

基于句法和语义的需求依赖关系自动获取

关 慧,吕 颖,贾成真

(沈阳化工大学 计算机科学与技术学院,辽宁 沈阳 110000)

摘 要:为了研究复杂系统中需求之间的依赖关系,针对需求文本固有的特点以及可能存在的上下文关系,结合各个词之间的词义以及特定语境下的语义知识,提出了一种基于句法和语义的需求依赖关系自动获取方法。在以自然语言描述的需求文本上,首先,通过对每条需求提取关键词,并结合领域知识对其添加领域词汇,对每个关键词进行词性标注并封装成集合来形式化需求;其次,通过对不同需求的关键词做语义分析,定义关联需求的语义词库表,将关系词库和集合中的词性特征结合起来判断两个需求之间的依赖关系,并给出抽取关系的方法,基于该方法设计并实现了一个系统,通过自动分析各个词之间的语义依存关系得出需求之间的依赖关系。最后,通过实例证明了该方法的可行性和有效性。

关键词:需求文本;依赖关系;语法词性;关系词库;语义;关键词集合

中图分类号:TP311.5

文献标识码:A

文章编号:1673-629X(2021)02-0020-07

doi:10.3969/j.issn.1673-629X.2021.02.004

Automatic Acquisition of Requirement Dependency Based on Syntax and Semantics

GUAN Hui, LYU Ying, JIA Cheng-zhen

(School of Computer Science and Technology, Shenyang University of Chemical Technology, Shenyang 110000, China)

Abstract: In order to study the dependency relationship between requirements in complex systems, based on the inherent characteristics of the requirement text and the possible contextual relationship, combining the meaning of each word and the semantic knowledge in a specific context, we propose a syntactic and semantic-based method for automatically acquiring requirement dependencies. On the requirement text described in natural language, first, by extracting keywords for each requirement and adding domain vocabulary to it with domain knowledge, each keyword is part-of-speech annotated and encapsulated into a set to formalize the requirement. Second, by performing semantic analysis on keywords of different requirements, we define a semantic vocabulary of related requirements, and combining the relational lexicon and the part-of-speech features in the collection, we determine the dependency relationship between the two requirements and give the method of extracting the relationship. Based on this method, a system is designed and implemented, and the dependency relationship between requirements is obtained by automatically analyzing the semantic dependency relationship between each word. Finally, it proves the feasibility and effectiveness of the proposed method according to an example.

Key words: requirements documents; dependency; grammatical part of speech; relational thesaurus; semantic; keyword collection

0 引 言

需求工程通过一种系统化的工程过程实现涉众需求的抽取、协商和文档化。作为需求工程输出制品的需求文档和需求规约是所有其他系统开发活动的重要基础^[1]。在软件生命周期的概念建模阶段,需求的常用表示方式有用例和基于自然语言的文本形式。基于自然语言的文本收集了各类需求信息,因为自然语言本身就有多个方面含义,并且对于自然语言的理解会由于各种因素的影响而不断改变,因此在处理自然语言

以及需求关系分析的过程中也离不开人的参与。但是这并不是说自动建模工具的辅助没有存在的必要。如果遇到一个比较复杂的系统时,需求文本的处理工作量是相当大的,为了能够尽可能缩短处理的时间,减轻需求分析师的工作负担,一些自动化工具或者系统就显得尤为必要。目前在UML模型的自动生成上已有一些相关的工作。利用UML建造系统时,在系统开发的不同阶段有不同的模型,并且这些模型的目的是不同的,例如用例图、类图和活动图等。这些方法的处

收稿日期:2020-03-30

修回日期:2020-07-30

基金项目:辽宁省自然科学基金项目(201602583)

作者简介:关 慧(1976-),女,博士,副教授,CCF会员(76281M),研究方向为软件安全性演化;通讯作者:吕 颖(1995-),女,硕士研究生,研究方向为软件安全。

理过程可以总结归纳为:首先从文本中获得特定的信息,然后定义一系列的转换规则,最后再通过这些规则将需求文本映射到对应的图形化模型上。

需求文本由于它独特的领域知识以及潜在的语义而不同于普通文本:(1)需求文本通常不仅包含需求制品(即所描述的目标、场景、功能需求、质量需求和约束),还包含一些重要的上下文关系。(2)需求文本由于它特定的领域知识,一些领域词汇被定义语法词性。如:“联系信息”一词在需求文本中被定义为名词。(3)自然语言描述的需求文本也有一些本质性的问题,如固有的二义性。

