

开放知识库构建技术的多维量化评价方法

陈新蕾 贾岩涛 王元卓 靳小龙 程学旗

(中国科学院计算技术研究所网络数据科学与技术重点实验室 北京 100190)

摘要 网络大数据时代的到来使得开放知识库的构建技术越来越受到学术界和工业界的关注。近年来,依托开放知识库构建技术而构建的知识库及相关应用层出不穷。然而,目前并没有一种较为统一、全面的开放知识库构建技术的多维量化评价方法。在已有工作的基础之上,提出了一套开放知识库构建技术的多维评价指标体系,综合考虑包括开放知识库的构建准确率、构建时间、构建规模在内的多个维度,并在此基础上提出了开放知识库构建技术的多维量化评价方法。实验表明,开放知识库构建技术的评价结果较传统的评价方法更为全面、合理,并且可以根据实际应用中维度的重要性差异产生不同角度的评价结果。

关键词 开放知识库评价, 多维, 量化

中图分类号 TP302.7 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.12.003

Multi-dimensional Quantitative Evaluation Method of Open Knowledge Base Construction Technology

CHEN Xin-lei JIA Yan-tao WANG Yuan-zhuo JIN Xiao-long CHENG Xue-qi

(CAS Key Lab of Network Data Science and Technology, Institute of Computing Technology,
Chinese Academy of Sciences, Beijing 100190, China)

Abstract With the coming of the era of network big data, the construction of open knowledge base attracts more and more attention from both academia and industry. In recent years, the applications based on open knowledge base construction technologies are emerging in an endless stream. However, there is no unified and comprehensive multi-dimensional quantitative evaluation method which evaluates the construction technique of the open knowledge base. Based on existing work, the multi-dimensional specification system of open knowledge base construction technique was put forward, which combines three dimensions, i. e., the accuracy, the construction time, and the scale of a knowledge base. Furthermore, a multi-dimensional quantitative evaluation method was proposed to evaluate a group of open knowledge base construction techniques. Experiments show the rationality and comprehensiveness of the proposed evaluation method, and it can deduce different results to meet the demand of different applications according to the importance of the dimensions.

Keywords Open knowledge base evaluation, Multi-dimension, Quantification

1 引言

人、机、物三元世界的高度融合引发了数据规模的爆炸式增长和数据模式的高度复杂化,世界已进入网络大数据时代^[1]。为了挖掘海量数据中蕴藏的大量知识,需要以知识库的形式将其组织起来。知识库主要由实体与实体间的关系构成,它作为对数据进行结构化和体系化的组织与关联,进而进行可视化展示的一项重要技术,近年来受到了学术界和工业界的广泛关注。其中,具有代表性的大规模知识库有微软公

司的 Probase,其核心概念数在 265 万以上^[2];谷歌公司的 Knowledge Vault 包含 4500 万个实体以及 2.7 亿个关系^[3];此外,还有一些比较知名的知识库以及建立在知识库之上的知识服务系统,如 NELL^[4], DBpedia^[5] 和 YAGO^[6] 等,以及国内的搜狗知立方、百度知心等。

由于互联网数据的多元化与爆炸式增长,仅针对在线百科等结构化或半结构化数据构建知识库已经远远不能满足需求,因此针对开放网络中的大量数据而构建的开放知识库应运而生。

收稿日期:2016-10-11 返修日期:2016-11-12 本文受国家重点研发计划(2016YFB1000902),国家自然科学基金项目(61173008,61100175,61232010,60933005,61402442,61572469,61572473),北京市自然科学基金青年基金项目(4154086)资助。

陈新蕾(1993-),女,硕士生,主要研究方向为开放知识库,E-mail:chenxinlei@software.ict.ac.cn;贾岩涛(1983-),男,博士,副研究员,主要研究方向为社会计算、开放知识库等;王元卓(1978-),男,博士,研究员,主要研究方向为社会计算、网络行为分析、信息安全等;靳小龙(1976-),男,博士,副研究员,主要研究方向为社会计算、网络性能建模与分析、多智能体系统等;程学旗(1971-),男,博士,研究员,主要研究方向为网络科学、网络与信息安全、互联网搜索与服务。

