《中文信息学报》稿件排版格式

文章编号：1003-0077（2011）00-0000-00

基于知识话题模型的文本蕴涵识别 *

任函 [1,4]，盛雅琦 [2,4]，冯文贺 [2,4]，刘茂福 [3,4]

（1.湖北工业大学计算机学院，湖北省武汉市 430068；2.武汉大学计算机学院，湖北省武汉市 430072；3. 武汉科技大学计算机科学与技术学院，湖北省武汉市 430065；4. 武汉大学湖北省语言与智能信息处理研究基地，湖北省武汉市 430072）

摘要：本文分析了现有基于分类策略的文本蕴涵识别方法的问题，并提出了一种基于知识话题模型的文本蕴涵分类识别方法。 其假设是：文本可看作是语义关系的组合，这些语义关系构成若干话题；若 T→H，说明 T 和 H 具有相似的话题分布，反之说明 T 和 H 不具有相似的话题分布。基于此，我们将 T 和 H 的蕴涵识别问题转化为相关话题的生成过程，同时将文本推理知识融入到抽样过程，由此建立一个面向文本蕴涵识别的话题模型。实验结果表明基于知识话题模型在一定程度上改进了文本蕴涵识别系统的性能。

关键词：文本蕴涵识别；话题模型；蕴涵分类；推理知识

中图分类号：TP391　　**文献标识码：**A

Recognizing Textual Entailment Based on Knowledge Topic Models

Han Ren[1,4]，Yaqi Sheng[2,4]，Wenhe Feng[2,4]，Maofu Liu[3,4]

(1.School of Computer Science, Hubei University of Technology, Wuhan, Hubei 430068, China; 2. School of Computer, Wuhan University, Wuhan, Hubei 430072, China; 3. College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, Hubei 430065, China; 4.Hubei Research Base of Language and Intelligence Information Processing, Wuhan University, Wuhan, Hubei 430072, China)

Abstract: This paper analyzes the problems of current entailment recognition approaches based on classification strategy and proposes an novel approach for recognizing textual entailment based on a knowledge topic model. The assumption in this approach is, if two texts have an entailment relation, they should share a same or similar topic distribution. Following the assumption, The approach builds an LDA model to estimate semantic similarities between each text and hypothesis, which will be the evidences for judging entailment relation. We also employ three knowledge bases to improve the precision of Gibbs sampling. Experiments show that knowledge topic model improves the performance of textual entailment recognition systems.

Key words: recognizing textual entailment ; topic model; entailment classification；Inference knowledge

1 引言

文本蕴涵识别（Recognizing Textual Entailment）是一个判断文本之间推理关系的任务，定义为：给定一个连贯文本 T（Text）和一个假设 H（Hypothesis），如果 H 的意义可以从 T 的意义中推断出来，那么就认为 T 蕴涵 H（即 H 是 T 的推断）[1]。作为一个挑战任务，文本蕴涵识别能够广泛应用于各类自然语言处理应用，如自动问答系统、多文档自动摘要、信息抽取、信息检索、机器翻译及自然语言理解(NLU)领域中的机器阅读等 [2,3] 。

在 RTE 挑战任务中，文本蕴涵识别问题可以看作是一个标准的二分类问题 [4]，即将需

* **收稿日期：**　　**定稿日期：**

基金项目：国家自然科学基金（61402341，61173062，61373108）；国家社会科学基金重大项目（11&ZD189）；中国博士后科学基金资助项目（2013M540594）

要识别的“文本—假设”对利用分类器进行分类，若假设与某一句子之间存在蕴涵关系，则归为“蕴涵”（Entailment）类；若不存在蕴涵关系或无法判断两者是否存在蕴涵关系，则归为“不蕴涵”（No Entailment）类。然而，蕴涵和非蕴涵两个类都比较庞杂，实例间的相似性难以保证，它们的区别性特征就不容易确定，所以据此建立的分类器性能不太理想。为改进分类器的性能，可以采取两种方法。第一种方法是利用语义相似度特征（如谓词论元结构）评估蕴涵关系，其假设为：语义结构重叠度越高，则 T 和 H 越有可能描述了相同的语义关系。然而这一假设仅考虑了局部语义关系是否一致，而缺乏总体语义关系的判断能力；并且现有语义分析方法仅能描述非常有限的语义关系（如谓词和相关论元的关系），使得蕴涵识别系统性能难以得到有效提升。第二种方法是引入合适的背景知识，利用这些知识描述难以从给定文本假设对中获得的推理关系，从而提高推理性能。然而，大多数分类方法的主要策略是相似度比较，难以有效利用外部知识识别可推理的关系。

