAFREC的良好效果。然而, 这些倾向并未充分体现在BAFREC的FSR上, 这是由于本文实现的FSR尚未考虑对三元组的分层分析。

(2) 有监督摘要系统

对于这些采用深度学习模型的摘要系统, 其用到的摘要特征并不明显, 尝试将其FSR结果与系统设计原理进行对照分析。

DeepLENS在3个评测集上都取得了高于ESA的F值。而在摘要特征上, DeepLENS的LFoP、GFoP、GFoV和IoPV特征的FSR在3个评测集上大多比ESA更接近FER, 即DeepLENS比ESA更准确地拟合了标准摘要的这些三元组级特征。在模型设计上, ESA将属性和取值以符号的形式进行处理, 而DeepLENS使用文本形式, 并借助预训练词嵌入来理解属性和取值的文本语义。对这一外部信息的利用可能使DeepLENS取得比ESA更好的效果。

ESA和DeepLENS的DoP和DoV特征的FSR总体上低于FER, 这一结果并不意外, 因为它们的设计中都未显式考虑摘要级特征。在ESBM-D和ESBM-L评测集上, DeepLENS的DoP特征的FSR接近或高于FER, 这是因为该方法在处理dct:subject和movie:actor等局部频度较高的属性时, 由于包含这些属性的大量三元组只有很少被选入标准摘要, 训练后的模型简单回避了这些属性。

5.4 案例分析

图1和图2给出了各摘要系统为两个实体描述生成的摘要, 这两个实体分别来自ESBM-D和ESBM-L评测集。RELIN系统将信息度最高的属性filmid(三元组t1a)选入了摘要, 然而该属性从未出现在标准摘要中。RELIN系统生成的摘要中还包含互相冗余的三元组(如三元组t11和t14, 三元组t1b和t1c)。CD系统避免了这种冗余, 但同样也将filmid(三元组t4e)选入了摘要。对于属性writer, CD选择了取值H. R. Christian(三元组t44), 而FACES-E系统则在信息度和取值流行度的平衡下选择了更优的取值Leigh Chapman(三元组t33)。DIVERSUM系统重视属性多样性, 却较忽视对取值的

筛选, 导致选入不理想取值 (如三元组 t2b) 或冗余取值 (如三元组 t21 和 t23)。MPSUM 系统同样重视属性多样性, 但选择的取值更易理解, 尽管依然存在冗余 (如三元组 t7a 和 t7b, t7d 和 t7e)。KAFCA 成功选到属性为 `release date` 的三元组 (t63), 该三元组的属性局部流行度较低且取值中包含局部流行度较高的词汇

“1981”。BAFREC 系统通过设定超参数的方式固定从元信息层面的三元组中选择两个, 从数据层面的三元组中选择 3 个。元信息层面三元组属性通常为 `rdf:type` 或 `rdfs:label`, 这些三元组 (如 t51 和 t55, t5a 和 t5e) 也确实常出现在标准摘要中。有监督的两个实体摘要系统 ESA 和 DeepLENS 生成的摘要质量更高, 尽管其中也存在一

RELIN ($F=0.200$) t11: <label, “King of the Mountain (film)”> t12: <writer, H.R. Christian> t13: <producer, Jack Frost Sanders> t14: <name, “King of the Mountain”> t15: <starring, Deborah Van Valkenburgh>	CD ($F=0.367$) t41: <producer, Jack Frost Sanders> t42: <director, Noel Nosseck> t43: <starring, Deborah Van Valkenburgh> t44: <writer, H.R. Christian> t45: <label, “King of the Mountain (film)”>	MPSUM ($F=0.167$) t71: <starring, Dennis Hopper> t72: <distributor, Universal Studios> t73: <writer, Leigh Chapman> t74: <editing, William Steinkamp> t75: <type, 1981 Films>
DIVERSUM ($F=0.333$) t21: <type, 1981 Films> t22: <starring, Deborah Van Valkenburgh> t23: <subject, 1981 films> t24: <writer, Leigh Chapman> t25: <director, Noel Nosseck>	BAFREC ($F=0.233$) t51: <type, movie> t52: <subject, American films> t53: <distributor, Universal Studios> t54: <starring, Dennis Hopper> t55: <label, “King of the Mountain (film)”>	ESA ($F=0.400$) t81: <director, Noel Nosseck> t82: <distributor, Universal Studios> t83: <type, movie> t84: <label, “King of the Mountain (film)”> t85: <name, “King of the Mountain”>
FACES-E ($F=0.333$) t31: <distributor, Universal Studios> t32: <type, movie> t33: <writer, Leigh Chapman> t34: <label, “King of the Mountain (film)”> t35: <release date, “1981-05-01”>	KAFCA ($F=0.200$) t61: <editing, William Steinkamp> t62: <gross, “2100000.0”> t63: <release date, “1981-05-01”> t64: <label, “King of the Mountain (film)”> t65: <subject, Universal Pictures films>	DeepLENS ($F=0.500$) t91: <release date, “1981-05-01”> t92: <director, Noel Nosseck> t93: <label, “King of the Mountain (film)”> t94: <distributor, Universal Studios> t95: <starring, Dennis Hopper>

