摘 要

词嵌入是近年来比较热门的一个研究方向，词嵌入的目的是通过低纬度的稠密向量表征出知识库中的实体和关系的语义信息，在有效存储信息的基础上，能够实现实体预测、链接预测等推理任务。目前，大多数的工作集中在通过优化评分函数、损失函数等来训练出一个更有效的翻译模型，使得训练出的翻译模型可以在meanrank、Hit@10的评分标准下取得更好的成绩。本文考虑到翻译模型训练数据与本体语义的相关性，利用本体推理研究的现有成果提出一种新的方法来提升构建出的翻译模型的性能。这种方法使用本体完成对训练集的语义表达，同时使用测试数据中的三元组集合构建出观察值的集合。通过这种方式，本文把优化翻译模型的问题，转化成本体研究中的溯因推理问题。溯因诊断是本体新知识获取的一个重要的推理方法，这种方法可以揭示出观察值没有被成功推理出的根本原因，并且可以提供出与原本体语义一致的解释集。解释集中的公理与本体中的公理具有相同的语法，在使用获得的解释集对缺陷本体进行完善修复后，新的本体不仅可以蕴含观察值，还因为关键性公理的补全，有着更好的获取新知识的能力。在本体的溯因诊断中，限制解释集的大小是必须考虑的一个关键因素。解释集过于强大的表达能力会导致解空间的无限膨胀，而过度地压缩解的空间，不仅会极大地削弱解释集的表达能力，也会影响解释集的对缺漏公理的覆盖能力。为了平衡解释集的表达能力与解空间的矛盾，本文根据决断集合的特点提出一个新的概念，决断集合模版。基本地，决断集合模版由决断集中获取，获取方法是把决断集中的公理进行变量化，解释集会从原本体以及决断集模版的集合中进行计算，然后实例化得来。同时，为了限制解释集的空间，本文会介绍一个新的概念，最小决断集模版。利用最小决断集模版，我们可以有效地限制解释集的空间大小。在本体的溯因诊断中，观察值的表达是另外一个需要考虑的因素，但是在词向量模型的训练集的数据中，三元组是它的唯一表达方式，因此本文将会使用简化的表达语言完成对观察值集合的表达，这样可以使得解释集的解空间缩小而不影响解释集的表达能力。本文的目标是通过利用本体的表达能力对词嵌入的训练数据进行表达，从而把优化词向量模型的问题转化成本体研究中的溯因诊断问题。通过利用决断集模版的语义能力，我们可以从模版和集合中实例化出目标解释集。然后，利用诊断出来的结果，我们可以对词向量的训练数据进行修正，从而达到优化训练出来的词向量模型的目的。最后，我们还设置了一组相关的实验，验证本文得出的解释集对训练数据的修正能力，以及对比不同方法下所得出

的词向量模型的表达能力和推理能力。

关键词：本体；词向量；溯因诊断

Abstract

Word embedding is a popular research direction in recent years. The purpose of word embedding is to express the semantic information of entities and relationships in the knowledge base through the dense vector of low latitude. Based on the effective storage of information, it can perform entity prediction, link Prediction and other reasoning tasks. At present, most of the work focuses on training a more efficient translation model by optimizing the scoring function, loss function, etc., so that the trained translation model can achieve better results under the score of meanrank and Hit@10. This paper takes into account the relevance of the translation model training data to the ontology semantics, and puts forward a new method to enhance the performance of the translated model by using the existing results of ontology reasoning research. This method uses the ontology to complete the semantic expression of the training set, and uses the set of triples in the test data to construct a set of observations. In this way, this paper transforms the problem of optimizing the translation model into the problem of abductive reasoning in ontology research. The retrospective diagnosis is an important reasoning method for the acquisition of new knowledge of ontology. This method can reveal the root cause that the observed value has not been successfully entailed, and can provide the explanation which is consistent with the original body semantics. The axiom of the explanation has the same grammar as the axiom in the ontology. After using the obtained explanation to perfect the defect ontology, the new ontology can not only contain the observation value, but also because of the completeness of the key axiom, The ability to acquire new knowledge. In the ontology diagnosis of the ontology, limiting the size of the explanatory set is a key factor that must be considered. Interpretation of the set of too strong expression will lead to the infinite expansion of the solution space, and excessive compression of the solution space, not only will greatly weaken the interpretation of the set of expression ability, will also affect the interpretation of the lack of coverage of the lack of ability. In order to balance the contradiction between the expression ability of the explanation and the solution space, this paper proposes a new concept based on the characteristics of the de-

cision set. Basically, the decision set template is obtained from the decision, and the acquisition method is to quantify the axioms of the decision set, and the explanation assembly is calculated from the original set and the set of decision sets, and then instantiated. At the same time, in order to limit the space of the interpretation set, this article will introduce a new concept, the smallest template. In the ontology diagnosis of the ontology, the expression of the observation is another factor to consider, but in the data set of the training set of the word vector model, the triplet is its only expression, so this article will use the simplified expression The language completes the expression of the set of observations, which allows the solution space of the interpretation set to be reduced without affecting the expression of the interpretation set. The goal of this paper is to express the training data of the word by using the expression ability of the ontology, so as to transform the problem of the optimization word vector model into the dying diagnosis problem in the ontology research. By using the semantic capabilities of the decision set template, we can extract the target set from the template and the set of examples. Then, using the results of the diagnosis, we can modify the training data of the word vector so as to achieve the purpose of optimizing the vector model of the training. Finally, we also set up a set of related experiments to verify the corrective ability of the explanatory set to be derived from the training data, and to compare the expression ability and reasoning ability of the word vector model obtained under different methods.

Keywords: Ontology; Word Vector; Abduction Diagnosis

目 录

第一章　引言

目前本体并没有统一的定义和固定的应用领域，斯坦福大学的Gruber给出的定义得到了许多同行的认可，即本体是共享概念模型的明确的形式化规范说明[?]。本体提出的最初目标是实现知识的共享、集成和重用。一个完善的本体可以澄清领域知识的结构，通过构建一个统一结构或者一个规范模型来减少概念和术语上的差异。同时，利用本体对需求解决的问题和任务进行规范化描述，可以提高需求分析、信息获取的效率，节约成本[?]。

1.1　本文的意义

进入信息时代，人们花费了大量的精力构建结构化的知识库。利用已有的知识库，进行表示学习的目标，是通过转化规则将现有的知识库的语义信息转化到低维的向量中进行表示，并利用转化后得到的低维向量学习出新的知识。但是，构建出的向量模型与现实世界观察值的不一致是一种出现频率较高的问题，这种不一致会降低向量模型推理结果的可信度，同时也意味着已构建的向量模型对语义信息表达的准确度有待提高。因此，如何找到产生这种不一致的原因，以及如何对已构建的向量模型进行修复，是目前研究的重点。然而，目前的研究工作，大多数集中在改进向量模型生成的算法上，但是对由知识库产生的问题，却还没有做较为深入的研究。知识库是向量模型的生成来源，一个不完整的或者是有缺陷的知识库，会对生成的模型产生重大的影响。与此同时，对本体的研究获得了较大的进展。本体作为语义网的核心，是一种清晰表达语义和知识共享的方式，也就是在特定领域之中那些存在着的对象类型或概念及其属性和相互关系，在本体中进行推理可以获得知识库中没有进行表达的知识，而其中，溯因诊断又是本体中一种重要的推理方式。溯因诊断是推理到最佳解释的过程，它是开始于事实的集合并推导出它们的最合适的解释的推理过程，溯因意味着生成假设来解释观察货结论。本文的主要目标，是利用本体语言对知识库与观察值进行表达，再在生成的本体中进行溯因诊断，利用针对词向量知识库改进的诊断方法，找出知识

库产生不一致的原因，并在此基础上，实现对知识库的修复，从而达到修正向量模型的目的。

1.2 研究现状

本体学习最早是由Alexander Madche和Steffen Staab 两人提出的，他们把本体学习描述成从数据中提取领域模型[?]。本体学习是一个广泛的概念，涉及的领域和学科众多，目前研究者们把本体学习分割成以下几个部分。在大多数情况下，对于每一层任务的研究都是建立在更底层的研究基础之上。目前对本体学习的研究集中在基础五层任务之中，包括术语提取、同义词识别、概念提取、概念层次关系提取、非分类关系的提取。

目前本体学习主要有两种分类方法，分别是按照信息输入源分类和按照构建的方法进行分类。

现阶段本体学习的信息输入源有三类，分别是基于结构化数据的本体学习，基于半结构化数据的本体学习和基于非结构化数据的。结构化数据主要包括关系数据库或面向对象数据库中的数据，半结构化数据是指具有隐含结构，但缺乏固定或严格结构的数据，目前web上也有不少半结构化数据，以RDF标注的网页也越来越多[?]。非结构化数据指的是纯文本无结构化标注的自然语言数据，文章、访谈甚至剧本都可以看做是本体学习的非结构化信息源。目前Web上数量最大、信息最多的就是这一类型的非结构化数据，同时这一类型数据也是价值含量最高的，所以目前的研究多集中于以非结构化数据作为信息源的本体学习[?]。

现有的本体构建方法有手工和自动两种。手工构建指的是有由领域的专家参与进行本体构建工作，本体的完全手工构建耗时、费力，容易出现倾向性错误，且难以及时地更新。另一种本体构建的方法是自动构建法，通过结合自然语言处理、机器学习、人工智能等多领域的研究成果在给定的信息输入源中自动构建出本体。自动构建法分为两种方式，一种是全自动，全自动方指的是在输入源的基础上构建出一个全新的本体。另一种是半自动的方式，半自动构建本体意味着在输入信息源的基础上还有一个该领域的初步本体模型，半自动构建本体的任务就

是从信息源中挖掘出更多的概念和关系用以扩充或者纠正原有的初步本体模型。本体的自动构建具有相当大的潜在价值，但是由于自动构建的难度较大，设计的学科领域众多，现时关于本体自动学习的研究尚未达到完全成熟的水平。[**?**]

本体的手工构建繁琐且难度较大，因此现有的手工构建的本体并不多，其中以普林斯顿大学认识科学实验室在心理学教授乔治?A?米勒的指导下建立和维护的英语字典wordnet最为出名。有别于普通意义上的字典，Wordnet包含了语义信息。Wordnet根据词条的意义将它们分组，每一个具有相近意义的词条组成了一个synset(同义词集)。Wordnet为synset提供了简短，概要的定义，并记录不同synset之间的语义关系。Wordnet发展到现在，已经具有了一定的完整性，涵盖额大多数的英语单词，缺点就是目前尚未支持英语以外的其它语言目前本体的自动构建并没有统一的步骤，得到大多数人认可的是由本体构建工具Protégé开发商斯坦福大学医学院生物信息研究中心提出的七步法[**?**]。七个步骤分别是：(1) 确定本体的专业和范畴

(2) 考察服用现有本体的可能性

(3) 列出本体中得重要术语

(4) 定义类和类的等级体系(完善等级体系可行的方法有：自顶向下法、自底向上法和综合法)

(5) 定义类的属性

(6) 定义属性的分面

(7) 创建实例

斯坦福大学医学院提出的七步法步骤清晰，合理，且经过实验验证在不同领域都有具有较好的适用性。在进行本体构建的时候并不一定要严格遵循以上七个步骤，可根据构建目标领域的需求进行调整以及适应性更改。以自然语言文本作为信息源的需要对文本进行预处理，在这方面自然语言处理的研究已经比较完善，由斯坦福大学开发的Standford Parser取得了不错的成绩。在中文处理方面，由国内复旦大学开发的复旦自然语言处理开发套件可以以较高的准确度完成分词、实体识别、依存句法树生成等任务。在术语提取方面，最主要的方法是由G.