为了解决上述问题,文中在自然语言表示的需求文本的基础上,施加语法限制,降低了自然语言描述需求存在的二义性等问题。通过定义语义词库,联系到需求文本的语义关系,实现了从大量的需求文本中自动抽取需求依赖关系的方法。基于该方法,设计并实现了一个自动从中文需求文本生成需求间依赖关系的原型系统,对需求依赖关系进行自动获取。最后,通过案例验证及对比,证明了方法的可用性和切实性。

1 相关工作

在需求自动建模开发工具上,李天颖等人^[2]基于依存文法,依次对自然语言进行分词、词性标注及语法分析,设计并实现了一个系统,从中文需求文本自动生成 i* 框架中的 SD(strategy dependency,策略依赖)模型,侧重抽取策略依赖关系信息,他们研究的同一个需求内部之间的依赖关系。Chetan Arora 等人^[3]应用 NLP 自动识别需求语句的组成短语,然后计算所有标识的成对相似度分数,输出句法和语义相似性函数,以判断需求之间存在关系,但他们并没有得出需求之间存在的关系类型。韩德帅等^[4]将可视化的 UML 与严格化的时间自动机相结合,但该方法侧重于软件的实

现阶段,不支持软件自适应的需求建模和形式化验证。

在定义需求依赖关系类型上,国内外学者均做了若干相关研究。罗术通^[5]依据 UML 的定义把需求定义为七种原子关系:调用、中断、唤起、修改、通知、控制、资源控制,其次,提出了需求簇的概念以及基于需求之间关系划分需求簇的方法。关系的两个主体是直接相互影响彼此,并不依赖于其他主体,然而,这种关系并没有形式上的语义。Arda Goknila 等人^[6]研究的重点是业务需求的演变所带来的需求变化,旨在通过使用形式来改进需求中的变更影响分析需求关系和需求变更类型的语义,他们确定了五种关系类型:要求、精炼、包含、部分细化和冲突。R. P. Gohil 等人^[7]提出一个分析其影响的框架,并用依赖图显示需求之间的依赖关系,把需求看成事件,在框架中自动添加需求依赖关系。Hassine 等人^[8]基于 UCM 模型分析需求变化的影响,定义 UCM 元素之间依赖关系,并通过功能依赖、时间依赖、包含依赖和组件依赖性对它们进行分组。

在关系抽取领域上,国内外已经有很多研究成果,应用到各种不同的领域。甘丽新等人^[9]主要利用依存句法分析中文实体关系抽取,彭博辉^[10]提出多通道卷积神经网络模型,用于多通道卷积神经网络的实体关系抽取,还有的学者如李丽双^[11]和田生伟^[12]等人将关系抽取应用到生物医学和事件因果关系方面上。目前尚没有一个方法是用于自动获取需求间的依赖关系的研究。

2 需求依赖关系自动获取方法

在同一系统中,两个需求之间可能存在直接或间接的依赖关系,需求依赖关系描述了需求之间的功能调用,也就是说,一个需求可以直接或间接地作用于另一个需求,并对之产生影响。文中提出了一种自动获取需求依赖关系的方法,其主要步骤如图 1 所示。

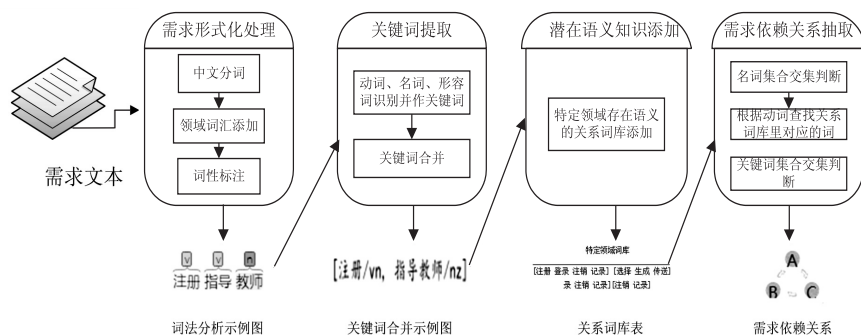


图1 需求关系自动获取方法框架

以自然语言描述的文本需求,往往不能被计算机理解,在实践中这已成为需求工程的一个显著的问题。为解决这一问题,文中需要对自然语言描述的需求文本进行处理,处理的第一步就是将需求形式化^[13]。将

