ISSN 1009-3044
Computer Knowledge and Technology 电脑知识与技术
Vol.10, No.32, November 2014
E-mail: eduf@dnzs.net.cn
http://www.dnzs.net.cn
Tel:+86-551-65690963 65690964

基于词向量空间的大规模中文语义网络构建与复杂性分析

曹茂元[1], 陈毅东[1*], 姚为龙[2], 杨雪娇[3]

(1. 厦门大学 智能科学与技术系, 福建 厦门 361005; 2. 中国科学院研究生院, 北京 100049; 3. 厦门大学 数学科学学院, 福建 厦门 361005)

摘要: 当前对于汉语语义层次的语言网络研究方法仅限于静态词典生成以及人工手动生成两种方法, 具有很大的局限性。对此, 该文从大规模语料库生成的语义空间出发, 结合语义空间丰富的语义信息和义类词典资源, 提出一种新颖的基于分布语义的语义网络构建策略, 并在此基础上探究了由不同性质的语义空间所构建的语义网络的统计特性。相比前人的方法, 该文提出的方法优势在于无需依赖人工标注, 支持大规模动态语料的网络自动构建。实验结果表明, 语义网络具有复杂网络两个典型的特性: 小世界效应和无标度特性。此外, 由于语义网络描述的是词之间最为本质的语义关系, 与不同文体中的措辞、使用习惯、风格等不存在直接的关系, 因此当语义网络节点到达一定规模时, 语义网络的某些统计特性可能会趋于一致。

关键词: 语义网络; 语义空间; 小世界; 无标度

中图分类号: TP391 **文献标识码**: A **文章编号**: 1009-3044(2014)32-7703-07

DOI:10.14004/j.cnki.ckt.2014.0998

复杂网络研究方法的出现使对语言网络进行大规模实证性研究成为可能[1]。语义研究是当前研究的热点,如何借助复杂网络方法研究语言的语义特性是一个十分关键的问题。唐璐、张永光等[2]在两个大型词典HowNet和WordNet基础上, 利用词典信息构建了两个语义网络。刘海涛[3]通过人工语义标注的语料, 构建了一个小型的语义网络, 借此探究语义网络的复杂特性。Steyvers和Tnenenbaum[4]利用WordNet、罗杰分类词典等资源分别构建了大规模英语语义网络, 并对其进行复杂统计分析。但现有的工作依赖手工标注或者完全借助词典(如WordNet)的方法来构建语义网络, 这些方法数据规模小, 移植拓展性差, 无法很好的说明问题。而分布语义是语义表示的重要方法, 由大规模语料所构建的语义空间里已经包含了可以计算的语义信息。其优势是不需要依赖人工标注, 可以从语料中获得大量语义表示。如果能从大规模分布语义空间中自动构建语义网络并应用复杂网络方法加以探究, 则将能很好地推动语义网络复杂特性方面的研究。该文将开展这方面的工作。该文主要关注中文的情况, 但相关的方法也可以扩展到其他语言。

1 网络统计参数及相关概念介绍

常用的复杂网络统计参数有平均最短路径、聚集系数、度分布和累计度分布[5]。分别介绍如下:

网络的平均路径长度$\langle d\rangle$定义为任意两个节点之间距离的平均值, 即:

$$\langle d\rangle = \frac{1}{\frac{1}{2}N(N+1)}\sum_{i\geq j} D_{ij}$$

其中, N为网络节点数, D_{ij}为节点i和j的最短路径长度。

假设节点i有k个邻居节点, 在k个节点之间最多有$k(k-1)/2$条边。而这k个节点之间实际存在的边数E和可能的总边数之比即为节点i的聚集系数C_i:

$$C_i = \frac{2E}{k(k-1)}$$

而整个网络聚集系数C为所有节点聚集系数C_i的平均值。

若网络的平均路径长度$\langle d\rangle \approx \langle d_r\rangle$, 且聚集系数$C \gg C_r$, 则称该网络具有小世界特性[6]。其中, $\langle d_r\rangle$、C_r是相同节点和边数下的随机网络的平均路径长度和聚集系数。

