

文章编号: 1003-0077 (2013) 00-0000-00

基于事件语义特征的中文文本蕴含识别

刘茂福^{1,2}, 李妍^{1,2}, 姬东鸿³

(1. 武汉科技大学计算机科学与技术学院, 武汉 430065, 中国; 2. 智能信息处理与实时工业系统湖北省重点实验室, 武汉 430065, 中国; 3. 武汉大学计算机学院, 武汉 430072, 中国)

E-mail: liumaofu@wust.edu.cn

摘要: 为了强化文本蕴含系统深层语义分析与推理能力, 本文提出了基于事件语义特征的中文文本蕴含识别方法。该方法基于事件标注语料生成事件图, 将文本间的蕴含关系转化为事件图间的蕴含关系; 利用最大公共子图的事件图相似度算法计算事件语义特征, 与统计特征、词汇语义特征和句法特征一起使用支持向量机进行分类, 得到初步实验结果, 再经过基于事件语义规则集合的修正处理得到最后的识别结果。实验结果表明基于事件语义特征的中文文本蕴含识别方法可以更有效地对中文文本蕴含关系进行识别。

关键词: 文本蕴含; 事件语义特征; 最大公共子图; 支持向量机

中图分类号: TP391 文献标识码: A

Event Semantic Feature Based Chinese Textual Entailment Recognition

LIU Maofu^{1,2}, LI Yan^{1,2}, JI Donghong³

(1. College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China; 2. Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan 430065, China; 3. School of Computer, Wuhan University, Wuhan 430072, China)

Abstract: In order to strengthen deep semantic analysis and inference of textual entailment, this paper proposes the method of event semantic feature based Chinese textual entailment recognition. The method generates event graphs base on event labeled corpus, and then the entailment recognition between text pairs can be changed to entailment recognition between event graphs. The event semantic feature can be computed based on max-common sub-graph. The event semantic feature combined with the surface statistical feature, lexical semantic feature and syntactic feature is used to classify textual entailment based on support vector machine and can obtain the preliminary experimental result, and the correction module based on event semantic rules handles preliminary experimental result to get the final experimental result. The experimental results show that the event semantic feature based Chinese textual entailment recognition is effective and efficient in Chinese textual entailment recognition.

Key words: Textual Entailment; Event Semantic Feature; Max-Common Sub-graph, Support Vector Machine

1 引言

文本蕴含^[1]通常定义为一个连贯文本 T 与一个假设文本 H 之间的一种关系, 即如果假设文本 H 的意义可以从文本 T 的意义中推断出来, 那么 T 蕴含 H, 可以记为 $T \rightarrow H$ 。语义表达多样性为自然语言的一个基本特征, 即同样的意义可以用多种不同的文本进行表达, 文本蕴含的研究目的之一就是解决语义表达多样性的问题。文本蕴含的识别实际上是语言基础研

基金项目: 国家自然科学基金 (61100133; 61173062); 国家社会科学基金重大项目 (11&Z189); 湖北省教育厅人文社科基金重点项目 (2011jyte126)

作者简介: 刘茂福 (1977-), 男, 博士, 教授, 研究方向为计算语言学; 李妍(1987-), 女, 硕士, 研究方向为自然语言处理; 姬东鸿 (1967-), 男, 博士, 教授, 研究方向为计算语言学

究，它在自然语言处理的很多应用中起着关键作用，如问答系统、多文档自动摘要、信息抽取、语义检索以及机器翻译评测等^[2]。

近年来，文本蕴含识别研究已经从早期的纯理论研究转换为较大规模理论研究和经验性平台建设，在自然语言处理领域备受关注。日本国立情报学研究所 NII (National Institute of Informatics) 组织的 NTCIR (NII Test Collection for IR Systems) 在 2011 年开始了文本蕴含识别 RITE (Recognizing Inference in TExt) 方面的评测任务。RITE 的目的是评测系统识别特定语句“关系”的能力。NTCIR 参与 RITE 评测的语言有日文 JA (Japanese)、简体中文 CS (Simplified Chinese) 以及繁体中文 CT (Traditional Chinese)，每种语言都包含两类子任务，即二分类 BC (Binary-Class) 子任务与多分类 MC (Multi-Class) 子任务。参与评测的系统需要辨识给定的两个文本之间的关系，输出二选一 (2-way) 或五选一 (5-way) 的关系标记 (Label)。图 1 是 BC 子任务和 MC 子任务的概览。