Salton和C. Buckley提出的名词索引的方法[**?**]。他们通过统计学的算法计算出文档中每个词的权重，通过设定阀值筛选出推荐的候选术语。同时有一部分的研究专注于以自然语言处理的方法提取术语，缺点是提取的精度受自然语言处理水平的影响较大。本体学习的另一项重要的任务是从信息源中提取出概念和概念之间的关系。大多数对概念之间关系的提取都使用了数据分析法以及或多或少的语义分析相关的方法。其中最具有代表性的是由D.Faure提出的以动词为核心的概念关系提取，他把实体的属性看成是概念之间的关系。缺点是以动词为核心的提取方法只能提取非分类关系，对于part-of,等层次关系提取成功率较低。

1.3 本文的工作

本文的主要工作是结合国内外已有的本体学习相关的研究成果，针对中文语法特点和语言习惯，从给定的领域语料库中构建出本体，并映射到以RDFS表示的模型当中，具体的步骤如下：(1) 利用复旦大学开发的fnlp中文自然语言处理开发库对输入的中文自然语言的领域语料库进行预处理，包括分词处理，实体名识别，依存句法树生成等。

(2) 从生成的依存句法树中提取出主谓宾结构，转化成知识三元组结构，构建出初步的RDF模型。

(3) 结合统计方法和与语义分析方法，在给定的语料库抽取出领域术语集。

(4) 分析出领域信息源中的概念，并在wordnet的帮助下识别出领域中的上下位关系(即RDFS中得子类关系)

(5) 利用获得的初步RDF模型和提取出来的领域术语及其关系生成该领域本体的RDFS模型。

1.4 论文结构简介

本文主要分五部分。第一部分是对待解决的问题进行简单的描述，描述了本体学习现在的研究进度，并对各个方面的研究水平给出评价，并概括全文的工作。第二部分是预备知识，简单阐释了了本文工作所需的基本知识。第三部分是

本文的核心部分，重点讲述了本文从中文自然语言文本构建出本体的详细方法。第四部分是实验部分，实验部分是本文对理论的验证。最后是致谢部分，以此感谢对写作本文予以帮助的人。

第二章　预备知识

2.1　知识库模型研究

知识库存储的是客观事实的元数据，构建知识库模型是知识表示学习的重要内容。构建知识库模型的主要任务，是寻找构建模型的方法完成对知识库中实体关系进行表达。在近年来，研究者对知识库模型进行了深入的研究，提出了多种知识库模型的构建方法，并对他们的研究成果进行了验证。

2.1.1　知识库与建模

首先我们先给出知识库的定义：

$$G = (E, R, S) \tag{2.1}$$

在这个式子中G表示知识库，其中

$$E = \{e_1, e_2, \cdots, e_{|E|}\} \tag{2.2}$$

E表示的由知识库中实体组成的集合，$\mid R \mid$表示在知识库中二元关系组成的集合。一般的，我们会用(h, r, t)表示一个表达实体间二元关系的三元组，其中$h \in E$，$r \in R$，$t \in E$，对于知识库中的所有h和t，我们把它称之为头实体和尾实体，r则称之为头实体和尾实体的关系。例如(特朗普，总统，美国)就是一个二元关系三元组的实例，它表达的意思是特朗普是美国的总统，这里的特朗普和美国就是头实体与尾实体，总统表示的是特朗普和美国之间的关系。在式子2.1中的S表示所有二元关系的三元组集合，记作$S \subseteq E \times R \times E$。对知识库$G$的建模可以看做是对$(E, R, S)$建立模型的研究。

目前构建知识库模型主要方法分为两大类，一类是语义网领域的本体构建方法，本体是一种概念模型的建模工具，构建本体虽然比较复杂，目前构造本体的方法大多基于手动或者是手动加半自动的方法，但是这种方法构建出来的本体可以对知识和语义进行表达，而且本体对于在人可读性与机器可读性中间找到了一

个平衡点，在两者的可读性上都要比以往的模型要友好。另一种构建知识库模型的方法是基于机器学习的知识库建模，目前基于机器学习方法的知识库模型有几个具有代表性的模型：距离模型、单层神经网络模型、能量模型、双线性模型、张量神经网络模型、矩阵分解模型和翻译模型[]。

距离模型是由(Bordes)早期提出的使用结构表示(structured embedding, SE)进行知识表示的方法，这种方法的主要思想是[]：

- 所有的实体都可以通过使用一个d-维的向量空间进行表达，把实体转化作多维向量空间中向量的操作记作“嵌入空间”。第i个实体会被转化成一个向量E_i，其中$E_i \in \mathcal{R}^d$。
- 在被嵌入的向量空间中，对于任意给定的关系类型，都存在一个特定的相似度衡量值，这个衡量值就是关系的向量空间。在判断头实体和尾实体是否满足该关系时，头实体和尾实体会被投射到关系所在的向量空间，在关系的向量空间中，头实体和尾实体的向量距离越近，就表示关系成立的可能性越大。作者对k^{th}个给定的关系上的实体对记作$R_k = (R_k^{lhs}, R_k^{rhs})$，其中$R_k = (R_k^{lhs}$和$R_k^{rhs})$都是一个$d \times d$的矩阵。因此，在该关系下$R_k = (R_k^{lhs}$和$R_k^{rhs})$的相似度计算如下：

$$S_k(E_i, E_j) = \| R_k^{lhs} E_i - R_k^{lhs} E_j \| \tag{2.3}$$

距离模型的建模方式利用实体在关系的向量空间上的投影进行相似度计算是合理的，但是在投影的时候，距离模型对关系中的头实体和尾实体采用了两个不同的投影矩阵，这种方法难以保证协同性，因此距离模型在链接预测等任务上没有表现出足够优秀的能力。

单层神经网络模型(single layer model, SLM)是Socher等人提出的新的神经网络模型，为了解决距离模型头尾实体投影矩阵不同带来的协同带来的关系刻画误差问题，单层神经网络模型对二元关系三元组使用了如下评分函数：

$$g(e_1, R, e_2) = u_R^T f(M_{r,1} l_h + M_{r,2} l_t) \tag{2.4}$$

在评分函数$g(e_1, R, e_2)$中，$M_{r,1}l_h$和$M_{r,2}l_t$是投影矩阵，u_R^T是关系R的表示向量，$f = tanh$。单层神经网络模型一定程度上优化了距离模型刻画关系不精准的问题，但是却带来了较大的计算量，降低了计算效率。

DISTMUL是Yang等人提出的一种简化的神经嵌入法，对于一个给定的由关系三元组(e_1, r, e_2)(定义e_1是在关系r下的主体，e_2是客体)知识库KB，DISTMUL的主要想法是对于正向的三元组，根据评分函数(能量函数)可以获得较高的评分(或者是较低的能量)，而不成立的关系三元组在相同的评分函数则只能得到相反的评分(能量)结果。在DISTMUL模型中，关系矩阵被限制为对角矩阵，这种限制虽然增加了矩阵生成的难度，但是这种对角关系矩阵不仅简化了模型的运算复杂度，也提高了模型的语义表达效果。利用DISTMUL模型还可以完成知识表示学习的子任务关系规则的挖掘。关系规则是知识库的一个重要部分，例如，给定一个事实：一个人是出生在纽约，并且纽约是美国的一个城市，那么这个人的国籍就是美国：

$$BornInCity(a, b) \wedge CityOfCountry(b, c) \Longrightarrow Nationality(a, c) \tag{2.5}$$

类似的逻辑规则可以帮助从知识库中获取新知识，而且它也可以优化知识的存储方式，利用规则的存储减少事实的存储，从而减小知识库的体积。更重要的，这些规则可以满足复杂的推理需求[]。DISTMUL模型挖掘关系规则的主要思想是: 如果二元关系三元组(e_1, r, e_2)是一个正向例的且二元关系r对应的翻译向量V满足$y_a + V - y_b \approx 0$，那么应该会有以下规则属性成立：从$y_a + V_1 \approx y_b$和$y_b + V_2 \approx y_c$得出$y_a + (V_1 + V_2) \approx 0$。DISTMUL在挖掘基数为2和基数为3的规则时取得了不错的效果，但是由于算法复杂度较高，在面对更高基数的规则挖掘时耗时增长严重。

2.1.2 翻译模型的研究

2013年由Mikolov领导的一支谷歌研究团队提供了一种对词的向量表示进行运算的方法，这种方法是将深度学习技术引入自然语言处理领域的一项核心技

术，Mikolov还提供了一个开源的Word2vec版本，这项技术使得自然语言处理多了一个新的研究方向。受到词向量特性的启发，Bordes等人提出了TransE模型。

TransE是一个为了学习出实体在低纬嵌入空间的基于能量的模型。在TransE中，关系会被认为是到嵌入空间的一个翻译[]。例如假设(h,l,t)是一个正向的二元关系三元组，那么在这个模型下尾部实体t在空间的嵌入应该约等于头部实体h和关系l的空间嵌入向量的和：

$$h + l \approx t \tag{2.6}$$

对于一个给定的由三元组(h,l,t)构成的集合S，其中$h,t \in E$(实体的集合)以及关系$l \in L$(关系集合)，以下算法的目标是学习关系和实体的嵌入向量。TransE根据能量框架定义三元组的能量等价于相似性测量$d(h+l,t)$，在TransE里d可以是曼哈顿距离或者是欧几里得距离。TransE对训练集数据的训练目标是最小化以下函数：

$$\mathcal{L} = \sum_{(h,l,t)\in S} \sum_{(h',l,t')\in S'_{(h,l,t)}} [\gamma + d(h+l,t) - d(h'+l,t')]_+ \tag{2.7}$$

在以上函数中，$[x]_+$表示$max(0,x)$。由于训练的时候需要生成反例(不成立的二元关系三元组)，这里的反例可以通过正向三元组提取。具体做法是取一正向实例(h,l,t)，从三元组中移除头部(或尾部)实体，使用实体集合中选择一个实体$h'(h' \in E)$组成损坏三元组(h',l,t)，其中损坏三元组满足$(h',l,t) \notin S_{(h,l,t)}$。所以对所有损坏三元组，有：

$$S'_{(h,l,t)} = \{(h',l,t \mid h' \in E)\} \cup \{(h,l,t') \mid t' \in E\} \tag{2.8}$$

TransE选择了随机梯度下降法(Stochastic Gradient Descent，SGD)作为优化方法，随机梯度下降法在进行训练的时候，并不需要对所有的和求梯度，因此随机梯度下降法也不需要在每次循环的时候更新所有的向量，而只需要对一个批次的向量进行求梯度计算就可以更新$theta$值。

TransE在实验中设置两组实验，分别验证模型在实体预测以及链接预测上的能力，与之对比的还有RESCAL，SE，SME(linear)/SME(bilinear)以及LFM等知

识库模型，从实验数据可以看出，TransE取得的效果非常优秀，在三组数据集的两种评价方法下，TransE都取得最好的成绩，值得注意的是在一个体量较大的数据集FB1M的测试中，TransE不仅完成了测试，对比以往的模型还获得了较大的提高。但是，TransE方法也有自身的缺点。假设有两组三元组分别为(上海，位于，中国)记为(h,l,t)以及(通州，位于，中国)记为(h,l,t')。根据TransE的损失函数，训练出来的模型会有$h+l\approx t$ 以及$h+l\approx t'$，所以我们有$t\approx t'$，也就是说在TransE的表达模型中，“上海”和“通州”的嵌入向量会近似相等，即使这两个实体具有较大的差异。也就意味着，TransE虽然在处理一对一关系的时候有着不错得性能，但是在面对一对多的关系的时候，TransE方法具有缺陷。

受TransE的启发，Wang等人在2014年时基于TransE提出了一种新的翻译模型，称之为TransH。TransH的提出就是为了解决在TransE中无法很好处理地一对多，多对多的关系。TransE在建立关系实体的向量空间的时候关系和实体都被嵌入到了平面空间，这个因素限制了TransE在处理多对多、多对一以及一对多的关系的表达能力。因此，为了解决这个问题，在TransH中关系被嵌入到了超平面空间。对于一个关系r，TransH用超平面w_r和在超平面上的向量d_r进行表示。特别地，对于一个二元关系三元组(h,r,t)，h和t将会被首先投影到平面w_r上，投影后的向量分别被记作$h_\perp$和$t_\perp$。在TransH中，如果(h,r,t)是一个正向三元组，$h_\perp$和$t_\perp$会被期望能够被超平面上的向量d_r联系起来。因此在TansH中的评分函数为：

$$|| h_\perp + d_r - t_\perp ||_2^2 \tag{2.9}$$

通过限制$|| w_r ||_2 = 1$，我们可以得到：

$$h_\perp = h - w_r^\top h w_r,\ \ t_\perp = t - w_r^\top t w_r \tag{2.10}$$

可以得到评分函数：

$$f_r(h,t) = || (h - w_r^\top h w_r) + d_r - (t - w_r^\top t w_r) ||_2^2 \tag{2.11}$$

损失函数为：

$$\mathcal{L} = \sum_{(h,l,t)\in S} \sum_{(h',l,t')\in S'_{(h,l,t)}} [\gamma + f_r(h,t) - f_{r'}(h',t')]_+ \tag{2.12}$$

2.2 本体知识库模型

2.2.1 描述逻辑与OWL

描述逻辑(description logic)是一种用于知识表示的逻辑语言和以其为对象的推理方法，主要用于描述概念分类及其概念之间的关系[]。描述逻辑是一阶逻辑的可决定性片段，但是描述逻辑又具有强的表达能力，描述逻辑在表达能力和推理能力之间取得了平衡。

描述逻辑语言$\mathcal{L}$包含以下三个集合，分别是一个由个体名组成的集合N_I，一个由概念名组成集合N_C以及一个由二元关系名组成的集合N_R。

定义 2.1 (描述逻辑词汇表) *描述逻辑词汇表$\mathcal{V}$是一个三元组(N_C, N_R, N_I)，其中N_C是概念名的集合，N_R是关系名的集合，N_I是个体名的集合。*

我们把$\mathcal{L}$ 的语义演绎表示为$\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$，其中$\Delta^{\mathcal{I}}$表示$\mathcal{I}$的域，是一个非空的个体集合，$\cdot^{\mathcal{I}}$是一个映射函数，这个映射函数可以完成概念名到$\Delta^{\mathcal{I}}$子集的映射，二元关系到$\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ 子集的映射以及个体到$\Delta^{\mathcal{I}}$元素的映射。定义如下[]：