现实中大多数网络的度分布都具有幂律形式, 即满足$p(k)\propto k^{-}\gamma$。其中$p(k)$为度为k的节点出现在网络中的概率。

而累积度分布表示的是度不小于k的节点的概率分布, 即:

$$P(k) = \sum_{k=k}^{\infty} P(k^{'})$$

若度分布为幂律分布, 即$P(k)\propto k^{-}\gamma$, 那么累积度分布符合幂指数为$\gamma-1$的幂律, 即:

$$P(k) \propto \sum_{k=k}^{\infty} k^{'-\gamma} \propto k^{-(\gamma-1)}$$

收稿日期: 2014-09-25

作者简介: 曹茂元(1989-), 男, 硕士, 主要研究方向为自然语言处理; 陈毅东(1977-), 男, 副教授, 工学博士, 主要研究方向为自然语言处理、机器翻译等。

如果一个网络的度分布服从幂律：$P(k) \propto k^{-\gamma}$，并且幂指数γ的值在2到3之间，则这样的网络为无标度网络[7]。

Harris提出语言学的分布假设[8]：两个词之间的相似度可由它们共现词的分布相似度近似，换而言之，即具有相似上下文的词具有相似语义。这里，我们对基于分布假设理论计算的相似度给出定义，称为分布语义相似度：

定义**1.1** 分布语义相似度，指在分布假设理论下，通过借助上下文共现分布的相似性对两个词相似性进行的度量。

从上文可知，分布语义相似度的计算是根据两个词语出现的上下文重叠程度计算它们之间的相似度，换而言之，上下文背景越相似，词的相似度就越大。目前对分布语义的表示、比较，采用的是基于向量空间模型的语义空间的方法[9]。由于语义空间内蕴含着丰富的语义信息，因此在语义空间的基础上构建语义网络是具有理论依据且十分有意义的。

2 语义网络构建算法

本文算法从语义空间出发，利用义类词典信息将语义空间转化为一个语义网络，基本框架如图1：

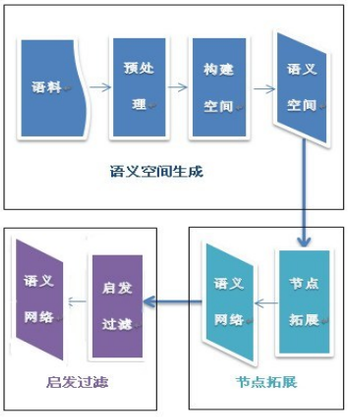


图1　基本框架

首先是语义空间的生成，该文分别从不同角度构建了两种基于Word-Context的语义空间，即词空间和依存空间；其次是节点拓展，由于语义空间提供了丰富的节点信息，因此构建基于动态文本的语义网络自然使用已有节点，从而只需进行拓展边的信息。最后，通过实验发现，由于语义空间所描述的分布语义相似度的复杂性，其所描述的语义关系混合了语义相似与语义相关两种关系，因此单单借助语义空间所生成的语义网络与实际手工标注的语义网络具有较大差异，对此本文提出2条启发式过滤规则来对初始语义网络进行边的过滤，从而得到更近似于人工生成的语义网络。

2.1 词空间与依存空间

本文从词共现关系构建的词空间和基于依存关系构建的依存空间两个角度介绍语义空间的生成。两种不同的语义空间构建方法区别在于上下文选择方式的不同，前者基于简单的共现链，而后者则引入依存信息。

2.1.1 词空间生成

多维空间类比(HAL)[10]是词共现语义空间的代表性方法，主要思想是利用一个窗口进行移动，窗口内的词即和目标词拥有共现关系，并考虑与目标词的前后关系构成共现矩阵。HAL的主要问题是高频度的特征维度对距离的度量的贡献与其所带的信息量不成比例。为了减少高频上下文的影响，该文采取了通过对频度进行规范化的COALS算法(correlated occurrence analogue to lexical semantic，COALS)[11]。算法如下：

首先定义空间生成算法里几个重要变量符号及函数，包括：

表1 词空间生成算法变量表

t	target，目标词，指将出现在语义空间中的词，不包括停用词；
b	basis element，基元，指在语义空间中成为维度坐标的单元；
$M[t][b]$	目标词t在维度b下的语义矩阵单元

