

文章编号: 1003-0077 (2011) 00-0000-00

基于神经网络的语义选择限制知识自动获取*

贾玉祥, 许鸿飞, 詹红英

(郑州大学 信息工程学院, 河南省 郑州市 450001)

摘要: 语义选择限制刻画谓语对论元的语义选择倾向, 对自然语言的句法语义分析有重要作用, 语义选择限制知识的自动获取也成为重要的研究课题。鉴于神经网络模型在自然语言处理的很多任务中都有出色的表现, 本文提出基于神经网络的语义选择限制知识获取模型, 设计了引入预训练词向量的单隐层前馈网络和两层 maxout 网络。在汉语和英语的伪消歧实验中神经网络模型取得了较好的效果, 优于基于隐含狄克雷分配模型。

关键词: 语义选择限制; 词汇知识获取; 神经网络; 伪消歧

中图分类号: TP391

文献标识码: A

Neural Network Models for Selectional Preference Acquisition

JIA Yuxiang, XU Hongfei, ZAN Hongying

(School of Information Engineering, Zhengzhou University, Zhengzhou, Henan, 450001, China)

Abstract: Selectional preference describes the semantic preference of the predicate for its arguments. It is an important lexical knowledge for the syntactic and semantic analysis of natural languages. Neural network models have achieved state-of-the-art performance in many natural language processing tasks. This paper proposes neural network models for selectional preference acquisition, including a one-hidden-layer feedforward network with pre-trained word vectors and a maxout network. In the pseudo-disambiguation experiments on both Chinese and English, neural network models outperform a LDA-based selectional preference acquisition model.

Key words: selectional preference; lexical acquisition; neural network; pseudo-disambiguation

1 引言

谓语动词对其论元有选择倾向性, 称为语义选择限制 (Selectional Preference, SP)。比如, “吃”这个动词的主语倾向于选择表示“人或动物”的名词, 宾语倾向于选择表示“食物”的名词。可以用函数 $sp_r(v, n)$ 表示语义选择倾向, v 表示谓语动词, r 表示论元类型, n 表示名词, sp 值为实数, 值越大, 表示 n 越适合充当 v 的论元 r 。比如, “苹果”比“石头”更适合充当“吃”的“宾语”。语义选择限制知识获取就是学习函数 $sp_r(v, n)$, 实现对任意 (v, r, n) 的打分。

语义选择限制知识对于分析句子语义有重要价值。比如:

1、判断句子是否合法。

人民的生活水平正在不断地改善。

此句不合法, 主谓搭配错误。可以利用“水平”与“改善”之间的选择倾向来判断是否合法。

他统一安排现场会的内容、时间和出席人员, 以及会议中应注意的问题。

此句不合法, 动宾搭配错误。可以利用“安排”与“问题”之间的选择倾向来判断是否合法。

*收稿日期: 2016-06-09

定稿日期: 2016-08-05

基金项目: 国家自然科学基金 (61402419); 国家社会科学基金 (14BYY096); 国家高技术研究发展计划 863 课题 (2012AA011101)、国家重点基础研究发展计划 973 课题 (2014CB340504)

2、识别隐喻表达^[1]。

山体滑坡 业绩滑坡

编织毛衣 编织梦想

根据在隐喻表达中具体动词倾向于选择抽象的名词作为主语或宾语这样的规律，可以识别出“业绩滑坡”与“编织梦想”为隐喻表达。

3、推测词义。

好莱坞特技车队**驾驶凯迪拉克**大秀车技。

爱她就请她**吃哈根达斯**。

通过“驾驶”倾向于选择“汽车”做宾语，可以推测“凯迪拉克”是一种汽车。通过“吃”倾向于选择“食物”做宾语，可以推测“哈根达斯”是一种食物。

语义选择限制知识可以用于自然语言处理的多个任务，比如：句法分析^[2]、语义角色标注^[3]、词义消歧^[4]、机器翻译^[5]等。语义选择限制是很多词汇知识库的重要组成部分，比如英语的 VerbNet、汉语的 HowNet 等。然而，手工构建的语义选择限制知识库很难满足自然语言处理的需求，因此要求从大规模语料中自动获取语义选择限制知识。