开放知识库是一个面向开放网络的、动态的、可增量的知识库^[7]。开放网络是指知识的来源多元化,包括来自互联网的非结构化多语言文本数据(如时事新闻、电子邮件、微博、社交媒体网站的帖子、即时通信以及可以转换成文本的信息等)、半结构化的在线百科知识库、机器可读的结构化语言知识库(如各类词库、专名库、主题词表、标注语料库等);动态是指领域知识和语言知识的要素之间的关系以及知识本身的属性是动态变化的;可增量是指知识的规模可以满足用户对于知识获取的实时性要求而进行动态的扩展。这些特点导致开放知识库的数据来源多样、规模巨大、变化迅速、质量参差不齐,从而对开放知识库构建技术的评价提出了巨大的挑战。

一部分传统的知识库评价工作针对知识库的内容质量,仅使用某些基本指标(例如准确率、召回率等)对知识库进行评价。尽管其中一部分的评价对象是开放知识库,但这些工作并没有根据开放知识库来源广、更新快、规模大的特点,结合开放知识库的构建过程,从多个维度对其进行评价。

另一部分传统的知识库评价工作侧重于关注知识库的应用价值。它们大多使用某个综合评价方法(例如层次分析法、模糊评价法、TOPSIS法等)^[8]从多个角度评价知识库的应用价值,并不侧重于通过知识库构建的结果对知识库构建技术进行多维、量化的评价。

本文的主要贡献为结合开放知识库的构建准确率、构建时间、构建规模等多个维度,提出一套针对开放知识库构建技术的多维评价指标体系,对开放知识库构建技术进行多维量化评价。具体如下:

(1)针对开放知识库来源广、更新快、规模大的特点,依据开放知识库构建过程中的命名实体识别(Named Entity Recognition)、属性/关系抽取(Slot Filling)、实体链接(Entity Linking)三大步骤分别计算开放知识库的构建准确率,并综合这些指标提出构建准确率这一概念。将开放知识库构建准确率与构建时间、构建规模相结合,评估开放知识库构建技术的总体效果,提出了一套开放知识库构建技术的多维指标体系,并依照此体系评价开放知识库的构建技术。

(2)为了更好地按开放知识库特点衡量知识库构建技术,方便不同构建技术之间的比较,基于开放知识库构建技术的多维指标体系,提出了开放知识库构建技术的多维量化评价方法,并使用此方法对国际权威评测的官方结果进行评价。

本文第2节介绍了知识库评价的部分相关工作;第3节描述了开放知识库构建技术的多维指标体系;第4节对开放知识库构建技术的多维量化评价方法进行了定义;第5节为实验;最后对本文进行总结。

2 相关工作

目前传统的知识库评价工作主要有两种:1)使用某些基准指标,针对所构建的知识库的质量(包括准确性、完备性等)进行评价;2)提出一套较完备的指标体系,利用综合评价方法对知识库进行评价,这类评价更多地关注知识库的应用价值。

知识库相关评测任务中的指标更多地关注知识库本身的质量。在由美国国家标准与技术研究院(National Institute of Standards and Technology, NIST)组织的文本分析会议的知识库填充(Text Analysis Conference Knowledge Base Population, TAC KBP)评测^[1]中,有一个被称为 Cold Start KB 的子任务,即冷启动构建知识库^[9]。此任务的输入为一批文档,这批文档来源于新闻与论坛,故可以视作开放知识库的构建任务。任务要求各参赛队伍经过命名实体识别、属性/关系抽取、实体链接这3个步骤构建各自的开放知识库。评价这些参赛队伍的开放知识库时,需分别统计这3个步骤的准确率 P 与召回率 R ,并计算出 F 值($F = \frac{2PR}{P+R}$),最终以 F 值大小确定各队伍的排名(但未考虑其他因素,如开放知识库的构建时间、构建规模等),进而得出评价的总指标。

另一种评价工作主要侧重于知识库的应用价值。其提出一套指标体系并使用综合评价方法对知识库的应用价值进行多方面评价。评价对象多为组织与存储学术机构数字知识资源的机构知识库,其主要目的是为作者、图书馆、用户服务,并且机构知识库的内容一般来源于各机构提供的结构化数据,故这些机构知识库评价方法相比质量而言更关注知识库的多方面应用价值,例如存储总量、全文率、版权问题、用户满意度、用户使用率^[10]等。