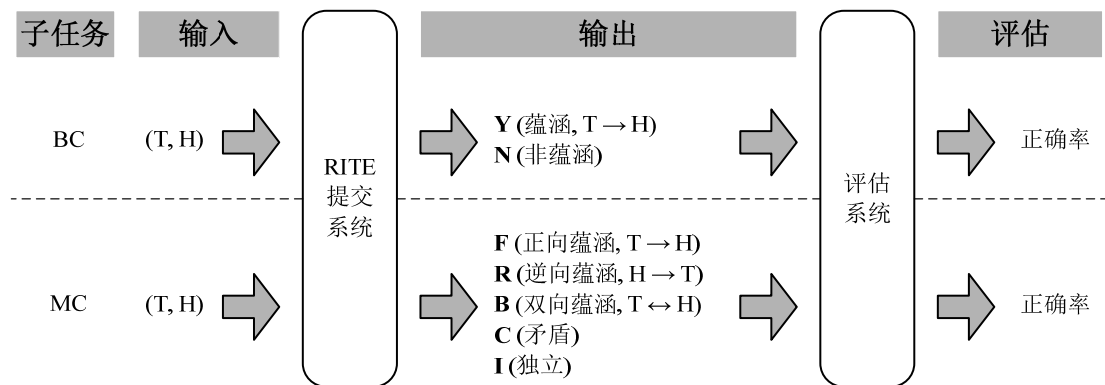


图 1 NTCIR-9 RITE 概览

蕴含关系识别通常采用有监督机器学习方法，也就是说，首先对大量已有正确蕴含关系标签的文本对进行训练，生成分类模型，然后由分类模型决定测试文本对是否具有蕴含关系。采用有监督机器学习方法的蕴含识别系统，通常为了提高正确率会混合多种特征。在 NTCIR-9 中，哈尔滨工业大学^[3]、复旦大学^[4]、国立台湾大学^[5]、北京邮电大学^[6]、武汉大学^[7]和台湾朝阳科技大学^[8]纷纷采用有监督机器学习方法解决文本蕴含问题，采用的特征包括统计特征、句法特征以及语义特征等。本文作者参与了 NTCIR-9 的 RITE 任务，并提交了基于统计与词汇语义特征的中文文本蕴含识别系统 WUST；而后的改进系统 WUST-SLS 在 WUST-ES 系统的基础上又增加了句法特征。

本文在统计特征、句法特征和词汇语义特征的基础上增加了事件语义特征。事件是一个由谓词以及与其相关的语义角色所构成。事件标注后的文本易于进行深层语义分析与推理，并能生成事件图，将文本间的语义关系转换为事件图之间的语义关系。本文采用基于最大公共子图的事件图相似度算法来计算事件语义特征，与统计特征、词汇语义特征和句法特征一起作为支持向量机的分类特征，得到初步实验结果，再经过基于事件语义规则集合的修正处理得到最后的识别结果。实验表明，基于事件语义特征的蕴含识别方法适用于中文文本蕴含关系识别。

2 WUST-ES 框架模型

WUST-ES 模型由预处理、事件语义特征计算、SVM 分类器和修正模块四个部分组成，具体的模型结构图如图 2 所示。

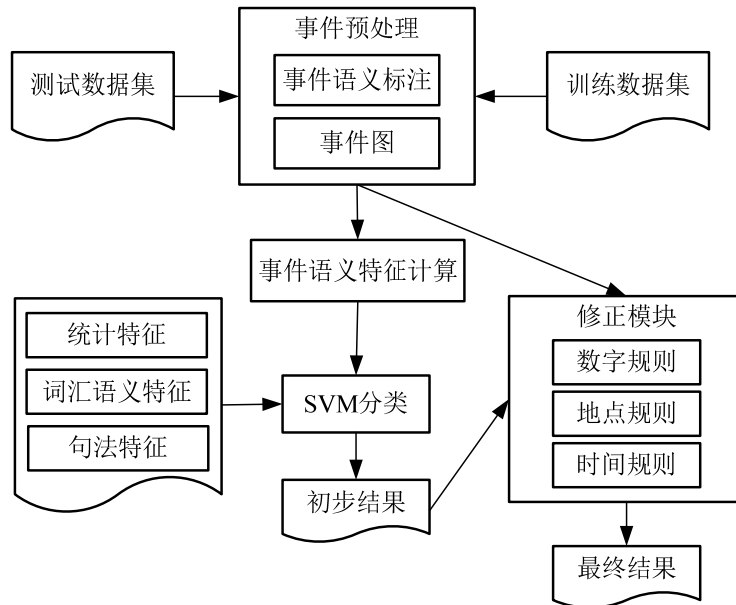


图2 WUST-ES 模型结构图

图2中，事件预处理模块对测试数据集和训练数据集进行事件语义标注，具体的标注方法见文章^[9]，表1列举了训练数据集中的一对文本的标注结果。表1中，T1可以分为三个事件T1_E1、T1_E2和T1_E3，H1则包含一个事件H1_E1。

