基于 FrameNet 框架关系的文本蕴含识别*

张 鹏[1]，李国臣[1]，李 茹[1,2]，刘海静[1]，石向荣[3]

[1]山西大学 计算机与信息技术学院，山西 太原 030006

[2]山西大学 计算智能与中文信息处理教育部重点实验室，山西 太原 030006

[3]中北大学 电子与计算机科学技术学院，山西 太原 030051

E-mail: zhangpeng0402@yahoo.cn

摘 要：文本蕴含识别是处理自然语言中广泛存在的同义异形现象的一种有效途径。本文基于 FrameNet 中框架及框架之间的八种关系，结合 WordNet 中词汇间的语义关系，提出了一种文本蕴含识别方法。在给定文本 T 和假设 H 中词元激起的框架基础上，该方法利用深度优先搜索，在 FrameNet 框架关系图中，查询 T 和 H 中框架之间的上下位关系；再使用 WordNet 中语义关系比较二者的框架元素是否一致或相似。实验对 RTE2007 中 50 个文本对进行了测试，达到了 73.07%的准确率，接近于 RTE2007 评测的最优结果。

关键词：文本蕴含识别；FrameNet；框架关系

Recognize Text Entailment Based on FrameNet Relations

Zhang Peng[1], Li Guochen[1], Li Ru[1,2], Liu Haijing[1], Shi Xiangrong[3]

[1] School of Computer & Information Technology, Shanxi University, Taiyuan 030006

[2] Key Laboratory of Ministry of Education for Computation Intelligence & Chinese Information Processing, Taiyuan 030006

[3] School of Electronics and Computer Science & Technology, North University of China, Taiyuan 030051

E-mail: zhangpeng0402@yahoo.cn

Abstract: This paper proposes a text entailment identification method using the frames and the frames relations in FrameNet that integrates the relevant knowledge with WordNet. The method finds the path between the frames evoked by the lexical units in text T and hypothesis H in the FrameNet Graph using depth-first Search Method in order to identify the hyponymy relationships between the frames; and then realizes the text entailment recognization through comparing the content of span which are filled the mapping FE slots. Our experiment are based on the part of the evaluation corpus of RET2007.The experimental results acquired 73.07% precision show that the precision of our method is consistent with the best results of RET2007 evaluation results in the task of Recognize Text Entailment.

Keywords: recognize text entailment; frameNet; frame-to-frame relations

1 引言

为了有效地处理自然语言中广泛存在的同义异形现象，近年来国外一些学者尝试使用文本蕴含（Text Entailment）[1]来为语言中纷繁复杂的同义表达建立模型。文本蕴含可以定义为：一个连贯的文本（Text）T 和一个被看作假设（Hypothesis）H 之间的一种语义包含关系。如果 H 的意义可以从文本 T 的意义中推断出来，那么就说 T 蕴含 H（即 H 是 T 的推断）。文本蕴含的研究对于自然语言处理中不同应用所需的语言表达多样性的推理识别有着重要意义。比如在多文本自动文摘中，从文本中省去的冗余句子或表达应该被摘要中的其他表达所蕴含；对于信息抽取，表达相同关系的不同文本之间也存在着蕴含关系。

识别文本蕴含（Recognizing Textual Entailment, RTE）是美国国家标准技术研究所（National Institute of Standards and Technology, NIST）举办的文本分析会议（Text Analysis Conference，TAC）

* 基金项目：国家自然科学基金（60970053）；山西省国际科技合作项目（2010081044）；山西省高校拔尖创新基金；山西省实验室开放基金（2009011059-4）；国家 863 计划项目（No.2006AA01Z142）。

中的一项评测，该评测已经举行了6年，构造了一定的文本蕴含推理模型和识别模型。Peter Clark and Phil Harrison[2]使用 WordNet 和 DIRT 推理规则库开发了一个基于推理的文本蕴含识别系统 BLUE。Debarghya Majumdar 和 Pushpak Bhattacharyya[3]通过分析文本 T 和假设 H 之间的词汇重叠来发现它们之间的蕴含关系。Alexander Volokh、Giinter Neumann 和 Bogdan Sacaleanu[4]提出了一种联合确定性依存句法分析和线性分类的鲁棒性文本蕴含识别方法。2009 年有 21 所科研院所参加 RTE-5 评测任务，其任务分为两类：3-ways 和 2-ways。在 3-ways 任务中最高准确率为 68.33%，平均准确率为 52.91%，在 2-ways 任务中最高准确率达到 73.5%，平均准确率为 61.52%。