神经网络可以实现函数 $sp_r(v,n)$ ，网络输入为代表 v 和 n 的值（如，词向量），输出为 $sp_r(v,n)$ 。神经网络模型在自然语言处理的很多任务上都取得了很好的效果，本文提出基于神经网络的语义选择限制知识获取模型，尝试不同的网络结构，并与基于 LDA (Latent Dirichlet Allocation) 的模型进行比较。本文的章节安排如下：第 2 节介绍相关研究工作，第 3 节介绍语义选择限制获取的神经网络模型，第 4 节给出实验结果与分析，第 5 节给出总结和展望。

2 相关研究

从训练语料中抽取句法搭配 (v,r,n) ，形成训练集，当然可以对搭配做一些过滤，如限制搭配必须至少出现多少次，或 v 、 n 至少出现多少次。对于训练集中出现过的搭配， $sp_r(v,n)$ 的计算比较直观，可以简单地用共现次数 $count(v,r,n)$ 或条件概率 $p(n|v,r)$ 来表示。关键是训练集中没有出现过的搭配 (v,r,n') 如何计算 $sp_r(v,n')$ ，即如何根据 v 的已知论元计算未知论元的 sp 值，称之为论元扩展。根据论元扩展是否使用语义分类体系可以将语义选择限制获取方法分为两大类：基于语义分类体系的方法与基于分布的方法（见表 1）。

表 1 SP 自动获取的代表性方法

来源	语义分类体系	方法	是否需要反例
Resnik, 1993	WordNet	KL 距离	否
Li and Abe, 1998	WordNet	最小描述长度 MDL	否
Clark and Weir, 2002	WordNet	假设检验	否
Ó S éaghdha, 2010	无	隐变量模型，LDA	否
Erk, 2010	无	向量空间模型，词语相似度	否
Tian et al., 2013	无	随机游走 Random Walk	否
Bergsma et al., 2008	无	SVM	是
Van de Cruys, 2014	无	神经网络	是

基于语义分类体系的方法。该方法借助语义分类体系（如 WordNet），计算谓语对论元

语义类的 sp 值, 那么对于未知论元, 只要它出现在某一个语义类中, 就可以给它一个 sp 值。对于语义类 sp 值的计算, Resnik^[6]使用一个基于 KL 距离的统计指标, Li 和 Abe^[7]基于最小描述长度 (Minimum Description Length, MDL) 模型, Clark 和 Weir^[8]则采用基于假设检验的方法。这类方法的优点是可以学习出关于语义类的选择限制知识, 易于人类理解, 便于集成到词汇知识库中。缺点是需要一个语义分类体系, 由于词典收词有限会导致论元覆盖率比较低, 并且不能很好地处理一词多义的问题。这类方法主要面向语言学研究 and 词汇知识库构建。

基于分布的方法。该方法不需要语义分类体系, 而是根据词语在语料中的分布来实现论元的扩展。隐变量模型^[9] (如 LDA) 是一种基于概率的模型, 隐变量可以看成一个个隐含的语义类, 把谓语句和未知论元联系起来。基于向量空间的模型^[10]利用大规模语料构建一个向量空间, 通过在该空间里计算未知论元和已知论元的相似度, 把谓语句和未知论元联系起来。Tian et al.^[11]通过在谓语句论元搭配图上的随机游走算法来解决未知论元 sp 值的计算问题。Bergsma et al.^[12]使用 SVM (Support Vector Model) 直接对论元进行二分类: 合适的论元和不合适的论元, 把分类器给论元的打分作为 sp 值。Van de Cruys^[13]首次使用神经网络进行选择限制知识的获取, 设计了一个单隐层前馈网络, 输入 v 、 n 的词向量表示, 输出 sp 值。基于分布的方法优点是不依赖语义分类体系, 论元覆盖率高, 对一词多义问题能更好地处理, 易于和其他自然语言处理任务结合。缺点是学习出的知识是词语层面的, 与语义类层面的知识相比, 概括性差, 不易于人类理解。这类方法主要面向自然语言处理, 也是 SP 获取的主流方法。

基于 SVM 和神经网络的方法把选择限制的学习当做机器学习的分类和回归问题, 属于有监督的学习, 需要正例和反例, 即合适的论元和不适合的论元。其他方法只需要正例, 训练集中都是正例, 以这些正例为种子, 通过语义分类体系或语料分布实现论元的扩展。反例并非由人工构建, 而是自动产生, 思想是使正例发生的可能性尽量大, 反例发生的可能性尽量小。例如, 对于每一个正例 (v, r, n) , 采用某一种策略将 n 替换为 n' , 只要 (v, r, n') 不在训练集中出现或某一个统计指标小于阈值 (如出现次数、互信息值), 则把 (v, r, n') 当做反例。