表1 事件语义标注结果

| |
|--|
| MC 训练集示例 |
| T1: 巩俐在1965年12月31日生于辽宁沈阳，祖籍山东济南并在此长大。 H1: 巩俐1965年生。 |
| T1与H1事件语义标注结果 |
| T1_E1: [巩俐 Th][在[1965年12月31日 T][生 EP]于[辽宁沈阳 L] |
| T1_E2: [巩俐 Th][祖籍 EP][山东济南 L] |
| T1_E3: [巩俐 Th][在[山东济南 L][长 EP][大 R] |
| H1_E1: [巩俐 Th][1965年 T][生 EP] |

事件语义标注完成之后，可以基于标注的事件语义，生成事件图，将文本相似度计算转化为图相似度计算，图3(a)和图3(b)为T1和H1的事件图。

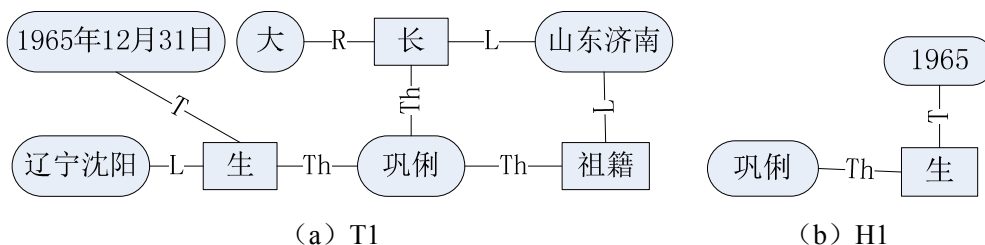


图3 T1与H1的事件图

图3中，矩形框为事件谓词，圆角矩形框为事件语义成分，其语义角色在边线上，事件谓词的数量即文本中包含事件的数量，T1由三个原子事件组成，H1只有一个原子事件。从图3可发现，即使T1和H1的文本结构不同，但H1的事件图基本上为T1事件图的一部分，且T1与H1正向蕴含；可以假设，如果两个事件图相似度越高，则文本对间具有蕴含关系的可能性就越大。

在生成的事件图基础上，使用基于最大公共子图的图相似度算法计算生成事件语义特征，将统计特征、词汇语义特征、句法特征与事件语义特征组合，利用 SVM 分类模块，生成图 2 中的“初步结果”；在“初步结果”的基础上进行修正，生成图 2 中的“最终结果”。

3 事件语义特征

基于事件图，将文本对之间的蕴含关系转换为事件图之间的蕴含关系。事件语义特征由基于最大公共子图的图相似度算法计算得到。

3.1 事件图结构

图结构表示为一个三元组，即 $G = (\text{Node}, \text{Edge}, \text{Weight})$ ，其具体计算步骤如下：

(1) 从事件标注文本对 (T, H) 生成的事件图中提取所有的节点组成节点集合 Node_T 和 Node_H ，集合 Node_T 和 Node_H 是由节点的内容词 w 组成，如公式 (1) 所示。

$$\text{Node} = \{w_1, w_2, \dots, w_n\} \quad (1)$$

(2) 从事件标注文本对 (T, H) 生成的事件图中提取所有边组成边集合 Edge_T 和 Edge_H ，集合 Edge_T 和 Edge_H 是一个三元组集合，每条边 E_{ij} 由两个节点 w_i 和 w_j 以及边语义角色标签 label_{ij} 组成，如公式 (2) 所示。

$$\text{Edge} = \{E_{12}, E_{13}, \dots, E_{ij}, \dots\}, E_{ij} = \langle w_i, w_j, \text{label}_{ij} \rangle \quad (2)$$

(3) 根据节点集和边集，求边的权重集合 Weight_T 和 Weight_H ，集合 Weight_T 和 Weight_H 是二元组集合，由边 E_{ij} 和边权重 $\text{wt}(E_{ij})$ 组成，如公式 (3) 所示，其中权重 $\text{wt}(E_{ij})$ 的计算见公式 (4)。

$$\text{Weight} = \{W_{12}, W_{13}, \dots, W_{ij}, \dots\}, W_{ij} = \langle E_{ij}, \text{wt}(E_{ij}) \rangle \quad (3)$$

$$\text{wt}(E_{ij}) = \text{freq}(w_i, w_j) / (\text{freq}(w_i) + \text{freq}(w_j) - \text{freq}(w_i, w_j)) \quad (4)$$

公式 (4) 中， $\text{freq}(w_i)$ 为节点 w_i 出现在文本中的频率， $\text{freq}(w_i, w_j)$ 为节点 w_i 和 w_j 在文本中的共现频率。