需求表示成由关键词组成的数学集合。首先对已有的分词工具 HanLP(<http://hanlp.com/>)进行改进,添加特定领域的语法词性对需求进行分词以及词性标注;其次提取需求文本中的关键词,通过依赖关键字中动

词、名词等词性的语义特征来判断不同需求之间的关系类型;最后,通过系统自动抽取需求依赖关系。

2.1 需求形式化处理

定义 1:需求是指用户解决某个问题或者达到某个目标所需要的条件或能力。每一个需求文本都由若干个需求组成,其中需求均被定义为以下形式的三元组:

$$R_i = \{ R. \text{content}, R. \text{kwset}, R. \text{relation} \}$$

其中, $R. \text{content}$ 是指需求文本的内容, $R. \text{kwset}$ 指提取需求中添加约束后提取的关键词集合, $R. \text{relation}$ 表示需求之间的依赖关系。关键词集合和依赖关系的定义见定义 2 和定义 3。

提取关键词首先需要分词,文中基于已有的工具 HanLP 进行中文分词及词性标注,但由于需求文本中存在大量的领域词汇,一些词汇在分词或词性标注的过程中可能会发生错误,会直接影响到关系抽取的准确性。例如:“注册一个新指导教师”和“系统应使学生能够检索学生和预订课程的讲师的联系信息”这两条需求语句中“指导教师”和“联系信息”会被注释为动名词,而需求文本中“指导教师”和“联系信息”均被定义为名词,这一差异会影响关系的抽取。为了解决该问题,文中在原有的分词工具上添加了领域词典,通过遍历领域词典,修改被错误标注的词性和分词结果,从而提高关系抽取的准确率。基于 3.1.1 节测试的文本主要添加的领域词汇有:指导教师(n)、联系信息(n)、指定学校(n)、分数评定(v)、评估结果(n)、信息传递设施(n)、课程评估(v)、完成项目(v)、解答疑问(v)。下面是对分词工具的改进方法之一:如 Custom-Dictionary.insert(“联系信息”,“n”)。

2.2 关键词提取

基于 2.1 节的分词及修改后的词性标注,文本中的需求可以用若干个短语组成,常见的短语有很多种,如:动词性短语、名词性短语、形容词性短语、副词性短语、连词性短语、介词短语以及动名词结构等。而提取出的关键词的共现程度可以被视为其语义相关性的度量,由于动词能够很好地反映需求中某些事件之间的关系类型,很多依赖关系通常可以通过动词来引发。同样判断两个需求潜在的语义关系时需要通过动词之间隐含的语义来引发,然而只依靠动词判断关系势必会造成很多不相关的需求也出现依赖关系,与研究结果差异太大,因此文中引入了通过名词的共现程度来考虑是否需要联系动词的语义特征。基于上述问题,文中主要保留动词(v)和名词(n)作为每个需求语句的关键词,并把每个需求封装成一个关键词集合来表示。

定义 2:关键词集合是指对需求进行关键词提取

之后由动词和名词词性所组成的集合。

$$R. \text{kwset} = \{ v_1, v_2, \dots, v_n, n_1, n_2, \dots, n_i \}$$

2.3 语义词库添加

文中重点在于需求关系特征的选择和抽取,以需求文本表示的需求通常被用于全文本分析,因此会忽略与特定领域相关的语义特征,同时词汇之间的语义信息也容易被忽略,这些潜在的问题给需求之间依赖关系的抽取任务带来了挑战。

为了解决上述问题,文中通过分析词汇之间的语义,提出自动抽取需求之间依赖关系的方法。通过定义具有语义关系的关系词库(如:有前提关系的词汇对),利用关键词集合中各个词的词义来联系两个需求之间的语义关系,结合领域相关的语义特征,通过关键字词性映射来抽取特定领域的需求依赖关系,更好地应用于抽取结果。此外,需求语义词库的建立是一个动态的过程,需要在更多了解用户的需求上修改或补充需求以及关系词库表和语义关系。

表 1 给出了针对文中测试的需求文本领域所添加的关系词库表。由于测试文本篇幅较小,因此提取出的关系词也有限。可以看出,关系词库知识的引入丰富了需求文本中词汇的语义,使得提取结果更为准确。