评价此类知识库时,常用的综合评价方法有 TOPSIS 法、层次分析法、模糊评价法、综合指数法等。刘雪梅等人^[10]首先用层次分析法构建了综合评价指标体系,并计算了指标权重,进一步筛选出了影响较大的指标,然后采用模糊评价法构建了评价模型。姚坤^[11]使用 TOPSIS 法,根据评价对象与最优方案、最劣方案的距离进行优劣相对性分析,对农业知识库进行了多维综合评价。

上述工作均从知识库的应用价值出发,从多个维度对其进行量化的评价,但并未从知识库本身的特点出发对知识库构建技术进行评价。

Adelman 与 Riedel 提出了从知识库需求、知识库本身(结构与内容)、硬件与软件、可用性、性能5个方面对知识库进行评价的框架^[12];孙培山等人^[13]提出了一个从效用与价值两个方面对知识库构建效果进行评价的框架,其中效用关注系统的应用方面,包括知识库质量、易用性、访问率等;而价值关注知识库产生的价值,包括客户满意度、工作效率等。但这两项工作仅提出了框架,未提出一种评价知识库的量化方法。

3 开放知识库构建技术的多维指标体系

根据开放知识库来源广、更新快、规模大这3个特点,本文衡量开放知识库构建技术的维度共有3个:构建准确率、构建时间与构建规模,其中构建准确率为所构建的开放知识库中实体与关系的准确率,分别从命名实体识别、属性/关系抽取、实体链接3个步骤进行衡量;构建时间为开放知识库从开始构建到构建完毕所花费的时间;构建规模为开放知识库构建完毕后的规模,可能包含指代规模、实体规模与关系规模。

¹⁾ 评测地址: <http://www.nist.gov/tac/2016/KBP/ColdStart/index.html>

这里的“指代”为实体指代(mention, 简称为指代), 特指命名实体识别的结果;“实体”特指实体链接后得到的实体(entity), 以示区分。例如使用命名实体识别工具从一篇文本中抽取出“巴拉克·奥巴马”和“奥巴马”, 这两个命名实体都称为指代; 而使用实体链接工具将“巴拉克·奥巴马”和“奥巴马”均链接到“奥巴马”, 此处链接到的“奥巴马”为实体。

开放知识库构建技术的多维评价指标体系的总体结构如图 1 所示, 其中实线为上下级关系, 虚线为命名实体识别、属性/关系抽取、实体链接这 3 个步骤的顺序关系。

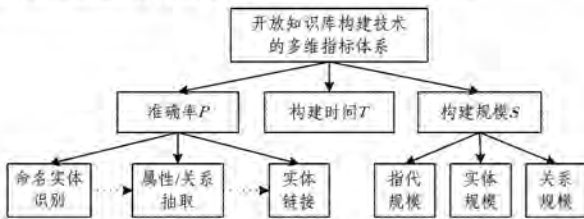


图 1 多维指标体系的总体结构图

3.1 构建准确率

开放知识库的构建主要分 3 个步骤:

- (1) 从开放网络中识别出命名实体, 作为实体的指代;
- (2) 从开放网络中抽取指代的属性与指代间的关系;
- (3) 将描述同一事物的指代链接到同一个实体。

这 3 个步骤的准确率构成了开放知识库构建准确率的基本指标, 最终通过这 3 项基本指标导出开放知识库构建准确率的总指标, 记作构建准确率 P 。

3.1.1 命名实体识别准确率

命名实体识别主要关注从开放网络中识别人物、组织、地点或特定领域的命名实体, 对命名实体识别准确率的评价主要关注所识别的命名实体中正确的比例。命名实体识别在早期的评测中作为独立的任务被关注, 例如 CoNLL¹⁾ (全称为 Conference on Computational Natural Language Learning, 国际著名的自然语言处理会议, 曾在 2002 年与 2003 年举办过两届命名实体识别评测)^[14]、SIGHAN²⁾ (针对中文)、EVALITA³⁾ (针对意大利语) 等。

从构建的开放知识库中随机抽取 θ_R 个命名实体样本(样本中每类命名实体的数量相等), 人工统计正确的个数, 例如对于 θ_R 中的命名实体“奥巴马”, 若被识别为“人物”则正确, 若被识别为“组织”则错误。因此命名实体识别的准确率公式如式(1)所示:

$$P_R = \frac{T_R}{\theta_R} \quad (1)$$

3.1.2 属性/关系抽取准确率

属性/关系抽取主要关注从开放网络中抽取出指代的属性与关系。其中属性一般描述指代的性质(例如身高、出生日期等), 关系一般描述指代之间的相互联系(例如父亲、毕业学校等)。对属性/关系抽取的评价主要关注抽取出的属性和关系中正确的比例。

