计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2017,26(10):29−35 [doi: 10.15888/j.cnki.csa.005979]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于字词联合的变体词规范化研究①

施振辉, 沙　瀛, 梁　棋, 李　锐, 邱泳钦, 王　斌

(中国科学院 信息工程研究所, 北京 100093)
(中国科学院大学, 北京 100049)

摘　要: 社交网络中的文本具有随意性和非正规性等特点, 一种常见现象是社交网络文本中存在大量变体词. 人们往往为了避免审查、表达情感等将原来的词用变体词替代, 原来的词成为目标词. 本文研究变体词的规范化任务, 即找到变体词所对应的初始目标词. 本文利用变体词所在文本的时间和语义, 结合变体词词性, 提出了一种时间和语义结合的方法获取候选目标词, 然后提出基于字词联合的词向量方法对候选目标词排序. 我们的方法不需要额外的标注数据, 实验结果表明, 相比于当前最好的方法在准确性上具有一定的提升, 针对与目标词存在相同的字的变体词其性能更好.

关键词: 变体词; 变体词规范化; 社交网络; 词向量; 字词联合训练

引用格式: 施振辉,沙瀛,梁棋,李锐,邱泳钦,王斌.基于字词联合的变体词规范化研究.计算机系统应用,2017,26(10):29–35. http://www.c-s-a.org.cn/1003-3254/5979.html

Research on Morph Normalization Based on Joint Learning of Character and Word

SHI Zhen-Hui, SHA Ying, LIANG Qi, LI Rui, QIU Yong-Qin, WANG Bin

(Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China)
(University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The text is informal in social networks. One of the common phenomena is that there are a lot of morphs in social networks. People are keen on creating morphs to replace their real targets to avoid censorship and express strong sentiment. In this paper we aim to solve the problem of finding real targets corresponding to their entity morphs. We exploit the temporal and semantic and POS constraints to collect target candidates. Then we propose a method based on joint character-word training to sort the target candidates. Our method does not need any additional annotation corpora. Experimental results demonstrate that our approach achieved some improvement over state-of-the-art method. The results also show that the performance is better when morphs share the same character as targets.

Key words: morph; morph normalization; social network; word embedding; joint character-word training

1　引言

变体词在社交网络中普遍存在, Huang[1]和 Zhang[3]首先提出了明确的变体词定义并对其进行了相关的研究. 本文主要研究变体词的规范化任务, 即结合语料的上下文或者背景知识找到变体词所指代的目标词. 例如, 新浪微博“小马哥如今已经不是小鲜肉了, 在岛内还可以闭门自赏, 出门了要适应自己的角色. ”, 其中“小马哥”是变体词, 变体词规范化任务是找到“小马哥”的目标词“马英九”.

研究变体词的规范化具有现实的意义. 在发现层面上, 能为下游的自然语言处理任务提供支撑, 可用于信息提取、语义的深层理解, 能帮助计算机自动化理解快速演化的社交媒体语言. 在生成层面上, 当我们掌握了变体词的生成技术后, 可以对文本进行自动的替

① 基金项目: 国家重点研发计划 (2016YFB0801003); 青年科学基金项目 (61402466)
收稿时间: 2017-01-10; 采用时间: 2017-02-13

换, 让文章更加有趣, 传播更广.

我们将变体词的规范化任务分为两个子任务来研究: 1) 变体词的候选目标词的获取任务; 2) 变体词的候选目标词的排序任务. 我们首先分析了变体词与目标词在时间、语义和多数据源上的分布等特征和关系.

对于变体词候选目标词的获取, 我们采用时间和语义结合的方法, 在多数据源上提取候选目标词. 利用变体词出现的时间和所在微博的语义分布, 从多个数据源 (本文以新浪微博、Twitter 和 Web 新闻为例) 中提取候选语料, 然后在候选语料中提取候选目标词. 此方法使得候选目标词集合的规模和覆盖率达到了比较好的平衡.

对于变体词候选目标词的排序, 我们采用基于神经网络的字词联合训练词向量的方法, 通过对变体词和候选目标词进行相似度计算得到候选目标词得分, 对候选目标词进行排序. 此方法的优势在于结合了变体词和目标词的上下文语义和字层面上的相似性.

