汉语阅读理解中词义判断题的解答研究

谭红叶[1,2] 武宇飞[1]
(山西大学计算机与信息技术学院 太原 030006)[1]
(山西大学计算智能与中文信息处理教育部重点实验室 太原 030006)[2]

摘 要 阅读理解任务是在给定的单篇文本上,要求计算机根据文本的内容对相应的问题作出回答。以北京语文高考阅读理解为背景,对其中的词义判断题进行了分析与研究,提出了一个基于支持度计算的解答框架,并尝试使用语言模型、点互信息与句子相似度 3 种方法来计算支持度。通过实验验证,3 种方法在真实数据集和自动构造的数据集上均有一定成效。其中,基于点互信息的支持度计算方法在真实数据集上表现最好,获得了 75%的选项正确率。

关键词 阅读理解,词义判断,支持度
中图法分类号 TP391 **文献标识码** A

Answering Word Sense Judgement Questions in Chinese Reading Comprehension

TAN Hong-ye[1,2] WU Yu-fei[1]
(School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China)[1]
(Key Laboratory for Ministry of Education of Computational Intelligence and Chinese Information Processing, Shanxi University, Taiyuan 030006, China)[2]

Abstract Read comprehension tasks require that computers answer relevant query according to the test context on a given single text. This paper researched judgment of word meaning with the background of reading comprehension in Beijing Chinese college entrance examination, proposed a framework based on support value, which was calculated by n-gram, PMI and sentence similar. The experimental results show that the three methods have good effect on real data and auto data. In all ways, support value based on PMI has the best performance on real data, with the accuracy reaching 75%.

Keywords Reading comprehension, Judgment of word meaning, Support value

1 引言

机器阅读理解要求根据单篇文档对给定问题做出回答,涉及复杂的推理及检索技术,是人工智能的研究范畴,也是判别计算机理解人类语言的标准之一[1]。由于阅读理解几乎涉及到自然语言处理领域中关于词语、句子、篇章的所有问题及技术,因此机器阅读理解问题较为困难。

随着 2011 年 IBM 超级电脑 Watson[2] 在智力竞赛节目中取得胜利,以及多个评测会议和智能答题项目的驱动,阅读理解目前已成为自然语言处理领域的研究热点之一,受到了国内外多家单位和研究者的高度关注。例如,2011 年 CLEF 开展了专门面向阅读理解的问答评测 QA4MRE[3],2012 年日本国立情报学研究所开展了“Todai 机器人”项目[1)],推动了人工智能领域中问答及阅读理解技术的研究。

2015 年我国开展了“基于大数据的类人智能关键技术与系统”项目的研究,旨在推进国内教育领域内的智能化进程,促进国内人工智能水平的提升。本文以该项目为背景,针对语文高考阅读理解中的词义判断选择题进行解答研究。我们提出了一个解答框架,尝试使用基于语言模型、基于点互信息和基于句子相似度的 3 种支持度计算方法,并通过支持度来对选项进行排序,从而选出正确答案。3 种方法在真实数据集上的选项正确率分别为 62.5%,75%,62.5%,在自动生成的数据集上的选项正确率分别为 67.5%,65.5%,64%。

2 相关研究

迄今为止,已有众多科研人员针对词语的理解进行了研究,主要包括词语语义相似度研究、词语情感倾向性分析、词语间的语义蕴含、词义消歧等。其中,词义消歧与本文的词义判断题较为相关,因此本文主要介绍词义消歧的研究进展。

词义消歧问题是给定一个含有多种词义的词语,在特定的语境中选择该词语的真实含义[4]。目前,词义消歧方法主要有无监督的词义消歧方法、有监督的词义消歧方法和基于词典的词义消歧方法等[5],其中有监督的词义消歧方法是最为有效的[6-7],因为此类方法将该问题看成一个分类任务,可以根据训练数据有效得知一个多义词所处的不同上下文和特定词义的对应关系,因此只要确定了上下文所属的类别,就可

1) http://21robot.org/

本文受国家自然科学基金项目(61673248),国家高技术研究发展计划(863 计划)项目(2015AA015407),国家自然科学基金青年项目(61100138,61403238,61502287),山西省回国留学人员科研项目(2013-022),山西省 2012 年度留学回国人员科技活动择优项目资助。