从构建的开放知识库中随机抽取 θ_E 个属性/关系抽取结果(样本中每类属性/关系的数量大致相等), 人工统计得到正确个数 T_E , 例如 θ_E 中有某个关系“巴拉克·奥巴马-配偶-迈克尔·奥巴马”, 则被视为正确; 但若为“巴拉克·奥巴马-兄弟姐妹-迈克尔·奥巴马”则被视为错误。由此可得到属性/关系抽取的准确率公式为:

$$P_E = \frac{T_E}{\theta_E} \quad (2)$$

3.1.3 实体链接准确率

不同来源或同一来源不同出处的不同指代可能代表同一事物, 这时需对描述同一事物的指代进行链接, 对实体链接的评价主要关注实体链接结果簇与标准答案结果簇的相似度。常用的指标包括 CEAF 和 B³^[15], CEAF 基于测试结果簇和标准答案簇的重合计算最优映射; 而 B³ 评估测试结果簇和标准答案簇的交集中每个实体所在簇的比例。若一个实体被分到多个簇, 则 CEAF 只考虑最大簇, 它比 B³ 对簇的规模更加敏感。故选用 CEAF 作为评价实体链接的指标。

例如, 对于如下的标准答案簇 K 和测试结果 R :

$$K = \underbrace{\{a, b, c\}}_{K_1} \underbrace{\{d, e, f, g\}}_{K_2}$$

$$R = \underbrace{\{a, b\}}_{R_1} \underbrace{\{c, d\}}_{R_2} \underbrace{\{f, g, h, i\}}_{R_3}$$

由于仅考虑 K 和 R 的最优重合(即重合元素最多), 因此令 K_1 与 R_1 对应, K_2 与 R_3 对应, 忽略 R_2 。使用 CEAF 计算得到的实体链接准确率为:

$$P_L = \frac{|K_1 \cap R_1| + |K_2 \cap R_3|}{|R_1| + |R_2| + |R_3|} = \frac{2+2}{2+2+4} = 0.50 \quad (3)$$

从开放知识库的不同来源中随机抽取一批文本, 人工识别所有可聚成簇的实体, 并构造标准答案簇 $K_{l_1}, K_{l_2}, \dots, K_{l_n}$ 。从构建的开放知识库中找出对应可聚成簇的实体及其构成的簇 $\theta_{l_1}, \theta_{l_2}, \dots, \theta_{l_n}$, 得到 K 和 θ 之间的最优重合, 使用式(3)计算实体链接准确率 P_L 。

3.1.4 构建准确率总指标

结合命名实体识别、关系/属性抽取、实体链接 3 个步骤评价知识库构建的准确率, 并对其进行排名。开放知识库构建准确率 P 的计算公式为:

$$P = \alpha P_R + \beta P_E + \gamma P_L \quad (4)$$

其中, α, β, γ 都是非负数。事实上, 根据开放知识库的特点可以得到这 3 个参数的其他约束条件。首先, 为了保证构建准确率的取值范围在 0~1 之间, 需满足 $\alpha + \beta + \gamma = 1$ 。其次, 由于命名实体识别作为开放知识库构建的起始环节, 其会影响属性/关系抽取的抽取结果, 而命名实体识别与属性/关系抽取共同影响实体链接的结果, 因此需满足条件 $\alpha > \beta > \gamma$ 。但由于这 3 个步骤的重要性没有过于明显的区分, 因此 α, β, γ 之间的差距不应过大。满足上述要求的参数设定都是合理的, 本文假设 α, β, γ 的取值分别为 0.4, 0.35, 0.25, 计算得到开放知识库的构建准确率 P 。

¹⁾ CoNLL 官方网站: <http://www.conll.org/>

²⁾ SIGHAN 官方网站: <http://www.sighan.org/>

³⁾ EVALITA 官方网站: <http://www.evalita.it/>

3.2 构建时间

开放知识库的构建时间记作 T ,指从开始构建到构建完毕所花费的时间,此时间有不同的计算方式,例如相关构建技术程序的总运行时间,或构建技术程序的总运行时间与人工参与时间的总和。但对多个开放知识库进行比较时,应当对构建时间的计算方式进行统一,比如均采用构建技术程序的总运行时间,或者均采用程序总运行时间与人工参与时间的总和。