表 1 含有语义关系的依赖关系词库表

关系类型	含有语义关系的词库
唤起关系	[注册 登录 注销 记录][选择 生成 传送][登录 注销 记录][注销 记录][批改作业 得到好分数 得到分数高]
冲突关系	[删除 添加 增补][添加 删除][一再 不再][决定 不决][低等 高等 上等][估计 确定][估量 确定 肯定]
调用关系	[选择 确定][判断 筛选][修改信息 告知是否 修改成功][完成功课 批改作业]
通知关系	[打印 反馈][提供 使用 发送 通知][提供 传递 通知][解答疑问 完成项目]
相似关系	[需要 希望][协助 帮助][好 高]

2.4 需求依赖关系抽取

2.4.1 依赖关系定义

需求之间存在多种依赖关系类型,比如:前提、相似、解释、影响、异常、冲突、业务相关、进化、价值、成本、精炼、包含、部分细化和不相关等多种关系^[14-16]。

需求之间的关系采用统一建模语言中定义的关系语义来描述,并且在特定的系统中,所抽取的需求关系被限制在一定的范围内。文中主要对以下 6 种需求关系做自动化抽取,分别是:相似关系(similar:s)、聚合关系(polymerize:p)、唤起关系(arouse:a)、调用关系(call:c)、通知关系(notice:n)及冲突关系(conflict:co)。

定义 3:需求依赖关系指的是一个系统中一个需

求与另一个需求之间的关系,也就是说,需求与需求之间可能存在直接或间接的依赖性关系。它被定义为:

$R. relation = \{R. s, R. p, R. a, R. c, R. n, R. co\}$ 。

通过系统对其进行抽取和验证。以上关系的非正式定义如下:

相似关系:如果 R_1 和 R_2 需求相同,则具有相似关系。

聚合关系:如果需求 R_2 是 R_1 的一部分,则 R_1 与 R_2 是聚合关系。

唤起关系:如果需求 R_2 需要在 R_1 后才能实现,则 R_1 和 R_2 有唤起关系。

通知关系:如果 R_1 在实现后允许 R_2 开始实现,则 R_1 和 R_2 有通知关系。

调用关系:如果 R_1 在实现的过程中需要先实现 R_2 ,即 R_1 在 R_2 之前实现,但 R_2 先于 R_1 完成实现,则 R_1 和 R_2 具有调用关系。

冲突关系:如果 R_1 和 R_2 不能同时实现,则 R_1 和 R_2 有冲突关系。

2.4.2 计算相似度

文中把需求划分为几部分短语再提取出关键词,通过关键词在任意两句的全部关键词集合中出现的频率,把词频表示成向量来计算余弦值得出相似度,这个方法分为以下三部分:(1)每条需求文本各取出若干个关键词,合并成一个集合,计算每个需求对于这个集合中的关键词的词频;(2)生成两个需求各自的词频向量(这个向量是 n 维的,形如: $A[A_1, A_2, \dots, A_n]$);(3)计算两个向量的余弦相似度,文中取余弦值大于 0.4 作为相似参考。常用的计算余弦值的公式为:

$$\cos\theta = \frac{a^2 + b^2 - c^2}{2ab} \quad (1)$$

已有数学家证明,计算余弦值的这种方法在计算 n 维向量中同样适用。在步骤 2 中把每个需求的词频分别用 n 维向量表示,如: $A = [A_1, A_2, \dots, A_n]$, $B = [B_1, B_2, \dots, B_n]$,那么 A 与 B 的夹角 θ 的余弦值为:

$$\cos\theta = \frac{A \bullet B}{|A| \times |B|} \quad (2)$$

通过这个公式可以求得两个需求之间的相似度,如: R_{13} :指导教师选择评估某个学校的作文, R_{14} :指导教师选择评估某个年级的作文。提取关键词得出 R_{13} [学校, 选择, 教师, 评估, 指导]; R_{14} [年级, 教师, 选择, 评估, 指导],词频向量 $A: [1, 1, 1, 1, 1, 0]$; $B [0, 1, 1, 1, 1, 1]$,通过公式计算余弦值为 0.8。