实验结果表明, 我们的方法是有效的, 比现有的最好的方法表现出一定的优势, 特别是在与目标词具有相同字的那些变体词上表现非常好, 准确率达到了 85%.

本文的主要贡献:

① 提出了一种时间和语义相结合的多数据源候选目标词获取方法;

② 提出了一种字词联合训练的候选目标词排序方法.

本文的结构安排如下: 第 2 节介绍了变体词规范化的相关工作, 第 3 节介绍了变体词规范化问题的定义, 第 4 节对变体词和目标词的特征与关系进行了分析, 并详细介绍了候选目标词获取方法和候选目标词排序方法, 第 5 节是实验验证部分, 最后是结论.

2 相关研究工作

变体词相关的概念和技术一直在不良文本过滤、社交媒体文本规范化等领域有所体现. 沙[4]的综述中总结介绍了变体词规范化的一般方法. 其中包括: 基于规则的方法, 如 Wong[5], Xia[6], 陈儒[7], Sood[8], Yoon[9] 等人的工作. 基于统计和规则的方法, 如 Wang[10,11], Choudhury[12], Han[13,14], Li[15]等人的工作. 然而, 上述的所有方法都不能很好的处理变体词规范化这一任务. 因为有些变体词是非常抽象的, 比如: 变体词“函数”的目标词是“杨幂”, 这是因为杨幂的名字中“幂”的意思是函数的幂. 而有些变体词比较具体, 如变体词“薛峦子”的目标词是“薛蛮子”, 这是因为“峦”和“蛮”在字形上非常相似. 对于那些根据目标词深层语义变形的变体词, 我们很难用规则和统计处理变体词规范化任务.

明确的变体词概念最早出现在 Huang[1]和 Zhang[2] 等人的论文中. Huang[1]等人最先研究了变体词规范化任务, 在论文中他提取了变体词和目标词的三类特征, 包括表面特征、语义特征和社交特征, 然后利用标注数据训练二分类模型, 通过学习排序的方法对候选目标词进行排序. 他们的方法需要人工提取大量的特征, 并且需要大量的标注数据用于模型训练. 而在 Zhang[2] 等人的文章中, 他们提出了一种端到端的变体词解码方法, 其中变体词的规范化任务是通过在大量语料中训练出词语的词向量, 然后计算变体词和候选目标词之间的相似度来进行候选目标词的排序. 他们的方法只考虑了词语的上下文, 忽略了变体词和目标词在字层面上的联系.

我们在变体词规范化任务上首先利用了字词联合[16,17]的词向量的方法, 综合考虑词语上下文和词语中的字. 我们的方法是利用神经网络训练出字词联合的词向量, 训练出变体词和目标词的相似度, 进而对变体词做规范化.

3 问题的定义

变体词规范化任务是根据给定输入的包含变体词的文本, 找到其中变体词的目标词.

形式化定义为: 已知文档集合 $D=\{d_1,d_2,\cdots,d_{|D|}\}$ 和输入文本集合 $T=\{t_1,t_2,\cdots,t_{|T|}\}$, 其中变体词 m 出现在 T 中的每一篇文本中, 即 $m\in t_k$. 我们的任务就是在文档 D 中找到候选目标词集合 $E=\{e_1,e_2,\cdots,e_{|E|}\}$, 使得集合 E 中的词按照与变体词相关性从大到小排序.

如图 1 所示, 变体词规范化任务的输入是一条微博, 包含变体词“小马哥”, 任务输出是变体词的候选目标词集合, 候选目标词按照与变体词相关性从大到小排序.

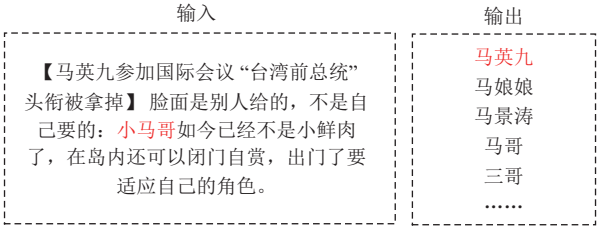

图 1 变体词规范化例子

4 基于字词联合的变体词规范化方法

变体词规范化任务是基于一个假设: 给定输入中我们已经知道了其中哪个词或者哪些词是变体词. 这一过程叫做变体词的识别, 变体词的识别不是本文的研究内容, 本文主要集中于在已知一个词为变体词的情况下, 发现此变体词所对应的目标词. 变体词规范化任务的输入是一条或者多条带有同一个变体词的微博, 输出是变体词的候选目标词集合, 按相关性大小先后排序. 图 2 是我们方法的一个总体流程图, 它由两个子任务组成.