谭红叶(1971-),女,博士,副教授,主要研究方向为自然语言处理、信息检索,E-mail:hytan_2006@126.com;武宇飞(1994-),男,硕士生,主要研究方向为中文信息处理,E-mail:598974237@qq.com(通信作者)。

以确定该词的词义类型[5]。但是，此类方法有着严重的数据稀疏问题，缺乏大规模高质量的标注语料。

随着知识库系统应用范围的不断扩大，越来越多的研究者开始利用基于知识库的方法进行词义消歧，这类方法通过外部知识和词汇的特定上下文推导出多义词的含义，可以有效解决数据稀疏问题。目前，较新颖的方法主要包括：基于定义重叠的方法，如 Pilehvar 等从单词词义比较到文本含义比较，在多个层次上综合考虑了语义相似性[8]；基于选择限制的方法，如卢文鹏等提出了基于依存适配度的知识自动获取词义消歧方法[9]；基于结构化知识的方法，如 Agirre 等提出了基于大量词汇知识库的随机散列的词义消歧算法[10]。

3　词语理解题分析

北京语文高考中的阅读理解材料有两类，分别为科技文和文学作品。所涉及的题型有选择题和问答题，其中选择题包括文意理解题、词语理解题、指代消解题等；问答题包括句子理解题、标题理解题、抽取概括题等。本文着重针对阅读理解选择题中的词义判断题进行研究。

我们将词语理解题分为两类：1）词义判断题，即给定某一词语释义，判断该释义在上下文中是否正确；2）词语解释题，即给定一个词语，结合原文给出其释义。这两类题目的难点主要表现在：要求解释的词语是词典中未出现的新词，或者该词在给定的阅读材料中被赋予了新的含义，即词语自身含义发生了转变。具体题目示例如表 1 所列。

表 1　词语理解题题目示例

类别	示例
词义判断	19. 下列词语在文中的意思，理解不正确的一项是（2 分） （2016 北京高考题《白鹿原上奏响一支老腔》） A. 太斩劲了：非常给力 B. 气韵弥漫：韵味充满（作品） C. 乡党：志同道合的同乡 D. 哗然：因惊讶和赞赏而沸腾 （本题答案为：C）
词语解释	19. 通读全文，用一句话简要表述作者所理解的“废墟”。（3 分）（2014 北京高考题《废墟之美》）

表 1 中，词义判断题给出了 4 个选项，其中选项 A 中的“斩劲”、选项 C 中的“乡党”在词典（现代汉语词典第七版）中尚未收录，属于新词；而且“斩劲”一词是四川、湖南、陕西等地的一种方言，在真实标准语料中很少出现。选项 B 中的“气韵弥漫”可以将该短语拆分，得到“气韵”与“弥漫”二词，二者在词典中的含义分别为“文章或书法绘画的意境或韵味”、“充满；布满”，从词形上看二者的含义与选项释义较为相近，可以利用简单的词形匹配来判断正误。选项 C 中的“乡党”错误的原因在于增加了修饰语“志同道合”，而对修饰语的判断较为困难，因为需要从相关的上下文中推理得出。选项 D 中的“哗然”一词在词典中的含义为“形容许多人吵吵嚷嚷”，但在阅读材料中被赋予新的含义（因惊讶和赞赏而沸腾），不仅语义发生了转变，而且情感极性也由贬义词变成褒义词。通过上述分析，可以看出：词义判断题涉及新词、方言词，而且词语会在指定语境中出现转义现象。因此，解答此题不是简单地利用知识库或词典就可以做到（如“太斩劲了”），还需要通过上下文中的语言信息来智能理解与推理。

表 1 中的词语解释题中要求通读全文，并对“废墟”做出解释。这里，“废墟”是文章标题词，该文作者想表达的“废墟”是指“含有历史文化背景、具有文物价值和美学内涵的建筑遗存”，而废墟在词典中的含义为“城市、村庄遭受破坏或灾害后变成的荒凉地方。”显然，废墟一词不仅被作者引申了含义，而且情感极性也由贬义词变成褒义词。