本文采用 FrameNet 的框架及其关系识别文本 T 和假设 H 所表达的语义场景之间的关系，结合 WordNet 的相关知识达到识别文本蕴含的目的。论文首先对 FrameNet 和在其上的一些研究做了简单的介绍，接着描述了本文采用 FrameNet 框架及其关系识别文本蕴含的方法，最后对实验及结果进行了分析，并对全文工作进行了总结和展望。

2 FrameNet 及其相关研究

FrameNet[5]（FN）是美国加州大学伯克利分校构建的一个基于框架语义学[6]（Frame Semantics）的词汇资源。框架语义学是 Fillmore 提出的研究词语意义和句法结构意义的一种理论方法，即试图以真实语料为基础，以经验主义方法，寻找语言和人类经验之间的紧密关系，并研究一种可行的描述方式来表示这种关系。

在 FrameNet 中框架（Frame）是用来描述一个事件或一个语义场景的一组概念。每个框架都包含了一系列被称为框架元素（frame elements, FEs）的语义角色，这些框架元素与描述事件或形态的词汇相对应。两个框架之间的语义关系用框架关系（Frame-to-frame Relations）来描述，不同框架的框架元素也依据框架关系相互映射在一起（FE-to-FE Mappings）。在 FrameNet 数据中共定义了八种框架关系，框架关系是两个框架之间的一种定向（非对称）关系。

近年来，FrameNet 受到国内外很多学者的关注，并基于 FrameNet 展开了一系列的研究。Jan Scheffczyk 和 Collin F.Baker[7]尝试使用 FrameNet 这一语义丰富的词汇资源结合领域本体进行推理。Ekaterina Ovchinnikova[8]等人提出了一种数据驱动和本体分析的方法来丰富和公理化 FrameNet 的框架关系使 FrameNet 能更加广泛地应用到自然语言处理中。在文本蕴含中，Aljoscha Burchardt 和 Anette Frank[9]提出一种利用 LFG 语法分析器结合 FrameNet 框架语义来识别文本蕴含的方法；Himanshu Shivhare、Parul 和 Anusha Jain[10]提出了一种使用 FrameNet 对文本 T 和假设 H 进行语义聚类的方法识别文本之间的蕴含关系。

3 文本蕴含识别

文本蕴含识别的内容是识别 H 的意义是否可以从 T 的意义中推断出来，本文使用两个蕴含模型，“框架蕴含识别”模型和“框架元素识别”模型，进行文本蕴含识别，分别用来实现对框架和框架元素之间的蕴含关系进行识别。模型如图 1 所示。

3.1 框架蕴含识别

框架蕴含识别旨在识别 T 和 H 所表述的语境是否相同，即比较词元激起的框架，两个框架之间如果存在蕴含关系则必须满足这样的条件：T 和 H 中由词元激起的框架相同或两者之间存在上下位关系。