(4) 将事件图 G 转化为三元组 $(\text{Node}, \text{Edge}, \text{Weight})$ ，如公式 (5) 所示。

$$G = \{g_1, g_2, \dots, g_k, \dots\}, g_k = \langle w_i, w_j, E_{ij}, W_{ij} \rangle \quad (5)$$

3.2 最大公共子图

根据图结构 G_T 和 G_H 求解最大公共子图 G_C 的步骤如下所示：

(1) 分别遍历图结构 G_T 和 G_H ，若 G_T 和 G_H 之间存在相同的节点集合，则将相同的节点集合作为最大公共子图的节点集合 Node_C ；判断节点相同，只需判断 $w_i = w_j$ ，其中 $w_i \in \text{Node}_T$ ， $w_j \in \text{Node}_H$ 。

(2) 如果集合 Node_C 中任两个节点之间存在一条边 E_{ij} ，且 E_{ij} 等于某一条既属于 Edge_T 又属于 Edge_H 的边，则将 E_{ij} 加入 G_C 边集合 Edge_C 中。

(3) 根据得到的 $Node_C$ 和 $Edge_C$ 计算 $Weight_C$, 其中 $Weight_C$ 中的二元组集合 W_{ij} 的计算见公式 (6)。

$$W_{ij} = \frac{\min(W_T(E_{ij}), W_H(E_{ij}))}{\max(W_T(E_{ij}), W_H(E_{ij}))}, E_{ij} \in Edge_C \quad (6)$$

公式 (6) 中, $W_T(E_{ij})$ 为边 E_{ij} 在图 G_T 中存在一条相等的边 E_{xy} 的权重 $w_t(E_{xy})$; 同理, $W_H(E_{ij})$ 为边 E_{ij} 在图 G_H 中存在一条相等的边 E_{ab} 的权重 $w_t(E_{ab})$ 。

以上求解最大公共子图的伪代码如下所示:

算法 3.1 求解最大公共子图

输入: $G_T = (Node_T, Edge_T, Weight_T)$

$G_H = (Node_H, Edge_H, Weight_H)$

输出: $G_C = (Node_C, Edge_C, Weight_C)$

//初始化 $Node_C$, $Edge_C$, $Weight_C$ 为空

1: $Node_C = Edge_C = Weight_C = \Phi$

// $w_i \in Node_T$ ($Node_T$ 含有 n 个节点)

// $w_j \in Node_H$ ($Node_H$ 含有 m 个节点)

//求 $Node_C$

2: **for** ($i = 0; i < n; i++$) {

3: **for** ($j = 0; j < m; j++$) {

4: **if** ($w_i == w_j$) $Node_C = \{w_i\} \cup Node_C$

5: }

6: }

//求 $Edge_C$ 与 $Weight_C$

7: **for** ($i = 0; i < n; i++$) {

8: **for** ($j = 0; j < m; j++$) {

9: **if** ($w_i, w_j \in Node_C \ \&\& \ E_{ij} \in (Edge_T \cap Edge_H)$) {

10: $Edge_C = \{E_{ij}\} \cup Edge_C$

11: $Weight_C = \{w_{ij}\} \cup Weight_C$

12: }

13: }

3.3 图相似度计算

根据求得的最大公共子图, 计算图相似度的如公式 (7) 所示。

$$sim(G_T, G_H) = \beta \frac{sizeof(Node_C)}{\max(sizeof(Node_T), sizeof(Node_H))} + (1 - \beta) \frac{\sum W_C(E_{ij})}{\max(sizeof(Edge_T), sizeof(Edge_H))} \quad (7)$$

其中，sizeof(Node_T)为G_T中节点数，sizeof(Edge_T)为G_T中边数，W_C(E_{ij})为图G_C中边E_{ij}的权重wt(E_{ij})；β为综合加权因子，当β=0.5时，图G_T和图G_H中节点和边对图相似度的影响程度相同，当β=0时，不考虑节点对图相似度的影响，当β=1时，不考虑边对图相似度的影响。本文β取值0.6，综合考虑了边和节点对图相似度的影响。

4 基于事件语义特征的修正

基于事件语义规则的修正目的是从标注文本对中提取时间、地点和数字等语义成分，修正被分类器误判为双向蕴含关系的矛盾关系。

4.1 数字语义特征

在SVM的分类结果中，绝大部分通过数字判断为矛盾的文本对都被判断为了双向蕴含，表2中列举了一些会被判断错误的典型例子。

表2 数字相关SVM输出错误文本对

| |
|--|
| 数值不同 |
| T2: 1976年蒙特利尔奥运会，拥有628家赞助厂商，赞助价值高达700万美元。 H2: 1976年蒙特利尔奥运会，拥有628家赞助厂商，赞助价值高达600万美元。 |
| 单位不同 |
| T3: 据《瞭望》杂志的估算，东海油田蕴含石油250亿吨。 H3: 据《瞭望》杂志的估算，东海油田蕴含石油250亿公斤。 |
| 范围不同 |
| T4: 重大创伤的受害者中，有近四分之一可能发展出创伤后压力症候或其部分症状。 H4: 重大创伤的受害者中，超过四分之一可能发展出创伤后压力症候或其部分症状。 |