本文提出一种基于知识话题模型的文本蕴涵识别模型。该模型的假设为：若两个文本片断具有蕴涵关系，则它们必然拥有相同或相似的话题。我们利用话题模型评估文本片断之间的相关程度，以此作为判断文本片断是否具有蕴涵关系的一个依据。我们利用文本假设对建立 LDA 模型，并利用多种方法评估文档的话题概率分布的相似性。针对因蕴涵知识缺乏而导致的话题生成错误的问题，我们引入背景知识以改进抽样精度，提高生成话题的性能，从而提高文本语义相关度的评估性能。实验表明，在分别采用话题模型和知识话题模型以后，系统的准确率逐渐提高，说明话题模型与背景知识结合，能够有效改进系统的性能。

本文第二部分简要介绍文本蕴涵相关工作，第三部分介绍知识话题模型，第四部分介绍了基于话题模型的文本蕴涵识别及基于知识话题模型的文本蕴涵识别的实验，第五部分对全文工作进行总结和展望。

2 相关工作

现有的识别文本蕴涵方法主要可以分为三大类：

一、基于分类策略的文本蕴涵识别。把识别文本蕴涵含当作一个分类问题，该策略将文本蕴涵形式化为一个两类（蕴涵和不蕴涵）或三类（蕴涵、矛盾和未知）的分类问题。根据已标注的训练实例，学习其中的特征并建立分类器。该方法主要抽取蕴涵对(T-H 对)的词汇特征，如词汇对齐特征、基于同义词林语义相似度特征、反义词特征等；句法结构特征，如依存图对齐特征、谓词-论元结构特征等；然后用类似支持向量机分类器进行分类。如基于 FrameNet 框架关系的文本蕴涵识 [5]，识别矛盾文本 [6]，用支持向量机和字符串相似识别蕴涵 [7]，基于事件语义特征的中文文本蕴含识别 [8]，基于知网的文本推理 [9]等。但就分类策略而言，其问题在于蕴涵和非蕴涵两个类都比较庞杂，实例间的相似性难以保证，它们的区别性特征就不容易确定，所以据此建立的分类器性能不太理想。

二、基于转换策略的文本蕴涵识别。是主要根据蕴涵规则，或者编辑距离来判断文本蕴涵。如 Kouylekov 和 Magnini[10]在任务 RTE-1 中应用编辑距离来识别文本蕴涵，Kouylekov [11]等人提出一个基于编辑距离识别文本蕴涵的开源框架 Edit Distance Textual Entailment Suite(EDITS)。然而这些编辑操作难以体现语义层面的转换，因此对于比较复杂的语义蕴涵关系难以准确识别。另一类方法利用自动抽取的推理规则来识别文本蕴涵，如复述规则 DIRT[12]，全局学习蕴涵规则 [13]，两层模型学习上下文相关推理 [14]等。然而，目前规则自动获取的性能还有待提高，其中一个重要的原因就是规则的歧义性。例如，对于推理模板 r:“X 打 Y”，以下两条规则都蕴含于它：

r1: X 玩 Y

r2: X 买 Y

当 X 为“我”，Y 为“球”时，r 与 r1 具有推理关系；而当 X 为“我”，Y 为“酱油”

时，r 与 r2 具有推理关系。显然，这两个模板所代表的意思完全不同。 出现该问题的原因是动作“打”具有多义性。该问题可以用以下两种方式解决：一是对蕴涵规则标注更多语义信息，然而目前缺乏相关资源和研究；二是在进行推理时选择合适的规则，然而做到这点并不容易，因为推理规则大多数是句法的转换，不带有语义信息，因此我们很难去判断究竟应该选择哪条规则才能使转换后的意义保持一致。