相对应的函数：

1) 上下文选择函数cont(t)决定了与目标词t同处一条共现链的词，即以目标词t为中心、大小为windowSize的窗口内的词。

2) 权值赋值函数n对共现词赋以权值。该文赋权方法是取其与目标词t间的间距词数目与窗口大小之差的绝对值。例如，目标词后面一个词的权值等于窗口大小，第二个词权值为窗口减1，依次类推直至为0。

3) 基元映射函数m决定共现词所在维度。该文取共现链终端的词，比如对共现链“呼吁/v——武装/n——分子/n”，则该共现链对应的维度即为终端词“分子/n”。

步骤一：原始矩阵建立简明算法如表2。

步骤二：矩阵变换

在对原始矩阵转换为规范化的矩阵前，需要将部分维度舍弃以保留所需的重要的维度，通常的做法是进行按频率递减排序，

表2 词空间原始矩阵建立算法

```
1   ∀basis element b:∀target t:
2   Initialize matrix cell M[t][b] with 0,frequecncy vector wordFreq with 0
3   for every target word t
4       Compute local context cont(t)
5       wordFreq(t)+= 1
6       For every word w in the cont(t)
7           b = μ(w)
8           if(notStopWord(b))Increment M[t][b] by path value v(w)
9       end
10  end
```

舍弃频数较低的维度;然后采用计算对数似然比[12]来对原始矩阵进行转换。

2.1.2 依存语义空间生成

上述词空间生成方法仅仅考虑到简单上下文里的分布统计信息,而忽略了重要的句法信息,而语义分析与句法分析有着重要的联系,因此,在语义空间的构建时考虑到句法信息是有意义的。该文句法分析采用哈尔滨工业大学信息检索研究室所发布的LTP句法分析模块,算法参考自文献[13]。算法如下:

首先对空间生成算法里的变量、上下文选择函数等进行定义:

表3 依存空间生成算法变量表

t	target,目标词,指将出现在语义空间中的词,不包括停用词;
b	basis element,基元,指在语义空间中成为维度坐标的单元;
M[t][b]	目标词t在维度b下的语义矩阵单元
p	path,依存关系路径(在依存树内)

以及对应的函数:

1) 上下文选择函数cont(t)决定了依存树的所有路径里用来表示目标词t的路径,其中每条路径起点为目标词t,路径长度受参数windowSize影响。比如,若windowSize取1,则仅考虑从目标词t出发且长度为1的路径。

2) 路径赋值函数n对路径赋以权值,从而能够考虑到语言信息。赋值函数有:根据路径上最显著句法关系赋值,比如在路径上存在"动宾关系",则赋以权值4;相同赋值,即n(p)=1;以及本文作者提出的结合路径句法关系和与路径长度比的赋值方法。为与COALS保证可比性,该文此处函数n同COALS算法。

3) 基元映射函数m决定了路径p所对应的语义空间的维。该文此处取路径终点所在词。比如对于路径"呼吁/v —— 释放/v —— 外交官/n"中,该路径所对应的维度为终点"外交官/n"。

步骤一:建立原始矩阵如表4。

表4 依存空间原始矩阵建立算法

```
1   ∀basis element b:∀ target t:
2   Initialize matrix cell M[t][b] with 0,frequecncy vector wordFreq with 0
3   for every target word t
4       Compute local context cont(t)
5       wordFreq(t)+= 1
6       For every path p in the cont(t)
7           b = μ(π)
8           if(notStopWord(b))Increment M[t][b] by path value v(π)
9       end
10  end
```

步骤二:变换矩阵

同COALS一样,需要对原始矩阵进行规范变化以避免频率偏差带来的影响,同样使用对数似然比进行变换。

2.2 节点拓展

利用语义空间自带的丰富的语义信息可以计算两个词之间的分布语义相似度,将相似度高于一定阈值的两个词连边,认为二者具有语义关系,从而将语义空间拓展成对应的语义网络。

对每一个当前进行拓展的新节点(拓展词)分配集合NewSet保存该节点拓展信息,集合OldSet保存已拓展词的历史信息。考虑

到复杂度以及作为基元(维度)的词的丰富语义信息,该文采用贪心思想进行节点的拓展来生成语义网络,即假定词*w1*的语义向量对应某基元的值大于某个阈值ε,则认为词*w1*与该基元存在语义关系,则将二者相连,并将拓展到的节点(基元)加入集合NewSet。在此假定下,继续按相同方法拓展基元直至无可再拓展基元,则认为该词*w1*拓展结束。为防止出现不连通图,即若出现NewSet和OldSet两集合不相交的情况,则以概率1/size(OldSet)将两个集合进行连边,否则计算拓展词与OldSet里非基元词的相似度进行连边。最后将NewSet并入OldSet中。