与词义判断题比较而言，词语解释题不仅需要对全文进行篇章结构分析，准确地查找包含词语含义的相关句，还需要对相关句进行融合来进一步生成答案。因此，我们从较为简单的词义判断题入手进行解答。

4　词义判断题解答

4.1　解答框架

词义判断题的一般形式为：给定文档与 4 个词语及其释义的选项，从选项中选出正确或者错误的释义。此类问题可以形式化描述为：

1）问题 Q 由题干 Q_S 与选项 $c_i \in C$ 构成，$i \in (1,2,3,4)$，选项 c_i 由被释义词 w_i 及其释义 E_i 构成。E_i 包含 m 个词，$E_i=\{w_{e_i1}, w_{e_i2}, \cdots, w_{e_ij}, w_{e_im}\}$。

2）文档 D 包含 k 个句子，$D=\{S_1, S_2, \cdots, S_k\}$。

3）根据文档 D，输出正确答案 c_i^*。

针对该问题，我们提出基于支持度（Support Value，SV）计算的策略来进行解答。支持度为文档 D 中相关句 S 对释义 c_i 的支持程度。具体解答策略如图 1 所示。

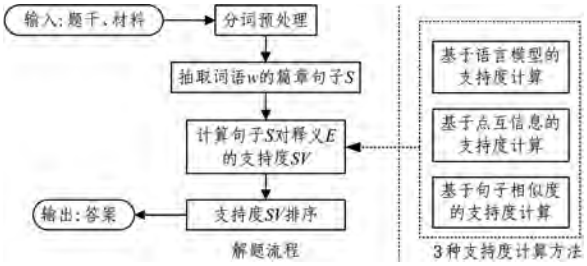

图 1　词义判断题的解答框架示意图

图 1 中，首先将题干与材料进行分词预处理，然后从文档 D 中找到每个选项释义词 w_i 所在的句子 S，并计算 S 对 w_i 的释义 E_i 的支持度 SV，最后依据各选项的 SV 进行排序，从而选出正确答案。本策略中支持度 SV 的计算最为重要，本文提出了 3 种计算支持度的方法，即基于语言模型、基于点互信息和基于句子相似度的支持度计算方法，分别在 4.2 节和 4.4 节对其详细介绍。

4.2　基于语言模型的支持度计算

基于语言模型的支持度计算基于一种直觉：如果词 w 的释义是正确的，则 S' 与 S 在语料中出现的可能性应该一致或接近。这里，S' 为将 S 中释义词 w 用其释义 $E_i=\{w_{e1}, w_{e2}, \cdots, w_{ej}, w_{em}\}$ 替换后的字符串。而语言模型恰好能够计算一个字符串在语料当中出现的可能性，因此可以通过比较二者的差值来获得支持度。支持度越低，二者的差值越小，释义正确的可能性就越大。

基于语言模型的支持度计算如式（1）所示：

$$SV=|P(S)-P(S')| \tag{1}$$

$$P(S')=P(w_1)P(w_2)\cdots P(w_{e1})P(w_{e2})\cdots P(w_{em})\cdots P(w_n) \tag{2}$$

$$P(S)=P(w_1),P(w_2),\cdots,P(w_n) \tag{3}$$

其中，$S=(w_1, w_2, \cdots, w_n)$ 为原文中包含被释义词 w 的句子。可以看出，$P(S')$ 和 $P(S)$ 的计算都是基于语言模型 $P(S)=P(w_1)P(w_2|w_1)P(w_3|w_1,w_2)\cdots P(w_n|w_1,w_2,\cdots,w_{n-1})$。

4.3 基于点互信息的支持度计算

通常，如果词 w 的释义是正确，那么释义 E 中的词语与 w 所在句子 S 中的词应该经常共现。而点互信息(Pointwise Mutual Information，PMI)可以判断两个事件的关联程度，因此我们提出基于点互信息的支持度计算方法来衡量两个句子的共现程度。

基于点互信息的支持度计算如式(4)所示：

$$SV=P(S,E)=\sum_{i=1,j=1}^{n,m}\log_2\frac{P(w_i,w_{ej})}{P(w_i)P(w_{ej})} \tag{4}$$