算法 1:计算相似度算法伪代码。

```
GetSimilarity(int[] wordA, int[] wordB) returns num
wordA; { *需求文本 A 的词频向量 * }
wordB; { *需求文本 B 的词频向量 * }
dividend; {需求文本 A、B 词频向量的乘积 * }
```

```
divisor1; {需求文本 A 的词频向量值 * }
divisor2; {需求文本 B 的词频向量值 * }
num; {需求文本 A、B 的相似度 * }
for i←0 to wordA.length
dividend←dividend+wordA[i] * wordB[i]
divisor1←divisor1+Math.pow(wordA[i], 2)
divisor2←divisor2+Math.pow(wordB[i], 2)
end for
num← dividend / ( Math.sqrt(divisor1) * Math.sqrt(divisor2) )
Return num
End GetSimilarity
```

2.4.3 关系抽取方法

若要判断直接关系(如相似等),通过关键词的共现程度采取集合运算即可;若两个需求之间存在间接依赖关系,那么在这两个需求中一定会出现两个动词,通过这两个动词判断出这两个需求是否存在上下文关系,如果不考虑动词之间所存在的语义关系,会让抽取结果不准确或遗漏。

因此,为了解决上述问题,文中采用基于潜在语义获取需求依赖关系的方法,通过关键词的共现程度联系动词的语义特征来判断需求依赖关系。基于 2.2 节提取关键词之后,关系抽取方法分为四步(如图 2):

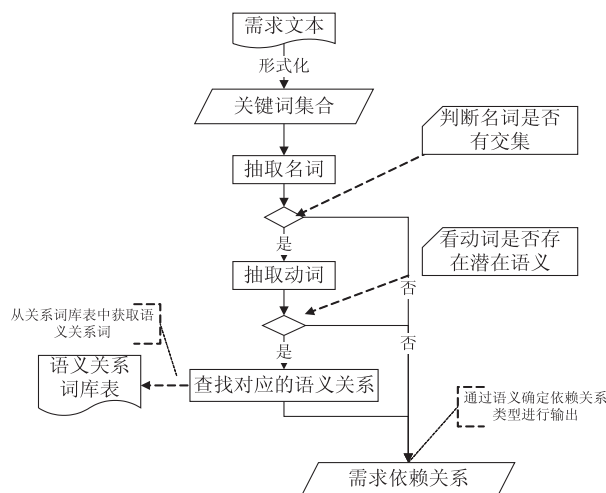


图2 依赖关系抽取方法

(1)抽取名词。将分词中的名词抽取放在集合 A_1 、 B_1 中;(2)判断名词有无交集。若集合有交集[名词有一致的部分-主语],则进行第 3 步;(3)抽取动词。抽取的动词放在集合 A_2 、 B_2 中;(4)判断关系。通过在自定义的潜在语义关系词典中,查找集合 A_2 中动词对应的关系词 C ;若关系词 C 存在集合 B_2 中的动词,则通过关系词判定另两个需求之间存在何种关系,最后确定依赖关系。为了证明文中方法的有效性,文献[5]中基于需求文本提到的依赖关系被用来进行对比实验,并对文献[5]的需求依赖关系做自动抽取。

算法 2:关系抽取算法伪代码。

```

Dependencies( StringxqTextA,String xqTextB ) returns result
xqTextA; { * 需求文本 A * }
xqTextB; { * 需求文本 B * }
List<Term> termA = NULL;
{ * 需求文本 A 的分词词组 * }
List<Term> ermB = NULL;
{ * 需求文本 B 的分词词组 * }
List<String> NListTermA= NULL;
{ * 需求文本 A 的名词集合 * }
List<String> NListTermB = NULL;
{ * 需求文本 B 的名词集合 * }
List<String> VListTermA= NULL;
{ * 需求文本 A 的动词集合 * }
List<String> VListTermB = NULL;
{ * 需求文本 B 的动词集合 * }
Boolean Nend = false; { * 名词是否匹配成功 * }
Stringresult = NULL; { * 需求依赖关系 * }
termA = HanLP. segment( xqTextA );
termB = HanLP. segment( xqTextB );
for i←0 to term[ A/B ]. size( )
{ * 通过词性标注 n 抽取 termA,termB 的名词集合 * }
if( term[ i ]. nature. startsWith( " n " ) = true ) then
NListTerm[ A/B ]. add( term[ i ]. word )
end if
end for
for j←0 to NListTermA. size( ) { * * }
if( NListTermB. contains( NListTermA[ j ] ) = true )
{ * 名词匹配成功的话获取动词,之后可以根据语义词库获
取需求依赖关系 * }
Nend=true;
for i←0 to term[ A/B ]. size( )
{ * 同理通过词性标注 v 抽取 termA,termB 的动词集合 * }
if( term[ i ]. nature. startsWith( " v " ) = true ) then
NListTerm[ A/B ]. add( term[ i ]. word )
end if
end for
Break;
end if
end for
if Nend == true { * 根据语义词库获取需求依赖关系 * }
for i ← 0 to VListTermA. size( )
for j ← 0 to VListTermB. size( )
if( ( VListTermA[ i ],语义词库 ). contains
( VListTermB[ j ] ) ) then
result ← 关系类型
end if
end for
end for
end if
Return result
End Dependencies

