

# 基于协同图排序的对比新闻自动摘要

黄小江 万小军<sup>†</sup> 肖建国

北京大学计算机科学技术研究所, 北京 100871; <sup>†</sup> E-mail: wanxiaojun@pku.edu.cn

**摘要** 采用协同图排序模型, 为两个可比的新闻话题自动生成对比摘要。利用一个话题内句子之间的相似性, 以及不同话题中句子之间的对比性, 采用迭代增强的方法, 同时计算两个话题中每个句子的重要程度, 并考虑信息的新颖程度, 选择适当的句子组成对比摘要。实验结果表明了该方法的有效性。

**关键词** 对比新闻摘要; 对比文本挖掘; 多文档摘要; 图排序

**中图分类号** TP391

## Comparative News Summarization Using Co-ranking Graph Model

HUANG Xiaojiang, WAN Xiaojun<sup>†</sup>, XIAO Jianguo

Institute of Computer Science and Technology, Peking University, Beijing 100871

<sup>†</sup> Corresponding author, E-mail: wanxiaojun@pku.edu.cn

**Abstract** This study proposes an approach of comparative news summarization using co-ranking graph model. The model makes use of the similarity between sentences within each topic and the comparativeness between sentences of different topics, and then calculates the saliences of sentences of both topics simultaneously using an iterative reinforcement approach. Experiment results show the effectiveness of the proposed approach.

**Key words** comparative news summarization; comparative text mining; multiple-document summarization; graph ranking model

随着互联网和 Web 媒体的发展, 我们能够轻易了解到世界各地发生的事件。如何从海量的数据中快速获得有用信息成为一个关键问题。新闻话题检测是一项广泛应用的重要技术, 它将相关的新闻报道聚集成新闻话题, 并提供简短的摘要描述。相关新闻推荐是另一项普遍应用的技术, 它根据当前浏览的新闻报道, 自动提供相似话题或相关话题的新闻链接。这些技术为挖掘新闻文档中更丰富的知识提供了机遇。我们可以从相关话题中挖掘发展趋势, 例如一段时间内的市场变化; 可以从相似事件中总结教训, 例如如何应对“颜色革命”、“阿拉伯之春”; 可以从相关报道中判断竞争者的优劣, 例如奥巴马和罗姆尼谁更适合当总统; 等等。然而, 为了做出准确的比较, 需要了解与参比话题相关的广泛知识, 因此人工对比分析往往需要耗费大量的时间和人

力。若能自动生成一段简短摘要, 概括话题之间的对比信息, 将能帮助读者更简单有效地分析新闻话题。

语义上, 一个比较阐述了两个或多个对象之间的共同点或不同点。它包含 3 个要素: 比较对象、比较点(对象所衡量和比较的属性)和比较结果(描述对象在比较点上的相对或绝对位置)。比较句是一种语法化的表述形式, 通过特殊的功能词和句法形式, 描述对象之间的比较关系。此外, 比较也可以通过一组文本片段(例如句子或段落)来表示, 每个片段描述一个对象在比较点上的值。例如:

- 1) 智利比海地富裕;
- 2) 海地是一个极度贫困的国家;
- 3) 智利是一个富裕国家。

其中, 句 1 是一个典型的比较句, 比较对象为“智

国家自然科学基金(61170166)、北京市科技新星计划(2008B03)和国家高技术研究发展计划(2012AA011101)资助

收稿日期: 2012-05-31; 修回日期: 2012-08-17; 网络出版时间: 2012-10-26 17:55

网络出版地址: <http://www.cnki.net/kcms/detail/11.2442.N.20121026.1755.021.html>

利”和“海地”，比较点为“经济/财富”，比较结果为智利大于海地。单独考虑句子 2 和 3 时，它们都不描述任何比较信息。但综合考虑两个句子，我们可以根据“贫困-富裕”推断两者之间的比较关系，得到与句子 1 相似的结果。

对比新闻摘要的目标是用简短可读的文本概括相关的新闻话题中重要的对比信息。近年来，对比自动摘要研究成为文本挖掘领域的一个热点问题，并已有研究者提出了一些模型和方法。但以往的研究主要关注于产品评论领域。一个重要原因是因为产品评论中的属性相对容易抽取，并且评论的对比形式相对简单，即“正面/负面”。在新闻报道中，新闻属性更加复杂，包括事件的时间、参与的人物、参与者的观点和行为等等。这些属性的表述方法各不相同，可以采用显式的描述，也可能隐含在其他描述之中。这些问题给新闻领域的对比摘要带来了巨大的挑战。