算法中函数定义如下:

1) 基元集合函数**biasis**(***w***,**e**):决定词*w*语义向量里值大于阈值ε的基元的集合,且集合内元素称之为词*w*的孩子。

2) 语义相似度计算函数**similarity**(***w1,w2)*** :计算词*w1*、词*w2*之间的分布语义相似度,该文采取余弦值作为相似度的度量。

3) 词之间连边函数 **addEdge**(***w1,w2)*** :将词w1,w2之间连上边

4) 随机连边函数 **RandomAddEdge**(***w***,***p***):以概率*p*将词*w*与已拓展节点相连。

算法描述如表**5**:

表5 语义网络构建算法

```
1   for every word w in semantic space
2       NewSet.clear();
3       Queue.add(w);
4       connected = false;
5       while(! queue.isEmpty())
6           t = Queue.poll();
7           NewSet.add(t);
8           bSet = biasis(t, ε);
9           for every element b in bSet
10              addEdge(t,b);
11              if(connected)
12                  if(!OldSet.conatains(b))
13                      OldSet.add(b);
14                      Queue.add(b);
15                  end if
16              else if(OldSet.contain(b))
17                  connected = true;
18                  for every word w2 in OldSet
19                      If(similarity(w,w2)> φ)addEdge(w,w2);
20                      end if
21                  end for
22                  OldSet.addAll(NewSet);
23              else
24                  NewSet.add(b);
25                  Queue.add(b)
26              end if
27          end for
28      end while
29      if(!connected)
30          RandomAddEdge(w,1/OldSet.size);
31          OldSet.addAll(NewSet);
32      end if
33  end for
```

按本节所提算法对语义空间进行边的拓展生成语义网络,但发现其与人工标注生成的语义网络结构存在较大的差异,主要原因是由于语义分布相似度描述的特性混合了相似性与相关性,因此产生了多余的、与语义分析相违背的连边,故需要对所生成的初始语义网络里不合理的边进行过滤,以生成更接近人工生成的语义网络。

2.3 过滤不合理的连接

本节首先对相似性和相关性给出定义及其度量方法,再据此提出2条启发式的过滤规则,实现对语义网络里不合理的连接进行过滤。

2.3.1 相似性与相关性

定义**1** 相似性是指两个词在不同的上下文中可以互相替换使用而不改变文本的句法语义结构的特性[14]。

其中，主要描述了近义、反义、上下位等关系[14]。比如“苹果”和“梨”就具有强相似性，在使用“苹果”的情景里多可以被“梨”替换而不影响句法语义结构，比如“吃苹果”和“吃梨”。相似性强弱经常使用义类词典进行量度，常用的方法为基于知网（HowNet）的相似性度量[14]。由于语义网络不包含虚词，故本文仅专注于实词在知网体系中相似度的计算。

语义相似性可借助词典资源的方式进行度量，而语义的相关性则是一个模糊的概念，没有明确的客观标准可以衡量。该文从语义分析角度给出一个语义相关性的定义：

定义**2** 语义相关性指的是在语义分析中，两个词能够组成部分-整体、属性-宿主、施事-事件、受事-事件、时间/地点-事件、材料-成品等关系的特性，是概念相关程度的刻画。

由该定义可知，“苹果”和“梨”的相关性较弱，而“苹果”和“吃”由于存在受事-事件关系，因此具有较强相关性。在基于语义分析构建语义网络的方法中，两个节点是否相连正是取决于二者是否具有较强的相关性，故在语义分析里将节点“苹果”和“吃”相连，体现二者之间的语义关系。