本文在神经网络^[13]中引入预训练的词向量来改进模型的效果, 并提出基于 maxout 网络的模型。作为实验对比的 LDA 模型^[14], 把每个谓语句动词 v 看做一篇文档, 文档内容由训练集中该动词特定论元 r 的所有名词 n 构成, 隐含主题 z 为名词语义类, 则语义选择倾向定义为公式 1, 其中参数 $p(n|z)$ 与 $p(z|v, r)$ 由模型训练得到。

$$sp_r(v, n) = p(n|v, r) = \sum_z p(n|z) p(z|v, r) \quad (1)$$

3 基于神经网络的 SP 获取模型

本文主要考察动词对宾语的语义选择限制, 因此论元类型 r 为动词宾语, 为了表述方便, 在后面的符号表达中将 r 省去。下面对选择限制获取的单隐层前馈网络、maxout 网络及模型训练方法作一介绍。

3.1 单隐层前馈网络

网络由输入层、隐藏层、输出层三层构成 (如图 1 所示)。输入层节点数为 $2N$, 隐藏层节点数为 H , 输出层节点数为 1。输入 \mathbf{x} 为动词 v 的词向量 \mathbf{v} 与宾语 n 的词向量 \mathbf{n} 的拼接 (concatenation), 即

$$\mathbf{x} = [\mathbf{v}, \mathbf{n}] \quad (2)$$

词向量的维度为 N , 因此 \mathbf{x} 的维度为 $2N$ 。动词和名词的词向量分开学习, 学习之前可以进行随机初始化或引入预训练的词向量, 在模型训练过程中通过反向传播进行更新。 \mathbf{W}_1 为输入层与隐藏层之间的权值矩阵, $\mathbf{W}_1 \in \mathbb{R}^{H \times 2N}$, \mathbf{b}_1 为偏置项, $\mathbf{b}_1 \in \mathbb{R}^H$, 隐藏层节点的激活函数为 \tanh , \mathbf{a}_1 为隐藏层节点的输出值, 见公式 3, $\mathbf{a}_1 \in \mathbb{R}^H$ 。 \mathbf{W}_2 为隐藏层与输出层之间的权值矩阵, $\mathbf{W}_2 \in \mathbb{R}^{1 \times H}$, 输出 y 为隐藏层输出的线性组合, 为一个实数, 见公式 4, 偏置项 \mathbf{b}_2 为实数, y 即为所求的动词 v 对宾语 n 的语义选择倾向 sp 。

$$a_1 = \tanh(W_1 x + b_1) \quad (3)$$

$$y = W_2 a_1 + b_2 \quad (4)$$

单隐层前馈网络结构简单，为了提高网络的性能，我们在词向量初始化时引入预训练的词向量，加入了额外的语言资源，并与随机初始化词向量的方法做对比。

3.2 两层 maxout 网络

Maxout^[15]可以看做神经网络中的一种的激活函数，与通常的非线性函数直接作用在输入的线性组合上以实现非线性变换不同，maxout 对每个输入进行 k ($k \geq 2$) 次线性组合，从中选择最大的值作为变换后的输出，达到非线性变换的目的。也就是说，对于一个输入，根据多组参数同时线性计算多组输出，然后取多组输出中对应位置的最大值构成最终输出。这样 maxout 会增加网络参数，但学习能力也得到了增强。理论上，maxout 可以拟合任意凸函数；在实践中，maxout 网络在多个机器学习任务上也取得了很好的效果。

我们对单隐层前馈网络结构进行调整，形成两层的 maxout 网络（如图 2 所示）。在原网络的隐藏层和输出层分别增加一倍的节点，形成两组节点 ($k=2$)，每一组节点个数与原网络该层节点个数相同，两组节点分别对所有输入进行线性组合，然后通过 max 函数选择两组节点中对应位置节点的最大值构成输出。公式 5 和公式 6 分别是隐藏层的输出 a_1 和输出层的输出 y ， $a_1 \in \mathbb{R}^H$ 且 y 是实数，表示动词 v 对宾语 n 的语义选择倾向 sp 。相比于单隐层前馈网络，maxout 网络增加了参数 W_1' 、 b_1' 、 W_2' 与 b_2' 。