本文中，我们提出了一个新的对比新闻摘要方法。一个好的对比摘要应当描述话题之间对比信息，同时这些所对比的属性在每个话题之内又反映了重要的信息。一个句子的对比性可以通过它与另一个话题中句子的对比程度计算，而一个句子的代表性可以通过它与本话题中句子的相似程度计算。据此，我们构建句子关系图，分别表示两个话题中句子之间的对比关系和代表关系。我们采用迭代的方法，利用句子之间的相互增强关系，同时计算两个话题中每个句子的重要性。最后考虑句子的重要性和新颖性，选择适当的句子组成摘要。实验结果显示了该模型的有效性。

## 1 相关工作

关于比较的研究在语言学 and 文学等领域由来已久，大量工作研究了比较的内涵、外延、形式和用法<sup>[1-2]</sup>。比较分析在互联网应用领域同样有着广泛应用。绝大多数电子商务系统提供了商品价格、属性比较功能。这些比较往往基于后台的结构化数据。近年来，从无结构和半结构数据中挖掘比较信息成为一个热点问题。一类工作着重于识别语法化的比较句，并从中抽取比较的要素<sup>[3-4]</sup>。另一类工作从单独的对象描述中抽取对象属性，再将其对应起来<sup>[5-7]</sup>。此外，还有一些研究关注于检测话题中的独特之处<sup>[8]</sup>，这可以认为是一种特殊的比较“包含/不包含”。

研究者提出了比较挖掘的多种具体形式。Liu 等<sup>[9]</sup>将用一个实数表示观点的极性和程度，并用柱状图展示两个产品的对比结果。Zhai 等<sup>[7]</sup>提出一个主题模型，通过词汇的概率分布表示对比主题。Witte 等<sup>[10]</sup>提出了对比摘要任务，挖掘文档中的共同和特有主题。对比摘要包含丰富的信息，同时又容易被人理解，因此受到广泛关注。到目前为止，对比摘要研究主要集中在产品评论领域<sup>[5-6,11]</sup>。Wan 等<sup>[12]</sup>为同一个话题的不同语言报道间生成差异性摘要。

自动文档摘要的目标是生成一段短的自然语言文本描述，概括原文本中的重要信息<sup>[13]</sup>。对比摘要是多文档摘要的一种特殊形式。传统的多文档摘要任务均衡考虑不同的信息，最大化摘要的信息覆盖量。有指导的摘要任务根据用户指定的主题，生成与之相关的文摘<sup>[10]</sup>。与之相比，对比摘要关注一种特殊类型的信息，而不是某个特定的主题。更新式摘要强调检测新文档中的新颖内容<sup>[14]</sup>，可以认为是对比摘要的一种特殊形式。目前主流的文档摘要技术采用抽取式的方法，识别原文中的重要片段(如句子)，再将其组合成文摘。片段重要性可以根据词频、位置、特征词等统计特征和语言学特征计算<sup>[15-16]</sup>，或通过机器学习的方法计算<sup>[17-18]</sup>。近年来，基于图的方法已应用在文档摘要领域，取得了良好的效果<sup>[19-22]</sup>。

## 2 问题定义

### 2.1 新闻话题对比

新闻话题指一个种子事件或活动，以及所有直接相关的事件和活动的新闻报道的集合。例如，关于地震救援的报道属于地震话题的一部分，而两篇关于不同地震的报道通常认为不属于同一个新闻话题。

一个新闻话题往往有与之类似的相关话题。对比相关话题，从中挖掘共同点和不同点，可以帮助读者更好地了解事实、总结经验、汲取教训。与产品评论对比不同，新闻话题对比涉及大量不同的方面。例如，新闻话题可能包括事件的起因和经过、参与者的态度和行为以及事件的其他细节。只要参与对比的话题中同时包含这些方面，那么它们都可以成为比较点。例如，海地地震话题和智利地震话题中，可以对比地震强度、人员伤亡数量、经济损失程度、当地政府救灾措施、国际援助等等。