其中，$P(w_i)$ 表示某一个词在给定语料下出现的概率，$P(w_i,w_{ej})$ 表示两个词语 w_i 与 w_{ej} 共同出现在一个句子的概率。SV 值越高，释义正确的可能性就越大。

4.4 基于句子相似度的支持度计算

我们认为如果词 w 的释义正确，则包含词 w 的句子 S 与 w 的释义 E 相似。相似度越高，则释义正确的可能性越大。基于这种假设，我们提出了基于句子相似度的支持度计算方法。相应的支持度计算如式(5)所示：

$$SV=\cos(V_S,V_E)=\frac{(V_S*V_E)}{|V_S|*|V_E|} \tag{5}$$

其中，V_S 和 V_E 分别是句子 S 和释义 E 的向量表示。

获取句子向量时，将句子 S(或释义句 E)中的词向量加权求和，得到相应的句子向量。句子向量 Vs 的具体计算如式(6)所示：

$$V_S=\sum_{w_i\in S}tf\_idf(w_i)\cdot V_{w_i} \tag{6}$$

其中，$tf\_idf(w_i)$ 表示每个词的权重，通过计算词在文档集(将原文中的每个句子与释义 E 分别视为短文档)中的 $TF\text{-}IDF$ 值获得。w 是句子中的词，V_w 是 w 对应的词向量(worEmbedding)。

5 实验结果及分析

5.1 数据集

本文提到的语言模型和 PMI 的相关参数计算均是基于散文、小说的文学作品(规模约为 8MB)。WordEmbedding 是基于 7 万余篇文学作品(规模约为 416MB)[1)]，利用 Google 开源的 word2vec 进行训练得到。

本文使用的测试集主要有两个：

1)真实数据集(RealData)。该数据集来源于北京语文高考试卷及相关模拟试卷，共 4 道词义判断题。

2)自动构造的数据集(AutoData)。真实数据集规模有限，并不能有效地验证本文所提方法。本文利用两种方法自动生成相关数据集，数据均来源于上述文学作品语料库。数据生成的主要思想为：第一步选词，即选定一篇文档，从中挑选词频为 1、多字词且在词典中是单义词的词语；第二步确定正确选项，即选出 4 个词语并赋予其词典释义作为正确选项；第三步确定错误选项，即从 4 个正确选项中挑选一个，将释义中的一个词语 w 用其相似度最低且字数相等的同义词 w' 替换或将被释义词 w 的释义用其相似度最低且字数相等的同义词 w' 的词典释义替换。自动构造的具体数据示例如表 2 所列。其中，“适当”和“对路”、“高速”和“劈手”是两对同义词。“和谐”的词典释义为“配合的适当”，方法 1“和谐”的释义是用“对路”将“适当”替换得到的；方法 2“高速”的释义是由同义词“劈手”的词典释义得到的。

表 2 自动构造的数据示例

方法	词与释义	词语所在的原句
1	和谐：配合的[对路]【适当】	在河流、沙漠和人三者之间，有了树，一切都会变得和谐起来。
2	高速：形容手的动作迅速，使人来不及防备【劈手】	世纪以高速活塞的方式在流放，情思以断续剪辑的模式在流淌。

5.2 实验结果及分析

本文用选项正确率 P 作为评价指标，该指标的计算如下：

$$P=\frac{Num\_correct}{Num\_sum}$$

其中，$Num\_correct$ 表示正确选项的个数，Num_sum 表示总选项个数。

具体计算时，如果一个词义判断题被正确解答，则认为所有选项判断正确；如果解答错误，则认为有两个选项判断错误。

为了更好地对比本文所提方法，我们实现了一个 Baseline：基于语言模型计算 S'(具体见 4.2 节中的定义)的概率，认为概率越大，释义准确率越高。这里，基于语言模型的支持度计算使用一元语言模型进行实验。实验结果如表 3 所列。

表 3 本文所提方法的正确率

(单位：%)

Method	RealData	AutoData
Baseline	62.5	64.5
基于语言模型的支持度计算	62.5	67.5
基于点互信息的支持度计算	75	65.5
基于句子相似度的支持度计算	62.5	64