三、基于深度分析和语义推理策略的文本蕴含识别。这种方法主要是采用传统的逻辑推理、自然逻辑推理、本体推理或语义特征等获得推理知识。其难点在于大量知识往往难以有效获取，没有足够的知识，对于深度推理及分析来说是比较困难的。

3 知识话题模型

本文提出一种基于话题模型的文本蕴涵识别方法，其假设是：文本可看作是语义关系的组合，这些语义关系构成若干话题；若 T 能推理出 H，说明 *T* 和 *H* 具有相似的话题分布，反之说明 T 和 H 不具有相似的话题分布。基于此，我们将蕴涵识别问题转化为话题的生成过程，由此建立一个面向文本蕴涵识别的话题模型。

另一方面，简单话题模型无法处理词义、句法结构等知识，而文本蕴涵识别需要利用各种外部知识，如词义关系、推理规则等，因此，如何在模型中应用外部推理知识，就成为面向文本蕴涵识别的话题建模的一个重要问题。Rubin 等将标签频率作为先验知识应用于话题模型，以改进多文档摘要的性能，并指出先验知识可以提高简单话题模型的性能 [15]；Chen 提出一种在多领域上进行话题建模的 LTM 模型，通过在不同领域中抽取各自的话题，对不同领域的话题进行融合，获得多领域之间的关联词汇对，并在 Gibbs 抽样过程中引入先验知识，来修正抽样结果 [16]。受此启发，我们提出了一种基于知识话题模型的文本蕴涵识别方法，在 Gibbs 抽样过程中引入词义关系和推理规则，以改进抽样结果。

3.1 KLDA 模型

知识话题模型（Knowledge-Based Topic Model, KBTM）利用先验知识指导话题建模，为话题生成提供了一种有监督的学习方法，能够在一定程度上改善话题生成的性能 [17]。我们对标准 LDA 话题模型进行修改，称之为 KLDA 话题模型。其构建算法流程如图 1 所示：

Algorithm 1：KLDA

```
E←KnowledgeMining
for   i = 1 to N do:
    W←Compute the words in each topic from i-1 sampling;
     GibbsSampling(D,E,W,1);
end for
```

图 1 KLDA 模型的的构建算法

KLDA 模型算法流程描述如下：首先对外部知识库进行知识挖掘，产生算法中所需要的先验知识，即得到一个基于外部知识的词汇关联矩阵 E；其次，对于第$i(\in [1,N])$次 Gibbs 抽样（*N* 为 Gibbs 抽样的迭代次数），计算在第 i-1 次 Gibbs 抽样后话题中的所有单词，得到每个话题中的所有单词，再用之前计算得到的先验知识对 Gibbs 抽样进行修正。

3.2 知识挖掘

文本蕴涵需要用到的外部知识，主要有外部的词语同义词、词语上下位关系、推理关系等。我们使用 WordNet 词典和蕴涵规则集 DIRT 计算词汇相关度，并建立词汇关联矩阵。算法描述如图 2 所示。

Algorithm 2： knowledgeMining

1: **for** all words w_i, $i \in [1,V]$ **in** voc **do**:
2: **for** the current w_j **in** voc **do**:
3: compute the similarity between w_i and w_j;
4: update the similarity matrix $E_{i,j}$;
5: **end for**
6: **end for**

图 2 知识挖掘算法

知识挖掘算法的基本过程如下：对于词汇表中的每对单词，通过外部资源计算词汇的语义相似度，更新相似度矩阵 E。通过该算法，我们可以得到一个基于外部知识库的词义相似度矩阵。

相似度矩阵 E 的详细计算过程如下。首先给出两个单词的词语相关度的计算公式，定义如下：

$$E_{k,w,w'} = \begin{cases} 1 & w = w' \\ avg(Lin(w,w'), DIRT(w,w')) & (w,w') \in WordNet, DIRT \\ 0 & otherwise \end{cases} \quad (1)$$