在SVM的判断结果中，表2中的三种句子均被判断错误，且因为计算出的句子相似度非常高，都被判断为了双向蕴含关系，经过分析发现，通常根据数字认定为矛盾的句子分为三种，一种是数值不同，例如T2中的“700万”和H2中“600万”的数值不同，导致了最后文本对矛盾；第二种是单位不同，例如T3中数字的单位“吨”和H3中数字的单位“公斤”不同，使句子对互相矛盾；最后一种为数字的范围，例如，T4中和H4中数字都是四分之一，但T4中是“近四分之一”而H4中是“超过四分之一”，于是T4和H4互相矛盾。经过上述分析，可以得到基于数字修正SVM分类结果的规则，记为**规则1**。

规则1: 如果文本对(T, H)中都含有数字信息且该文本对被SVM判断为双向蕴含，那么检查文本T和文本H中所有事件中数字的数值、单位和范围，若其中有一项或多项不同时，则将SVM输出标签更正为矛盾关系。

4.2 地点语义特征

对于被分类为双向蕴含的涉地点语义成分文本对，主要是两种形式的文本结构，表 3 列举了这两种文本结构。

表 3 涉地点的分类错误文本对

| |
|--|
| 地点不同 |
| T5: 2005 年 7 月 7 日伦敦地铁爆炸事件的地点是利物浦街站附近。 H5: 2005 年 7 月 7 日伦敦地铁爆炸事件的地点是利物浦市附近。 |
| 地点所在事件不同 |
| T6: 袁隆平是中国杂交水稻育种专家，中国工程院院士。他生于北平，祖籍江西德安，现在居住在湖南长沙。 H6: 袁隆平是中国杂交水稻育种专家，中国工程院院士。他生于北平，祖籍湖南长沙，现在居住在江西德安。 |

在表 3 中，T5 和 H5 涉地点事件的谓语动词都是“爆炸”，并考虑到因为句子相似度高，被 SVM 判断为双向蕴含，则认为 T5 和 H5 要表达同一个事件；但是实际上事件地点不同，则 T5 和 H5 矛盾。T6 和 H6 都含有三个地点，分别为“北平”、“湖南长沙”和“江西德安”，虽然表面上看，T6 和 H6 相似，但是 T6 和 H6 中“湖南长沙”和“江西德安”的谓语动词不同，即其所处事件不同，则事件表达的意义不同，因此 T6 和 H6 矛盾。根据以上两种文本类型，可以总结出**规则 2**。

规则 2: 如果文本对 (T, H) 中都含有地点信息且该文本对被 SVM 判断为双向蕴含，那么检查文本 T 和文本 H 中所有事件的地点和对应的谓语动词，若 T 和 H 中存在一对事件谓语动词相同但事件谓词动词对应地点不同，则更改为矛盾关系。

4.3 时间语义特征

在文本结构类似的文本对中，若时间不同，可以直接判断为矛盾，但是在 SVM 判断中，因为句子相似度过高，导致错误判断为双向蕴含，表 4 列举了这样的例子。

表 4 时间相关 SVM 分类错误文本对

| |
|--|
| 时间不同 |
| T7: 2008 年北京奥运会即第二十九届夏季奥林匹克运动会于 2008 年 8 月 8 日 20 时开幕。 H7: 2008 年北京奥运会即第二十九届夏季奥林匹克运动会于 2008 年 8 月 8 日 22 时开幕。 |

表 4 中，文本 T7 中的时间为“2008 年 8 月 8 日 20 时”，文本 H7 的时间为“2008 年 8 月 8 日 22 时”，时间不同，则文本对相互矛盾。根据这种特征，设计**规则 3**。

规则 3: 如果文本对 (T, H) 中都含有时间信息且该文本对被 SVM 判断为双向蕴含，那么检查文本 T 和文本 H 中所有事件的时间语义成分并比较，若不同，则修正为矛盾关系。

5 实验结果

实验语料来自于 NTCIR-9 RITE 任务，其中训练集与测试集的文本对数都是 407，系统的整体性能使用平均准确率 (Accuracy) 指标来进行度量，如公式 (8) 所示。

$$Accuracy = \frac{1}{\# pair} \sum [output label is correct] \quad (8)$$

公式 (8) 中, #pair 指 RITE 测试集中文本对的数目。

针对具体的文本蕴含关系的类别, 该文给出每一类别的准确率 (Precision)、召回率 (Recall) 和 F-measure 来进行度量, 其具体计算如公式 (9)、(10) 和 (11) 所示。