2.3.2 过滤规则

节点拓展算法里，对两个词的是否有语义关系的判断是通过计算两个词的分布语义相似度得到的。但在实际研究中，分布语义所描述的两个词的相似度往往并不是单一类型的语义特性，即具有高分布语义相似度的两个词可能具有较强的语义相似性(定义1)，也可能具有较强的语义相关性(定义2)。以读者语料库生成的、基于依存语法所生成的语义空间为例，若单纯考虑分布语义相似度高的点进行连边，则与“皮帽/n”和“戴/v”的连边生成的网络如图2：

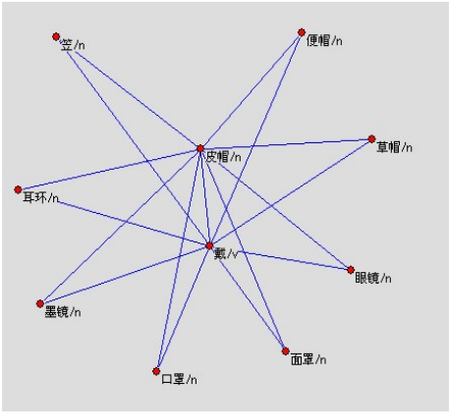


图2 “皮帽/n”与“戴/v”连接示意图

可以看到，虽然将相关性强的“皮帽/n”和“戴/v”连线，但同时也将相似性强、在一般构建的语义网络中不会连线的“皮帽/n”和“草帽/n”等连线。因此，我们提出了如下启发式规则，该规则借助知网相似性信息过滤强相似性的词对边，仅保留强相关性的词对边：

启发规则**1** 若词*w1*,词*w2*具有高分布语义相似度(大于阈值ε),且语义的相似性较弱（语义的相似性强弱的度量基于词*w1*和*w2*在知网里的相似性刻画，小于δ），则认为词*w1*,词*w2*具有较强的相关性，反之则相关性较弱。

以上文例子为例，“草帽/n”与“口罩/n”的相似性量化值为0.32，而“草帽/n”与“戴/v”的相似性度量则为0.04，因此在过滤规则下，“草帽/n”与“口罩/n”之间的边应该舍弃。运用该启发规则，过滤后的网络如图3：

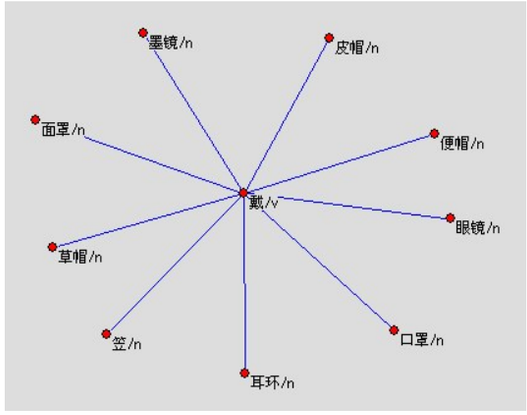


图3 “皮帽/n”与“戴/v”连接示意图

正如前文所言，分布语义的计算基于上下文的重叠程度，因此具有高分布语义相似度的词之间可能会由于具有高重叠的上下文从而产生传递的现象，即词A和词B具有高分布语义相似度，词B和词C也具有高分布语义相似度，可能就有词A和词C具有高分布语义相似度。如图4，“水汪汪/a”除了与“眼睛/n”具有高分布语义相似度外，又与“眨/v”、“瞎/a”等也具有高分布相似度，这正

是由于后者与中间词“眼睛/n”的高相似度所造成的传递现象。而又由于双方与中间词所描述语义相关关系不同,故二者也会具有弱相似性。因此需在启发式规则1外新增规则2。

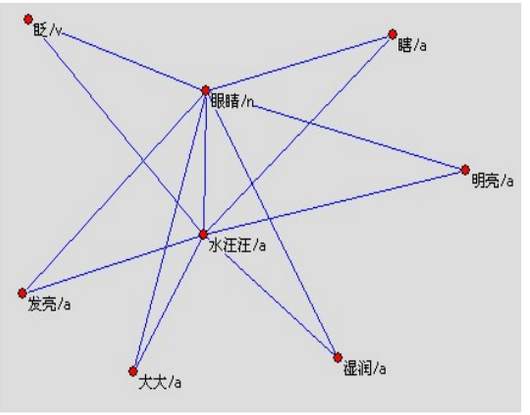


图4 “眼镜/n”与“水汪汪/n”连接示意图

启发规则**2** 若词*w1*、*w2*、*w3*间两两符合规则1,则过滤掉分布语义相似度最低的边。

运用启发规则2后,生成网络如图5:

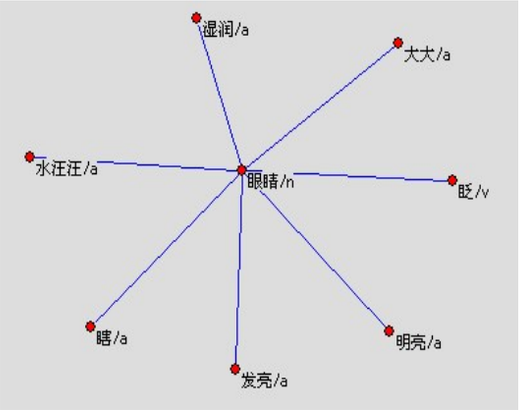


图5 “眼镜/n”与“水汪汪/n”连接示意图

由上分析得出,通过对语义空间语义信息和义类词典知识的综合使用,能够对词之间的语义相关关系进行有效挖掘,从而达到句法分析所无法达到语义分析效果。

3 实验与结论

为了探究不同的语义空间对最终生成的语义网络的影响,该文分别在词空间和依存空间上进行了实验。此外,该文对不同文体(小说、散文、新闻、综合性文选)的语义网络在统计特征上的差异性以及其复杂特性做了详细的探究。

对于不同文体,该文分别采集了以下几个数据集:

综合性文选:《读者》语料库,包含词数约655万,分别用DZ1,DZ2来表示利用该语料词空间和依存空间生成的网络;

新闻:《人民日报》2000年上半年语料库,包含词数约630万。分别用XW1,XW2来表示利用该语料词空间和依存空间生成的网络;

散文:以当代著名散文家余秋雨等作家为主的散文集,包含词数约434万。分别用SW1,SW2来表示利用该语料词空间和依存空间生成的网络;

小说:选取了我国长篇小说茅盾文学奖第三届至第八届的23部获奖作品,词数约647万,分别用XS1,XS2来表示利用该语料词空间和依存空间生成的网络。

对文档的预处理首先是进行过滤非法字符、无效字符等,其次是进行分句。最后最重要的是分词及词性标注,以及为生成依存空间还需进行句法分析,该文实验采取的工具是在哈尔滨工业大学信息检索研究室所发布的LTP词性标注及句法分析模块基础上进行了包装。其次,对停用词的处理,由于语义网络节点为实词,因此本文将除了三大实词(名词、形容词、动词)之外的词均作为停用词处理,通过词性即可分辨,效率更高。另外,对属于三大实词但所带语义信息极少或不适合在语义网络作为节点的特殊词,比如“要/v”、“是/v”、“有/v”等,也作为停用词处理。

本文在上述4种语料基础上一共生成8个语义空间,维度均为1500维。为增强实验的可比性,该文生成的网络保持相同的节点数,即取8个原始语义空间中最小词数。对词的选择方法是将所有词按词频排序,超过限制范围的词将不予考虑。实验结果如表6:

表6　语义网络统计特性

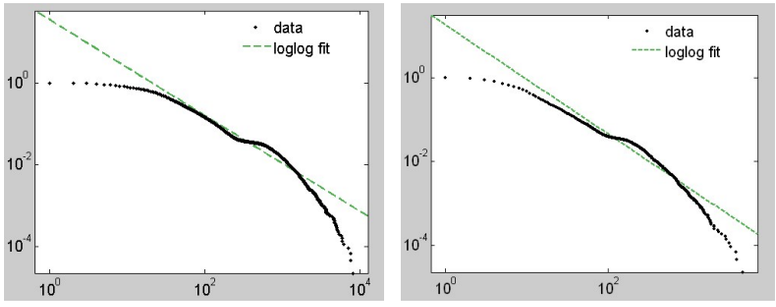
网络类型		N	E	<k>	C	<d>	γ	C_r	<d_r>
词空间	DZ1	44070	603466	27	0.066	3.48	2.30	0.0003	4.38
	XW1	44070	1190930	54	0.098	3.16	2.17	0.0006	3.84
	XS1	44070	1013059	46	0.089	3.17	2.23	0.0005	3.96
	SW1	44070	838694	38	0.087	3.25	2.25	0.0005	4.10
依存空间	DZ2	44070	1836067	83	0.110	2.85	2.17	0.0009	3.58
	XW2	44070	1740410	79	0.111	2.96	2.14	0.0009	3.62
	XS2	44070	1608450	73	0.104	2.95	2.18	0.0009	3.66
	SW2	44070	1319273	60	0.102	3.01	2.20	0.0006	3.78
	XWLB	5903	22018	7.4	0.079	3.95	2.49	0.0011	4.55