公式(1)表明，如果词 *w* 和 *w'*相同，则它们的相似度为 1，否则利用 WordNet 和 DIRT 计算两个词的相似度。若以上条件均不满足，则两个词的相似度为 0。$avg(Lin(w,w'), DIRT(w,w'))$计算方法如下：

$$avg(Lin(w,w'), DIRT(w,w')) = \frac{Lin(w,w') + DIRT(w,w')}{\#n} \quad (2)$$

公式(2)计算的是 *w*，*w'*在 WordNet 和 DIRT 中的平均相似度，#n 表示 *w*，*w'*在知识库出现的个数，若 *w*，*w'*出现在 WordNet 和 DIRT 中，则#n 为 2。若 *w*，*w'*在 WordNet 出现，则#n 取值为 1。若 *w* 或 *w'*未出现在任何资源中，则公式值为 0。

$Lin(w,w')$计算方法为：若 w 和 w'属于 WordNet，则计算两个词语之间的 Leacock-Chodorow 相似度 [18]。$DIRT(w,w')$计算方法为：若 DIRT 规则包含 w 和 w'的关系，则认为两个词的 DIRT 相似度为 1，否则为 0。例如，对于 *buy* 和 *acquire*，存在规则“X *buy* Y”→“X *acquire* Y”，则认为两个词的 DIRT 相似度为 1。

3.3 Gibbs 抽样

KLDA 模型中的 Gibbs 抽样过程利用外部先验知识修正 LDA 模型抽样的精度问题。其主要的思想是，在抽样过程中，对于每一个单词 w 的兴趣抽样分布，是与需要分配的话题中词语有关。即可以通过比较单词 w 与话题中词语的相关度，来衡量单词 w 是否可以分配到话题桶中。

我们对标准 Gibbs 抽样公式 [19]进行修改，引入先验外部知识，具体公式如下：

$$P(z_i = k | z_{\neg i}, E) \propto \eta \frac{n_{m,\neg i}^{(k)} + \alpha_k}{\sum_{k=1}^{K} n_{m,\neg i}^{(t)} + \alpha_k} \times \frac{n_{k,\neg i}^{(t)} + \beta_t}{\sum_{t=1}^{V} n_{k,\neg i}^{(t)} + \beta_t} + (1-\eta)\frac{1}{V'}(\sum_{w'=1}^{V'} E_{k,w',w_i}) \quad (3)$$

其中，η是松弛变量，权衡主题下词频和基于知识库的语义相似度，w_i是在当前主题 z_i下抽样的单词，$n_{m,\neg i}^{(k)}$是当前主题在文本 *m* 中出现的次数，$n_{k,\neg i}^{(t)}$是单词 w_i出现在主题 z_i的次数，α、β是 Dirichlet 超参数，V 是词汇表单词的个数，V'是主题 z_i中单词的个数，*w'*是主题 z_i中的单词，E 是词汇关联矩阵。修改后 Gibbs 抽样的基本算法如图 3 所示：

Algorithm 3：Gibbs Sampling
While not finished **do**
compute the words in each topic
for all documents $m \in [1, M]$ **do**
for all words $n \in [1, N_m]$ in document m **do**
for the current assignment of k to a term t for word wm,n **do**:
decrement counts and sums：$n_m^{(k)}-1$，n_m-1，$n_k^{(t)}-1$，n_k-1
multinomial sampling according to eq.3-3:
sample topic index $k \sim p(z_i \mid z_{\neg i}, w)$
use the new assignment of $z_{m,n}$ to the term t for word $w_{m,n}$ to:
increment counts and sums：$n_m^{(k)}+1$，n_m+1，$n_k^{(t)}+1$，n_k+1
end for
end for
end for
end while

图 3 Gibbs 抽样算法步骤

修改后的 Gibbs 抽样算法基本过程如下：在每次抽样之前，计算每个主题中所包含的单词，即统计主题中词频计数大于 0 的所有词语。对于第 m 篇文档的第 n 个单词，把它分配到主题 k 的概率是由词频和外部知识共同决定的，即还受到该词与主题中所有词的相关度决定。然后重新对单词进行主题的划分。当迭代收敛后，输出模型的参数。