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (11)$$

上式中, TP (True Positives) 和 FP (False Positives) 分别指测试集中该类所有正例被正确和错误分类的数目; 而 FN (False Negatives) 则指测试集中该类所有负例被错误分类的数目。

本文中的基于统计与词汇语义特征的中文文本蕴含关系识别系统标记为 WUST, 该系统利用 BC 和 MC 任务的训练集对分类器进行训练。在两个任务的训练中, 惩罚因子 C 的 \log 值都为 10, 而核函数因子 γ 的 \log 值皆为 -9; BC 任务的训练正确率为 76.9%, MC 任务训练正确率为 57.74%。WUST 系统给出了对 BC 和 MC 任务测试集的分类结果, RITE 组织方最终的评测结果如表 5 所示。

表 5 系统的 BC 和 MC 任务评测结果

| BC | 准确率 | MC | 准确率 |
|----------------------|---------------|------------------------|---------------|
| UIOWA-CS-BC-01 | 0.9750 | UIOWA-CS-MC-01 | 0.8919 |
| UIOWA-CS-BC-03 | 0.9631 | UIOWA-CS-MC-02 | 0.8919 |
| UIOWA-CS-BC-02 | 0.9361 | UIOWA-CS-MC-01 | 0.8870 |
| ICRI_HITSZ-CS-BC-03 | 0.7764 | ICRI_HITSZ-CS-MC-03 | 0.6413 |
| FudanNLP-CS-BC-02 | 0.7617 | ICRI_HITSZ-CS-MC-02 | 0.6241 |
| ICRI_HITSZ-CS-BC-02 | 0.7568 | ZSWSL-CS-MC-02 | 0.6192 |
| FudanNLP-CS-BC-01 | 0.7469 | WHUTE-CS-MC-02 | 0.6093 |
| WHUTE-CS-BC-03 | 0.7371 | III_CYUT_NTHU-CS-MC-02 | 0.5897 |
| NTU-CS-BC-01 | 0.7346 | FudanNLP-CS-MC-02 | 0.5848 |
| WHUTE-CS-BC-02 | 0.7322 | WHUTE-CS-MC-01 | 0.5823 |
| WUST-CS-BC-01 | 0.7248 | WUST-CS-MC-01 | 0.5823 |
| Baseline | 0.7617 | Baseline | 0.5315 |

表 5 中, UIOWA 系统结果明显高于其它系统, 最主要的原因是 UIOWA 系统使用了众包 (Crowdsourcing) 技术, 有人工参与其中。WUST 和其它很多参与者的系统在 BC 任务的表现不如基准测试 (Baseline), 其中最可能的原因是基准测试使用的是字符重叠度特征^[2], 而考虑到是对中文进行处理, WUST 系统自然而然的使用了词重叠度特征。

WUST-ES 系统是基于 WUST-SLS 系统和 WUST 系统的改进版, WUST-SLS 系统在 WUST 系统的基础上增加了句法特征, 而 WUST-ES 系统在 WUST-SLS 系统基础上增加了事件语义特征和基于事件语义规则的修正模块。WUST-ES 系统实验效果最为理想, 与 WUST 系统和 WUST-SLS 系统结果比较图如图 4 所示。

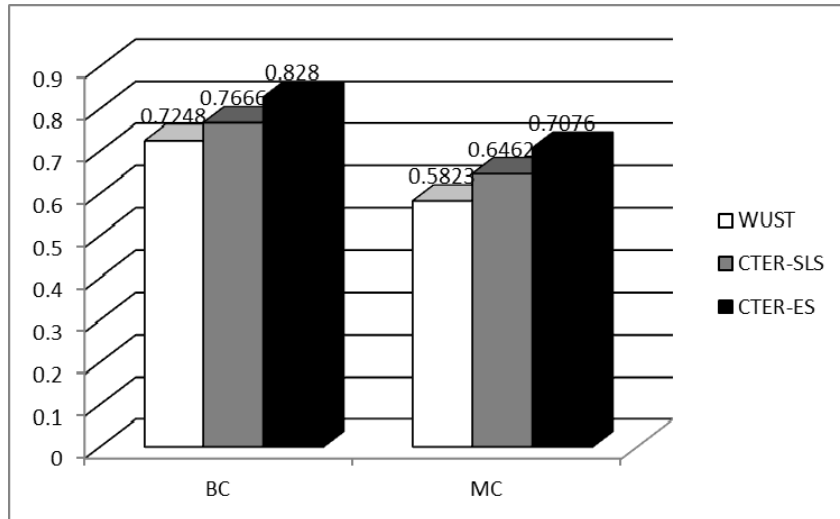


图 4 WUST、WUST-SLS 与 WUST-ES 的对比评测结果

图 4 中，WUST-ES 实验系统的效果最优，主要是采用了事件语义特征的结果，两个文本即使文本顺序不同，但是同一个事件的事件图是一样的，增加了蕴含关系判断的效果。且 WUST-ES 实验系统为了提高矛盾关系的判断效果，增加了基于事件语义规则的修正模块，弥补了 WUST-SLS 系统和 WUST 系统上的不足，提升了系统性能。