由表 3 可知，基于点互信息的支持度计算方法在 RealData 数据集上的正确率最高，表明释义句与词语所在句子的关联程度较高；而该方法在 AutoData 数据集上的正确率有所下降，可能是因为词典释义中包含太多的无用信息，例如“昂首阔步：仰起头，迈着大步向前。形容精神振奋，意气昂扬”。在 AutoData 数据集上基于语言模型的支持度计算方法的正确率最高，并且比在 RealData 数据集上的正确率提高了 5%，一方面说明在一个上下文片段中使用词语与使用其释义是较为接近的，另一方面可能是因为自动构造的数据集适合使用语言模型。

令我们意外的是，基于句子相似度的支持度计算方法的正确率在 AutoData 数据集上表现最差。此方法包含了较为丰富的上下文语境信息，但是其正确率却最低。原因可能是：1)释义只是表达了词语所在句子的一部分语义信息，并不能完全与词语所在句子的语义完全吻合；2)释义中包含了无用信息，如“昂首阔步：仰起头，迈着大步向前”中的“仰起头，迈着大步向前”。

结束语 本文初步对高考阅读理解中的词义判断题进行了解答研究，提出了一个解题框架，尝试用 3 种方法来计算支持度，并根据支持度大小对选项进行排序，从而选出正确答案；同时，我们尝试为此类题目自动生成数据，所提方法在两种数据集上获得了一定的正确率。本文提出的支持度计算方法为高考语文阅读理解的问答题解答提供了支持，推动了阅读理解的研究。虽然我们在此类型题目上取得了一定的进展，但主要是从词匹配的角度来研究。下一步我们将重点从语义推理与蕴含的角度来研究，以提升该类题型的正确率。

(下转第 90 页)

1) 从“散文吧(http://www.sanwen8.cn/)”网站上爬取

表 4 三维 10 点航迹规划比较实验

地图	算法	迭代次数	航迹长度	全连通时间/s	解决 TSP 时间/ms	总运行时间
稀疏型	动态规划法	280	1445.0	1.034	7240.0	8.274/s
	贪心算法	280	1789.5	1.068	0.267	1.068/s
	贪心 MB-RRT* 算法	71	1974.1	—	—	0.169/ms
中等型	动态规划法	350	1476.0	2.169	7300.2	9.469/s
	贪心算法	350	1628.8	1.596	0.315	1.596/s
	贪心 MB-RRT* 算法	80	1837.5	—	—	0.282/ms
密集型	动态规划法	500	1534.0	2.724	5989.0	8.718/s
	贪心算法	500	1574.5	3.395	0.188	3.395/s
	贪心 MB-RRT* 算法	107	1779.2	—	—	0.328/ms

结束语 本文通过将 MB-RRT* 算法与贪心策略结合设计了一种贪心 MB-RRT* 算法，用于解决无人机多点航迹规划问题，并利用二维实验证明了该算法在解决无人机多点航迹规划问题上的可行性；由于无人机真实作业环境为三维空间，因此继续利用三维实验来证明该算法对于无人机在真实环境中执行多点巡航任务时的航迹规划效率优势。

本文所提算法得到的航迹质量与最优航迹的差距仅在于航迹长度，其他质量（如光滑度、可行性等）是一致的；并且本文所展示的数据为贪心 MB-RRT* 算法在得到第一个可行解后的航迹长度，在此之后，可以通过后续降采样优化对该长度进行缩短，拉近所得航迹质量与最优航迹的差距，可以预见，该后续优化所需要的时间将少于构建全连通图所需时间，特别是在三维空间中的时间优势更明显。另外，由于无序多点航迹规划问题为 NP 复杂度的问题，对于更为复杂的环境以及所需访问航点数量更多的航迹规划任务，如果仍采用全连通图的方法来得到最优航迹，时间代价将是巨大的。因此，通过牺牲部分路径质量来提升航迹规划效率的做法是可行的。

除了可以继续优化航迹长度外，未来可以在全局静态规划的基础上加入局部动态规划方法，使得能够对动态障碍物进行有效避障。

参 考 文 献

[1] 何雨枫，曾庆化，王云舒，等. 室内微型飞行器实时路径规划算法研究[J]. 电子测量技术，2014，37(2)：23-27.