注：网络XWLB为刘海涛人工生成的语义网络[18]。

对相同节点、不同的语义空间转化成语义网络，总体上来说，依存空间最终生成的语义网络边数更加繁多，平均度更高，这是因为词空间相比于依存空间更稀疏，节点之间互相表示能力弱于依存空间；对于相同的空间类型、不同文体生成的语义网络，边数大体有XW≥XS≥SW，在词空间中DZ1边数接近于SW1居于末位，而在依存空间中DZ2却接近于XW2居于首位。

由表6可以看出，生成的8个语义网络均具有较小的平均最短路径（$<d>\approx<d_r>$）和较大的聚集系数（$C >> C_r$），表明虽然基于不同角度生成语义空间，但由语义空间转换成的语义网络均符合小世界特性。因此，可以得出语义网络是一种小世界网络的结论。这与刘海涛[3]手工生成的语义网络的研究结论是一致的。此外，依存空间生成的语义网络较词共现空间生成的语义网络有更明显的聚集现象，这很可能是由于词共现空间本身的稀疏性导致较少的边被加入，因此节点的平均度<k>和聚集系数远远小于依存空间生成的语义网络。另外，相同的语义空间类型，利用不同的文体语料所生成的网络在小世界特性上表现出十分相似的统计特性。由此可以认为，由于语义网络描述的是词之间最为本质的语义关系，与不同文体中的措辞、使用习惯、风格等不存在直接的关系，因此当语义网络节点到达一定规模时，语义网络的统计特性可能会趋于一致。

将8个语义网络累计度分布曲线进行线性拟合（其中读者语料生成的语义网络的累计度分布曲线拟合情况见图6），得出其均服从幂律分布，且幂律指数在2到3之间（见表6），表明语义网络是一种无标度网络。



1) DZ2　　(b) DZ1

图6　ＤＺ语义网络累计度分布拟合曲线

4 总结及展望

语义网络介于句法网络和概念网络之间，是人类知识的高级表示。而当前对语义网络的研究仅有人工手动生成与使用义类词典资源两种方式，对进行大规模语义网络研究有很大的局限性。由于语义空间内蕴含着大量准确而丰富的语义信息，因此本文提出了基于语义空间和义类词典资源结合的语义网络生成算法，能够对大规模语料进行语义网络复杂特性的探究，网络节点更加丰富，更能体现语言在真实文本中的动态特性。实验结果发现：基于语义空间生成的语义网络符合小世界和无标度特性；当语义网络节点到达一定规模时，语义网络的某些统计特性可能会趋于一致；一定规模下，不同方式生成的语义空间对最终生成的语义网络的某些统计特性不会造成重大的影响。未来的工作主要有：分布语义受训练文本的影响较大，也仅能表现出现在文本内的语义，因此在一个更大规模语料上进行本文的研究是必要的；当前对语言网络的研究还多局限于总体宏观统计特性，在未来的研究工作中应该关注于复杂网络局部所表现出的特性，比如社区发现等。

（下转第7711页）

1.2 学生功能模块

学生登录前就能看到系统的公告。公告中显示的内容有:答疑教师的信息、每周固定的答疑时间、有无临时的时间调整变动、系统设定的每次最多允许登录进答疑教室提问的人数、学生最晚可以进入答疑教室的时间等。