与 ICRI_HITSZ 系统相比，WUST-ES 的 BC 与 MC 任务评测结果已经优于 ICRI_HITSZ 系统；当然，同采用众包技术的 UIOWA 系统还有一定差距。本文后面的内容主要针对 WUST-ES 系统进行分析，以期发现 WUST-ES 系统优劣的原因，有利于系统的后续改进。

WUST-ES 系统的混淆矩阵如图 5 (a) 和 (b) 所示。

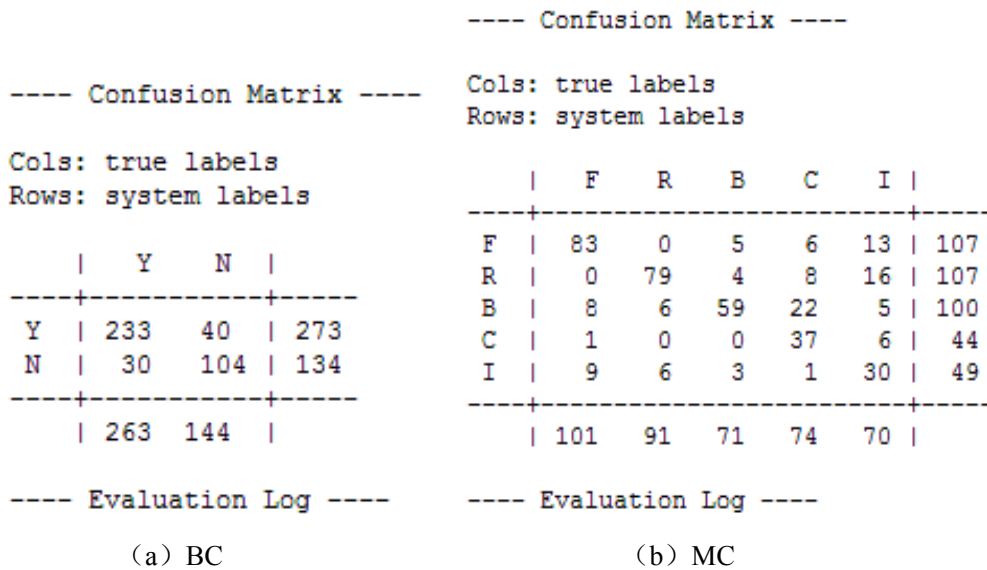


图 5 WUST-ES 系统输出的混淆矩阵

使用混淆矩阵计算每个类别的准确率、召回率和 F-measure，BC 任务的结果在表 6 中，MC 任务的结果见表 7。

表 6 WUST-ES 系统 BC 任务

| 类标签 | 准确率 | 召回率 | F-measure | 正确率 |
|-----|--------|--------|-----------|--------|
| Y | 0.8535 | 0.8859 | 0.8694 | 0.8280 |
| N | 0.7761 | 0.7222 | 0.7482 | |

表 6 中，WUST-ES 系统正确率比 WUST-SLS 系统正确率高 6%左右，比 WUST 系统正确率高 10%左右，说明 WUST-ES 能够较好地识别 2-way 文本蕴含关系。WUST-ES 系统蕴

含关系识别的 F-measure 达到 86.94%，说明 WUST-ES 系统对蕴含关系识别的有效性；WUST-ES 系统非蕴含关系识别的 F-measure 为 74.82%，相比 WUST-ES 对蕴含关系判断的高效性稍有不足。

表 7 WUST-ES 系统 MC 任务

| 类标签 | 准确率 | 召回率 | F-measure | 正确率 |
|----------|--------|--------|-----------|--------|
| F | 0.7757 | 0.8217 | 0.7981 | 0.7076 |
| R | 0.7383 | 0.8681 | 0.7980 | |
| B | 0.59 | 0.8309 | 0.6901 | |
| C | 0.8409 | 0.5 | 0.6271 | |
| I | 0.6122 | 0.4286 | 0.5042 | |