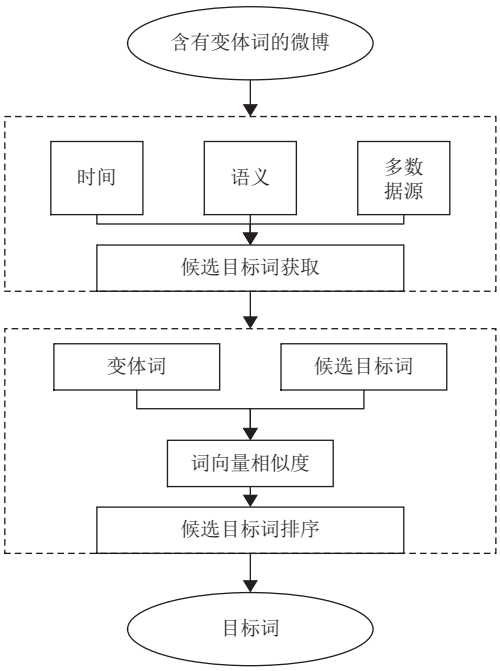

图 2 变体词规范化流程图

① 候选目标词的获取: 对于每一个变体词 m, 找到一个候选词集合 $E=\{e_1,e_2,\cdots,e_{|E|}\}$. 首先, 根据给定的含变体词的微博, 我们提取出变体词出现的时间, 根据这个时间分布, 我们筛选出用于提取候选目标词的语料 $D1$. 其次, 我们将输入的微博看作一篇篇的文档, 通过计算多源语料 $D1$ 中的文档与输入文档之间的话题相似度, 在 $D1$ 中抽取出与输入微博比较相关的语料作为语料 $D2$. 然后在语料 $D2$ 上我们利用中文分词、词性标注、名词检测等工具, 选出候选目标词集合 $E=\{e_1,e_2,\cdots,e_{|E|}\}$.

② 候选目标词的排序: 对候选的目标词集合 E 进行排序. 根据变体词和目标词在词和字层面上的相似性, 利用神经网络训练出字词联合的词向量来计算变体词和候选目标词的相似度, 进而对集合 E 进行排序.

4.1 变体词与目标词的特征分析

4.1.1 时间关系

我们随机选取了 100 个变体词与目标词对, 在时间上对变体词和目标词进行了分析. 如图 3, 变体词“咆哮教主”和目标词“马景涛”在新浪微博中会在同一天共现. 由此我们推断变体词和目标词在时间上具有高度一致性.

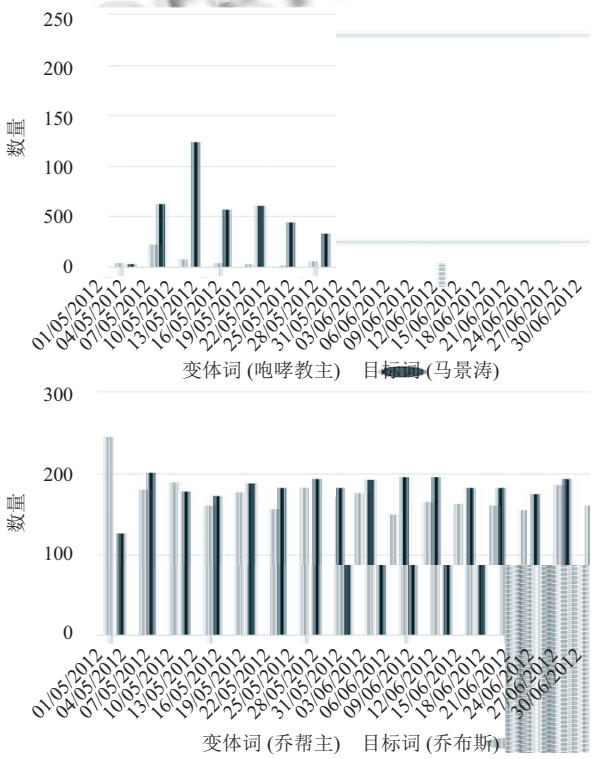

图 3 变体词和目标词在新浪微博的时间分布

4.1.2 话题关系

无论什么原因形成的变体词, 它们的最终目的都是为了替换目标词. 如图 4, 两条微博谈论的都是“美国、中国、外交”等话题, 其中人民日报称呼美国总统为“奥巴马”, 而今日华尔街称呼其为“奥观海同志”, 就是用“奥观海同志”这个变体词替换了目标词“奥巴马”. 由此我们推断变体词和目标词在话题上具有相似性和相关性.