把 FrameNet 中的框架看作是节点，以连接两个框架之间的上下位语义关系为有向边，得到 FrameNet 框架关系图 G=(V, E)，如图 2 所示。

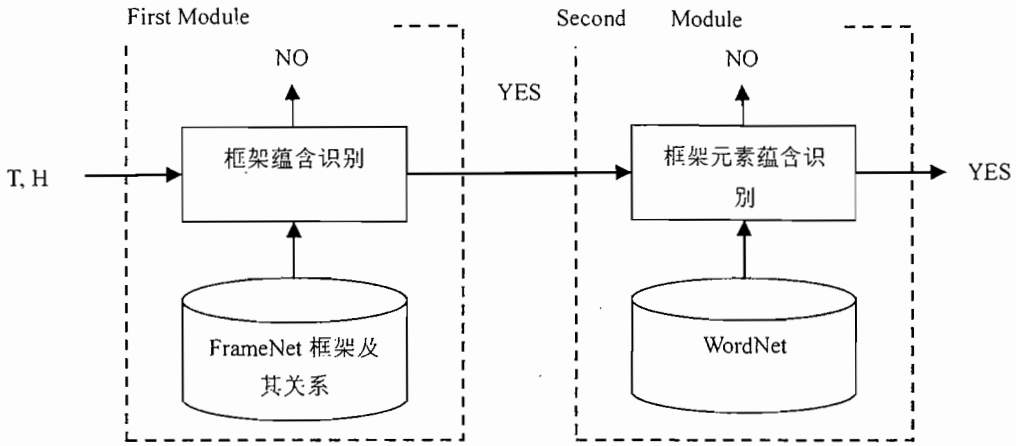

图1 文本蕴含识别模型

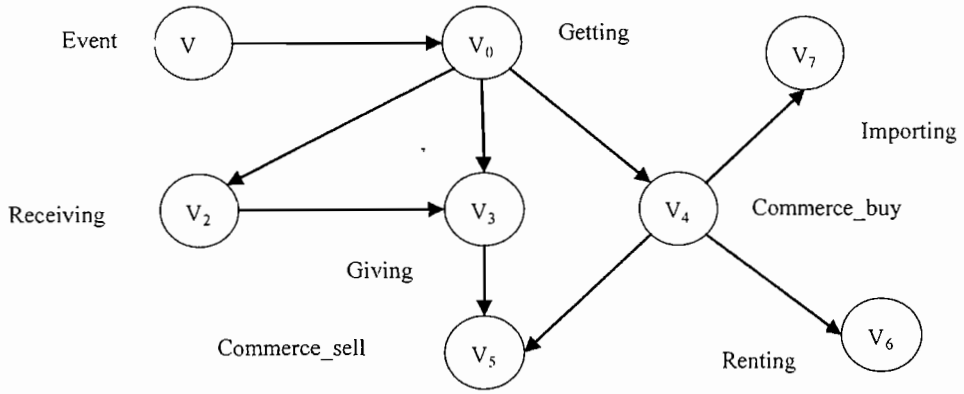

图2 部分框架关系图

识别框架之间的蕴含关系按照以下几步进行：

Step1：初始化，VT 是以 T 中的框架为起始节点 v_0 遍历查找到的所有框架节点的集合，设 VT={ v0}，E=Φ；

Step2：从 VT 中的节点 v 为出发点，利用深度优先算法搜索 FrameNet 框架关系图，对遍历到的每个节点 $v' \in V$ 且 $v' \notin VT$ 进行标记，并添加到 VT 中，直到找到 H 中的框架节点为止，考虑到算法的执行效率，搜索允许的最大路径为 5。

3.2 框架元素蕴含识别

框架之间的蕴含识别只能够识别 T 和 H 所描述的语义场景之间的关系，识别文本蕴含还需要对填充相应框架元素的语块进行比较，具体步骤如下：

Step1：提取两个框架中依据框架关系相互映射的 FE 内容；

Step2：对 step1 中提取的 FE 进行比较，通过词汇重叠判断内容是否一致或相似；

Step3：对 step2 中不一致的内容，利用 WordNet 中的语义关系进行识别；

Step4：正确识别，重复 step1~step3，比较下一对 FE，直到 FE 比较完或内容不同为止。

3.3 实例分析

例 1 是 2007 年 RTE-3 评测中的一个（T, H）文本对，其中加粗并带有下划线的单词就是激起

框架的词元。图 3 是对例 1 进行蕴含识别的图形示例说明。

例 1 <id="46" entailment="YES"><t>British Airways Ltd … rapidly acquired Hillman's Airways, adopted …Gatwick Airport.</t> <h>Hillman's Airways was sold to British Airways.</h>

如图 3 所示，T 中词元 acquired 激起 Getting 框架，语块 British Aireways 和 Hillman's Aireways 分别填充框架元素 Recipient 和 Theme。在 H 中词元 sold 激起 Commerce_sell 框架，语块 British Aireways 和 Hillman's Aireways 分别填充框架元素 Goods 和 Buyer。在 FrameNet 的框架关系图中按图搜索，可得到从 Getting 到 Giving，再到 Commerce_sell 的一条路径。框架 Getting 与 Giving 之间存在“Perspective_on”关系，框架 Commerce_sell 又继承于 Giving，根据这种上下位关系的传递，可认为框架 Getting 与 Commerce_sell 之间有蕴含关系。然后根据框架元素之间的映射关系，比较填充框架元素 Recipient 与 Buyer、Theme 与 Goods 的语块内容，判定 T 蕴含 H。