学生点击学生入口,输入用户名、密码后登录系统。

学生最早可以排队进入答疑教室的时间,一般为答疑开始前几个小时。即答疑教室没有开放前学生登录的话系统会显示答疑时间未到。为了避免考试前答疑人数急剧增多,学生扎堆登录进答疑教室的问题,设计一个不固定的开放时间,可以是答疑开始前的1至4小时的任意时刻。在答疑教室开放后,只要未达到规定的人数,学生就会登录成功。一旦登录成功就进入了排队等待状态,教师正式开始答疑前系统每隔几分钟自动发送确认在线的问题,题目为简单的整数加法,要求在短时间内正确回答,不回答或回答错误自动从排队的队伍里剔除。如果学生登录时答疑教室内的人数已满,系统会自动提示:本次登录人数已达到最大值,下次答疑时间请早点登录。

教师正式开始答疑后,进入答疑教室的学生可以提问,可以看到教师的回答,也能看到其他同学的问题与教师的解答。就如同现实生活中的教师答疑一样。

1.3 教师功能模块

教师点击教师入口,输入用户名、密码后登录系统。

1)进入答疑教室答疑。答疑时能看到当前在线人员情况。正常情况下教师应坚守岗位一直到答疑时间结束。答疑过程中如果教师有紧急的事情要处理,可通知学生后离开。类似正常上课时教师生病或有更高级别的事情必须马上去完成。当然为了保证答疑的时间与质量,管理方可对教师的迟到、早退等做出相应的规定。

2)教师申请答疑时间临时调整,管理员审核后发布。类似于现实生活中的调课、停课、补课。

2 结束语

仅靠教师的自觉与奉献是远远不够的,学校应制定一些奖励制度与激励机制,给从事答疑解惑的教师物质上与精神上的支持。比如将教师的答疑工作量核算为正常上课的教学工作量。这样系统会更加稳定高效地运行。

参考文献:

[1] 毛养红.在线答疑系统设计与实现[J].华南理工大学,2010.

[2] 韩璐.基于MVC模式的在线答疑系统设计与实现[D].大连:辽宁科技大学,2012.

(上接第7709页)

参考文献:

[1] 刘海涛.语言网络:隐喻,还是利器? [J].浙江大学学报:人文社会科学版,2011,41(2):169-180.

[2] Tang L, Zhang Y G, Fu X. Structures of semantic networks: How do we learn semantic knowledge[J]. Journal of Southeast University (English Edition), 2006, 22(3):413-417.

[3] 刘海涛.汉语语义网络的统计特性[J].科学通报,2009,54(16):2781-1785.

[4] Steyvers M, Tenenbaum J B. The large-scale structure of semantic networks: statistical analyses and a model of semantic growth[J]. Cognitive Science: A Multidisciplinary Journal, 2005,29(1): 41-78.

[5] 汪小帆,李翔,陈关荣.复杂网络理论及其应用[M].北京:清华大学出版社,2006.

[6] Watts D J, Strogatz S H. Collective dynamics of 'small-world' network[J].Nature,1998, 393(6648):440-442.

[7] Barab A L, Albert R. Emergence of scaling in random networks[J].Science, 1999,286(5439):509-512.

[8] Harris Z S. Distributional structure[M]. Springer Netherlands, 1970.

[9] Jurgens D, Stevens K. The S-Space package: An open source package for word space models[C]//Proceedings of the ACL 2010 System Demonstrations. Association for Computational Linguistics, 2010: 30—35.

[10] Burgess C, Cottrell G. Symposium at the cognitive science society conference : using high - dimensional semantic spaces derived from large text corpora[C]//Proceedings of the Cognitive Science Society. Hillsdale, NJ: Erlbaum Publishers, 1995:13-14.

[11] Rohde D L T, Gonnerman L M, Plaut D C. An improved model of semantic similarity based on lexical co-occurrence[J]. Communications of the ACM, 2006,8:627-633.

[12] Pado S, Lapata M. Dependency-based construction of semantic space models[J]. Computational Linguistics, 2007, 33(2): 161-199.

[13] 刘群,李素建.基于《知网》的词汇语义相似度计算[C]//第三届汉语词汇语义学研讨会论文集.台北,2002,7:59-76.

[14] Agirre E, Alfonseca E, Hall K, et al. A study on similarity and relatedness using distributional and WordNet-based approaches[C]//Proceedings of Human Language Technologies: The 2009 Annual Conference of the North America Chapter of the ACL. Association for Computational Linguistics, 2009:19-27.