4.1.3 变体词与目标词在多数据源上的分布

变体词一般是在不规范的文本中出现, 如新浪微博, 因为用户在发表微博时有很高的自由度. 而目标词

通常在正规的文本中出现, 如新闻, 因为新闻一般用于正式的场合, 需要表述的清晰明确. 由此, 变体词和目标词在不同的数据源中分布不同.

如表 1 所示, 变体词“呆丸”在新浪微博中大量存在, 而在 Web 新闻中因为不规范而不出现; 另外一些目标词因为敏感、审查等原因, 如目标词“陈光诚”等, 在新浪微博中极少出现甚至不出现.

图 4 变体词和目标词在话题上的分布

表 1 变体词和目标词在不同数据源中分布

变体词	目标词	新浪微博		Twitter		Web新闻	
		变体词	目标词	变体词	目标词	变体词	目标词
呆丸	台湾	467	52842	1	157	0	1409
盲人	陈光诚	3056	1	70	2081	1034	1607

4.2 候选目标词的获取

这一子任务的目标是获取到给定变体词 m 的候选目标词集合 $E=\{e_1,e_2,\cdots,e_{|E|}\}$. 这些候选目标词可能出现在新浪微博、Twitter 和 Web 新闻中. 然而, 如果我们把语料库中的全部词作为变体词的候选目标词, 我们得到的候选目标词集合会非常大, 这是因为包含了大量的无关词语, 这些词语就是规范化系统的噪音, 导致整个规范化系统的效率很低. 如果候选目标词集合过小, 很有可能导致变体词的目标词不在候选集合中, 无法将变体词正确规范化. 所以找到合适的候选目标词集合是这一子任务的难点和关键.

为了解决上述难点, 我们考虑了以下 3 个方面: 1) 变体词和目标词在时间上具有高度一致性; 2) 变体词和目标词所在的文本在话题上具有相似性和相关性; 3) 有些变体词和目标词在不同的数据源中分布不同.

如图 5, 首先, 我们根据给定的含变体词的微博, 我们提取出变体词出现的时间, 根据这个时间分布, 我们筛选出用于提取候选目标词的语料 $D1$. 其次, 我们将输入的微博看作一篇篇的文档, 通过计算多源语料 $D1$ 中的文档与输入文档之间的话题相似度, 在 $D1$ 中抽取出与输入微博比较相关的语料作为语料 $D2$. 然后在语料 $D2$ 上我们利用中文分词、词性标注、名词检测等工具, 选出候选目标词集合 E.

4.3 候选目标词的排序

这一子任务的目标是给变体词 m 的候选目标词集合 $E=\{e_1,e_2,\cdots,e_{|E|}\}$ 按照与变体词相关性从大到小进行排序. 变体词是目标词的一种变形, 变体词形成的原因多种多样. 根据 Zhang[3] 的文章, 我们进一步将变体词的成因归纳为 2 类: 1) 基于拼音、字形等规则变形; 2) 基于词语深层语义的变形. 所以我们综合考虑以下两方面因素: 1) 变体词和目标词在上下文中的联系; 2) 变体词和目标词在字层面上的联系. 我们采用了一种基于神经网络的字词联合[20, 22]词向量方法对候选目标词进行排序.

如图 6, 我们通过字词联合方法训练词向量的时候, 不仅考虑了文本中词语的上下文, 还考虑了组成词语的字. 最后通过训练到的词向量, 我们对变体词和候选目标词进行相似度计算, 以此来对候选目标词进行排序.