定义 2.2 *一个演绎$\mathcal{I}$是二元组$(\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$，其中$\Delta^{\mathcal{I}}$被称为域，$\cdot^{\mathcal{I}}$是一个从$N_I$到$\Delta^{\mathcal{I}}$的函数。*

- $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
- $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
- $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$
- $(C_1 \sqcap C_2)^{\mathcal{I}} \subseteq C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}}$

在本文中，我们考虑表2.1中的语言片段。

一个DL本体$\mathcal{O} = (\mathcal{T}, \mathcal{A})$包括两个部分，一个是TBox $\mathcal{T}$，它描述术语知识与应用领域相关的背景知识，处理概念的定义，由有限个公理构成，其中有引入

表 2.1: 语法以及语义表

Constructor	Syntax	Semantics
top concept	$\top$	$\triangle^{\mathcal{I}}$
bottom concept	$\bot$	$\emptyset$
conjunction	C $\sqcap$ D	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
existential restriction	$\exists r.C$	$\{X \in \triangle^{\mathcal{I}} \mid \exists y \in \triangle^{\mathcal{I}}\text{: } (x,y) \in r^{\mathcal{I}} \wedge r \in C^{\mathcal{I}}\}$
general concept inclusion	$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
role inclusion	$r_1 \circ \cdots \circ r_k \subseteq r$	$r_1^{\mathcal{I}} \circ \cdots \circ r_k^{\mathcal{I}} \subseteq r^{\mathcal{I}}$

新概念名和角色名称的公理，有断言包含关系的公理以及断言觉得可传递角色或功能性角色公理[面向Web 的个性化语义信息检索技术]。另一个是断言知识的集合ABox $\mathcal{A}$，它描述的是TBox词汇表中的个体断言,包括概念类的成员元素，二元关系的成员元素二元组以及二元关系的等价关系。在本文中，断言知识部分我们仅考虑二元关系的断言，形如$r(a,b)$，其中r是$\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ 中的一个二元关系实例，a和b是$\Delta^{\mathcal{I}}$中的一个个体。表2.2展示了TBox与ABox的语法以及语义。

表 2.2: DL示例公理的语义

Syntax	Semantics
$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
$C \equiv D$	$C^{\mathcal{I}} = D^{\mathcal{I}}$
$C(a)$	$a^{\mathcal{I}} \in C^{\mathcal{I}}$
$r(a,b)$	$\langle a^{\mathcal{I}}, b^{\mathcal{I}} \rangle \in r^{\mathcal{I}}$

一个$\cdot^{\mathcal{I}}$如果能够满足本体$\mathcal{O}$中的所有公理，那么这个$\cdot^{\mathcal{I}}$就被称为本体的一个模。因此我们有以下定义：

定义 2.3 **(一致)** 对于一个DL本体O，如果它有至少一个模，那么这个本体就会被称为是一致的，记作$\mathcal{O} \nvDash \bot$。相反地，如果一个DL本体$\mathcal{O}$没有至少一个模，那么这个本体就会被称为不一致的，记作$\mathcal{O} \vDash \bot$。

2.2.2 本体决断集

本体包含TBox与ABox，因此本体具有从已有的知识库中获取新知识的能力，这种本体中没有，但是可以通过推理出来得到的公理就叫做蕴含。一个本体

可能会有一个或多个蕴含，在本体的开发中，能够从蕴含逆推出在推理过程相关的公理具有重大的现实意义，这些公理的集合我们称之为决断集。

定义 2.4 (决断集) 令$\mathcal{O}$为一个一致的DL本体，且$\mathcal{O} \models \alpha$，其中$\alpha$是蕴含。对于$\mathcal{O}$的一个子集$\mathcal{O}'$，如果对于$\mathcal{O}'$的所有子集$\mathcal{O}''$满足$\mathcal{O}'' \not\models \alpha$且$\mathcal{O}' \models \alpha$，那么$\mathcal{O}'$就是本体$\mathcal{O}$ 中对于α的一个决断集。

例 2.1 考虑一个一致的DL本体$\mathcal{O}$，其中TBox包含以下三条公理:

$$(1)\ \mathsf{Girl} \sqsubseteq \mathsf{Female}$$

$$(2)\ \mathsf{Female} \sqsubseteq \mathsf{Person}$$

$$(3)\ \exists \mathsf{giveBirth.Person} \sqsubseteq \mathsf{Female}$$

ABox包含以下两条公理:

$$(1)\ \mathsf{Female}(Mary)$$

$$(2)\ \mathsf{giveBirth}(Lily, Mary)$$

蕴含为:

$$\mathsf{Female}(Lily)$$

2.3 描述逻辑溯因诊断

2.3.1 论域词汇表与术语断言

2.3.2 溯因诊断

逻辑研究的是基于规则的推理方式，目前的研究中把推理的方式分为三类，分别是演绎、归纳和溯因推理。演绎推理是最常用的推理方式，演绎推理根据已

有的前提事实以及规则，得出结论。对于相同的输入，如果严格按照规则进行运算，演绎推理具有相同的输出，具有恒真性(truth-preserving)。归纳推理则是在已知事实的集合中寻找共同特性，推导出更多事实或同类事实的性质[论语用推理的逻辑属性]。它的推理格式形如以下形式：

a. 所有已知的A为B。

b. 因此，A为B。

溯因推理是推理方式中的第三种方式，溯因推理的方式与前两种推理方式有着本质的区别，溯因推理又称作反绎推理，是推理到最佳解释的过程。一般的，它是开始于事实的集合并推导出它们的最合适的解释的推理过程。术语溯因（abduction）意味生成假设来解释观察或结论。因为需要生成假设来解释观察或结论，因此溯因推理会在进行解释的过程中为前提事实增加新的知识使得前提事实与解释的并集可以通过演绎推理的方式演绎出观察值或结论。在描述逻辑本体中，溯因推理是一种重要的推理方式。在描述逻辑本体的构建过程中，本体由于构建不够完善，会经常性出现本体无法蕴含观察值的现象。因此，这个时候就需要进行利用溯因推理的方法，已构建的本体进行诊断，找出本体不完善的原因，并在找到的原因的基础上，提出解释对本体进行修复。这种找出原因并提出解释的推理方法就叫做溯因诊断。溯因诊断对修复本体有着重要的意义。我们对术语断言的溯因诊断问题进行了如下两个定义[ABox Abduction in the Description Logic ALC]：

定义 2.5 (术语断言溯因诊断问题) 令$\mathcal{L}_K$和$\mathcal{L}_Q$为两个DL本体，$\mathcal{K} = (\mathcal{T}, \mathcal{A})$是一个$\mathcal{L}_K$中的一个知识库，$\Phi$是一个在$\mathcal{L}_Q$中的术语断言集合。当且仅当$\mathcal{K} \nvDash \Phi$且$\mathcal{K} \cup \Phi \nvDash \bot$的时候，二元组$\langle \mathcal{K}, \Phi \rangle$被称作术语断言的溯因诊断问题。

定义 2.6 (术语断言溯因诊断解释) 令$\mathcal{L}_S$为一个DL本体，且A为一个多个在$\mathcal{L}_S$中的术语断言集合。对于一个溯因诊断问题$\langle \mathcal{K}, \Phi \rangle$，当且仅当$\mathcal{K} \cup A \vDash \Phi$我们把$A$称为可接受解释。更多的，我们把$A$称为：

1. (一致) 当且仅当$\mathcal{K} \cup A \nvDash \bot$。

2. (非平凡) 当且仅当$\mathcal{K} \nvDash \Phi$。

3. (最小) 当且仅当不存在对于$\langle \mathcal{K}, \Phi \rangle$问题的解释$B$，其中$B$是$A$的实例化子集。我们说$B$是$A$的实例化子集当且仅当存在一个重命名映射$\rho$:$N_I^*(B) \mapsto N_I^*(A)$，其中$N_I^*(B)$和$N_I^*(A)$是来自于$A$和$B$的个体名且不出现于$\mathcal{K}$，使得$A \vDash \rho B$。但是对于所有的$\varrho$:$N_I^*(A) \mapsto N_I^*(B)$满足$B \nvDash \varrho A$。满足以上条件的，我们称$A$是问题$\langle \mathcal{K}, \Phi \rangle$的最小解释。

术语断言的溯因诊断需要计算出由一条或多条术语断言的集合，这些集合需要满足最小集的条件。当这些集合被加入到本体中的时候，它需要保持本体的一致性，并且可以使得更新后的本体能够蕴含(⊨)含观察值。$\mathcal{O} \vDash \alpha$表示对于所有满足本体$\mathcal{O}$的模，都可以使得$\alpha$成立。

2.4 定义

定义 2.7 (可扩展公理) 对于本体的一条公理，如果公理中的一个或多个的二元关系或者个体被二元关系变量或者个体变量替换，则这条公理会被称作可扩展公理。一条可扩展公理会被称为可全扩展公理如果这条公理的所有二元关系和个体都被二元关系和个体变量替换。

定义 2.8 (替换) 对于一条可扩展公理或者一个由可扩展公理构成的集合E，替换是一个从E中的二元关系变量或个体变量到其它二元关系变量或者二元关系以及个体变量或个体的映射。其中，如果该替换把所有二元关系变量映射到二元关系以及所有个体变量映射到个体，我们就把这个替换称作实例化替换。

定义 2.9 (解释) 给定一个一致的本体$\mathcal{O}$以及一个二元关系实例α，$\mathcal{O} \nvDash \alpha$, 并且$\mathcal{O} \cup \{\alpha\}$ 是一致的，那么假如存在一个公理的集合$\mathcal{E}$使得$\mathcal{O} \cup \mathcal{E} \models \alpha$，$\mathcal{E} \nvDash \alpha$并且$\mathcal{O} \cup \mathcal{E}$ 是一致的，那么我们称这个集合$\mathcal{E}$为在本体$\mathcal{O}$中对α的解释。

定义 2.10 ($\subseteq_{ds}$-*minimal* 解释) 一个解释$\mathcal{E}$会被称为$\subseteq_{ds}$-*minimal* 如果这个$\mathcal{E}$满足以下条件：不存在这样一个解释$\mathcal{E}'$使得$\mathcal{E}' \subseteq_{ds} \mathcal{E}$ 且$\mathcal{E} \not\subseteq_{ds} \mathcal{E}'$，其中$\mathcal{E}' \subseteq_{ds} \mathcal{E}$ 表示存在一个$\mathcal{E}'$的差异化替换θ，使得$\mathcal{E}'\theta \subseteq \mathcal{E}$。

定义 2.11 (基于决定集模版的解释) 对于一个给定的一致本体$\mathcal{O}$，一个基于$role(X,Y)$的模版$\mathcal{P}$以及一个观察值α其中$\mathcal{T} \nvDash \alpha$且$\mathcal{O} \cup \{\alpha\}$是一致的，对于$\alpha$在$\mathcal{O}$的解释$\mathcal{E}$会被称为基于决定集模版的解释，如果这个解释满足以下四个条件：

- (非平凡) $\mathcal{E} \nvDash \alpha$
- (一致) $\mathcal{O} \cup \mathcal{E}$ 是一致的
- ($\subseteq_{ds}$ $-$*minimal*) 不存在一个对于$alpha$在$\mathcal{O}$中的解释$\mathcal{E}'$使得$\mathcal{E}' \subseteq_{ds} \mathcal{E}$ 且$\mathcal{E} \not\subseteq_{ds} \mathcal{E}'$。
- (可容许) 存在一个决定集模版$\mathcal{J}_p \in \mathcal{P}$以及$\mathcal{J}_p$的一个差异化替换$\theta$使得$\alpha = role(X\theta, Y\theta)$，$\mathcal{E} \subseteq \mathcal{J}_p\theta$且$\mathcal{J}_p\theta \in Jst(\alpha, \mathcal{O} \cup \mathcal{E})$

定义 2.12 (诊断问题) 我们把$\mathcal{P} = (\mathcal{T}, \mathcal{A}, \alpha)$称作一个诊断的问题实例，其中本体$\mathcal{O} = \mathcal{T} \cup \mathcal{A}$是一个一致的描述逻辑本体。对于问题$\mathcal{P}$的一个解释$\mathcal{E}$满足：$\mathcal{T} \cup \mathcal{A} \cup \mathcal{E} \models \alpha$ 且$\mathcal{T} \cup \mathcal{A} \cup \mathcal{E} \nvDash \bot$。

定义 2.13 (决定集) 对于一个一致的描述逻辑本体$\mathcal{O}$，且$\mathcal{O} \models \alpha$($\alpha$是一个推论)，$\mathcal{O}$的子集$\mathcal{O}'$会被称作在本体$\mathcal{O}$中对于$\alpha$的决定集如果$\mathcal{O}' \models \alpha$，且对于所有的$\mathcal{O}'' \subset \mathcal{O}'$ 都有$\mathcal{O}'' \nvDash \alpha$。