4 实验

文本设计了两个实验，一是基于 LDA 模型的文本蕴涵识别系统，二是基于 KLDA 的文本蕴涵识别系统，分别利用 RTE-8 数据集进行测试。

文本蕴涵识别系统框架采用我们在 RTE-5 中使用的系统 [20]。系统首先对文本进行预处理，包括词根还原、词性标注、句法分析、语义分析和命名实体识别，然后基于预处理结果构建四类特征，包括字串特征、结构特征、语言学特征和基于 LDA 的语义相关度特征。其中，我们采用前三类特征中的部分特征作为本系统中的特征；同时，我们利用话题模型计算语义相关度特征。在学习阶段，文本-假设对利用这些特征生成特征向量，放入分类器进行学习。

4.1 任务及数据集介绍

实验选取 RTE-8 测评任务进行测试，RTE-8 测评针对教育 NLP 领域中的学生答案进行分析，它是文本蕴涵在教育领域的一个应用，通过判断学生答案是否蕴涵标准答案来判别学生答案是否正确。它的测评主任务分为五分类(5-way task)，三分类(3-way task)和二分类(2-way task)三个子任务。每个子任务的数据集，给出了一个问题 Q(Question)，和该问题的标准答案 RA(Reference Answer)以及学生答案 A(Answer)，任务中把学生答案 A 当做 T(Text)，把问题的标准答案当做 H(Hypothesis)，然后对该蕴涵对 T-H 进行蕴涵判断。从而判断学生答案是否正确。文本主要做二分类任务。

数据：使用的数据集分两部分：一是 Beetle 数据集，该数据集是从 BEETLE II 教育辅导系统中获取的标注语料，数据集包括高中电学知识，二是 Science Entailments 语料库(SciEntsBank)，该语料库中包含了 16 个不同科学领域的知识，如物理学、生命科学等。提供三个测试集，第一个测试集为 Unseen answers(UA)测试集，在该测试集中，提供的问题和标准答案与训练集相同，但学生答案不同。第二个测试集为 Unseen questions (UQ)测试集，

该测试集中问题、标准答案以及学生答案均与训练集不同，但和训练集处于同一领域范围。第三个测试集为 Unseen domains (UD)，测试集随机选取三个与训练集不同的领域，从选取的领域中获得问题、标准答案和学生答案。例 1 中的句子来自 SciEntsBank 语料库中的 2way 训练集，其中 SA1 标记为 correct，SA2,SA3 标记为 incorrect。

例 1：

Question: You used several methods to separate and identify the substances in mock rocks. How did you separate the salt from the water?

RA: The water was evaporated, leaving the salt.

SA1: Let the water evaporate and the salt is left behind.

SA2: You just get water and the smashed mock rock and put the smashed rock and water together.

SA3:I do not know.

Beetle 语料库中训练集、Unseen answers(UA)测试集、Unseen questions(UQ)测试集分别有 3941 对、439 对、819 对文本，SciEntsBank 语料库中训练集、Unseen answers(UA)测试集、Unseen questions(UQ)测试集、Unseen domains(UD)测试集分别有 4969 对、540 对、733 对、4562 对文本。

4. 2 实验结果

本文分别在 SciEntsBank 和 Beetle 数据集上做了实验，在 SciEntsBank 数据集上，把训练样本和测试样本总共 10804 对文本对作为 KLDA 的数据集，同样的在 beetle 数据集上，共有 5199 对文本对作为 KLDA 的数据集。设定 α 的初始值为 50 除以主题数， β 初始值为 0.1，松弛变量η取值在 0.1-0.9 之间变化，取主题数 k 在 20-100 之间进行变化，迭代次数设为 2000，训练 KLDA 模型，最终分别得到了训练集和测试集文本 T 和假设 H 的主题分布；接着，分别计算训练集和测试集中文本 T 和假设 H 主题的 KL 距离；最后，把计算得到的 KL 距离值作为一维特征加入系统中训练。KLDA 模型主题数对实验结果的影响如图 4 所示，图中展示的是在 SciEntsBank 数据集上的实验效果。从图 4 中可以看出，话题数对系统实验最终结果的影响并不是很大，分析原因，是由于本实验把主题相似度做为一个特征，与其他特征融合到系统中，所以主题数对结果影响并不会很大。