3.3 构建规模

开放知识库的构建规模 S 由以下 3 种情况构成。

(1)仅包含总量规模,即开放知识库中实体与实体间关系的总数。此情况下构建规模即为开放知识库中的实体与关系总数 S_C 。

(2)包含实体规模与关系规模。将实体规模记作 S_E ,关系规模记作 S_R 。

(3)同时包含指代规模、实体规模、关系规模。一般情况下指代与实体的比值越大,说明实体链接中平均每个簇所含指代的个数就越多,实体链接效果越好,因此在这种情况下令指代-实体规模的比为: $S_M = \frac{\text{指代个数}}{\text{实体个数}}$,关系规模记作 S_R 。

若只考虑总量规模(即情况(1)),则将构建准确率、构建时间、总量规模视作 3 个维度来评价开放知识库。

我们在开放知识库构建技术的多维量化评价中将情况(2)与情况(3)中的两项规模视作与构建准确率和构建时间并列,共将其作为 4 个维度对开放知识库进行评价。

4 开放知识库构建技术的多维量化评价方法

为了简化各方面均有不确定差异的开放知识库的比较,我们采用对开放知识库各维度进行排名的方式对若干开放知识库进行评价。首先确定各个维度的重要性,并在各个维度上对开放知识库进行排名,最后根据这两项计算各个开放知识库的总分数,分数从高到低的顺序即为开放知识库的排名。

由于各维度的权重与维度的重要性正相关,因此提出了根据相对重要性确定维度权重的方法:由评判者确定各项维度的重要性,其中每两项维度之间均需相互比较。例如比较构建准确率 P 和构建时间 T ,并且认为构建准确率比构建时间 T 明显重要,则可以得到 P 相对于 T 的标度是 5,反之, T 对于 P 的标度为 $\frac{1}{5}$ 。具体标度确定方法如表 1 所列。

表 1 重要性标度

标度	含义
1	两个要素具有相同重要性
3	前者比后者稍重要
5	前者比后者明显重要
7	前者比后者强烈重要
9	前者比后者极端重要
2,4,6,8	上述相邻判断的中间值
倒数	后者比前者的重要性标度

假设共有 n 个维度,则可以构造如下重要性标度矩阵,其中 $a_{ii} = 1, a_{ij} \cdot a_{ji} = 1(1 \leq i \leq n, 1 \leq j \leq n)$;

$$\begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & & \vdots \\ a_{j1} & \cdots & a_{jj} & \cdots & a_{jn} \\ \vdots & & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{ni} & \cdots & a_{nn} \end{bmatrix}$$

根据重要性标度矩阵,求得每个维度的重要度 W_i ,即矩阵中每个对应行的所有元素乘积的 $\frac{1}{n}$ 次方:

$$W_i = \left(\prod_{j=1}^n a_{ij} \right)^{\frac{1}{n}} \tag{5}$$

对 W_i 进行归一化,得到 W_i^0 :

$$W_i^0 = \frac{W_i}{\sum_{j=1}^n W_j} \tag{6}$$

假设共 m 个开放知识库参与比较,需计算第 k 个开放知识库的分数 $Score_k$ 。先在每个维度上对开放知识库进行排名,其中第 i 个维度的排名为 $Rank_{ki}$ 。然后根据式(7)计算第 k 个开放知识库的分数:

$$Score_k = \sum_{i=1}^4 W_i^0 \cdot \frac{m - Rank_{ki}}{m} \tag{7}$$

得到每个开放知识库的分数之后对其进行排序,便可以评价这 m 个开放知识库的构建技术。

目前,假设从构建准确率 P 、构建时间 T 、实体规模 S_E 和关系规模 S_R 4 个维度(分别由 1-4 代替)对开放知识库进行评价。假设重要性顺序依次为:构建时间 > 准确率 > 实体规模 = 关系规模,满足如下条件:

- 1)构建时间比构建准确率稍重要;
- 2)构建时间比实体规模与关系规模稍重要;
- 3)构建准确率比实体规模与关系规模稍重要;
- 4)实体规模与关系规模同等重要。