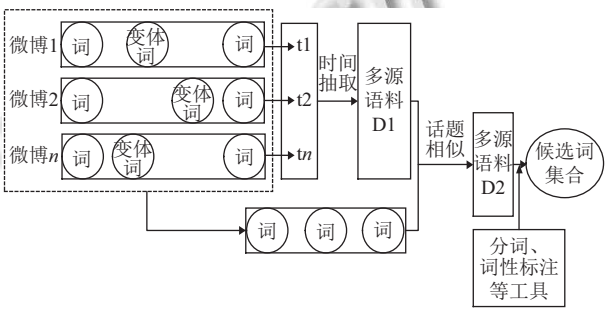

图 5 候选目标词获取框架

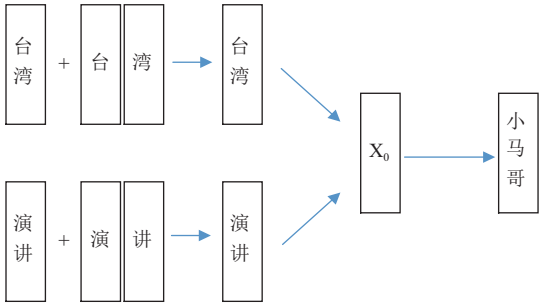

图 6 字词联合训练词向量方法

如图 7 和图 8, 在词向量的训练过程中, CBOW 方法只考虑了词语的上下文, 字词联合方法在 CBOW 方法上进行了改进, 使用词本身的向量以及组成这个词的各个字向量的平均值表示这个词的语义.

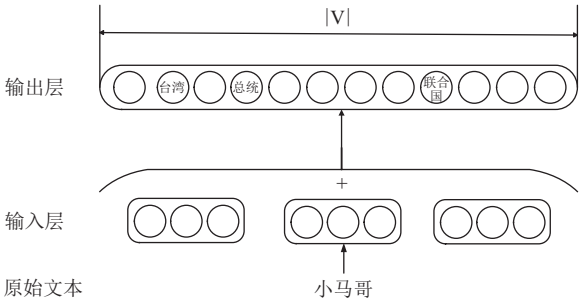

图 7 CBOW 方法神经网络结构图

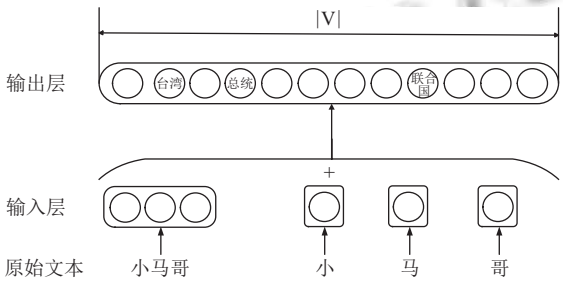

图 8 字词联合神经网络结构

5 实验验证

5.1 数据集

主要使用了 2 个数据集: 第一个数据集来自于 Huang[1]的论文, 其中包括: 1546988 条 2012 年 5 月 1 日到 6 月 1 日的新浪微博数据消息, 收集了 25003 条同样时间段的 Twitter 中文数据推文, 以及 66559 篇新闻文档, 它们来自于新浪微博和 twitter 中的链接, . 其中标注了 450 对变体词和目标词. 第二个数据集是我们另外根据标注好的变体词和目标词, 我们通过关键词搜索, 在 Twitter 中爬取了 337113 条 2015 年 1 月 1 日到 6 月 1 日的中文数据推文, 用于验证我们方法的有效性. 另外, 我们在已有的 450 个标注数据上新增了 225 个标注数据, 来源于中国大陆网络语言列表[18].

5.2 候选目标词的获取

我们认为, 当候选目标词集合 E 中包含变体词 m 的目标词时, 候选目标词的获取是正确的. 我们选取了 557 对变体词和目标词, 在新浪微博、Twitter 和 Web 新闻中通过我们的候选目标词获取方法进行了实验, 分析了获取候选目标词的正确率和时间的关系.

如图 9, 根据覆盖率时间曲线, 在新浪微博中我们设置的时间窗口为 1 天, 在 Twitter 中设置为 3 天, 在 Web 新闻中设置为 1 天, 结合这三种数据源, 此时候选目标词集合的规模和正确率能达到一个较好的平衡. 另外我们发现, 只利用新浪微博和 Twitter 语料候选目标词集合的正确率就达到一个比较好的效果.