```

3 系统实现和实例验证

3.1 实验验证和相关工作对比

3.1.1 实验验证

为证明该方法的可行性,文中采用如下评测标准:准确率 P 、召回率 R 和 F1 度量值,其中准确率 P 用来表示抽取结果中正确标注为给定关系类型所占的比率,召回率 R 表示相对于测试数据集抽取结果中正确给出关系类型所占的比率,F1 度量值表示度量实验的准确性。针对需求依赖关系抽取的结果,具体评价公式为:

$$P = (\text{抽取结果中正确标注为给定关系类型的需求对的个数}) / (\text{结果中标有存在给定关系类型的需求对总个数}) \quad (3)$$

$$R = (\text{抽取结果中正确标注为给定关系类型的需求对的个数}) / (\text{测试数据集中所有存在关系的需求对总个数}) \quad (4)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (5)$$

文中测试了三个数据集来验证方法的可行性,分别是文献[2]、文献[5]和文献[6]中的需求文本,结果如表 2 所示。

表 2 实验结果 %

需求文本	准确率 P	召回率 R	F1 值
文本一 ^[2]	70.83	68	69.39
文本二 ^[5]	81.8	100	89.98
文本三 ^[6]	60	50	54.55
均值	70.88	72.67	71.31

如表 2 所示,无论准确率还是召回率的均值均在 70% 以上,且文本一和文本二的准确率和召回率均较高,说明该方法是可行的。文本三的召回率只有 50%,并不高,说明当前系统在关键词的语义上处理得并不是很准确,主要原因在于文本三的需求文本采用文献[6]的英文需求文本翻译而来,在一些词汇和句式上处理不准确,如一个英文单词可能对应多个汉语释义,在系统提取关键词对比时匹配不到相同的关键词而引起误差。

3.1.2 相关工作对比

笔者分别对三个领域的需求文本做了实验并与专家抽取的关系作对比,分别统计需求和需求间关系识别结果,其对比数据如表 3 所示。为了更直观地表示文本一需求之间的依赖关系,图 3 给出了文本一案例对比的需求依赖关系图。

由于文中是基于中文需求,和国外相关工作使用同一数据集作对比时有一定偏差,同义词太多,系统无法正确做出判断,导致了一些关系抽取的误差。但从实验结果来看,文中方法准确率较高,有一定的可用