对照表 1 可得到如下形式的重要性标度矩阵(重要性数值不唯一,可以自行设定,但在使用同一套重要性标度的前提下,最终开放知识库的排名是一致的):

$$\begin{bmatrix} 1 & \frac{1}{2} & 2 & 2 \\ 2 & 1 & 3 & 3 \\ \frac{1}{2} & \frac{1}{3} & 1 & 1 \\ \frac{1}{2} & \frac{1}{3} & 1 & 1 \end{bmatrix}$$

例如,根据式(5)计算构建准确率 P 的重要性标度:

$$W_P = \left(1 \times \frac{1}{2} \times 2 \times 2 \right)^{\frac{1}{4}} \approx 1.189$$

同理可得另外 3 项的重要性标度分别为 2.060, 0.639, 0.639。

根据式(6)计算构建准确率 P 归一化之后的权重:

$$W_P^0 = \frac{1.189}{1.189 + 2.060 + 0.639 + 0.639} \approx 0.263$$

同理可得构建时间、构建的实体规模、关系规模 3 项的权重分别为 0.455, 0.141, 0.141。

将参与比较的开放知识库分别从每个维度进行排名,然后将计算所得权重代入式(7)计算每个开放知识库的分数,最后依照此分数对开放知识库进行排名。

5 实验

本实验的目的是根据维度重要性设定的差异,针对同一批开放知识库分别产生不同角度的评价结果。

本实验采用的数据来源于 TAC KBP 评测,此评测由权威组织 NIST 主办,已连续举办多年,每年均会吸引一批国内外著名高校、研究所以及公司参加,例如斯坦福大学、卡内基梅隆大学、微软公司、IBM 公司等。

由于官方仅在 2014 年 TAC KBP 评测的 Cold Start 任务中公布了各队伍所构建知识库的指代、实体、关系规模,而在其他年份的评测中仅公布了各队伍的构建准确率,因此本文采用 2014 年的官方数据进行实验^[4]。采用本文所提出的评价方法对参与 2014 年 TAC KBP 评测 Cold Start 任务的 4 支队伍进行评价,它们分别来自于 BBN 科技公司、印度理工学院(Co_Link)、约翰霍普金斯大学(hltcoe)、纽约大学(NYU)。

Cold Start 任务期限为两周,每支队伍的构建准确率与构建规模均已知(包含指代规模、实体规模、关系规模,故此处为 3.3 节中的情况(3)),但无法获知每个队伍的具体用时,故在此默认所有队伍的构建时间一致。可获取各个开放知识库的构建准确率 P 、指代-实体规模 S_M 和关系规模 S_R ,各队伍的具体数据如表 2 所列,其中规模单位为千。

表 2 各队伍的开放知识库构建结果

	构建准确率	指代规模	实体规模	关系规模
BBN	0.5449	2009	223	325
Co_Link	0.3333	2225	2225	2225
hltcoe	0.5821	2070	306	284
NYU	0.6945	1727	228	164

由表 2 可以得到它们的排名,结果如表 3 所列。

表 3 各队伍不同维度的排名

	P	T	S_M	S_R
BBN	3	—	1	2
Co_Link	4	—	4	1
hltcoe	2	—	3	3
NYU	1	—	2	4

假设最重视构建准确率,令:

- 1) 构建准确率比实体规模稍重要;
- 2) 构建准确率比关系规模稍重要;
- 3) 实体规模比关系规模稍重要。

则重要性标度矩阵具有如下形式:

$$\begin{bmatrix} 1 & 2 & 3 \\ \frac{1}{2} & 1 & 2 \\ \frac{1}{3} & \frac{1}{2} & 1 \end{bmatrix}$$

根据该矩阵和式(5)、式(6)计算出 3 个维度的权重分别为 0.540, 0.297, 0.163。由式(7)可计算得到 BBN, Co_Link, hltcoe 和 NYU 的分数分别为 0.43925, 0.12225, 0.385 和 0.5535, 排名为 NYU>BBN>hltcoe>Co_Link。

现在假设提升对实体规模与关系规模的关注,降低对构建准确率的关注,令:

- 1) 实体规模与关系规模均比准确率稍重要;
- 2) 实体规模与关系规模同等重要。

则可以得到如下矩阵:

$$\begin{bmatrix} 1 & \frac{1}{3} & \frac{1}{3} \\ 3 & 1 & 1 \\ 3 & 1 & 1 \end{bmatrix}$$