$$a_1 = \max(W_1 x + b_1, W_1' x + b_1') \quad (5)$$

$$y = \max(W_2 a_1 + b_2, W_2' a_1 + b_2') \quad (6)$$

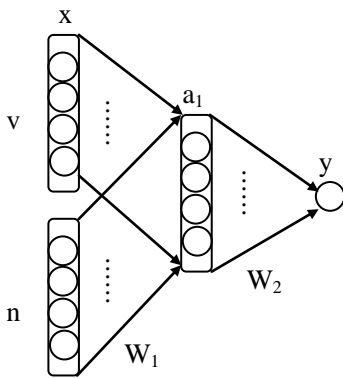


图 1 单隐层前馈网络结构

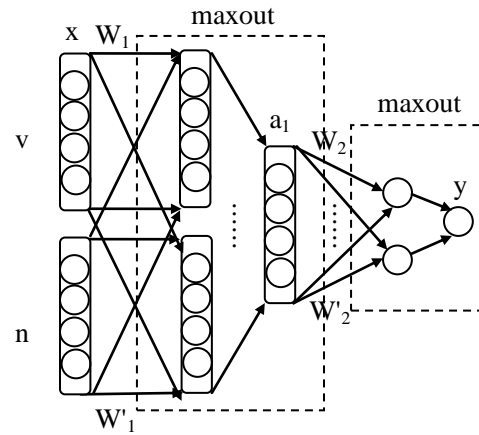


图 2 两层 maxout 网络结构

3.3 网络训练

语义选择限制知识获取的任务要求通过大规模无标注语料进行神经网络参数的训练。Collobert 和 Weston^[16]提出一种基于无标注语料训练神经网络语言模型的方法，即将语料中正常出现的 ngram 序列视为正例，随机替换其中某一个词后得到的序列为反例。与之类似，把训练集中出现的搭配 (v, n) 视为正例，而把 n 随机地替换为 n' 后形成的搭配 (v, n') 视为反例。我们关心的是模型对正例和反例打分的大小关系，期望正例的打分至少要比反例的打分大 l (l 的取值可以调整，这里取 $l=0.1$)，于是对于每一对搭配 (v, n) 定义以下排序目标函数

$$\sum_{n' \in J} \max(0, l - g[(v, n)] + g[(v, n')]) \quad (7)$$

表示正例 (v, n) 与所有反例的分差之和。其中 J 为名词词表， $g[\cdot]$ 是模型的打分。可见，当 $g[(v, n)] - g[(v, n')] \geq l$ 时，max 取值为 0。否则，max 取值大于 0。模型训练的目标就是

使公式 7 的值最小。

我们在训练时，总是给每一个正例随机生成一个反例，作为一对训练样本，由网络输出正例的打分和反例的打分，计算目标函数关于模型参数的导数，借助反向传播更新模型参数及输入词向量。

4 实验与分析

实验考察中英文两种语言中动词对宾语的选择限制情况，模型包括随机初始化词向量的单隐层前馈网络、引入预训练词向量的单隐层前馈网络、两层 maxout 网络及基于 LDA 的模型，使用伪消歧 (pseudo-disambiguation) [17] 的方法进行评价。

4.1 模型的实现

单隐层前馈网络输入词向量维数 $N=50$ ，输入层节点个数为 100，隐藏层节点个数 $H=50$ 。两层 maxout 网络使用一个 100 维输入 50 维输出的 maxout 网络和一个 50 维输入 1 维输出的 maxout 网络串联组成，每个 maxout 网络中都包含两个线性层。

实验在 Torch7 平台上进行，使用 mini-batch 梯度下降方法进行训练。每采样一个正例，随机生成一个反例，构成一个正反例对，每批抽取 512 个正反例对，每 256 批为一个周期，训练过程中记录每个周期内误差的均值，如果连续 20 个周期内没有出现更小的误差，认为网络收敛，停止训练。网络参数的初始学习速率为 0.00025，词向量的初始学习速率为 0.000125，开始训练时先以初始学习速率训练 8 个周期，来预热模型，使之快速的收敛到一定位置，然后学习速率开始随训练周期数下降 (初始学习速率/当前训练的周期数，未包含预热的 8 个周期)。