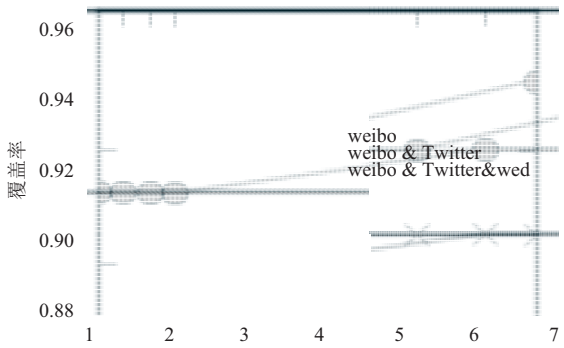

图 9 候选目标词时间覆盖率曲线

本文采用设置时间窗口减少候选目标词的规模, 相比于未设置时间窗口, 候选目标词的规模降低了近 20 倍. 如图 10 所示, 未设置时间窗口时平均每个变体词的候选目标词规模平均为 121590 个, 而按上述设置时间窗口时平均每个变体词的候选目标词的规模为 6131 个. 另外, 我们从图 9 中能得出结论, 变体词的候选目标词的覆盖率达到了 95%, 说明在设置时间窗口的情况下, 候选目标词的损失量仍然很小.

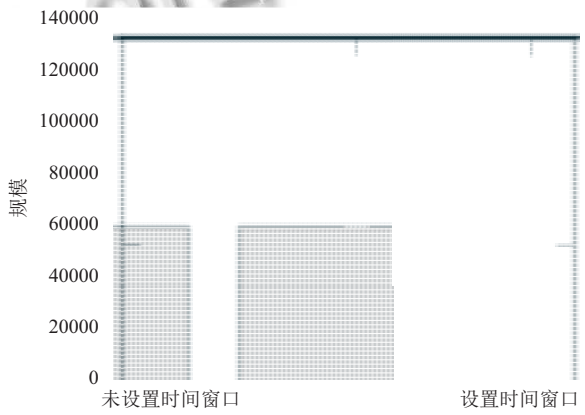

图 10 候选目标词规模与时间窗口的关系

5.3 候选目标词的排序

在 Huang[1]提供的数据上, 我们通过字词联合方法训练词向量, 计算变体词和候选目标词之间的余弦相似度来对候选目标词进行排序. 其中训练词向量时我

们设置的窗口大小为 5, 词向量维度为 300, 训练迭代次数为 15 次.

为了评价我们的方法, 我们采用了和 Huang[1]相同的评价指标, 即$Acc@k = C_k/Q$, 其中C_k指的是返回的前 k 个候选目标词中变体词正确规范化的个数, Q 指的是输入的查询的变体词总数. 我们认为当返回的前 k 个候选目标词中包含了变体词的真实目标词, 那么此时变体词规范化是正确的.

如图 11 所示, 曲线 Huang 13 和 Zhang 15 分别是 Huang[1]和 Zhang[2]的方法, cwe_all 是在 675 对变体词和目标词上的规范化准确率, cwe_part 是在 327 对与目标词存在相同的字的变体词上的规范化准确率. 我们可以得出结论, 在与目标词有相同字的那类变体词的规范化任务上, 本方法要优于当前最好的方法. 当 k>9 的时候, 我们的方法在数据集上要优于当前最好的方法, 当 k<9 时, 我们的方法表现不如当前最好的方法, 可能的原因字词联合训练词向量时字向量的权重偏大.

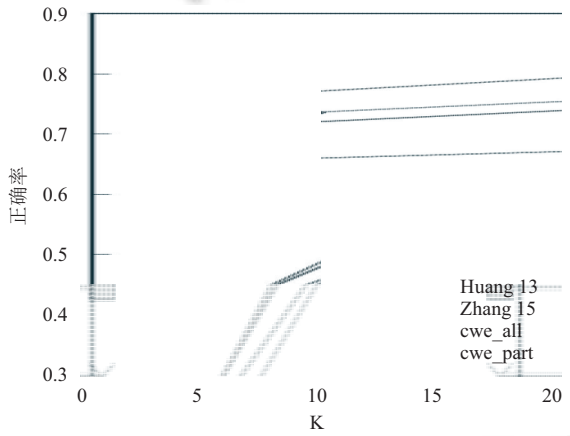

图 11 变体词规范化准确率

另外为了说明语料对规范化任务的影响, 我们在自己爬取的数据集和维基百科数据集上进行了实验, 其中一些参数设置同上述实验. 如图 12 所示, 在维基百科数据上训练出的词向量在变体词规范化这样任务上无法得到较好的结果, 而在 Twitter 数据集上达到一个较好的效果. 这是因为维基百科数据集中的文本是规范文本, 绝大多数变体词不在其中, 导致与候选目标词的相似度计算不准确. 而 Twitter 数据集是我们通过关键词采集的, 包含变体词和目标词及其上下文, 因而能得到较好的效果.