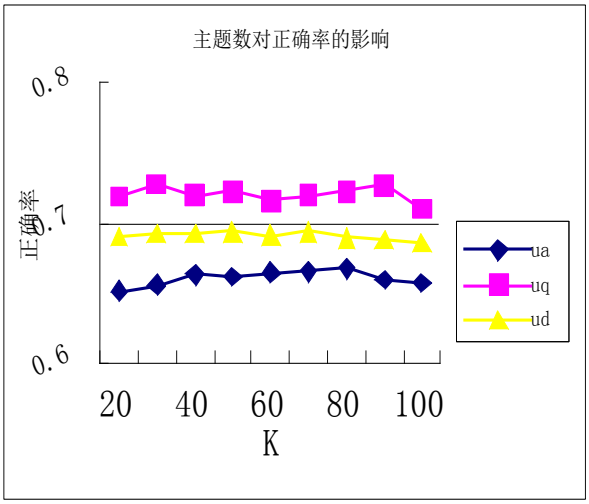

图 4 主题数对正确率的影响

松弛变量η对实验结果的影响如图 5 所示。图中展示的是在主题数 k=20 的条件下，在 SciEntsBank 数据集上的实验效果。从图中可以看出，松弛变量η设的越大，也就是外部知识

的权重越大，实验效果越好，这符合本文的实验预期。

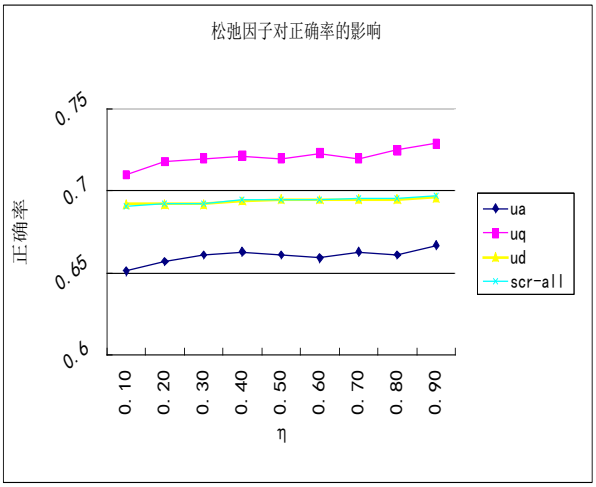

图 5 松弛因子对正确率影响

系统整体实验测评结果如表 1 至表 3 所示。基准系统 baseline 在 sys-lda 基础上去掉了话题模型特征。各表中 SoftCardinality-run1、ETS-run1 和 CU-run1 展示的是参赛队伍前三名的结果，Mean 显示的是全部参赛队伍平均性能 Lexical overlap baseline 显示的是 RTE-8 中基于词汇重叠的基准系统。

表 1 系统正确率评测

		Beetle			SCIENTSBANK				Overall	
Task	System	UA	UQ	ALL	UA	UQ	UD	ALL	ALL	Rank
2way	baseline	0.784	0.759	0.768	0.659	0.690	0.688	0.686	0.700	-
	sys-lda	0.794	0.767	0.776	0.661	0.723	0.694	0.694	0.705	
	sys-klda	0.802	**0.775**	**0.784**	0.667	0.729	0.697	0.698	**0.714**	-
	SoftCardinality-run1	0.781	0.667	0.707	**0.724**	**0.745**	**0.726**	**0.716**	0.715	1
	ETS-run1	**0.811**	0.741	0.765	0.722	0.711	0.698	0.702	0.713	2
	CU-run1	0.786	0.718	0.742	0.656	0.674	0.693	0.687	0.697	3
	Mean	0.738	0.658	0.686	0.678	0.639	0.644	0.647	0.654	-
	Lexical overlap baseline	0.797	0.740	0.760	0.661	0.674	0.676	0.674	0.690	-