根据该矩阵和式(5)、式(6)计算出 3 个维度的权重分别为 0.143, 0.429, 0.429。由式(7)可计算得到 BBN, Co_Link, hltcoe 和 NYU 的分数分别为 0.572, 0.32175, 0.286 和 0.32175, 排名为 BBN>Co_Link=NYU>hltcoe。

实验表明在规定维度的重要性之后,各队伍的排名有一定的改变,并且表 3 的各维度排名说明这个结果是符合常理的。这说明本文所提出的开放知识库构建技术的多维量化评价方法可以通过指定各维度的重要性,得到期望的评价结果。

官方仅对各队伍结果的准确率与召回率进行了统计,仅采用 F 值对各队伍的结果进行排名,排名为 BBN>NYU>hltcoe>Co_Link。但假设更看重开放知识库的构建准确率,比如,事实上 NYU 的准确率更高,仅通过 F 值进行评价并不能凸显 NYU 在开放知识库构建的准确性上的优势。而我们提供了一种可以根据实际应用中维度重要性的差异而产生从不同角度评价结果的评价方法。通过对某些维度赋以更高的重要性,可以得到更加合理的结果。

结束语 由于开放知识库具有来源广、更新快、规模大的特点,因此非常需要在短时间内快速构建具有一定质量的开放知识库,故需要一种针对开放知识库构建技术的多维量化评价方法。本文综合开放知识库构建过程中的命名实体识别、属性/关系抽取、实体链接三大步骤来计算开放知识库构建的准确率,并在考虑开放知识库构建时间与构建规模的基础上评估知识库构建的总体效果,提出了一套开放知识库构建技术的多维指标体系,并基于此体系提出了开放知识库构建技术的多维量化评价方法。实验表明此方法较传统的评价方法更为全面、合理,并且可以根据实际应用中维度重要性设定的差异,产生不同角度的评价结果。本文方法仍有以下 3 方面需要改进:1)将领域概念抽取^[17]、知识推理^[18]等步骤的准确率融入知识库构建的准确率指标,完善本文所述方法;2)针对不同行业领域的具体应用,进一步确保本文所述指标体系与评价方法的准确性;3)进一步优化对构建知识库各步骤准确率的计算方法。

参考文献

[1] LI G J, CHENG X Q. Research Status and Scientific Thinking of Big Data[J]. Bulletin of Chinese Academy of Sciences, 2012, 27(6):647-657. (in Chinese)
李国杰,程学旗.大数据研究:未来科技及经济社会发展的重大战略领域[J].中国科学院院刊,2012,27(6):647-657.

[2] WU W T, LI H S, WANG H X, et al. Probase: a probabilistic taxonomy for text understanding[C]//Proceedings of the 2012 ACM SIGMOD International Conference on Management of Da-