随机初始化词向量把词向量元素赋值为 $[0,1]$ 之间的随机浮点数。中文的预训练词向量采用 word2vec [18] 在 Chinese Gigaword 上训练出的词向量，词向量维数为 50，模型采用 Skip-gram，上下文窗口大小为 5，负采样，负样本个数为 5。英文预训练词向量采用 Glove (<http://nlp.stanford.edu/projects/glove/>) 在 Wikipedia 2014 与 English Gigaword 5 上训练出的词向量，词向量维数为 50。

作为比较的 LDA 模型采用 GibbsLDA++ (<http://gibbslda.sourceforge.net/>) 来实现，中文模型隐含主题个数设为 200，迭代训练 2000 次；英文模型隐含主题个数也设为 200，迭代训练 1000 次。其他参数均采用缺省设置。

4.2 评价方法

语义选择限制获取模型对动宾搭配打分，合理的搭配得分要高于不合理的搭配，伪消歧方法就是利用这种思想对选择限制获取模型进行评价。对于测试集中的每一个正例 (v,n) ，构造反例 (v,n') ，如果 $sp_r(v,n) > sp_r(v,n')$ ，则判断正确，记做 correct；如果 $sp_r(v,n) = sp_r(v,n')$ ，则记做 tie；否则，判断错误。正确率 *accuracy* 的计算公式如下：

$$accuracy = (\#correct + \#tie * 0.5) / \#allsamples * 100\% \quad (8)$$

中文测试数据使用 98 年 1 月的人民日报语料，从中抽取动宾搭配，并通过人工校对选取 1289 个作为正例，包括 365 个动词和 379 个名词。从 Penn Treebank 中选取动宾搭配 2500 个作为英文测试数据中的正例。

表 2 正反例举例

verb	pos	pre	post	rand	verb	pos	pre	post	rand
办理	案件	组织	理论	质量	kidnap	mailman	mailer	make	kind
参加	会议	情况	公司	观众	open	door	course	film	loss
担任	总理	朋友	家庭	业务	reject	offer	business	interest	change

反例中替代词 n' 的选择有不同的策略，比如随机选择 (按训练集中名词的真实分布)、选择词频相近的词等。我们采用三种选择策略，分别为：名词按词频降序排列，选择直接前

驱词 (pre); 名词按词频降序排列, 选择直接后继词 (post); 随机选择 (rand)。表 2 分别给出中英文测试集中正反例的例子。verb 是动词 v , pos 是正例中的 n , pre、post 与 rand 是三种策略下的反例中的 n' 。可见, 正例的搭配较合理, 而反例大多不合理, 但也可能存在合理的情况, 比如 “reject change”, 因此要进行更准确的评价, 还需要对测试集中自动生成的反例进行过滤或人工校对。

中文训练语料为人民日报 1995、1997-2005 共十年的语料, 使用 NiuParser^[19]进行分词、词性标注、依存句法分析, 抽取动宾搭配, 选择动词出现 30 次以上、名词出现 5 次以上的搭配。去掉测试集中出现的搭配 (包括正例和反例), 最后得到动宾搭配 8,877,091 对 (去重后搭配 1,709,168 对)。

英文训练语料为 English Gigaword 语料中的 Agence France-Presse (AFP, 2001-2010 年) 和 New York Times (NYT, 2001-2010 年), 从中抽取动宾搭配, 选择动词出现 30 次以上、名词出现 5 次以上的搭配。去掉测试集中出现的搭配 (包括正例和反例), 最后分别得到动宾搭配 25,004,475 对 (去重后搭配 1,912,264 对)、28,133,369 对 (去重后搭配 3,274,509 对)。

从训练数据中去掉测试样本中的所有搭配, 包括正例和反例, 这样做是为了保证所有的测试样本对模型来说都是没有见过的, 以使实验结果更能反映模型的泛化能力, 将重点放在谓语动词对未见论元的选择限制的预测上。

4.3 实验结果及分析

中英文语义选择限制获取结果分别见表 3、表 4 (粗体表示测试集上的最好结果)。方案 SHL 为单隐层前馈网络, 词向量随机初始化。方案 SHL+wv 为单隐层前馈网络, 词向量初始化为预训练词向量。方案 Maxout+wv 为两层 maxout 网络, 词向量初始化为预训练词向量。方案 LDA 为基于 LDA 的选择限制获取模型。方案 wv+cos 利用预训练的词向量直接计算 v 和 n 的相似度, 作为 sp 值, 相似度计算公式使用夹角余弦 $\cos(v, n)$, 该方案用作 baseline 与其他方案相比较。为了考察语料规模的影响, 我们在中文人民日报语料上做了三个实验, 分别把 10 年 (x10)、2 年 (x2, 2004、2005 年)、1 年 (x1, 2005 年) 的语料作为训练集。英文 nyt 训练集中的样本数多于 afp。