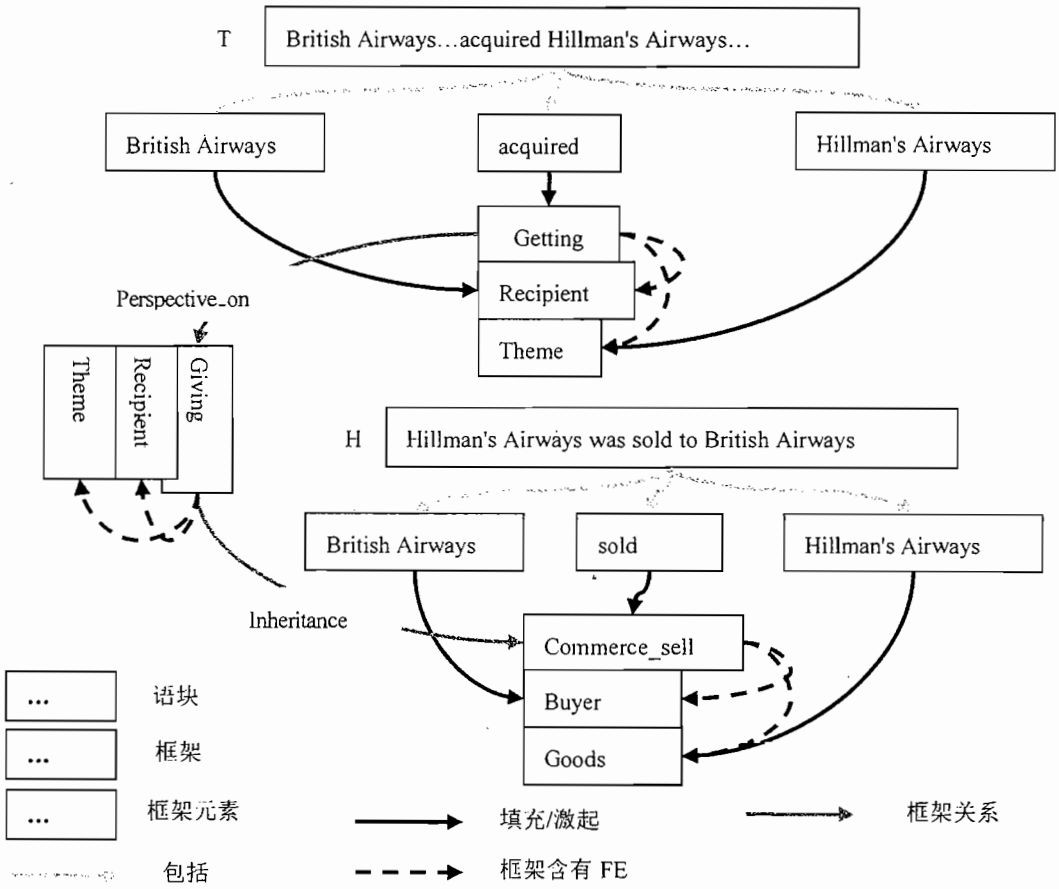

图 3　基于 FrameNet 框架及其关系识别文本蕴含示例图

4　实验及结果分析

实验选取的语料是 2007 年 RTE-3 评测语料中的前 50 个<T, H>文本对，用 RTE 评测任务的评测标准进行评测，其结果如表 1 所示。表 2 是对识别的各种语料的分布进行说明。

文本蕴含识别正确分为两种情况，一种是识别出文本之间有蕴含关系，称为正确肯定，如例 2 所示；另一中是识别出文本之间没有蕴含关系，称之为正确否定，如例 3 所示。

表1 实验结果

Precision:	73.07%
Recall:	38%
F-score:	49.98%

表2 实验明细

RTE2007	All	Recognize	Recognize="TRUE"
All	50	26	19
Entailment="YES"	26	8	8
Entailment="No"	24	18	11

例2 <id="25" entailment="YES"><t>Born in Kingston-upon-Thames, Surrey, Brockwell played … the 19th century.</t><h>Brockwell was born in Surrey.</h>

例3 <id="34" entailment="NO"><t>A Revenue Cutter, the ship was named for Harriet Lane,… Buchanan's White House hostess.</t><h>Harriet Lane owned a Revenue Cutter.</h>