定义 2.14 (决定集模版) 一个由扩展公理组成的集合$\mathcal{J}_p$会被称为在本体$\mathcal{O}$中对于$role\,(x,y)$的决定集模版如果$\mathcal{J}_p$满足以下两个条件:(1) 存在一个替换θ使得$\mathcal{J}_p\theta \in$ *Jst* $(role\,(X\theta, Y\theta)\,, \mathcal{O})$ (2)对于所有的实例化替换σ存在$\mathcal{J}_p\sigma \models role\,(X\sigma, Y\sigma)$

命题 2.15 给定一个一致的本体$\mathcal{O}$, 一个在$\mathcal{O}$中对于$role(X,Y)$的决定集模版$\mathcal{P}$，以及一个观察值$role(A,B)$，其中$\mathcal{O} \nvDash role(A,B)$ 且$\mathcal{O} \cup \{role(A,B)\}$是一致的，那么对于观察值$role(A,B)$在本体$\mathcal{O}$中由$\mathcal{P}$映射得到的解释集合$\mathcal{S} = \{\mathcal{J}_2\theta \mid \mathcal{J}_p \in \mathcal{P}$, $(\mathcal{J}_1, \mathcal{J}_2) \in$ *bipart*$(\mathcal{J}_p)$,且θ是一个在$\mathcal{J}_1 \cup \{role(X,Y)\}$上的替换使得$X\theta = A, Y\theta = B, \mathcal{J}_1\theta \subseteq \mathcal{O}, \mathcal{J}_2\theta \nvDash role(A,B)\}$。

命题 2.16 一个对于观察值α在本体$\mathcal{O}$中的$\subseteq_{ds}$-*minimal* 解释$\mathcal{E}$同时也是一个*subset-minimal*的解释。

命题 2.17 给定一个一致的本体$\mathcal{O}$, 一个在$\mathcal{O}$中对于$role(X,Y)$的决定集模版$\mathcal{P}$，以及一个观察值$role(A,B)$，其中$\mathcal{O} \nvDash role(A,B)$ 且$\mathcal{O} \cup \{role(A,B)\}$是一致的，那么对于观察值$role(A,B)$在本体$\mathcal{O}$中由$\mathcal{P}$映射得到的解释集合$\mathcal{S} = \{\mathcal{J}_2\theta \mid \mathcal{J}_p \in \mathcal{P}, (\mathcal{J}_1, \mathcal{J}_2) \in bipart(\mathcal{J}_p)$,且$\theta$是一个在$\mathcal{J}_1 \cup \{role(X,Y)\}$上的替换使得$X\theta = A, Y\theta = B, \mathcal{J}_1\theta \subseteq \mathcal{O}, \mathcal{J}_2\theta \nvDash role(A,B)\}$。

例 2.2 **(诊断)** 考虑以下训练集：

1) $(Mike, /isFatherOf, Josan)$

2) $(Mike, /isParentOf, Lily)$

3) $(Hill, /isFriendOf, Peter)$

关系路径: 观察值:

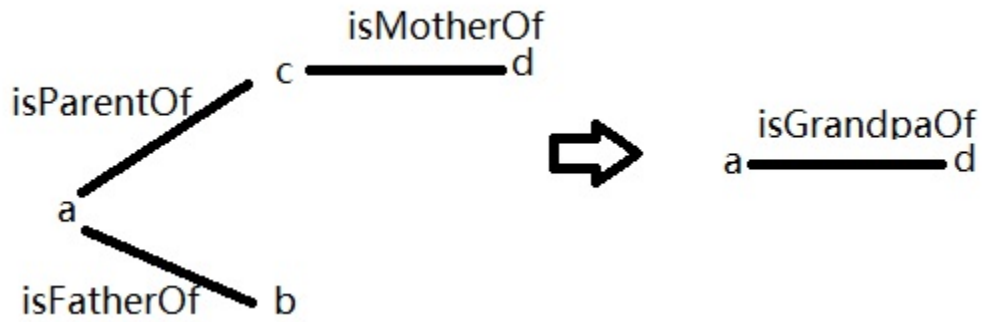

图 2.1: 关系路径

$$(Mike, /isGrandpaOf, Peter)$$

对于例中的训练集，关系路径以及观察值，我们使用本体语言对其进行表达，分别得到TBox $\mathcal{T}$,ABox $\mathcal{A}$以及观察值α。

其中$\mathcal{T}$为：

$$isFatherOf\,(a,b) \wedge isParentOf\,(a,c) \wedge isMotherOf\,(c,d) \rightarrow isGrandpaOf\,(a,d)$$

$\mathcal{A}$为:

1) *isFatherOf*$(Mike, Josan)$

2) *isParentOf*$(Mike, Lily)$

3) *isFriendOf*$(Hill, Peter)$

观察值α为:

isGrandpaOf$(Mike, Peter)$

使用二分法对$\mathcal{T}$中的公理进行划分，分别得到以下公理集：

1) J_1: $isFatherOf\,(a,b) \wedge isParentOf\,(a,c)$$J_2$: $isMotherOf\,(c,d)$

2) J_1: $isParentOf\,(a,c) \wedge isMotherOf\,(c,d)$$J_2$: $isFatherOf\,(a,b)$

3) J_1: $isFatherOf\,(a,b) \wedge isMotherOf\,(c,d)$$J_2$: $isParentOf\,(a,c)$

4) J_1: $isMotherOf\,(c,d)$$J_2$: $isFatherOf\,(a,b) \wedge isParentOf\,(a,c)$

5) J_1: $isFatherOf\,(a,b)$$J_2$: $isParentOf\,(a,c) \wedge isMotherOf\,(c,d)$

6) J_1: $isParentOf\,(a,c)$$J_2$: $isFatherOf\,(a,b) \wedge isMotherOf\,(c,d)$

首先，我们使用第1个公理集中的J_1 在$\mathcal{A}$中进行实例化，得出以下替换$\mathcal{M}$:

$$\{a \to Mike, b \to Josan, c \to Lily, d \to Peter\}$$

根据得出的替换$\mathcal{M}$,对公理集中的J_2执行实例化，因此得到的第一个解释$\mathcal{E}_1$为：

$$\{\mathit{isMotherOf}(Lily, Peter)\}$$

第 三 章　基于决断集模板进行溯因诊断

本体中的知识库以及规则是本体可以进行推理的前提。目前的本体构造主要由手工或者是半手工构建。由于本体的结构复杂，信息量大，因此本体的构建是一个长期的过程。在本体的构建过程中，构建的本体与观察值的不一致是一个常会发生的问题。因此，找到问题的原因，对观察值提出合理解释成为本体推理中的一个重要任务，这类任务也被称作是溯因推理问题。

3.1　解释与决定集

溯因推理的一个重要目的就是需要找出合理的解释对观察值的进行解释。一般的，这个解释不应该导致诊断本体的不一致，同时，结合本体中的背景知识，新构建出的本体能够蕴含观察值，且新构建的本体能够保持一致的特性。因此，我们对解释有以下定义：

定义 3.1 **(解释)** 给定一个一致的本体$\mathcal{O}$以及观察值α，$\mathcal{O} \nvDash \alpha$, 并且$\mathcal{O} \cup \{\alpha\}$ 是一致的，那么假如存在一个公理的集合$\mathcal{E}$使得$\mathcal{O} \cup \mathcal{E} \models \alpha$，$\mathcal{E} \nvDash \alpha$并且$\mathcal{O} \cup \mathcal{E}$ 是一致的，我们称这个集合$\mathcal{E}$为在本体$\mathcal{O}$中对α的解释。

为了能够满足本体对蕴涵值的推理需求，解释在本体的溯因诊断中通常会以一种表达能力较高的DL语言比如$\mathcal{SROIQ}$。表达能力高的语言虽然可以满足本体对观察值的推理需求，但是较高的表达能力会带来另外一个问题，解释的空间会无限增大。为了能够尽可能提高解释的表达能力的同时能够限制解释的空间大小，本文利用模板来实现对解释的空间进行限制。在本体的推理中，决断集是一个重要的概念，对于一个一致的本体O以及观察值α，我们定义$Jst(\alpha, \mathcal{O})$为本体$\mathcal{O}$对观察值α的决断集的集合。因为决断集对蕴含值具有推理的合理性，同时决断集满足本体对观察值的解释需要是最小集合的约束，因此直观地本体对观察值的解释也会遵循相应的模板。

在定义决断集模板之前，我们首先需要可扩展公理，对于本体的一条公理，

如果公理中的一个或多个的二元关系或者个体被二元关系变量或者个体变量替换，则这条公理会被称作可扩展公理。更多地，一条可扩展公理会被称为可全扩展公理如果这条公理的所有二元关系的个体都被二元关系中的个体变量替换。同时，可全扩展公理会保留本体中的$\top$，$\bot$以及个体的不变。在本体中，公理的类型种类多，由多条公理组成的决断集会产生数量难以接受的模板，因此我们需要限制模板的数量，也就意味着我们需要使用尽可能少的模板来表达尽可能多的决断集，同时，对于每个生成的模板，应该要有一种映射的方式使得决断集与之相对应。这种映射的方式我们称之为可扩展公理的替换，可扩展公理的替换能够把扩展公理的变量个体映射到个体名或者是个体变量。一个替换会被称作实例化替换如果这个替换能够把所有的个体变脸映射到个体名。

一般地，我们都需要这样的限制，对于每个由决断集产生的模板，都存在一个实例化的替换，使得这个由决断集产生的模板被映射到一个决断集中。但是这样的限制依旧会产生公理数不受限制的模板，考虑以下例子：

例 3.1 (解释) 令决断集$\mathcal{J}$是一直本体$\mathcal{O}$中对于二元关系断言$r_y(e_1, e_2)$的决断集，J_p是一个由决断集生成的决断集模板：

$$\mathcal{J} = \{r_M \circ r_N \sqsubseteq r_K, r_M(e_1, e_\mu), r_N(e_\mu, e_2)\}$$

$$\mathcal{J}_p = \{r_M \circ r_N \sqsubseteq r_K, r_M(e_{x_1}, e_{y_1}), \cdots, r_M(e_{x_n}, e_{y_n}), r_N(e_m, e_n)\}$$

从$\mathcal{J}$和$\mathcal{J}_p$可以看出，$\mathcal{J}_p$是一个公理基数没有上限的决断集模板，因为存在这样一个映射θ：

$$\theta = \{e_{x_i} \mapsto e_1, e_{y_i} \mapsto e_\mu, e_m \mapsto e_\mu, e_n \mapsto e_2 \mid 1 \leq i \leq n\}$$

由于这个θ满足条件所有的个体变量被映射到个体实例，因此这个映射θ是一个实例化替换使得$\mathcal{J}_p\theta = \mathcal{J}$。

为了避免出现决断集模板基数无限增大的情况，我们需要的决断集模板的替换做出进一步限制。在实例化替换的基础上，我们提出差异化实例替换。差异化实例替换不仅需要满足条件所有的个体变量都被映射到个体实例，还需要满足条件对于所有的不相同变量，被映射后的个体也不相同。直观地，一个差异化实例替换会是变量到个体间的一一映射。同时，一般地一个决断集模板需要能够生成至少一个决断集，然而为了保证最后生成的解释的合理性，我们需要决断集模板的所有差异化实例替换都只生成决断集，因此我们需要限制所有的对于$r(e_x, e_y)$决断集模板，在差异化实例替换的映射下能够维持蕴含$r(e_x\theta, e_y\theta)$这一特性。因此我们对决断集模板做出如下定义：

定义 3.2 (决定集模版) 一个由扩展公理组成的集合$\mathcal{J}_p$会被称为在本体$\mathcal{O}$中对于$r(e_x\, e_y)$的决定集模版如果$\mathcal{J}_p$满足以下两个条件:(1) 存在一个替换θ使得$\mathcal{J}_p\theta \in Jst\left(r\left(e_x\theta, e_y\theta\right), \mathcal{O}\right)$ (2)对于所有的差异化实例替换σ存在$\mathcal{J}_p\sigma \models r\left(e_x\sigma, e_y\sigma\right)$。

继续考虑例3.1中的决定集模板:

$$\mathcal{J}_p = \{r_M \circ r_N \sqsubseteq r_K, r_M(e_{x_1}, e_{y_1}), \cdots, r_M(e_{x_n}, e_{y_n}), r_N(e_m, e_n)\}$$

根据以上定义，在任一个一致的本体$\mathcal{O}$中，$\mathcal{J}_p$不是一个符合定义的对于二元关系断言实例$r_K(e_1, e_2)$的决断集模板。因为对于决定集模板$\mathcal{J}_p$的差异化实例替换θ,存在:

$$\{r_M \circ r_N \sqsubseteq r_K, r_M(e_{x_1}\theta, e_{y_1}\theta), r_N(e_m\theta, e_n\theta)\} \in Jst(r_M(e_{x_1}, e_{y_1}), \mathcal{O})$$

同时:

$$\{r_M \circ r_N \sqsubseteq r_K, r_M(e_{x_1}\theta, e_{y_1}\theta), r_N(e_m\theta, e_n\theta)\} \subset \mathcal{J}_p\theta$$