值得注意的是，参与比较的新闻话题数量并不限制于两个。我们可以比较3个或更多的相关话题，例如在发生在突尼斯、埃及、利比亚、叙利亚等地的“阿拉伯之春”革命等。严格意义上，比较点应该包含在所有的话题之中。因此参与比较的话题越多，比较点往往越少。更宽泛的定义允许比较点只出现在话题集合的一个子集中。在极限情况下，多话题对比退化为若干组两个话题之间对比的组合。在本文中，我们关注于两个话题之间的对比。多个话题之间的对比留待未来进一步研究。

除了对比不同事件之外，还可以对比一个连续事件的不同时间阶段。例如，对比2011年3月之前和之后的利比亚局势，可以发现由北约空袭导致的巨大变化。此外，还可以对比不同报道者对同一个话题的报道，例如阿根廷媒体和英国媒体对马岛战争的报道，从而分析其视角和立场差别。同一话题不同时期、不同版本的对比分析更为强调差别和矛盾，而不同话题间的对比则同时注重其共性和差别。本文中，我们关注于不同话题之间的对比。

## 2.2 对比新闻摘要

对比新闻摘要概况两个可比话题之间的共同点和不同点。摘要由两部分文本组成，其中每一部分则关注于一个话题，同时两部分都讨论两个话题中共有的、可比较的方面。表1展示了关于汶川地震和玉树地震的一个对比摘要。左栏介绍汶川地震，而右栏介绍玉树地震。两段文本都讨论了发生时间、地点、震级、人员伤亡和救援情况等。

形式化的，令  $t_1$  和  $t_2$  为两个可比的新闻话题，每个话题  $t_i$  由文档集合  $D_i$  描述。对比摘要任务要求抽取两段文本  $B_1 = S_{11} \cup S_{12} \dots \cup S_{1n}$ ,  $B_2 = S_{21} \cup S_{22} \cup \dots \cup S_{2n}$ ，其中  $S_{i,j} \subset D_i$  ( $i=1, 2$ ) 为一组关于话题  $t_i$  在  $j$  方面的代表性句子集合。换言之， $S_{1j}$  和  $S_{2j}$  组成了两个话题在  $j$  方面的一个对比。同时，摘要的长度(即字/词数)不能超过最大长度  $L$ 。

对比摘要应当满足下列条件：首先，摘要应当关注于新闻话题之间对比信息，即两者之间共有的方面；其次，摘要中应当尽可能包含重要的、具有代表性的信息，而不应当过分关注话题中的细枝末节；第三，摘要应当在长度限制内尽可能包含更丰富的信息；最后，摘要应当具有良好的语言质量，即应当连贯，并且容易被人理解。

## 3 对比自动摘要方法

### 3.1 概述

对比摘要生成的一种朴素方法是抽取比较句组成摘要。但是，在新闻报道中，比较句并不常见，而恰好涉及给定的对比话题的比较则更为罕见。因此，我们选择两个话题中讨论相关方面的句子组成对比摘要。在对比摘要中，句子应当能够反映话题之间的比较，同时又能反映话题的重要内容。因此，句子的重要性应当满足以下两个假设。

对比性假设：如果一个句子与对比话题中的其他重要句子具有强对比性，则这个句子重要。

代表性假设：如果一个句子与本话题中的其他重要句子具有强相似性，则这个句子重要。

给定两个可比话题，句子之间存在两类不同的关系。

对比关系(话题间关系)：反映两个不同话题之间句子组成一个对比的可能性，即两个句子所讨论的事件方面之间的相关程度。

代表关系(话题内关系)：反映一个话题内部句子之间的信息相关程度，即一个句子可以代表另一个句子的可能性。通常根据句子之间的内容相似性计算。

图1给出了两类关系之间的图示。算法首先构造反映上述关系的图，然后基于关系图迭代计算句子的权重。最后根据关系图和权重，综合考虑句子重要性和信息新颖性，选择适当的句子组成摘要。

表 1 “汶川地震”和“玉树地震”的对比摘要

Table 1 Comparative summary about “Wenchuan earthquake” vs. “Yushu earthquake”