另外我们发现一个有趣的现象, 如图 1 所示, 变体词“小马哥”的目标词是“马英九”, 通过我们的方法输出的排序的候选目标词集合中, “马英九”的另一个变体词“马娘娘”排名同样靠前. 由此我们推断, 同一目标词的不同变体词在语义上是相似的. 故此我们可以借助变体词识别方法来发现不同变体词, 并且通过多个变体词对应目标词来进一步提升我们规范化任务的准确率. 具体的做法是, 我们先在语料上将需要规范化的变体词都识别出来, 即先进行变体词的识别操作, 然后在语料上通过字词联合方法训练出词向量. 接下来需要在本文的方法上进行以下两个方面的修改: 1) 在变体词 m 的候选目标词获取中, 我们不仅需要获取候选目标词 E, 还需要获取这一变体词的目标词的其他可能变体词 m’. 2) 在变体词的候选目标词排序中, 我们通过计算变体词词向量和候选目标词之间的相似度对候选目标词进行排序, 找到目标词的其他可能变体词 m’. 接下来通过本文的变体词规范化方法, 获取变体词 m’的候选目标词集合 E’, 按可能性大小排序. 最后对 E 和 E’做交集得出最后的变体词 m 的候选目标词集合.

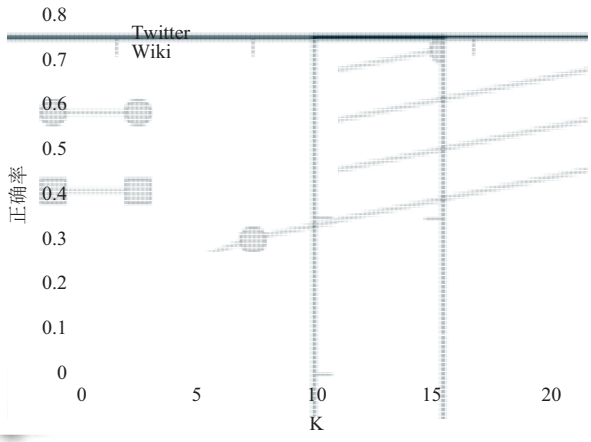

图 12 不同数据源的变体词规范化准确率

如图 13, 分别是设置了时间窗口和未设置时间窗口的规范化系统的正确率曲线, 从图上我们可以推断出, 设置合适的时间窗口, 能大大提高我们的规范化系统的正确率.

5.4 实验分析

通过上述在候选目标词的获取和候选目标词的排序进行的实验, 我们总结下了我们方法的优势: 1) 在候选目标词的获取上, 我们结合时间和语义在多数据源上提取候选目标词, 合理设置时间窗口, 降低了候选目标词集合的规模, 又保证了候选语料中目标词的覆盖率. 2) 在候选目标词的排序上, 我们结合变体词和目标词在字和词两个层面上的语义和词形上的相似性, 采

用字词联合词向量法进行相似度计算, 提升了排序准确性. 当较新的变体词未能被分词器识别出时, 我们通过字向量拼接词向量方法, 仍能进行规范化任务而不需要实时更新词向量模型.