因此可以得出：

$$\mathcal{J}_p\theta \notin Jst(r_M(e_{x_1}\theta, e_{y_1}\theta), O)$$

同时，$\mathcal{J}_p$也不是$r_M(e_{x_1}, e_{y_k})$的决断集模板当$k > 1$。这是因为$\mathcal{J}_p\sigma \nvDash r_M(e_{x_1}\sigma, e_{y_k}\sigma)$，其中：

$$\sigma = \{e_{x_i} \mapsto e_1, e_{y_i} \mapsto e_2 \mid 1 \leq i \leq n, i \neq k\} \cup \{e_{x_k} \mapsto e_2, e_{y_k} \mapsto e_1\}$$

为了保证决断集能够被映射到决断集空间上，我们使用从决断集上生成决断集模板的方法来获取决断集模板。我们在生成的决断集的基础上，对二元关系中的个体进行变量替换，且对于不相同的个体名，我们使用不同的个体变量替换。我们对用不相同变量替换不相同个体的过程记作$lift(S, A, B)$。$lift(S, A, B)$表示对公理集合中的所有个体名，我们把A映射为变量X, B映射为变量Y, 其它不相同的个体分别映射到不同的个体变量。

命题 3.3 令$\mathcal{O}$为一致本体，$r_M(e_1, e_2)$ 是$\mathcal{O}$的一个蕴涵值，且$\mathcal{J}$是在$\mathcal{O}$中对于蕴含值$r_M(e_1, e_2)$的一个决断集。那么$lift(\mathcal{J}, e_1, e_2)$就是一个在$\mathcal{O}$中对于$r_M(e_1, e_2)$的一个决断集模板。

证明：(1)因为在决断集$\mathcal{J}$与$lift(\mathcal{J}, e_1, e_2)$之间具个体变量到个体实体的一对一映射，因此这里存在一个$lift(\mathcal{J}, e_1, e_2)$的差异化实例替换$\theta$使得$X\theta = e_1$，$Y\theta = e_2$，且$lift(\mathcal{J}, e_1, e_2) \cdot \theta = \mathcal{J}$。(2)令$\sigma$为$lift(\mathcal{J}, e_1, e_2)$的一个差异化实例替换，因此对于$lift(\mathcal{J}, e_1, e_2) \cdot \sigma$ 必然存在一个到$lift(\mathcal{J}, e_1, e_2) \cdot \theta$个体变量之间的映射$\rho$从而使得$X\theta = X\sigma$，$Y\theta = Y\sigma$且$lift(\mathcal{J}, e_1, e_2) \cdot \theta \cdot \rho = lift(\mathcal{J}, e_1, e_2) \cdot \sigma$。又因为$lift(\mathcal{J}, e_1, e_2) \cdot \theta \models r_M(X\theta, Y\theta)$，因此必然有$lift(\mathcal{J}, e_1, e_2) \cdot \theta \cdot \rho \models r_M(X\sigma, Y\sigma)$，考虑到$lift(\mathcal{J}, e_1, e_2) \cdot \theta \cdot \rho = lift(\mathcal{J}, e_1, e_2) \cdot \sigma$，因此可知对于所有的差异化实例替换$\sigma$都有$lift(\mathcal{J}, e_1, e_2) \cdot \sigma \models r_M(X\sigma, Y\sigma)$，命题成立。

要使得对于观察值α以及一致本体$\mathcal{O}$中的溯因解释$\mathcal{E}$遵循决断集模板，我们做出如下限制：$\mathcal{E}$是观察值α在$\mathcal{O} \cup \mathcal{E}$中的决断集的子集，其中这个决断集是由决断集模板通过差异化实例替换计算得来。我们把这个解释$\mathcal{E}$称作可接受解释：

定义 3.4 给定一个一致本体$\mathcal{O}$，一个可蕴含$r_M(X, Y)$决断集模板的集合$\mathcal{P}$以及观察值α，其中$\mathcal{O} \not\models \alpha$，解释$\mathcal{E}$ 如果满足存在一个决断集模板$\mathcal{J}_p \in \mathcal{P}$以及一个在决断集模板$\mathcal{J}_p$上的差异化实例替换$\theta$使得$\alpha = r_M(X\theta, Y\theta), \mathcal{E} \subseteq \mathcal{J}_p\theta$以及$\mathcal{J}_p\theta \in Jst(\alpha, \mathcal{O} \cup \mathcal{E})$，我们就把这样的解释称作可接受解释。

下面给出一个可接受解释例子：

例 3.2 考虑以下本体$\mathcal{O}$：

$$\{r_M \circ r_N \sqsubseteq r_K, r_M(e_1, e_3)\}$$

决断集模板$\mathcal{J}_p$为：

$$\{r_M \circ r_N \sqsubseteq r_K, r_M(A, e_w), r_N(e_w, B)\}$$

观察值α为：

$$r_K(e_1, e_2)$$

根绝决断集模板$\mathcal{J}_p$，我们可以得到$\mathcal{E} = r_M(e_1, e_3), r_N(e_3, e_2)$是一个对于观察值$r_K(e_1, e_2)$在$\mathcal{O}$中的可接受解释，因为存在一个差异化实例替换：

$$\theta = \{A \mapsto e_1, e_w \mapsto e_3, B \mapsto e_2\}$$

满足$r_o(e_1, e_2) = r_o(X\theta, Y\theta)$，$\mathcal{E} \subseteq \mathcal{J}_p\theta$且$J_p\theta \in Jst(r_o(e_1, e_2), \mathcal{O} \cup \mathcal{E})$。

给定一个一致本体$\mathcal{O}$以及一个观察α，其中$\mathcal{O} \not\models \alpha$，且$\mathcal{O} \cup \{\alpha\}$是一致的，那么本体的使用者更通常地会考虑以下三个特性：非平凡(i.e., $\mathcal{E} \not\models \alpha$)，一致(i.e. $\mathcal{O} \cup \mathcal{E}$是一致的)以及子集最小的(i.e., $\mathcal{O} \cup \not\models \alpha$ 对于所有的$\mathcal{E}' \subset \mathcal{E}$)。我们称满足以上三个条件的解释为合理解释。可以看出，在例子3.2中得到的解释$\mathcal{E} = \{r_M(e_1, e_3), r_N(e_3, e_2)\}$不是一个合理的解释。因为存在一个$\mathcal{E}$的子集$\mathcal{E}' = \{r_N(e_3, e_2)\}$)满足$r_o(e_1, e_2) = r_o(X\theta, Y\theta)$，$\mathcal{E}' \subseteq \mathcal{J}_p\theta$且$J_p\theta \in Jst(r_o(e_1, e_2), \mathcal{O} \cup \mathcal{E}')$。所以，我们结合以上条件定义如果一个可接受的解释满足非平凡，一致以及子集最小这三个条件，那么我们就称它为合理解释。

例 3.3 继续考虑例3.2，本体$\mathcal{O}$以及$\mathcal{O} \cup r_K(e_1, e_2)$也是一致的。正如前文提到的，$\mathcal{E}$并不是一个合理的解释，因为对于本体$\mathcal{O}$ 以及观察值$r_K(e_1, e_2)$，$\mathcal{E}$并不是一个最小集使得$r_o(e_1, e_2) = r_o(X\theta, Y\theta)$，$\mathcal{E} \subseteq \mathcal{J}_p\theta$且$J_p\theta \in Jst(r_o(e_1, e_2), \mathcal{O} \cup \mathcal{E})$。

相对的，$\mathcal{E}' = \{r_N(e_3, e_2\})$是根据模板集合$\mathcal{P}$在本体$\mathcal{O}$中对观察值$\alpha$的一个合理解释，使得$\mathcal{E}' \not\models r_K(e_1, e_2)$，$\mathcal{O} \cup \mathcal{E}'$是一致的且对于所有的$\mathcal{E}'' \subset \mathcal{E}'$不存在$\mathcal{O} \cup \mathcal{E}'' \models r_K(e_1, e_2)$。

在本体的诊断中，由于解释需要遵循决断集模板，而且不存在$\mathcal{J}_p\theta \subset \mathcal{O}$的情况，因此并非所有的变量个体都会被映射到本体$\{\mathcal{O} \cup \alpha\}$的个体集合中。在本文中，我们称这些变量为解释变量。由于解释变量的存在，如果想保证诊断的方法能够更加高效，一种考虑是只计算无法被其它解释通过实例化替换后得到的解释。然而解释变量的允许却会导致解释公理集合基数的变大甚至无法限制，考虑以下解释$r_o(A, A)$，这个解释可以由解释$\mathcal{E} = \{r_o(A, e_{x_1}), r_o(e_{x_2}, e_{x_3}), \cdots, r_o(e_{x_{n-1}}, e_n)\}$通过映射$\theta = \{e_{x_i} \mapsto A \mid 1 \leq i \leq n\}$获得。因此对于解释的替换，我们提出一个新的概念，叫做解释差异化替换。解释$\mathcal{E}$的差异化替换会可以把$\mathcal{E}$的解释变量映射到不同的解释变量或者是一个没有在$\mathcal{E}$中已经出现的变量。在前文提到的例子中，θ并不是一个解释的差异化替换，因为e_{x_i}被映射到的变量A是解释$\mathcal{E}$中已经存在的个体实例，所以θ不是解释的差异化替换。因此我们提出一个新的子集概念：

定义 3.5 ($\subseteq_{\mathsf{ds}}$-*minimal* **解释**) 一个解释$\mathcal{E}$会被称为$\subseteq_{\mathsf{ds}}$-*minimal* 如果这个$\mathcal{E}$满足以下条件：不存在这样一个解释$\mathcal{E}'$使得$\mathcal{E}' \subseteq_{\mathsf{ds}} \mathcal{E}$ 且$\mathcal{E} \not\subseteq_{\mathsf{ds}} \mathcal{E}'$，其中$\mathcal{E}' \subseteq_{\mathsf{ds}} \mathcal{E}$ 表示存在一个$\mathcal{E}'$的差异化替换θ，使得$\mathcal{E}'\theta \subseteq \mathcal{E}$。

$\subseteq_{\mathsf{ds}}$-*minimal*解释是一个比子集最小更强的概念，以下定理说明$\subseteq_{\mathsf{ds}}$-*minimal*是子集最小的充分条件：

命题 3.6 一个对于观察值α在本体$\mathcal{O}$中的$\subseteq_{\mathsf{ds}}$-*minimal* 解释$\mathcal{E}$同时也是一个*subset-minimal*的解释。

证明：对于所有的$\mathcal{E}' \subset \mathcal{E}$，我们有$\mathcal{E}' \subseteq_{\mathsf{ds}} \mathcal{E}$且$\mathcal{E}' \not\subseteq_{\mathsf{ds}} \mathcal{E}$。所以，如果不存在对于观察值$\alpha$在本体$\mathcal{O}$中的解释$\mathcal{E}'$使得$\mathcal{E}' \subseteq_{\mathsf{ds}} \mathcal{E}$且$\mathcal{E}' \not\subseteq_{\mathsf{ds}} \mathcal{E}$，那么也不会存在$\mathcal{E}' \subset \mathcal{E}$使得$\mathcal{O} \cup \mathcal{E}' \models \alpha$，所以命题成立。

验证解释$\mathcal{E}$是否是$\subseteq_{\mathsf{ds}}$-*minimal*不需要考虑解释$\mathcal{E}$的所有子集组合排列的情况。实际上我们只需要考虑n个更小的解释即可。这里的n等于$\mathcal{E}$中公理数的基数以及个体实例的总和。在介绍这两种方法之前，我们需要先介绍两种集合概念。一种是公理集S的最近变量集$lifts_1(S)$，最近变量集$lifts_1(S)$使用变量替换一个在S中已经存在的个体实例。例如：对于$S = \{r_o(A, B)\}$存在两个最近变量集分别是$\{r_o(e_x, B)\}$和$\{A, r_o(e_x)\}$。然后，我们定义一个最大真子集，最大真子集有且仅有S中的某一公理外的所有公理。我们把最大真子集记作$subs_1(S)$。以下命题展示了验证解释$\mathcal{E}$是否是一个$\subseteq_{\mathsf{ds}}$-*minimal*解释的方法。

命题 3.7 对于观察值α在本体$\mathcal{O}$中的解释$\mathcal{E}$是一个$\subseteq_{\mathsf{ds}}$-*minimal*解释当且仅当$\mathcal{E}$满足以下条件：$\mathcal{O} \cup \mathcal{E}' \not\models \alpha$对于所有的$\mathcal{E}' \in subs_1(\mathcal{E}) \cup lifts_1(\mathcal{E})$。