图 1 各实体摘要系统为 ESBM-D 评测集的实体 `dbr:King_of_the_Mountain_(film)` 生成的摘要

RELIN ($F=0.267$) t1a: <filmid, “41408”> t1b: <date, “1946-02-06”> t1c: <initial release date, “1946-02-06”> t1d: <page, http://www.freebase.com/...33 > t1e: <performance, performance #35123>	CD ($F=0.267$) t4a: <page, http://www.freebase.com/...33 > t4b: <editor, Harry W. Gerstad> t4c: <initial release date, “1946-02-06”> t4d: <writer, Ethel Lina White> t4e: <filmid, “41408”>	MPSUM ($F=0.333$) t7a: <label, “The Spiral Staircase”> t7b: <title, “The Spiral Staircase”> t7c: <writer, Ethel Lina White> t7d: <date, “1946-02-06”> t7e: <initial release date, “1946-02-06”>
DIVERSUM ($F=0.267$) t2a: <actor, Dorothy McGuire> t2b: <performance, performance #35119> t2c: <writer, Ethel Lina White> t2d: <director, Robert Siodmak> t2e: <editor, Harry W. Gerstad>	BAFREC ($F=0.333$) t5a: <type, film> t5b: <music contributor, Roy Webb> t5c: <actor, George Brent> t5d: <made-, Robert Siodmak> t5e: <label, “The Spiral Staircase”>	ESA ($F=0.500$) t8a: <director, Robert Siodmak> t8b: <writer, Ethel Lina White> t8c: <type, film> t8d: <label, “The Spiral Staircase”> t8e: <page, http://www.freebase.com/...33 >
FACES-E ($F=0.300$) t3a: <music contributor, Roy Webb> t3b: <actor, George Brent> t3c: <made-, Robert Siodmak> t3d: <producer, Dore Schary> t3e: <director, Robert Siodmak>	KAFCA ($F=0.300$) t6a: <editor, Harry W. Gerstad> t6b: <page, http://www.freebase.com/...33 > t6c: <music contributor, Roy Webb> t6d: <producer, Dore Schary> t6e: <type, film>	DeepLENS ($F=0.533$) t9a: <type, film> t9b: <director, Robert Siodmak> t9c: <initial release date, “1946-02-06”> t9d: <producer, Dore Schary> t9e: <title, “The Spiral Staircase”>

图 2 各实体摘要系统为 ESBM-L 评测集的实体 `film:41408` (即电影 “The Spiral Staircase”) 生成的摘要

些冗余(如三元组t84和t85)。

6 结束语

本文尝试对实体摘要系统进行解释性评测,从摘要特征维度对摘要系统的效果进行分析。通过计算6个常用摘要特征的FER,对3个评测集上的标准摘要进行分析,总结出标准摘要的一些典型特点:标准摘要常包含较多样的属性和取值,并倾向于频度较高的属性;在取值为实体的三元组中,倾向于较流行的实体取值;在取值为类型或字面量的三元组中,倾向于信息度较大即较低频的内容。这些发现为未来实体摘要系统的设计改进提供了思路。本文通过计算11个系统的FSR,对这些系统生成摘要的效果进行了分析,将FSR与FER对比,细粒度地揭示了各系统的部分优缺点:非监督实体摘要系统在设计中显式固定一组摘要特征,往往在不同评测集上表现迥异,泛化能力不足,并且常缺失一些有用摘要特征;有监督实体摘要系统利用深度神经网络建模,在一定程度上克服了上述不足,避免了人工特征选择,但现有模型仍较简单,对多样性等摘要级特征的代表能力不足。

未来工作可对本文提出的评测指标及其实现进行扩展。首先,本文仅实现了6个常用摘要特征,增加新的摘要特征有助于开展更全面的分析。其次,由于本文提出的评测指标依赖于标准摘要,增加新的评测集有助于对摘要系统的泛化能力开展更全面的检验。

参考文献:

[1] NOY N, GAO Y Q, JAIN A, et al. Industry-scale knowledge graphs: lessons

and challenges[J]. Communications of the ACM, 2019, 62(8): 36-43.