表 2 系统加权 F1 评测

Task	System	Beetle		SCIENTSBANK			Rank
		UA	UQ	UA	UQ	UD	
2way	baseline	0.797	0.768	0.626	0.699	0.696	-
	sys-lda	0.797	0.771	0.678	0.708	0.698	-
	sys-klda	0.803	0.769	0.680	0.729	0.701	-
	SoftCardinality-run1	0.782	0.652	**0.722**	**0.745**	**0.712**	1
	ETS-run1	**0.810**	0.732	0.714	0.703	0.694	2
	CU-run1	0.786	0.704	0.623	0.658	0.666	3
	Mean	0.725	0.623	0.659	0.621	0.626	-
	Lexical overlap baseline	0.797	0.735	0.635	0.653	0.665	-

表 3 系统宏平均评测

Task	System	Beetle		SCIENTSBANK			Rank
		UA	UQ	UA	UQ	UD	
2way	baseline	0.782	0.750	0.611	0.683	0.664	-
	sys-lda	0.782	0.754	0.629	0.672	0.662	-
	sys-klda	0.790	0.756	0.622	0.714	0.668	-
	SoftCardinality-run1	0.774	0.635	**0.715**	**0.737**	**0.705**	1
	ETS-run1	**0.802**	0.720	0.705	0.688	0.683	2
	CU-run1	0.778	0.689	0.603	0.638	0.673	3
	Mean	0.714	0.607	0.645	0.602	0.610	-
	Lexical overlap baseline	0.788	0.725	0.617	0.630	0.650	-

从表 1 可以看出基于 KLDA 的文本蕴涵识别系统在 beetle 和 ScientsBank 数据集上，总体正确率达到 71.4%，在所有参赛队伍中排名第二，尤其是在 beetle 的 UA 数据集上，正确率达到了 80.2%，在所有参赛队伍中排名第二；在 beetle 的 UQ 数据集上 sys-klda 系统正确率比第一名队伍高出 1%，说明知识话题模型的部分性能优于 RTE-8 最优参赛队伍的系统性能。

基于话题模型的系统在 Beetle 和 SciEntsBank 数据集上的总体正确率分别高于基准系统 0.8%和 0.8%，两个数据集的总体宏平均和加权宏平均也有所提高，说明基于话题模型的特征能够提供有效的相关性度量，从而改进文本蕴涵系统的识别性能。另一方面，基于知识话题模型的系统在 Beetle 和 SciEntsBank 数据集上的总体正确率分别高于基于话题模型的系统 0.8%和 0.4%，两个数据集的总体宏平均和加权宏平均也有所提高，说明外部推理知识能够有效提高话题模型的准确率。另一方面，在分别采用话题模型和知识话题模型以后，系统的准确率逐渐提高，说明话题模型与现有特征相结合，能够稳步提高系统的性能。

5 总结

本文提出一个基于知识话题模型的文本蕴涵识别系统。该模型通过话题模型建立词汇相关性评估特征，用于改进基于分类的文本蕴涵识别系统的性能。为改进话题模型构建缺乏知识指导的问题，本文提出通过在抽样过程中引入文本蕴涵所需的背景知识，来修正抽样的结果的方法，以提高话题生成的精度。实验表明，知识话题模型能够有效改进基于分类的文本蕴涵识别系统的性能。

本文引入了外部推理和词义知识，一定程度对话题模型的产生起到修正作用。但是该模型的时间复杂度较高，如何减少时间复杂度是今后研究的一个要点。其次，如何在引入更多的外部知识，如语义结构变换规则，以进一步提高话题生成性能，是今后研究的另一个主要问题。

参考文献

[1] Dagan I, Glickman O, Magnini B. The PASCAL recognising textual entailment challenge[M]//Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment. Springer Berlin Heidelberg, 2006: 177-190.