证明：令$\mathcal{E}$是一个在本体$\mathcal{O}$中对于观察值α的解释，且$\mathcal{E}$满足$\mathcal{O} \cup \mathcal{E}' \not\models \alpha$对于所有的$\mathcal{E}' \in subs_1(\mathcal{E}) \cup lifts_1(\mathcal{E})$。令$\mathcal{E}'$为公理组成的集合，$\theta$是$\mathcal{E}'$的一个差异化替换使得$\mathcal{E}'\theta \subset \mathcal{E}$。因此必然存在$\mathcal{E}'' \in subs_1(\mathcal{E})$使得$\mathcal{E}'\theta \subseteq \mathcal{E}''$。因为$\mathcal{E}'' \cup \mathcal{O} \not\models \alpha$，所以我们有$\mathcal{E}'\theta \cup \mathcal{O} \not\models \alpha$，从而$\mathcal{E}' \cup \mathcal{O} \not\models \alpha$。令$\mathcal{E}'$是一组公理集合以及$\theta$是$\mathcal{E}$的一个差异化替换使得$\mathcal{E}'\theta = \mathcal{E}$并且$\theta$满足存在至少一个实例使得$\mathcal{E}'$中的解释变量被映射到$\mathcal{E}$个体实例中。因此，必然存在一个$\mathcal{E}'' \in lifts_1(\mathcal{E})$以及一个$\sigma$ 使得$\mathcal{E}'\sigma = \mathcal{E}''$。因为$\mathcal{E}'' \cup \mathcal{O} \not\models \alpha$，所以我们又有$\mathcal{E}' \cup \mathcal{O} \not\models \alpha$。因此对于所有满足$\mathcal{E}' \subseteq_{\mathsf{ds}} \mathcal{E}$且$\mathcal{E} \not\subseteq_{\mathsf{ds}} \mathcal{E}'$的$\mathcal{E}'$我们有$\mathcal{E}' \cup \mathcal{O} \not\models \alpha$。所以$\mathcal{E}$是一个对于观察值$\alpha$在$\mathcal{O}$中的$\subseteq_{\mathsf{ds}}$-*minimal*解释。

通过使用$\subseteq_{\mathsf{ds}}$-*minimal*，我们获得了比子集最小更强的对解释的限制能力。因此，结合上文提到的限制条件，我们得到一类新的解释集合，称为基于决断集模板的解释，定义如下：

定义 3.8 **(基于决定集模版的解释)** 对于一个给定的一致本体$\mathcal{O}$，一个基于$r_o(X, Y)$的模版$\mathcal{P}$以及一个观察值α其中$\mathcal{T} \not\models \alpha$且$\mathcal{O} \cup \{\alpha\}$是一致的，对于$\alpha$在$\mathcal{O}$ 的解释$\mathcal{E}$会被称为基于决定集模版的解释，如果这个解释满足以下四个条件：

- (非平凡) $\mathcal{E} \not\models \alpha$
- (一致) $\mathcal{O} \cup \mathcal{E}$ 是一致的

- ($\subseteq_{\mathsf{ds}}$ *−minimal*) 不存在一个对于$alpha$在$\mathcal{O}$中的解释$\mathcal{E}'$使得$\mathcal{E}' \subseteq_{\mathsf{ds}} \mathcal{E}$ 且$\mathcal{E} \not\subseteq_{\mathsf{ds}} \mathcal{E}'$。
- (可容许) 存在一个决定集模版$\mathcal{J}_p \in \mathcal{P}$以及$\mathcal{J}_p$的一个差异化替换$\theta$使得$\alpha = r_o(X\theta, Y\theta)$，$\mathcal{E} \subseteq \mathcal{J}_p\theta$且$\mathcal{J}_p\theta \in Jst(\alpha, \mathcal{O} \cup \mathcal{E})$

计算基于决断集模板的解释是一个多项式时间内的计算过程。根据观察值$r_o(A, B)$以及决断集模板集合$\mathcal{P}$计算在本体$\mathcal{O}$中的基于决断集模板解释步骤如下：对于所有在$\mathcal{P}$中的决断集模板$\mathcal{J}_p$，我们使用二分法对决断集模板$\mathcal{J}_p$进行切分，其中一部分公理的集合我们把它看做是本体$\mathcal{O}$的对应公理集合，另一部分我们把它看做是我们需要求得的基于决断集模板的解释。给定一个决断集模板$\mathcal{J}_p$，我们定义$\mathcal{J}_p$的二分结果为二元组$(\mathcal{J}_1, \mathcal{J}_2)$，其中$\mathcal{J}_1 \cap \mathcal{J}_2 = \emptyset$且$\mathcal{J}_1 \cup \mathcal{J}_2 = \mathcal{J}_p$。我们把$\mathcal{J}_p$二分后的二元组集合记作$B^*(\mathcal{J}_p)$。同时，对于一个由多条扩展公理的集合$S$我们把从个体变量映射到个体实例的一一映射操作记作$inst(S)$。因此，我们有如下命题：

命题 3.9 给定一个一致的本体$\mathcal{O}$，一个观察值$r_o(A, B)$且$\mathcal{O} \not\models r_o(A, B)$，以及蕴含$r_o(X, Y)$的决断集模板的集合$\mathcal{P}$，那么在本体$\mathcal{O}$中基于决断集模板集合$\mathcal{P}$对于观察值$r_o(A, B)$ 的基于决断集模板的解释的集合$\mathcal{S} = \{inst(\mathcal{J}_1\theta) \mid \mathcal{J}_p \in \mathcal{P}, (\mathcal{J}_1, \mathcal{J}_2) \in B^*(\mathcal{J}_p)$ 且θ 是$\mathcal{J}_2 \cup r_o(X, Y)$的一个差异化实例替换使得$X\theta = A$，$Y\theta = B$，$\mathcal{J}_2\theta \subseteq \mathcal{O}$, 同时$inst(\mathcal{J}_1\theta) \not\models r_o(X, Y)$，$\mathcal{O} \cup inst(\mathcal{J}_1\theta)$是一致的，且$\mathcal{J}_2\theta \cup inst(\mathcal{J}_1\theta) \in Jst(r_o(A, B), \mathcal{O} \cup inst(\mathcal{J}_1)\theta)$,以及不存在一个在本体$\mathcal{O}$对于观察值$r_o(A, B)$的解释$\mathcal{E}'$满足$\mathcal{E}' \subseteq_{\mathsf{ds}} inst(\mathcal{J}_1\theta)$以及$inst(\mathcal{J}_1\theta) \not\subseteq_{\mathsf{ds}} \mathcal{E}'\}$。

证明：(1)令$\mathcal{E}$是集合$\mathcal{S}$中的一个元素，存在一个或多个的$\mathcal{J}_p \in \mathcal{P}$，$(\mathcal{J}_1, \mathcal{J}_2) \in B^*(\mathcal{J}_p)$以及对于$\mathcal{J}_2 \cup \{r_o(X, Y)\}$的差异化实例替换$\theta$使得$\mathcal{E} = inst(\mathcal{J}_1\theta), X\theta = A, Y\theta = B, \mathcal{J}_2\theta \subseteq \mathcal{O}, \mathcal{E} \not\models r_o(A, B)$，$\mathcal{O} \cup \mathcal{E}$是一致的。$\mathcal{J}_2\theta \in Jst(r_o(A, B), \mathcal{O} \cup \mathcal{E})$，并且不存在一个在$\mathcal{O}$中对于$r_o(A, B)$的解释$\mathcal{E}'$满足$\mathcal{E}' \subseteq_{\mathsf{ds}} \mathcal{E}$且$\mathcal{E} \not\subseteq_{\mathsf{ds}} \mathcal{E}'$。由于$\mathcal{E}$是由$\mathcal{J}_1\theta$通过实例化替换而来，因此一定存在一个$\mathcal{J}_p$的实例化替换$\theta'$使得$\mathcal{J}_1\theta' = \mathcal{E}$，$\mathcal{J}_2\theta' = \mathcal{J}_2\theta, X\theta' = X\theta = A$以及$Y\theta' = Y\theta = B$。通过定义可知，我们

有$\mathcal{J}_p\theta' \models r_o(X,Y)$。由于$\mathcal{J}_p\theta' = \mathcal{J}_2\theta \cup \mathcal{E}$并且$\mathcal{J}_2\theta \subseteq \mathcal{O}$，我们有$\mathcal{J}_p\theta \subseteq \mathcal{O} \cup \mathcal{E}$所以$\mathcal{O} \cup \mathcal{E} \models r_o(A,B)$，$\mathcal{E}$是在本体$\mathcal{O}$中对于观察值$r_o(A,B)$的一个解释。另外，$\mathcal{E}$满足基于决断集模板解释的四个特性：非平凡(i.e., $\mathcal{E} \not\models \alpha$)，一致(i.e. $\mathcal{O} \cup \mathcal{E}$ 是一致的)以及$\subseteq_{\mathsf{ds}}$ $-minimal$(不存在这样一个解释$\mathcal{E}'$使得$\mathcal{E}' \subseteq_{\mathsf{ds}} \mathcal{E}$ 且$\mathcal{E} \not\subseteq_{\mathsf{ds}} \mathcal{E}'$，其中$\mathcal{E}' \subseteq_{\mathsf{ds}}$ $\mathcal{E}$ 表示存在一个$\mathcal{E}'$的差异化替换θ，使得$\mathcal{E}'\theta \subseteq \mathcal{E}$)，以及可容许(存在一个决定集模版$\mathcal{J}_p \in \mathcal{P}$以及$\mathcal{J}_p$的一个差异化替换$\theta$使得$\alpha = r_o(X\theta, Y\theta)$，$\mathcal{E} \subseteq \mathcal{J}_p\theta$且$\mathcal{J}_p\theta \in Jst(\alpha, \mathcal{O} \cup \mathcal{E})$)，所以$\mathcal{E}$是一个在$\mathcal{O}$中对于$r_o(A,B)$ 的一个基于决断集模板的解释。(2)令$\mathcal{E}$为在本体$\mathcal{O}$中基于决断集模板$\mathcal{P}$对观察值$r_o(A,B)$的解释。由于$\mathcal{E}$是一个可接受的解释，因此必定存在一个决断集模板$\mathcal{J}_p \in \mathcal{P}$以及一个$\mathcal{J}_p$的差异化实例替换$\theta$使得$X\theta = A, Y\theta = B, \mathcal{E} \subseteq \mathcal{J}_p\theta$以及$\mathcal{J}_p\theta \in Jst(r_o(A,B), \mathcal{O} \cup \mathcal{E})$。由于$\mathcal{E} \subseteq \mathcal{J}_p\theta$以及$\mathcal{J}_p\theta \subseteq$，所以必定存在一个$\mathcal{J}_p$的二分$(\mathcal{J}_1, \mathcal{J}_2)$使得$\mathcal{E} = \mathcal{J}_1\theta$以及$\mathcal{J}_2\theta \subseteq \mathcal{O}$。令$\theta'$是一个$\mathcal{J}_2 \cup \{r_o(A,B)\}$使得$X\theta' = X\theta, Y\theta' = Y\theta$以及$\mathcal{J}_2\theta' = \mathcal{J}_2\theta$，所以$X\theta'A = A, Y\theta' = B, \mathcal{J}_2\theta' \subseteq \mathcal{O}$以及$\mathcal{E}$是$inst(\mathcal{J}_1\theta')$的重名等价解。又一次$\mathcal{E}$是非平凡的，我们有$inst(\mathcal{J}_1\theta') \not\models r_o(A,B)$。$\mathcal{E}$是一致的，因此$\mathcal{O} \cup inst(\mathcal{J}_1\theta')$是一致的。$\mathcal{J}_2\theta \cup \mathcal{E} \in Jst(r_o(A,B), \mathcal{O} \cup \mathcal{E})$，我们有$\mathcal{J}_2\theta' \cup inst(\mathcal{J}_1\theta') \in Jst(r_o(A,B), \mathcal{O} \cup inst(\mathcal{J}_1\theta'))$。又$\mathcal{E}$是$\subseteq_{\mathsf{ds}}$ $-minimal$的，因此不存在一个在本体$\mathcal{O}$对于观察值$r_o(A,B)$的解释$\mathcal{E}'$ 使得$\mathcal{E}' \subseteq_{\mathsf{ds}} inst(\mathcal{J}_1\theta')$并$inst(\mathcal{J}_1\theta') \not\subseteq_{\mathsf{ds}} \mathcal{E}'$

在上文中，我们已经证明了基于决定集模板的解释可以在一个有限的空间中形成。

算法一最终可以对输入的问题求解出所有基于决断集模板的合理解释。

例 3.4 (诊断) 考虑以下诊断问题，其中观察值为:

$$(Mike, Nationality, China)$$

蕴含值$r_{Nationality}(Mike, China)$的决断集:

Input: Triples $\mathcal{A}$ in training data set, observation $r_o(A, B)$, Justification Pattern Set $\mathcal{P}$ for $r_o(X, Y)$
Output: Justification Pattern based Explanations $\mathcal{S}$ for observation $r_o(A, B)$ in $\mathcal{A}$ w.r.t $\mathcal{P}$