表 7 中，WUST-ES 系统正确率比 WUST-SLS 系统正确率高 6%左右，比 WUST 系统正确率高 12%左右，说明 WUST-ES 判断 5-way 文本蕴含关系效果较好。WUST-ES 系统正向蕴含关系判断 F-measure 达到 79.81%，逆向蕴含关系判断达到 79.80%，说明 WUST-ES 系统能够高效地识别正向蕴含关系和逆向蕴含关系；WUST-ES 系统双向蕴含关系识别的 F-measure 为 69.01%，矛盾关系识别的 F-measure 为 62.71%，独立关系识别的 F-measure 为 50.41%。

相比 WUST-ES 对正向蕴含关系和逆向蕴含关系判断的高效性，WUST-ES 对其它类语义关系的判断稍有不足，尤其是矛盾关系判断和独立关系。通过分析图 4 (b) 可以发现，WUST-ES 系统将 22 个矛盾关系判断成了双向蕴含关系，说明 WUST-ES 系统将一部分矛盾和双向蕴含混淆了；除此之外，WUST-ES 系统独立关系判断并不理想，说明独立关系也需要重点研究。

6 结论

以 NTCIR-9 的 RITE 任务为背景，本文设计并实现了基于简体中文文本的蕴含识别系统。系统基于 SVM 构建分类器，以统计特征、词汇语义特征、句法特征以及事件语义特征生成特征文件，其中词汇语义特征是基于同义词词林、反义词表、否定词表、知网等语义资源计算的，本文将 RITE 问题作为单纯的分类问题进行处理。实验表明，基于事件语义特征的中文文本蕴含识别方法对中文文本蕴含识别的高效性。

该系统存在两点不足，首先是混淆了部分矛盾和双向蕴含关系，其次是用于独立关系判断的特征存在不足，导致独立关系判断效果不佳；未来研究将致力于解决这两个问题。

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[J]. 中文信息学报, 2010, 24 (2): 3-13
- [2] Hideki Shima, Hiroshi Kanayama, Cheng-Wei Lee, et al. Overview of NTCIR-9 RITE: Recognizing Inference in TExt [C]// National Institute of Informatics. The 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access. Tokyo, Japan: National Institute of Informatics, 2011, 291-301
- [3] Yaoyun Zhang, Jun Xu, Chenlong Liu, et al. ICRC_HITSZ at RITE: Leveraging Multiple Classifiers Voting for Textual Entailment Recognition [C]// National Institute of Informatics. The 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access. Tokyo, Japan: National Institute of Informatics, 2011, 325-329
- [4] Ling Cao, Xipeng Qiu, Xuanjing Huang. FudanNLP at RITE 2011: a Shallow Semantic Approach to Textual Entailment [C]// National Institute of Informatics. The 9th NTCIR Workshop Meeting on Evaluation

- of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access. Tokyo, Japan: National Institute of Informatics, 2011, 335-338
- [5] Hen-Hsen Huang, Kai-Chun Chang, Haver II J.M.C. et al. NTU Textual Entailment System for NTCIR 9 RITE Task [C]// National Institute of Informatics. The 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access. Tokyo, Japan: National Institute of Informatics, 2011, 349-352
- [6] Ranxu Su, Sheng Shang, Pan Wang, et al. ZSWSL Text Entailment Recognizing System at NTCIR-9 RITE Task [C]// National Institute of Informatics. The 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access. Tokyo, Japan: National Institute of Informatics, 2011, 394-399
- [7] Han Ren, Chen Lv, Donghong Ji. The WHUTE System in NTCIR-9 RITE Task [C]// National Institute of Informatics. The 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access. Tokyo, Japan: National Institute of Informatics, 2011, 373-378
- [8] Shih-Hung Wu, Wan-Chi Huang, Liang-Pu Chen, et al. Binary-class and Multi-class Chinese Textual Entailment System Description in NTCIR-9 RITE [C]// National Institute of Informatics. The 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access. Tokyo, Japan: National Institute of Informatics, 2011, 422-426
- [9] Maofu Liu, Yan Li, Donghong Ji, et al. Atomic Event Semantic Roles and Chinese Instances Analysis [C]// Donghong Ji & Guozheng Xiao (Eds.). Chinese Lexical Semantics. Berlin, Heidelberg: Springer-Verlag, 2013, 110-121
- [10] Maofu Liu, Yan Li, Yu Xiao, et al. WUST SVM-Based System at NTCIR-9 RITE Task [C]// National Institute of Informatics. The 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access. Tokyo, Japan: National Institute of Informatics, 2011, 318-324
- [11] Malakasiotis P. Paraphrase recognition using machine learning to combine similarity measures[C]// Association for Computational Linguistics. Proceedings of the ACL-IJCNLP 2009 Student Research Workshop. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009, 27-35.