- [2] GUNARATNA K. Semantics-based summarization of entities in knowledge graphs[D]. Dayton: Wright State University, 2017.
- [3] THALHAMMER A. Linked data entity summarization[D]. Karlsruhe: Karlsruher Institut für Technologie, 2017.
- [4] GUNARATNA K, THIRUNARAYAN K, SHETH A, et al. Gleaning types for literals in RDF triples with application to entity summarization[C]//Proceedings of the ICWE 2016. Cham: Springer, 2016: 85-100.
- [5] LIU Q X, CHENG G, GUNARATNA K, et al. Entity summarization: state of the art and future challenges[J]. arXiv preprint, 2019, arXiv:1910.08252v1.
- [6] LIU Q X, CHENG G, GUNARATNA K, et al. ESBM: an entity summarization BenchMark[C]//Proceedings of the ICWE 2020. Cham: Springer, 2020: 548-564.
- [7] CHENG G, TRAN T, QU Y Z. RELIN: relatedness and informativeness-based centrality for entity summarization[C]//Proceedings of the ICWE 2011. Cham: Springer, 2011: 114-129.
- [8] SYDOW M, PIKUŁA M, SCHENKEL R. The notion of diversity in graphical entity summarisation on semantic knowledge graphs[J]. Journal of Intelligent Information Systems, 2013, 41(2): 109-149.
- [9] GUNARATNA K, THIRUNARAYAN K, SHETH A. FACES: diversity-aware entity summarization using incremental hierarchical conceptual clustering[C]//Proceedings of the 29th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2015: 116-122.
- [10] XU D Y, ZHENG L, QU Y Z. CD at ENSEC 2016: Generating characteristic and diverse entity summaries[C]//Proceedings of the ESWC 2016. [S.l.:s.n.], 2016.
- [11] THALHAMMER A, LASIERRA N, RETTINGER A. LinkSUM: using link analysis to summarize entity data[C]//Proceedings of the ICWE 2016. Cham:

- Springer, 2016: 244–261.
- [12] KROLL H, NAGEL D, BALKE W T. BAFREC: balancing frequency and rarity for entity characterization in linked open data[C]//Proceedings of the 1st International Workshop on Entity Retrieval. [S.l.:s.n.], 2018.
- [13] POURIYEH S, ALLAHYARI M, KOCHUT K, et al. ES-LDA: entity summarization using knowledge-based topic modeling[C]//Proceedings of the 8th International Joint Conference on Natural Language Processing. [S.l.:s.n.], 2017: 316–325.
- [14] POURIYEH S, ALLAHYARI M, KOCHUT K J, et al. Combining word embedding and knowledge-based topic modeling for entity summarization[C]//Proceedings of the 12th International Conference on Semantic Computing. Piscataway: IEEE Press, 2018: 252–255.
- [15] WEI D J, GAO S Y, LIU Y X, et al. MPSUM: entity summarization with predicate-based matching[J]. arXiv preprint, 2020, arXiv:2005.11992.
- [16] KIM E, CHOI K S. Entity summarization based on formal concept analysis[C]//Proceedings of the EYRE 2018. [S.l.:s.n.], 2018.
- [17] WEI D J, LIU Y X, ZHU F Q, et al. ESA: entity summarization with attention[J]. arXiv preprint, 2019, arXiv: 1905.10625.
- [18] LIU Q X, CHENG G, QU Y Z. DeepLENS: deep learning for entity summarization[J]. arXiv preprint, 2020, arXiv: 2003.03736.
- [19] THALHAMMER A, KNUTH M, SACK H. Evaluating entity summarization using a game-based ground truth[C]//Proceedings of the ISWC 2012. Heidelberg: Springer, 2012: 350–361.
- [20] LANGER P, SCHULZE P, GEORGE S, et al. Assigning global relevance scores to DBpedia facts[C]//Proceedings of the 2014 IEEE 30th International Conference on Data Engineering Workshops. Piscataway: IEEE Press, 2014: 248–253.
- [21] WAITELONIS J, SACK H, BOBIC T. FRanCo—a ground truth corpus for fact ranking evaluation[C]//Proceedings of the 1st International Workshop on Summarizing and Presenting Entities and Ontologies. [S.l.:s.n.], 2015.
- [22] RDF Working Group. W3C: resource description framework (RDF)[R]. 2014.
- [23] STOILLOS G, STAMOU G, KOLLIAS S. A string metric for ontology alignment[C]//Proceedings of the ISWC 2005. Heidelberg: Springer, 2005: 624–637.

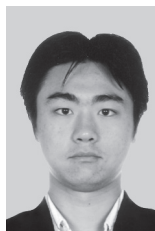
作者简介



刘庆霞 (1990–), 女, 南京大学计算机软件新技术国家重点实验室博士生, 主要研究方向为数据摘要和智能问答。



李俊宥 (1996–), 男, 南京大学计算机软件新技术国家重点实验室硕士生, 主要研究方向为数据摘要和强化学习。



程龚 (1984-), 男, 博士, 南京大学计算机软件新技术国家重点实验室副教授, 主要研究方向为语义搜索、数据摘要、智能问答。

收稿日期: 2020-02-03

通信作者: 程龚, gcheng@nju.edu.cn

基金项目: 国家重点研发计划资助项目 (No. 2018YFB1004300); 国家自然科学基金资助项目 (No. 62072224)

Foundation Items: The National Key Research and Development Program of China (No. 2018YFB1004300), The National Natural Science Foundation of China (No. 62072224)