$\mathcal{S} \leftarrow \emptyset$
for *each* $\mathcal{J}_p$ *in Justification Pattern Set* $\mathcal{P}$ **do**
  $\mathcal{T}_p \leftarrow TBox\ \mathcal{T}$ in $\mathcal{J}_p$;
  $\mathcal{O} \leftarrow \mathcal{T}_p \cup \mathcal{A}$;
  if $\mathcal{O}$ *is not consistent* **then**
    continue;
  end
  if $\mathcal{O} \cup r_o(A, B)$ *is not consistent* **then**
    continue;
  end
  $\mathcal{B}^* \leftarrow bipart(\mathcal{J}_p)$;
  for *each* $(\mathcal{J}_1, \mathcal{J}_2) \in \mathcal{B}^*$ **do**
    $\mathcal{D} \leftarrow \{\theta | \theta$ is a differentiated substitution for $\mathcal{J}_p$ such that $X\theta = A, Y\theta = B, \mathcal{J}_1\theta \subseteq \mathcal{O}\}$;
    for *each* $\theta \in \mathcal{D}$ **do**
      if $\mathcal{J}_2\theta \models r_o(A, B)$ **then**
        continue;
      end
      if $\mathcal{J}_2\theta$ **or** $\mathcal{J}_2\theta \cup \mathcal{O}$ *is not consistent* **then**
        continue;
      end
      $\mathcal{S}_e \leftarrow \{\mathcal{E}' \mid \mathcal{E}' \in subs_1(\mathcal{J}_2\theta) \cup lifts_1(\mathcal{J}_2\theta)\}$;
      $isMinimal \leftarrow true$;
      for *each* $\mathcal{E}' \in \mathcal{S}_e$ **do**
        if $\mathcal{E}' \cup \mathcal{O} \models r_o(X, Y)$ **then**
          $isMinimal \leftarrow false$;
          continue;
        end
      end
      if $isMinimal$ **then**
        $\mathcal{S} \leftarrow \mathcal{S} \cup \mathcal{E}'$
      end
    end
  end
end

Algorithm 1: 基于决断集模板诊断算法

1) $r_{BornInCity} \circ r_{CityInProvince} \sqsubseteq r_{BornInProvince}$

2) $r_{BornInProvince} \circ r_{ProvinceInCountry} \sqsubseteq r_{Nationality}$

3) $r_{BornInCity}(Mike, Guangzhou)$

4) $r_{CityInProvince}(Guangzhou, Guangdong)$

5) $r_{ProvinceInCountry}(Guangdong, China)$

训练集：

1) $(Mike, \textit{BornInCity}, Hangzhou)$

2) $(Beihai, \textit{CityInProvince}, Guangxi)$

3) $(Zhejiang, \textit{ProvinceInCountry}, China)$

4) $(Guangxi, \textit{ProvinceInCountry}, China)$

根据例子，首先使用本体语言对训练集进行表达，得到ABox $\mathcal{A}$：

1) $r_{BornInCity}(Mike, Hangzhou)$

2) $r_{CityInProvince}(Beihai, Guangxi)$

3) $r_{ProvinceInCountry}(Zhejiang, China)$

4) $r_{ProvinceInCountry}(Guangxi, China)$

然后可得本体$\mathcal{O}$：

1) $r_{BornInCity} \circ r_{CityInProvince} \sqsubseteq r_{BornInProvince}$

2) $r_{BornInProvince} \circ r_{ProvinceInCountry} \sqsubseteq r_{Nationality}$

3) $r_{BornInCity}(Mike, Hangzhou)$

4) $r_{CityInProvince}(Beihai, Guangxi)$

5) $r_{ProvinceInCountry}(Zhejiang, China)$

6) $r_{ProvinceInCountry}(Guangxi, China)$

易得，$\mathcal{O}$与$\mathcal{O} \cup r_{Nationality}(Mike, China)$都是是一致的。又，通过决断集可以得到蕴含$r_{Nationality}(A, B)$的基于决断集模板$\mathcal{J}_p$:

$$r_{BornInCity}(A, C)$$

$$r_{CityInProvince}(C, D)$$

$$r_{ProvinceInCountry}(D, B)$$

观察值α为:

$$r_{Nationality}(Mike, China)$$

对决断集模板进行二分划分$(\mathcal{J}_1, \mathcal{J}_2)$，使得$\mathcal{J}_1 \cap \mathcal{J}_2 = \emptyset$且$\mathcal{J}_1 \cup \mathcal{J}_2 = \mathcal{J}_p$。

1) J_1= $\{r_{BornInCity}(A, C), r_{CityInProvince}(C, D)\}$, J_2=$\{r_{ProvinceInCountry}(D, B)\}$

由于不存在一个差异化实例替换θ使得$\mathcal{J}_1\theta \subseteq \mathcal{O}$，因此划分1)无法生成合理解释。

2) J_1 $\{r_{BornInCity}(A, C)\}$, J_2=$\{r_{CityInProvince}(C, D), r_{ProvinceInCountry}(D, B)\}$

根据划分2)，可以得到一个θ使得$\mathcal{J}_1\theta \subseteq \mathcal{O}$，其中$\theta = \{A \mapsto Mike, C \mapsto Hangzhou, B \mapsto China\}$，所以我们设$\mathcal{E}_2 = \mathcal{J}_2\theta = \{r_{CityInProvince}(Hangzhou, e_1), r_{ProvinceInCountry}(e_1, China)\}$，又因为$\mathcal{O} \cup \mathcal{J}_2\theta$是一致的，$\mathcal{O} \cup \mathcal{J}_2\theta \models \alpha$，且对于所有的$\mathcal{E}' \in subs_1(\mathcal{E}_2) \cup lifts_1(\mathcal{E}_2)$都有$\mathcal{O} \cup \mathcal{E}' \not\models \alpha$，所以$\mathcal{E}_2$是一个合理的解释。

3) J_1= $\{r_{BornInCity}(A, C), r_{ProvinceInCountry}(D, B)\}$, J_2=$\{r_{CityInProvince}(C, D)\}$

根据划分3)，可以得到一个θ使得$\mathcal{J}_1\theta \subseteq \mathcal{O}$，其中$\theta = \{A \mapsto Mike, C \mapsto Hangzhou, B \mapsto China, D \mapsto Zhejiang\}$，所以我们设$\mathcal{E}_3 = \mathcal{J}_2\theta = \{r_{CityInProvince}(Hangzhou, Zhejiang)\}$，又因为$\mathcal{O} \cup \mathcal{J}_2\theta$是一致的，$\mathcal{O} \cup \mathcal{J}_2\theta \models \alpha$，且对于所有的$\mathcal{E}' \in subs_1(\mathcal{E}_2) \cup lifts_1(\mathcal{E}_3)$都有$\mathcal{O} \cup \mathcal{E}' \not\models \alpha$，所以$\mathcal{E}_3$是一个合理的解释。

4) J_1=$\{r_{ProvinceInCountry}(D, B)\}$, J_2= $\{r_{BornInCity}(A, C), r_{CityInProvince}(C, D)\}$

根据划分4)，可以得到一个θ使得$\mathcal{J}_1\theta \subseteq \mathcal{O}$，其中$\theta = \{A \mapsto Mike, B \mapsto China, D \mapsto Zhejiang\}$，所以我们设$\mathcal{E}_4 = \mathcal{J}_2\theta = \{r_{BornInCity}(Mike, e_1), r_{CityInProvince}(e_1, Zhejiang)\}$，又因为$\mathcal{O} \cup \mathcal{J}_2\theta$是一致的，$\mathcal{O} \cup \mathcal{J}_2\theta \models \alpha$，且对于所有的$\mathcal{E}' \in subs_1(\mathcal{E}_4) \cup lifts_1(\mathcal{E}_4)$都有$\mathcal{O} \cup \mathcal{E}' \not\models \alpha$，所以$\mathcal{E}_4$是一个合理的解释。

5) J_1=$\{r_{CityInProvince}(C, D), r_{ProvinceInCountry}(D, B)\}$, J_2 $\{r_{BornInCity}(A, C)\}$

根据划分5)，可以得到一个θ使得$\mathcal{J}_1\theta \subseteq \mathcal{O}$，其中$\theta = \{A \mapsto Mike, C \mapsto Beihai, B \mapsto China, D \mapsto Guangxi\}$，所以我们设$\mathcal{E}_5 = \mathcal{J}_2\theta = \{r_{BornInCity}(Mike, Beihai)\}$，又因为$\mathcal{O} \cup \mathcal{J}_2\theta$是一致的，$\mathcal{O} \cup \mathcal{J}_2\theta \models \alpha$，且对于所有的$\mathcal{E}' \in subs_1(\mathcal{E}_5) \cup lifts_1(\mathcal{E}_5)$都有$\mathcal{O} \cup \mathcal{E}' \not\models \alpha$，所以$\mathcal{E}_5$是一个合理的解释。。

6) J_2=$\{r_{CityInProvince}(C,D)\}$, J_1= $\{r_{BornInCity}(A,C), r_{ProvinceInCountry}(D,B)\}$

根据划分6)，可以得到一个θ使得$\mathcal{J}_1\theta \subseteq \mathcal{O}$，其中$\theta = \{A \mapsto Mike, C \mapsto Beihai, B \mapsto China, D \mapsto Guangxi\}$，所以我们设

$$\mathcal{E}_6 = \mathcal{J}_2\theta = \{r_{BornInCity}(Mike, Beihai), r_{ProvinceInCountry}(Guangxi, China)\}$$

又$\mathcal{O} \cup \mathcal{J}_2\theta$是一致的，$\mathcal{O} \cup \mathcal{J}_2\theta \models \alpha$。但是因为存在一个$\mathcal{E}' \in subs_1(\mathcal{E}_6) \cup lifts_1(\mathcal{E}_6)$使得$\mathcal{O} \cup \mathcal{E}' \models \alpha$，所以$\mathcal{E}_6$不是一个合理的解释。

综上所述，可以得到可接受的解释为$\{\mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4, \mathcal{E}_5, \mathcal{E}_6\}$，经验证，其中合理的解释为$\{\mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4, \mathcal{E}_5\}$。

3.2 RDFS构建框架

第 四 章　实验与分析

在上两章的理论中，我们提出了基于决断集模板进行溯因诊断的方法，因此在本章中我们进行了两组实验来对上述理论进行验证。在第一组实验中，我们主要观察基于决断集模板的解释的基数生成解释的数量的影响以及对系统运算时间的影响。在第二组实验中我们重点验证生成的解释对知识库模型的诊断修复效果。

4.1　实验工具与环境

在本文中，我们实现了算法1中的算法逻辑，并把程序命名成Pattern Based Abduction(PBA)。PBA是一个用Java写成的程序，其中我们使用了本体语言编辑库OWL API来进行本体的生成、编辑与持久化存储，本体语言的推理机我们使用的是Pellet API，我们利用Pellet API完成的主要任务包括蕴含检测，蕴含寻找以及一致性检测。我们使用的数据库系统是Mysql，利用Mysql我们主要完成以下几个任务：第一是完成数据的存储。这里的数据主要包括本体中的ABox，训练集的数据和测试数据，实验完成后生成的解释集合也会存储在相同数据库的不同表中。使用Mysql完成的第二个主要任务是寻找差异化实例替换θ。通过把决断集模板编译成sql语句后，我们可以根据数据库中的ABox计算出相对应的差异化实例替换θ。本文实验中用到的实验数据如下表：

表 4.1: 实验工具

工具	描述
开发语言	Java
JDK	7u121-2.6.8-1ubuntu0.14.04.3
数据库	Mysql(vesion 5.7)
本体编辑库	OWL API(vesion 3.5.1)
本体推理机	Pellet API(version 2.3.1)

本次实验的数据，我们使用了两组数据，它们都来自FreeBase，分别是FB15K和FB40K。因为我们的问题输入还有一项是观察值，因此这些数据我们没有直接使用。我们从FB15K的有效三元组中抽取了1295组作为我们求问题的输入，也就是观察值，同时也会从训练集中把这1295组三元组移除，以避免影响实验结果。它们的具体数据如下表[ptranse]：

表 4.2: FreeBase数据集相关数据

Dataset	#Rel	#Ent	#Train	#Valid	#Observation
FB15K	1,345	14,951	483,142	50,000	59,071
FB40K	1,336	39,528	370,648	67,946	96,678

我们所有的实验运行在同一台机器上，机器的配置如下：

表 4.3: 实验环境

环境	描述
CPU	1.4 GHz Intel Core i5
RAM	4GB
OS	OS X Yosemite 10.10.3
Java Heap Space	4GB

4.2 程序框架

系统PBA采取了分模块设计，所有的模块如图4.1所示，其中数据解析模块负责对输入的数据集进行解析，并结合本体编辑推理模块完成对数据集的语义表达。本体编辑推理模块除了完成对数据集语义表达的任务以外，还需要完成本体编辑存储以及推理任务，其中推理模块需要在解释生成的过程中对生成的本体和解释进行一致性检测，并验证临时本体是否能够蕴含蕴含值(或观察值)。决断集模板编译模块主要完成寻找差异化实例替换θ的任务。寻找差异化实例替换θ的任务可以在内存中通过搜索的方式找到，并且计算速度较快，但是因为计算机资源有限，而数据的增大有可能导致计算任务的无法完成，所以我们会把在内存中完