实验结果显示: (1) 预训练词向量比随机初始化词向量的结果有大幅度的提升。(2) Maxout 网络比单隐层前馈网络的结果又有所提升。(3) 引入预训练词向量的神经网络模型结果优于 LDA 模型。(4) 直接使用词向量计算相似度的方案 wv+cos 是一个很强的 baseline, 超过了 LDA 模型的结果, 但是神经网络模型仍然能取得更好的结果。(5) 模型可以在较小规模的训练集上取得不错甚至更好的结果, 因此不必一味增加训练集的规模。

中英文的测试结果相差很大, 中文的结果明显好于英文。原因分析如下: (1) 中文测试数据和训练数据都来自人民日报语料, 而英文的测试数据 Penn Treebank 语料来源于 Wall Street Journal, 不同于训练语料 Agence France-Presse 与 New York Times, 这种语料的差异是典型的领域迁移问题, 会带来性能的损失。(2) 中文测试集中的正例和反例均经过了人工验证, 而英文测试集没有经过人工验证, 随机选择 Penn Treebank 中出现的动宾搭配作为正例, 生成的反例只要没有出现在 Penn Treebank 中即可。这样可能会造成英文出现 “正例不正, 反例不反” 的情况, 即反例虽然在测试语料中不存在, 但有可能在原始训练语料 (剔除测试样本之前) 中出现的很多, 甚至比正例还多, 从而使测试结果变差。如表 5 所示, post 策略下三个反例在原始训练集 afp 中出现的次数均高于正例。在选择测试样本的时候可以使用共现频率、互信息等统计指标进行过滤 (如, 要求反例不在训练集中出现), 来保证样本的质量。

表 3 中文语义选择限制获取实验结果

	pre			post			rand		
	x10	x2	x1	x10	x2	x1	x10	x2	x1
SHL	47.54	47.09	48.98	49.69	50.63	48.03	50.87	49.36	48.97
SHL+wv	84.67	82.78	85.46	84.15	84.15	84.31	85.61	85.06	86.25
Maxout+wv	86.32	87.50	85.77	87.30	85.41	86.04	87.84	89.90	86.88
LDA	75.31	83.23	81.50	74.45	83.25	80.88	68.60	77.58	75.36
wv+cos	86.98			85.59			85.93		

表 4 英文语义选择限制获取实验结果

	pre		post		rand	
	afp	nyt	afp	nyt	afp	nyt
SHL	49.49	50.29	49.15	53.77	49.88	51.25
SHL+wv	65.67	67.00	65.99	65.78	67.98	67.19
Maxout+wv	65.67	67.54	66.31	67.18	67.27	66.53
LDA	52.19	52.21	52.94	52.55	47.34	46.66
wv+cos	64.22		64.39		59.63	

表 5 正反例 (post) 在原始训练集 (afp) 中的共现次数

v	n	count(v,n)	n'	count(v,n')
boost	revenue	553	investment	844
maintain	attitude	102	balance	348
postpone	sale	39	plan	229

5 总结与展望

本文提出基于神经网络的语义选择限制自动获取模型,包括引入预训练词向量的单隐层前馈网络和两层 maxout 网络模型,在中文和英文两种语言上的伪消歧实验中,两个模型的结果都优于 LDA 模型,引入预训练词向量的神经网络模型也可以取得比基于词向量的词语相似度方法更好的结果。

深度神经网络在语义选择限制知识获取上的应用还比较初步,我们下一步会对模型参数进行调优,深入分析实验结果,对模型进行改进,尝试其他的神经网络结构。在模型的对比方面,我们只是与前期工作中的 LDA 模型作了比较,由于数据获取及实验细节方面还存在一些问题,尚未与其他模型进行全面的比较,这将是下一步的工作重点。我们对中英文语义选择限制知识的获取进行了初步的实验,研究跨语言的语义选择限制知识获取,探索不同语言中的语义选择限制规律也是将来的工作之一。

致谢

感谢 Zhenhua Tian 分享的英文数据:从 Agence France-Presse 语料、New York Times 语料及 Penn Treebank 中提取的动宾搭配。感谢北京大学计算语言学研究所的李天时提供的中文预训练词向量。