表 3 相关工作对比

	依赖关系	文中自动化系统抽取结果	对应文献中人工抽取结果
文本一 ^[2]	通知关系	$(R_{11}, R_{14}) (R_{11}, R_{15}) (R_5, R_{12}) (R_5, R_{15})$	$(R_{11}, R_{14}) (R_{11}, R_{15}) (R_{11}, R_{12}) (R_{11}, R_{13})$
	唤起关系	$(R_{14}, R_8) (R_{14}, R_9) (R_{15}, R_8) (R_{15}, R_9)$	$(R_{14}, R_8) (R_{14}, R_9) (R_{15}, R_8) (R_{15}, R_9) (R_{13}, R_8) (R_{13}, R_9)$
	冲突关系	(R_{14}, R_{15})	(R_{14}, R_{15})
	调用关系	$(R_4, R_{14}) (R_{10}, R_{14}) (R_{10}, R_{15}) (R_4, R_{15}) (R_3, R_{15})$ $(R_{12}, R_{14}) (R_{12}, R_{15}) (R_{16}, R_{17}) (R_{14}, R_{15})$	$(R_3, R_7) (R_2, R_{16}) (R_{15}, R_{16}) (R_4, R_{15}) (R_3, R_{15})$ $(R_{12}, R_{14}) (R_{12}, R_{15}) (R_{16}, R_{17})$
	聚合关系	$(R_{13}, R_{14}) (R_{15}, R_{13}) (R_{12}, R_{13})$	$(R_{13}, R_{14}) (R_{15}, R_{13})$
	相似关系	$(R_4, R_5) (R_4, R_{10}) (R_8, R_9)$	$(R_4, R_5) (R_4, R_{10}) (R_8, R_9) (R_{12}, R_{13})$
文本二 ^[5]	通知关系	(R_{18}, R_{19})	(R_{18}, R_{19})
	唤起关系	$(R_1, R_2) (R_1, R_3) (R_1, R_4) (R_2, R_4) (R_3, R_4) (R_{13}, R_{16}) (R_{14}, R_{16})$	$(R_1, R_4) (R_2, R_4) (R_3, R_4) (R_{13}, R_{16}) (R_{14}, R_{16})$
	调用关系	$(R_{10}, R_{11}) (R_{12}, R_{11})$	$(R_{11}, R_{10}) (R_{11}, R_{12})$
文本三 ^[6]	冲突关系	(R_5, R_6)	(R_5, R_6)
	聚合关系	(R_{61}, R_{62})	(R_{61}, R_{62})
文本一、二、三	需要关系	$(R_7, R_{18}) (R_7, R_8) (R_4, R_{18}) (R_4, R_8)$	$(R_7, R_{18}) (R_7, R_{16}) (R_7, R_{25}) (R_7, R_{24}) (R_4, R_8)$
	平均效率/s	5.052	工程量大,以小时为单位

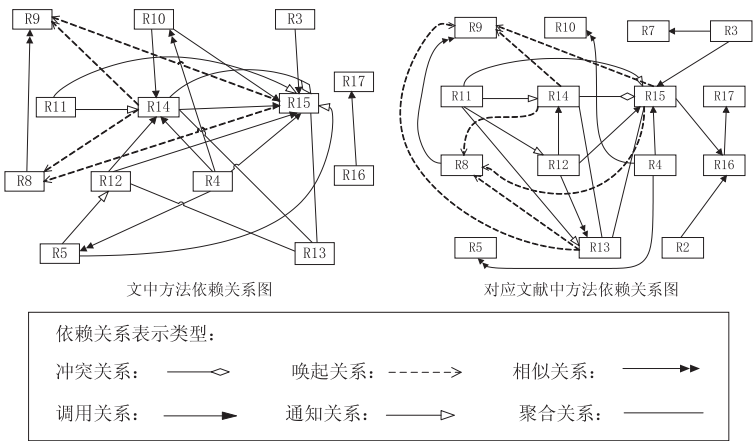


图 3 文本一需求依赖关系对比

性。同时,表 4 给出了文中方法在不同关系类型抽取上的性能分析数据。从表 3 的效率以及表 4 的数据来看,文中在加入领域词汇以及定义语义词库这两大特征后,其效率以及各方面性能百分比都较高,从一定程度上证明了文中方法的可行性及可靠性。图 4 表示在了不同领域上文中方法抽取结果的数据分析。

表 4 文中方法在不同关系类型抽取上的性能 %

关系类型	准确率 P	召回率 R	F1 值
通知关系	60	60	60
唤起关系	81.82	81.82	81.82
互斥关系	100	100	100
调用关系	63.64	70	66.67
相似关系	100	75	85.71
聚合关系	75	100	85.71

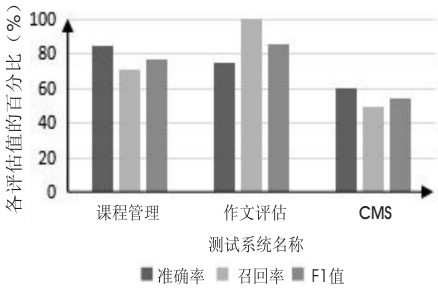


图 4 不同领域文中方法的数据分析

3.2 系统实现

使用 servlet+jsp 技术设计并实现了一个针对中文自然语言需求文本的依赖关系抽取系统,对系统中用到的方法中的各个步骤进行了支持。该系统主要由需求文本添加、文本相似度、文本查询以及需求依赖关系显示 4 个模块组成,想要查询的需求对象与哪些需求

有关并获取需求之间的关系类型都可以在点击此需求 时显示,页面如图 5 所示。

序号	名称	描述	操作
1	R1	注册一个新指导教师	删除
2	R2	指导教师登录	删除
3	R3	指导教师注销	删除
4	R4	指导教师信息记录	删除
5	R5	添加一个作文方面评语	删除
6	R6	删除一个作文方面评语	删除