成差异化实例替换θ的这种方式称为不可扩展的。在PBA系统中的决断集模板编译模块，我们采用把决断集模板编译成结构化查询语言来查找差异化实例替换θ的方法。利用数据库的持久化存储特性，使我们的系统达到可扩展的要求，虽然计算速度相比在内存中搜索的方法要慢，但是这种方法降低了数据集大小的要求，使我们可以完成使用所有收集到的数据进行测试。最后一个模块是解释生成模块，解释生成模块实现了上文中列出的解释生成算法，根据决断集模板，观察值和训练集生成合理解释集。

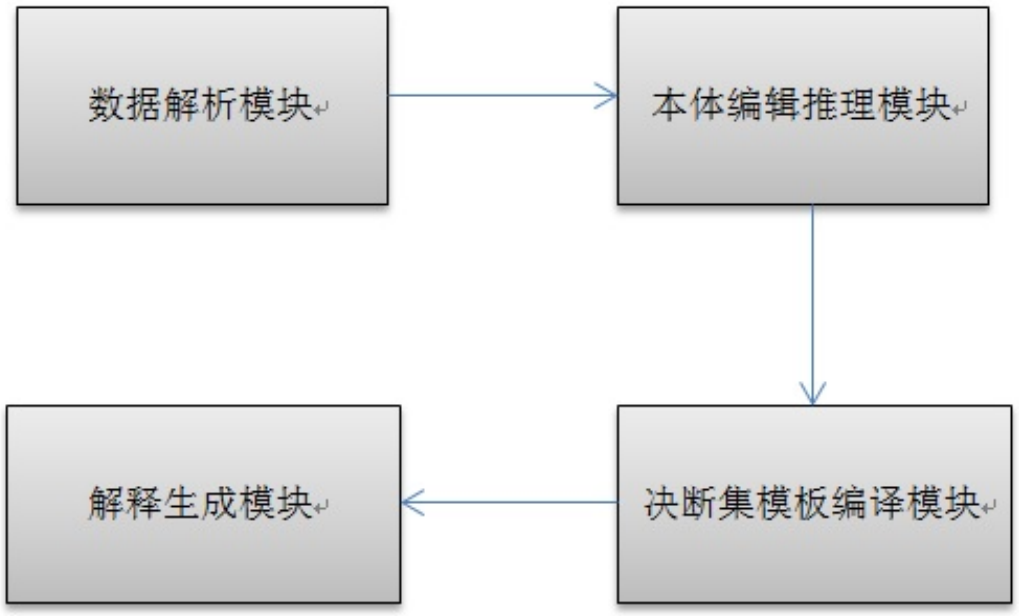

图 4.1: PBA系统模块

4.3 解释基数对生成解释的效率以及数量对比实验

解释的生成是本文中的重要工作，因此生成的解释基数与生成解释的数量以及生成解释所需要的时间是评价解释生成工作的重要参考指标。

4.3.1 实验设计

本次实验使用的数据集依旧是来自FreeBase的FB15K。解释基数对生成解释的效率以及数量对比实验对决断集模板的生成并不需要保证决断集模板的语义，

因此我们根据训练集的实体字典集合训练集的二元关系字典集随机生成了5,000个观察值，根据生成的观察值我们利用脚本配合本体推理机Pellet生成了10,000个决断集模板。在实验的时候，我们按照解释基数的限制设置了十组实验，每组实验分别对应基数为1 10的解释。对于每组实验，我们记录PBA生成的解释的数量以及生成的解释所需要的时间，最后根据实验结果得出我们的结论。

4.3.2 实验结果与数据分析

左图的数据是我们进行解释基数对生成解释的效率以及数量对比实验的实验结果，从图中可以看到，随着基数大小限制的增加，基于决断集模板生成的解释的个数也有所增加，但是基于决断集模板生成的解释个数的增长速率却相对平缓，得到了有效的控制，其中的原因包括：1)解释变量的允许。由于解释的生成是从决断集模板实例化而来，由$\mathcal{O} \cup \alpha$生成的差异化实例替换θ不能保证结解释中所有的个体实例被覆盖，因此解释变量的允许是必须的也是有效的手段在保证语义的同时又限制解释的数量。2)由于我们的解释需要满足$\subseteq_{ds}$-*minimal*的限制，因此最终生成的解释的集合中可被替换生成的解释都被去掉，达到了减小解释数量的目的的同时又保证了所有的解释能够被表达。

在右图我们可以看到在解释基数的变化下，生成每个解释的平均时间的变化情况。可以看到，随着解释基数的增加，生成每个合理解释的平均时间也开始逐渐增加。从图中的数据可以看到，基于决断集模版的解释在基数为1的情况下，对于所有的决断集模板和观察值，解释都可以在较短的时间内完成计算。一个原因是随着基数的增加，每次生成解释的时候需要进行检测一致性的解释数量有所增加，而检测$\subseteq_{ds}$-*minimal*特性的耗时也会相应增加，因此生成的解释平均时间就随着解释基数的上升而上升。

4.4 决断集模板解释修复知识库模型对比实验

本实验主要考虑的问题是基于决断集模板诊断求解得出的解释是否能够对知识库模型有一个较好的修复作用。知识库的训练集具有强的语义表达能力，如果

对于一个观察值知识库模型无法正确地对它进行表达，我们尝试利用决断集模版对它进行修复，修复的方式就是利用求解得到的基于决断集模版的解释对训练进行补充，对比诊断前后知识库模型对观察值的表达能力，我们可以验证解释集对知识库模型的修复能力。同时，由于基数的不同会对解释的修复能力产生影响，所以我们也对不同基数的解释的修复能力进行了比较。

4.4.1 实验设计

我们对诊断结果的评价方法由[3]中的方法调整而来。在利用现有的构造知识库模型的方法构造出向量模型之后，我们使用观察值集作为测试集来对模型进行测试。测试的时候我们对于测试集中的三元组(称为“有效三元组”)进行拆散操作，例如对于三元组(h,r,t)，我们首先移除三元组中的实体h，然后使用实体集合中的实体构造新的三元组，这些三元组我们称为“损坏三元组”，然后我们在把有效三元组和损坏三元组放在一起进行计算，分别计算这些三元组的能量函数，最后根据能量得分对它们进行由高到低排列。我们关注的指标有两个，一个是$MeanRank$，表示测试集中有效三元组的平均排名，平均排名越高代表模型表达该三元组的能力越强。另一个指标是$hits@10$，它表示测试集中有效三元组进入排名前百分之10的比率，同$MeanRank$，$hits@10$的比率也是越高越好。测试过程会进行两次，一次是诊断前，第二次是诊断后，对比诊断前后的模型性能，我们就可以验证诊断结果对知识库模型的修复效果。

我们按照基数对实验中计算出的解释进行划分，直观上地，在其它条件不变的情况下，基数越低的解释意味着训练集的数据中与决断集模板的重合度越高，所以可以非常合理的推测基数越低的解释在语义上觉有更高的可信度。解释的生成条件虽然已经在定义中进行了限制，但是解释的数量仍然较大，因此我们对基数较低的解释赋予了更高的计算优先级，这样我们可以优先计算出语义可信度较高的解释。我们从有效三元组中抽取了500个作为观察值，根据观察值我们又利用脚本对FB15K的数据进行处理，根据实体关系的限制筛选出二元关系链的断言集合，然后我们请了语义网领域内的专家对数据进行复查，人工确认了其中正确的

决断集。我们对处理后的数据进行转化，得到了决断集模版。我们用收集到的训练集、决断集模版以及观察值作为PBA系统的输入进行了决断集模板解释修复知识库模型对比实验。

4.4.2 实验结果与数据分析

表 4.4: 实验环境

Metric	$No\ Exp$				$Exp.C \leq 1$				$Exp.C \leq 2$			
	$MeanRank$		$Hit@10$		$MeanRank$		$Hit@10$		$MeanRank$		$Hit@10$	
	Raw	Filter	Raw	Filter	Raw	Filter	Raw	Filter	Raw	Filter	Raw	Filter
TransE	C2d	c3	C4d	C4a	C4a	C4a	C4a	C4a	C4a	C4a	C4a	C4a
TransH	C2d	c3	C4d	C4a	C4a	C4a	C4a	C4a	C4a	C4a	C4a	C4a
PTransE(add)	C2d	c3	C4d	C4a	C4a	C4a	C4a	C4a	C4a	C4a	C4a	C4a
PTransE(mul)	C2d	c3	C4d	C4a	C4a	C4a	C4a	C4a	C4a	C4a	C4a	C4a
PTransE(RNN)	C2d	c3	C4d	C4a	C4a	C4a	C4a	C4a	C4a	C4a	C4a	C4a
TransR	C2d	c3	C4d	C4a	C4a	C4a	C4a	C4a	C4a	C4a	C4a	C4a
CTransR	C2d	c3	C4d	C4a	C4a	C4a	C4a	C4a	C4a	C4a	C4a	C4a

在本次实验中我们用来验证诊断效果的知识库模型包括以下：TransE(Bordes et al., 2013)，TransH (Wang et al., 2014)，TransR (Lin et al., 2015)以及PTransE (Lin et al., 2015)。模型的实现代码来自`https://github.com/mrlyk423/relation_extraction`，由Lin等人实现。在表4.3中记录了我们的实验数据，在表的左边列出了我们使用的知识库模型。在表的上方$No\ Exp$，$Exp.C \leq 1$以及$Exp.C \leq 2$表示我们对每一种知识库模型进行了三次实验，第一次是没有在没有对数据集进行诊断操作之前运行程序的结果。在第二次实验中，我们把数据集，本体以及观察值作为PBA系统的输入进行诊断并求解。我们的系统PBA可以根据给出的决断集模板以及数据集计算出所有解释，但是我们可以合理地假设基数越小的解释会有更易于理解，同时也会更加地直观，因此在后两次的实验中我们调高了基数低的解释的计算优先级。

由于基数较大的解释不仅计算较为费时，而且这些解释不易于理解，容易丧

失语义上的可读性，因此在实验中我们进行对比的解释基数分别为1和2。

第 五 章　总结与展望

5.1　工作总结

现今阶段，语义网的高速发展以及它的潜在价值获得了越来越多人的关注，作为语义网的重要组成部分，本体的构建已经成了许多研究的重点。在国外，本体构建的研究已经取得了不少的进展，例如Delia Rusu 等人提出的基于Treebank Prser的Triplet提取算法可以比较准确的从语句中提取术语之间的关系了。遗憾的是，中英文之间的差别影响了这些学术成果的可复用性，因此，国内的相关研究人员开始对本体的自动和半自动构建进行研究，但是目前进展缓慢。基于现有的研究成果，本文从中文语料库中进行领域术语识别、分类关系和非分类关系提取，实现了从中文语料库中初步构建出以领域本体的RDFS模型。同时我们还分别比较了利用统计学方法和语义分析法对领域术语进行识别，经过试验寻找出不同方法的使用场景。在进行关系抽取的时候，我们利用代词和零代词的消解提取出隐藏的关系，提高了抽取的准确率和覆盖率。

5.2　不足与改进方向

在本文的本体构建中，由于中文短语组合的复杂性，尚不能对以短语形式出现的领域本体术语进行比较准确的提取。术语的缺少会导致关系抽取数量的减少，影响了构建出的本体的全面性。在未来的研究中，主要由两个方面需要进行改进。一个是对上下位关系的识别，目前的方法依赖于中文wordnet等语义词典的帮助，而目前中文语义词典的覆盖尚不全面，遗漏关系的现象比较严重。同时，在本次实验的基础上，通过对语言模型的应用抽取出本体中规则的推倒，将使构建出的本体的可用性提高到一个新的层次。

致　谢

不知不觉就走完了大学本科四年的历程。在四年的大学生活中，我遇到了很多优秀的老师，他们在传授知识上尽心尽力，学士渊博的他们也是我学习的榜样，他们治学严谨，对知识渴求，对学生负责，让我走进知识的殿堂，对此我要对他们表示衷心的感谢。要特别感谢万海老师的悉心指导，从大一开始上万老师的课，万老师对我的严格要求为今后的学业道路打下坚实的基础。在保研之后，老师就开始对我进行了科研方面的相关指导，带领我提早开始进行研究生的工作，多次在我存有疑惑的时候耐心的替我解决问题，并且总能够在最及时的时候给予回复，在此要特别感谢。

要感谢我的舍友们，在撰写本文时多次进入熬夜状态，晚上的灯光和键盘声多少对他们的休息产生了影响，感谢他们给予的理解。

还要感谢我的爸爸妈妈，多年的养育之恩终于让我长大成人，他们在我低谷的时候给予我关怀与帮助，在我开心的时候一起分享快乐，这份来自家庭的感动一直是我前进的最大动力。

最后要感谢和我一起度过这四年的朋友和同学，因为你们让我这四年更加的珍贵。

参考文献

附　录

中英文词汇对照表

Resource Description Framework(RDF)	资源描述框架
Resource Description Framework Schema(RDFS)	资源描述框架模式
DARPA Agent Markup Language(DAML)	DARPA代理标记语言
Ontology Inferrence Layer(OIL)	本体建模语言
Web Ontology Language(OWL)	网络本体语言
Extensible Markup Language(XML)	可扩展标记语言

原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。