参考文献

- [1] 贾玉祥, 俞士汶. 语义选择限制的自动获取及其在隐喻处理中的应用[C]. //第四届全国学生计算语言学研讨会(SWCL 2008), 2008, 90-96.
- [2] Guangyou Zhou, Jun Zhao, Kang Liu, et al. Exploiting Web-Derived Selectional Preference to Improve Statistical Dependency Parsing[C]. //Proceedings of ACL2011, 2011, 1556-1565.
- [3] Shumin Wu and Martha Palmer. Can Selectional Preferences Help Automatic Semantic Role Labeling?[C]. //Proceedings of Lexical and Computational Semantics (*SEM 2015), 2015, 222-227.
- [4] Diana McCarthy and John Carroll. Disambiguating nouns, verbs, and adjectives using automatically acquired selectional preferences[J]. Computational Linguistics, 2003, 29(4): 639-654.
- [5] 唐海庆, 熊德意. 基于选择偏向性的统计机器翻译模型[J]. 北京大学学报 (自然科学版), 2016, 52(1): 127-133.

- [6] Philip Resnik. Selection and Information: A Classed-Based Approach to Lexical Relationships[D]. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA, 1993.
- [7] Hang Li, and Naoki Abe. Generalizing case frames using a thesaurus and the MDL principle[J]. Computational Linguistics, 1998, 24(2): 217-244.
- [8] Stephen Clark and David Weir. Class-based probability estimation using a semantic hierarchy[J]. Computational Linguistics, 2002, 28(2): 187-206.
- [9] Diarmuid O S éaghdha. Latent variable models of selectional preference[C]. //Proceedings of ACL2010, 2010, 435-444.
- [10]Katrin Erk, Sebastian Pado, and Ulrike Pado. A Flexible, Corpus-driven Model of Regular and Inverse Selectional Preferences[J]. Computational Linguistics, 2010, 36(4): 723-763.
- [11]Zhenhua Tian, Hengheng Xiang, Ziqi Liu, et al. A Random Walk Approach to Selectional Preferences Based on Preference Ranking and Propagation[C]. //Proceedings of ACL2013, 2013, 1169-1179.
- [12]Shane Bergsma, Dekang Lin, and Randy Goebel. Discriminative Learning of Selectional Preference from Unlabeled Text[C]. //Proceedings of EMNLP2008, 2008, 59-68.
- [13]Tim Van de Cruys. A Neural Network Approach to Selectional Preference Acquisition[C]. //Proceedings of EMNLP2014, 2014, 26-35.
- [14]贾玉祥, 王浩石, 咎红英, 等. 汉语语义选择限制知识的自动获取研究[J]. 中文信息学报, 2014, 28(5): 66-73.
- [15]Ian J. Goodfellow, David Warde-farley, Mehdi Mirza, et al. Maxout networks[C]. //Proceedings of ICML2013, 2013, 1319-1327.
- [16]Ronan Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning[C]. //Proceedings of ICML2008, 2008, 160-167.
- [17]Nathanael Chambers and Dan Jurafsky. Improving the use of pseudo-words for evaluating selectional preferences[C]. //Proceedings of ACL2010, 2010, 445-453.
- [18]Tomas Mikolov, Kai Chen, Greg Corrado, et al. Efficient estimation of word representations in vector space[DB]. arXiv preprint, arXiv:1301.3781, 2013.
- [19]Jingbo Zhu, Muhua Zhu, Qiang Wang, et al. NiuParser: A Chinese Syntactic and Semantic Parsing Toolkit[C]. //Proceedings of ACL-IJCNLP 2015, 2015, 145-150.

贾玉祥（1981—），男，博士，讲师，主要研究领域为自然语言处理。E-mail: ieyxia@zzu.edu.cn



许鸿飞（1994—），男，硕士研究生，主要研究领域为自然语言处理与深度学习。E-mail: hfzunlp@foxmail.com



咎红英（1966—），女，博士，教授，主要研究领域为自然语言处理。

E-mail: iehyzan@zzu.edu.cn



作者联系方式：姓名 地址 邮编 电话（最好手机） 电子邮箱

贾玉祥，河南省郑州市科学大道 100 号郑州大学信息工程学院，450001， 15890056535， ieyxia@zzu.edu.cn