[2] Androutsopoulos I, Malakasiotis P. A Survey of Paraphrasing and Textul Entailment Methods[J]//Journal of Artificial Intelligence Research, 2010, 38(1): 135-187.

[3] Dagan I, Dolan B. Recognizing textual entailment: Rational, evaluation and approaches[J]//Natural Language Engineering, 2009, 15(4): i-xvii.

[4] O.Dzikovska M, D.Nielsen R, Brew C, et al. SemEval-2013 Task 7: The Joint Student Response Analysis and 8th Recognizing Textual Entailment Challenge[C]// Second Joint Conference on Lexical and

Computational Semantics, 2013: 263-274.
[5] 张鹏，李国臣，李茹等. 基于 FrameNet 框架关系的文本蕴含识别[J]. 中文信息学报，2012，26(2)：46-50.
[6] De Marneffe M C, Rafferty A N, Manning C D. Finding Contradictions in Text[C]//ACL. 2008, 8: 1039-1047.
[7] Malakasiotis P, Androutsopoulos I. Learning textual entailment using SVMs and string similarity measures[C]//Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing. Association for Computational Linguistics, 2007: 42-47.
[8] 刘茂福，李妍，姬东鸿.基于事件语义特征的中文文本蕴含识别[J]. 中文信息学报，2013，27(5)：129-136.
[9] 石晶，戴国忠.基于知网的文本推理[J]. 中文信息学报，2006，20(1)：76-84.
[10] Kouylekov M, Magnini B. Recognizing textual entailment with tree edit distance algorithms[C]//Proceedings of the First Challenge Workshop Recognising Textual Entailment. 2005: 17-20.
[11] Kouylekov M, Negri M. An open-source package for recognizing textual entailment[C]//Proceedings of the ACL 2010 System Demonstrations. Association for Computational Linguistics, 2010: 42-47.
[12] Lin D, Pantel P. Discovery of inference rules for question-answering[J]. Natural Language Engineering, 2001, 7(4): 343-360.
[13] Berant J, Dagan I, Goldberger J. Global learning of typed entailment rules[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, 2011: 610-619.
[14] Melamud O, Berant J, Dagan I, et al. A Two Level Model for Context Sensitive Inference Rules[C]// Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, 2013:1331-1340.
[15] T.N. Rubin, A. Chambers, P. Smyth, and M. Steyvers. Statistical topic models for multi-label document classification[M]//Arxiv preprint arXiv, 2011:1107.2462.
[16] Chen Z, Liu B. Topic Modeling using Topics from Many Domains, Lifelong Learning and Big Data[C]//Proceedings of the 31st International Conference on Machine Learning (ICML-14), 2014: 703-711.
[17] Chang J, Boyd-Graber J, Chong W, Gerrish S and Blei D M. 2009. Reading Tea Leaves: How Humans Interpret Topic Models. In Proceedings of the 23rh Annual Conference on Neural Information Processing Systems, Vancouver, Canada.
[18] Leacock C and Chodorow M. Combing Local Context and WordNet Similarity for Word Sense Identification[M]//In C.Fellbaum, editor, An Electronic Lexical Database, MIT Press, 1998:265--283.
[19] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]//The Journal of machine Learning research, 2003, 3: 993-1022.
[20] Ren H, Ji D and Wan J. WHU at TAC 2009: A Tri-categorization Approach to Textual Entailment Recognition[C]//In Proceedings of the Fifth PASCAL Challenges Workshop on Recognizing Textual Entailment, Gaithersburg, Maryland, USA, 2009.

任函（1980—），男，博士，主要研究领域为自然语言处理。Email: hanren@whu.edu.cn；

盛雅琦（1991—），女，硕士，主要研究领域自然语言处理。本文通讯作者。 Email: shmilysyq@ whu.edu.cn；

冯文贺（1976—），博士，博士后（在站），主要研究领域为理论语言学、计算语言学。Email: wenhefeng@gmail.com。

刘茂福（1977—），男，博士，教授，主要研究领域为自然语言处理。Email: liumaofu@wust.edu.cn。