序号	名称	描述	相似度	关系
1	R24	团队是由学生使用信息传递设施邀请其他学员在同一课程中创建的。	10.66%	调用关系
2	r17	管理员需要重新设置项目成员并被告知是否修改成功	50%	调用关系

图 5 系统实现界面

4 结束语

文中基于领域知识和语义相结合的方法实现了对需求依赖关系的自动化抽取,首先对已有的分词工具进行改进,其次提取关键词,对关键词集合里的词性作集合判断,联系潜在的语义判断需求依赖关系,最后通过系统证明方法的可用性,不仅对定义的几种关系进行自动获取,同时对参考文献中定义的关系也进行了实现并与之结果进行比较得出该方法的优点与不足。该方法的优势在于:(1)基于关键词,不同关系上对关键词处理可能不同,使结果更可信;(2)添加了一些领域词汇,消除了自然语言固有的二义性;(3)建立语义词库表。根据用户需求的特点,设计关系词库表,定义各个词汇之间的语义关系,使结果更准确;(4)自动抽取关系。以便于在需求量巨大的时候可以节省人力物力,具有更高的效率。不足之处在于包含语义的词库添加上属于人工添加并且仅仅是基于特定领域上,测试集越多,需要的人力资源也就越大,未来将在这方面以及需求变更上做进一步研究。

参考文献:

- [1] 骆 斌. 需求工程-软件建模与分析[M]. 北京:高等教育出版社,2013.
- [2] 李天颖,刘 璘,赵德旺,等. 一种基于依存文法的需求文本策略依赖关系抽取方法[J]. 计算机学报,2013,36(1): 54-62.
- [3] ARORA C, SABETZADEH M, GOKNIL A, et al. Change impact analysis for natural language requirements: an NLP approach[C]//International conference on requirements engineering. Ottawa, ON, Canada: IEEE, 2015.
- [4] 韩德帅,杨启亮,邢建春. 一种软件自适应 UML 建模及其形式化验证方法[J]. 软件学报,2015,26(4): 730-746.
- [5] 罗术通,张长海,金 英,等. 通过需求依赖判定横切关注点[J]. 吉林大学学报:工学版,2011,41(4): 1065-1070.
- [6] GOKNIL A, KURTEV I, DEN BERG K V, et al. Change impact analysis for requirements: a metamodeling approach [J]. Information & Software Technology, 2014, 56(8): 950-972.
- [7] GOHIL R P, BHATTACHARYA S, CHAUHAN R. Requirements change impact analysis using event based traceability [J]. Lecture Notes on Software Engineering, 2016, 4(3): 162-168.
- [8] HASSINE J, RILLING J, HEWITT J, et al. Change impact analysis for requirement evolution using use case maps[C]//International workshop on principles of software evolution. Lisbon, Portugal: IEEE, 2005: 81-90.
- [9] 甘丽新,万常选,刘德喜,等. 基于句法语义特征的中文实体关系抽取[J]. 计算机研究与发展,2016,53(2): 284-302.
- [10] 彤博辉,付 琨,黄 宇,等. 基于多通道卷积神经网络的实体关系抽取[J]. 计算机应用研究,2017,34(3): 689-692.
- [11] 李丽双,郭 瑞,黄德根,等. 基于迁移学习的蛋白质交互关系抽取[J]. 中文信息学报,2016,30(2): 160-167.
- [12] 田生伟,周兴发,禹 龙,等. 基于双向 LSTM 的维吾尔语事件因果关系抽取[J]. 电子与信息学报,2018,40(1): 200-208.
- [13] 唐 聃,白宁超. 自然语言处理理论与实践[M]. 成都:电子工业出版社,2018.
- [14] 温赵欣,关 慧,贾成真. 基于需求依赖关系识别横切关注点[J]. 电子技术与软件工程,2018(20): 64-66.
- [15] 邵 飞. 基于依赖关系的软件需求建模与优先级评估方法研究[D]. 武汉:武汉大学,2017.
- [16] WHITTLE J, WIJESEKERA D, HARTONG M, et al. Executable misuse cases for modeling security concerns [C]//ACM/IEEE international conference on software engineering. Leipzig, Germany: IEEE, 2008: 121-130.