[2] KAN E M，SIEN H J，PING Y S，et al. An evolutionary algorithm for multiple waypoints planning with B-spline trajectory generation for Unmanned Aerial Vehicles (UAVs) [C]// International Conference on Computational Problem-Solving. IEEE，2011：77-81.

[3] LAVALLE S M，KUFFNER J J. Randomized Kinodynamic Planning[J]. IEEE International Conference on Robotics & Automation，1999，1(5)：473-479.

[4] MELCHIOR N A，SIMMONS R. Particle RRT for Path Planning with Uncertainty[C]// IEEE International Conference on Robotics & Automation. 2007：1617-1624.

[5] FAIGL J. On Self-Organizing Map and Rapidly-Exploring Random Graph in Multi-Goal Planning [M]. New York：Springer International Publishing，2016.

[6] TRIHARMINTO H H，PRABUWONO A S，ADJI T B，et al. UAV Dynamic Path Planning for Intercepting of a Moving Target：A Review [C]// RoboWorld Congress. Springer Berlin Heidelberg，2013：206-219.

[7] MARTIN S R，WRIGHT S E，SHEPPARD J W. Offline and Online Evolutionary Bi-Directional RRT Algorithms for Efficient Re-Planning in Dynamic Environments[C]// IEEE International Conference on Automation Science and Engineering. IEEE，2007：1131-1136.

[8] WANG J W，DAI G M，XIE B Q，et al. A Survey of Solving the Traveling Salesman Problem[J]. Computer Engineering & Science，2008，30(2)：72-74.

[9] 来学伟. 贪心算法在 TSP 问题中的应用[J]. 许昌学院学报，2017，36(2)：41-44.

（上接第 74 页）

参 考 文 献

[1] 赵红红. 汉语阅读理解问答题解答研究[D]. 太原：山西大学，2016.

[2] FERRUCCI D，LEVAS A，BAGCHI S，et al. Watson：Beyond Jeopardy![J]. Artificial Intelligence，2013，199-200(3)：93-105.

[3] PEÑAS A，HOVY E H，FORNER P，et al. Overview of QA4MRE at CLEF 2011：Question Answering for Machine Reading Evaluation[C]// Proceedings of the Cross Language Evaluation Forum 2011 Labs and Workshop，Notebook Papers. 2011：303-320.

[4] 卢志茂，刘挺，李生. 统计词义消歧的研究进展[J]. 电子学报，2006，34(2)：333-343.

[5] 宗成庆. 统计自然语言处理第二版[M]. 北京：清华大学出版社，2013.

[6] 杨陟卓. 基于上下文语境的词义消歧方法[J]. 计算机应用，2015，35(4)：1006-1008.

[7] CHAN Y S，NG H T. Scaling Up Word Sense Disambiguation via Parallel Texts[C]// The Twentieth National Conference on Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence Conference. Pittsburgh，Pennsylvania，USA，DBLP，2005：1037-1042.

[8] PILEHVAR M T，JURGENS D，NAVIGLI R. Align，Disambiguate and Walk：A Unified Approach for Measuring Semantic Similarity[C]// Meeting of the Association for Computational Linguistics. 2013.

[9] LU W P，HUANG H Y. Word Sense Disambiguation Based on Dependency Fitness with Automatic Knowledge Acquisition[J]. Journal of Software，2013，24(10)：2300-2311.

[10] AGIRRE E，SOROA A. Random walks for knowledge-basedword sense disambiguation[M]. Cambridge：The MIT Press，2014.

[11] MANNING C D，RAGHAVAN P，SCHUTZE H. Introduction to Information Retrieval [M]. 王斌，译. 北京：人民邮电出版社，2010.

[12] 吴军. 数学之美第二版[M]. 北京：人民邮电出版社，2014.

[13] WOODS A M. Exploiting Linguistic Features for Sentence Completion[C]// Meeting of the Association for Computational Linguistics. 2016：438-442.

[14] LIU T，CUI Y，YIN Q，et al. Generating and Exploiting Large-scale Pseudo Training Data for Zero Pronoun Resolution[J]. arXiv. org/abs/1606. 01603.