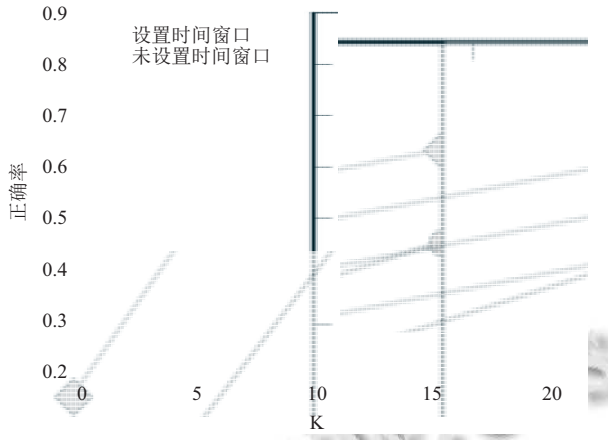

图 13 时间窗口与正确率曲线

6 结论

本文首先主要通过字词联合的词向量技术来解决变体词规范化任务. 首先在分析了变体词和目标词在语义和词形上的异同点的基础上, 分析了变体词规范化任务的挑战. 利用大量未标注数据, 通过时间和语义结合获取候选目标词集合, 并通过字和词两个层面上语义和词形的结合对候选目标词进行排序来解决变体词规范化任务. 下一步工作包括利用同一个目标词的多个变体词之间的关联来进一步提高变体词规范化的准确性.

参考文献

1 Huang HZ, Wen Z, Yu D, *et al*. Resolving entity morphs in censored data. Proc. of the 51st Annual Meeting of the Association for Computational Linguistics. Sofia, Bulgaria. 2013. 1083–1093.

2 Zhang BL, Huang HZ, Pan XM, *et al*. Context-aware entity morph decoding. Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics. Beijing, China. 2015. 586–595.

3 Zhang BL, Huang HZ, Pan XM, *et al*. Be appropriate and funny: Automatic entity morph encoding. Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers). Baltimore, Maryland, USA. 2014. 706–711.

4 沙灞, 梁棋, 王斌. 中文变体词的识别与规范化综述. 信息安全学报, 2016, 1(3): 77–87.

5 Wong KF, Xia Y. Normalization of Chinese chat language. Language Resources and Evaluation, 2008, 42: 219–242. [doi: 10.1007/s10579-008-9067-7]

6 Xia YQ, Wong KF, Li WJ. A phonetic-based approach to Chinese chat text normalization. Proc. of the 21st International Conf. on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistic. Sydney, Australia. 2006. 993–1000.

7 陈儒, 张宇, 刘挺. 面向中文特定信息变异的过滤技术研究. 高技术通讯, 2005, 15(9): 7–12.

8 Sood SO, Antin J, Churchill EF. Using crowdsourcing to improve profanity detection. AAAI Spring Symposium Series. 2012. 69–74.

9 Yoon T, Park SY, Cho HG. A smart filtering system for newly coined profanities by using approximate string alignment. Proc. of 2010 IEEE 10th International Conference on Computer and Information Technology (CIT). Bradford, UK. 2010. 643–650.

10 Wang A, Kan MY, Andrade D, *et al*. Chinese informal word normalization: An experimental study. Proc. of the 6th International Joint Conference on Natural Language Processing. Nagoya, Japan. 2013.

11 Wang AB, Kan MY. Mining informal language from chinese microtext: Joint word recognition and segmentation. Proc. of the 51st Annual Meeting of the Association for Computational Linguistics. Sofia, Bulgaria. 2013. 731–741.

12 Choudhury M, Saraf R, Jain V, *et al*. Investigation and modeling of the structure of texting language. International Journal of Document Analysis and Recognition, 2007, 10(3-4): 157–174. [doi: 10.1007/s10032-007-0054-0]

13 Han B, Cook P, Baldwin T. Automatically constructing a normalisation dictionary for microblogs. Proc. of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Jeju Island, Korea. 2012. 421–432.

14 Han B, Baldwin T. Lexical normalisation of short text messages: Makn sens a # twitter. Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Portland, Oregon. 2011, 1: 368–378.

15 Li ZF, Yarowsky D. Mining and modeling relations between formal and informal chinese phrases from web corpora. Proc. of the Conference on Empirical Methods in Natural Language Processing. Honolulu, Hawaii. 2008. 1031–1040.

16 Chen XX, Xu L, Liu ZY, *et al*. Joint learning of character and word embeddings. Proc. of the 24th International Conference on Artificial Intelligence. Buenos Aires, Argentina. 2015. 1236–1242.

17 来斯惟. 基于神经网络的词和文档语义向量表示方法研究 [博士学位论文]. 北京: 中国科学院自动化研究所, 2016. 1.

18 中国大陆网络语言列表. https://zh.wikipedia. org/wiki/中国大陆网络语言列表. [2016-12].