通过对识别错误的 <T, H>文本对进行分析，发现 FrameNet 本身的一些不足对实验结果有较大影响，主要是两方面因素，一是词元覆盖率不高，另一个是框架关系的缺失。

例4<id="11" entailment="YES"><t>… Super Nintendo Entertainment System release of the game as Final Fantasy III </t> <h>Final Fantasy III is produced by …Entertainment System.</h>

例5< id="40" entailment="YES"><t>Robinson's garden style can be seen today at Gravetye Manor, West Sussex, …Robinson's time.</t> <h>Gravetye Manor is located in West Sussex.</h>

例 4 是由框架缺失造成识别错误的例子，T 中谓词 release 激起框架“Releasing”，该框架在 FrameNet 中的解释是“释放”，与 T 中的词汇含义不同，T 中 release 的释义是“发布”，有一定“生成”的含义，但 FrameNet 中 release 并不能激起“Creating”框架或相关框架，所以无法识别文本 T 和 R 之间的蕴含关系。

识别错误的<T, H>文本对中有 42%是由于框架关系的缺失导致的，如例 5 所示。T 中谓词 seen 激起“Perception_experience”框架，H 中谓词 located 激起“Being_located”框架，在 FrameNet 框架关系图中找不到从“Perception_experience”到“Being_located”的路径，因此判定文本 T 和 H 之间是非蕴含关系。而这与现实语境是有出入的，比如“我在上海看见了东方之珠”这句话中就蕴含了“东方之珠坐落在上海”的意思。

5 总结

文本蕴含对于自然语言处理中不同应用所需的语言表达多样性的推理研究有着重要意义。本文使用了 FrameNet 和 WordNet 中的语义关系，提出了一种文本蕴含识别方法，并用该方法对 RTE2007 语料中前 50 个文本对进行了测试，达到了 73.07%的准确率，这表明，FrameNet 框架及其关系对于文本蕴含识别任务是有帮助的。相比于基于规则的或者基于词汇概率的文本蕴含识别方法，本文提出的基于语义词典中语义关系的文本蕴含识别方法更加逼近人类理解蕴含关系的心智过程，并进一步提高蕴含识别的准确率。

本文提出的文本蕴含识别方法也存在一些不足：（1）目前，该方法只针对文本和句子中由动词词元激起的框架进行蕴含识别，而实际上名词、形容词也能够激起框架，所以，本文下一步将扩大框架的研究范围；（2）FrameNet 中存在词元覆盖率不高，有些框架关系缺失的情况，这些都导致了本文的方法不能适用于某些语料，影响了实验结果的精度。接下来的工作中，我们将探索完善 FrameNet 中框架间关系，并采用机器学习的方法完成对整个 RTE2007 评测语料的实验。

参 考 文 献

[1] 袁毓林, 王明华. 文本蕴含的推理模型与识别模型. 中文信息学报. 2010. 3.

[2] Peter Clark, Phil Harrison. An Inference-Based Approach to Recognizing Entailment. In Proceedings of Text Analysis Conference (TAC), 2009.

[3] Debarghya Majumdar, Pushpak Bhattacharyya. Lexical Based Text Entailment System for Main Task of RTE6. In Proceedings of Text Analysis Conference (TAC), 2010.

[4] Alexander Volokh, Günter Neumann, Bogdan Sacaleanu. Combining Deterministic Dependency Parsing and Linear Classification for Robust RTE. In Proceedings of Text Analysis Conference (TAC), 2010.

[5] FrameNet. http://framenet.icsi.berkeley.edu.

[6] C. J. Fillmore. Frame semantics and the nature of language. Annals of the New York Academy of Sciences, 1976.

[7] J. Scheffczyk, C. F. Baker, S. Narayanan. Ontology-based reasoning about lexical resources. In Proc. of OntoLex 2006: Interfacing Ontologies and Lexical Resources for Semantic Web Technologies, Genoa, Italy, 2006.

[8] Ekaterian Ovchinnikova, Laure Vieu, Alessandro Oltranari. Data-Driven and Ontological Analysis of FrameNet for Natural Language Reasoning. 2009.

[9] Aljoscha Burchardt, Anette Frank. Approaching Textual Entailment with LFG and FrameNet Frames. In Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing 2006.

[10] Himanshu Shivhare, Parul, Anusha Jain. Himanshu Shivhare、Parul 和 Anusha Jain. In Proceedings of Text Analysis Conference (TAC), 2010.