- ta. Scottsdale, 2012; 481-492.
- [3] DONG X, GABRILOVICH E, HEITZ G, et al. Knowledge vault: A web-scale approach to probabilistic knowledge fusion [C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, 2014; 601-610.
- [4] CARLSON A, BETTERIDGE J, KISIEL B, et al. Toward an Architecture for Never-Ending Language Learning [C]//AAAI. Atlanta, 2010.
- [5] AUER S, BIZER C, KOBILAROV G, et al. Dbpedia: A nucleus for a web of open data [M]. The Semantic Web. Berlin Heidelberg, Springer, 2007; 722-735.
- [6] BIEGA J, KUZEY E, SUCHANEK F M. Inside YAGO2s: A transparent information extraction architecture [C]//Proceedings of the 22nd International Conference on World Wide Web. Riode Janeiro, 2013; 325-328.
- [7] WANG Y Z, JIA Y T, LIU D W, et al. Open Web Knowledge Aided Information Search and Data Mining [J]. Journal of Computer Research and Development, 2015, 52(2): 456-474. (in Chinese)
王元卓, 贾岩涛, 刘大伟, 等. 基于开放网络知识的信息检索与数据挖掘 [J]. 计算机研究与发展, 2015, 52(2): 456-474.
- [8] JIA P, LI X B, WANG J X. Comparison of several kinds of typical comprehensive evaluation methods [J]. Chinese Journal of Hospital Statistics, 2008, 15(4): 351-353. (in Chinese)
贾品, 李晓斌, 王金秀. 几种典型综合评价方法的比较 [J]. 中国医院统计, 2008, 15(4): 351-353.
- [9] Cold Start Knowledge Base Population at TAC 2015 Task Description [OL]. [2015-7-14]. <http://www.nist.gov/tac/2015/KBP/ColdStart/guidelines.html>.
- [10] LIU X M, DAO K Q. Construction of Evaluation Model for Institutional Repository Basing on Fuzzy Comprehensive Evaluation [J]. Information Research, 2015(5): 22-24, 28. (in Chinese)
刘雪梅, 刀克群. 基于模糊综合评价法的机构知识库评价模型构建 [J]. 情报探索, 2015(5): 22-24, 28.
- [11] YAO K. The comprehensive evaluation system for agricultural knowledge base based on multiple dimension QoS [D]. Xianyang: Northwest A&F University, 2015. (in Chinese)
姚坤. 基于多维 QoS 的农业知识库综合评价系统 [D]. 咸阳: 西北农林科技大学, 2015.
- [12] ADELMAN L, RIEDEL S L. Handbook for evaluating knowledge-based systems: Conceptual framework and compendium of methods [M]. New York: Springer Science & Business Media, 2012; 317-318.
- [13] SUN P S, FAN Z P, CHEN X, et al. Framework and Process for Evaluating the Construction Effectiveness of Knowledge Base [J]. Journal of Northeastern University (Natural Science), 2010, 31(9): 1361-1364. (in Chinese)
孙培山, 樊治平, 陈曦, 等. 评价知识库构建效果的框架与流程 [J]. 东北大学学报(自然科学版), 2010, 31(9): 1361-1364.
- [14] TJONG K, ERIK F, FIEN D. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition [C]//Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003. Edmonton, 2003; 142-147.
- [15] PRADHAN S, LUO X, RECASENS M, et al. Scoring Coreference Partitions of Predicted Mentions: A Reference Implementation [J]. ACL, 2014(2): 30-35.
- [16] TAC 2014 Cold Start [OL]. [2014-11-17]. <http://www.nist.gov/tac/protected/2014/TAC2014-workshop-notebook/agenda.htm>.
- [17] CHEN X L, PANG L, JIA Y T, et al. Word Vector-based Recognition for Unstructured Text Domain Concepts [J]. Journal of Shanxi University (Natural Science Edition), 2015, 38(4): 553-559. (in Chinese)
陈新蕾, 庞琳, 贾岩涛, 等. 基于词向量的开放文本领域概念识别方法 [J]. 山西大学学报(自然科学版), 2015, 38(4): 553-559.
- [18] ZHAO Z Y, JIA Y T, WANG Y Z, et al. Link Inference in Large Scale Evolutionable Knowledge Network [J]. Journal of Computer Research and Development, 2016, 53(2): 492-502. (in Chinese)
赵泽亚, 贾岩涛, 王元卓, 等. 大规模演化知识网络中的关联推理 [J]. 计算机研究与发展, 2016, 53(2): 492-502.

(上接第 16 页)

- Data Freshness [J]. IEEE Transactions on Computers, 2014, 63(4): 979-994.
- [6] LOCKE D, BESTAVORS A, LIN K J, et al. Real-time database: Read-world requirements [C]//Real-Time Database Systems: Issues and Applications. Kluwer Academic, 1997; 83-91.
- [7] RAMAMRITHAM K. Real-time database [J]. Distributed and Parallel Database, 1993, 1(2): 199-266.
- [8] STANKOVIC J A, SON S H, HANSSON J. Misconceptions about real-time databases [M]//Real-Time Database Systems. 2002; 9-16.
- [9] HOS J, KUO T W, MOK A K. Similarity-based load adjustment for real-time data-intensive applications [C]//Proceedings of IEEE Real-Time Systems Symposium, 1997; 144-153.
- [10] XIONG M, RAMAMRITHAM K. Deriving deadlines and periods for real-time update transactions [J]. IEEE Transactions on Computers, 2004, 53(5): 567-583.
- [11] BAI T, LI G H, LIU Y S. A sensor transaction scheduling algorithm for maintaining real-time data temporal validity [J]. Journal of Central South University of Technology, 2011, 18(6): 2068-2073.
- [12] 周本海, 姚大鹏. CPS 中基于预留的实时调度算法研究 [C]//全国 Petri 网理论与应用学术年会, 2013.
- [13] 白天. 数据的时间一致性维护策略 [M]. 长沙: 中南大学出版社, 2015; 38-50.