话题	汶川地震	玉树地震
摘要	北京时间5月12日14时28分，位于北纬31度、东经103.4度的四川省汶川县发生里氏7.8级地震，造成重大人员伤亡和重大财产损失。截至13日7时，四川汶川县地震已造成四川、甘肃、陕西、重庆、云南、山西、贵州、湖北8省市共11921人遇难，倒塌房屋50余万间。至15日上午，参与救援行动的解放军和武警官兵、公安民警、干部群众和医务工作者等救援人员已从灾区抢救出伤员6万多人。	4月14日7时49分，青海省玉树藏族自治州玉树县发生7.1级地震，给当地人民群众生命财产造成严重损失。截至北京时间15日上午9时，发生在中国青海玉树的地震已经造成617人遇难，313人失踪，9110人受伤，其中，970人伤势严重。截至17日，已调集至玉树震灾现场的各类救援人员达到15000余人，累计搜救营救被困群众17000人。

### 3.2 关系图构建

给定两个话题  $t_1, t_2$ , 及话题下的句子集合  $S_1, S_2$ , 我们可以构建 3 个无向图  $G_{11}, G_{22}$  和  $G_{12}$ , 分别表示  $t_1$  内句子间的代表关系、 $t_2$  内句子间的代表关系以及  $t_1, t_2$  句子之间的对比关系。

#### 3.2.1 话题内代表关系图

以话题  $t_1$  为例,  $S_1$  中每个句子  $s_{1i}$  对应于  $G_{11}$  中的一个顶点。若两个句子  $s_{1i}, s_{1j}$  之间的相似度大于 0, 则  $s_{1i}$  和  $s_{1j}$  之间存在一条边, 边的权重即为两者的相似度。本文中, 我们通过两个文本向量夹角的余弦计算相似度。每个句子用一个词汇向量表示:  $\mathbf{s}_{1i} = \langle w_{1i}^1, \dots, w_{1i}^m \rangle$ , 其中每项  $w_{1i}^k$  对应于一个词  $k$ , 其权重通过  $tf_{1i}^k \cdot isf^k$  计算。 $tf_{1i}^k$  为词  $k$  在  $s_{1i}$  中出现的频率,  $isf^k$  为词  $k$  在背景语料中的逆句子频率。由此, 我们可以用一个邻接矩阵  $\mathbf{U} = [U_{ij}]_{|S_1| \times |S_1|}$  表示  $G_{11}$ :

$$U_{ij} = \begin{cases} \frac{\mathbf{s}_{1i} \cdot \mathbf{s}_{1j}}{|\mathbf{s}_{1i}| \times |\mathbf{s}_{1j}|}, & \text{若 } i \neq j, \\ 0, & \text{若 } i = j, \end{cases}$$

其中  $|\mathbf{s}_{1i}|$  表示向量  $\mathbf{s}_{1i}$  的模。类似的, 我们可以定义  $G_{22}$ , 以及对应的邻接矩阵  $\mathbf{V} = [V_{ij}]_{|S_2| \times |S_2|}$ :

$$V_{ij} = \begin{cases} \frac{\mathbf{s}_{2i} \cdot \mathbf{s}_{2j}}{|\mathbf{s}_{2i}| \times |\mathbf{s}_{2j}|}, & \text{若 } i \neq j, \\ 0, & \text{若 } i = j, \end{cases}$$

#### 3.2.2 话题间对比关系

话题之间的句子关系反映两个句子的对比程度。两个可对比的句子讨论相关的新闻话题, 但在语义上存在差异。例如:

“梅西获得了 2010 年世界足球先生称号。”

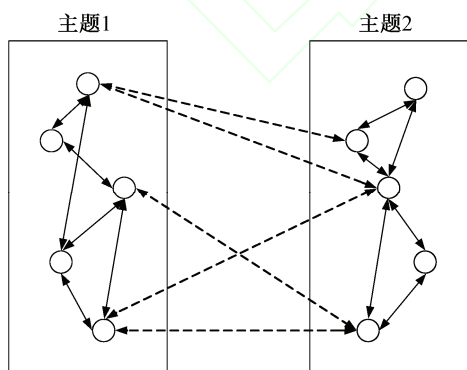


图 1 句子关系图示

Fig. 1 Illustration of the relations between sentences

“C.罗纳尔多是 2009 年世界足球先生。”

上述两个句子比较了不同年份的“世界足球先生”获得者。但除比较点“世界足球先生”之外, 句子的其他部分, 包括比较对象和比较结果并不相同。

可以注意到, 比较点一般为名词或动词词性, 比较对象通常为名词或专有名词, 而比较结果以及其他修饰成分则往往为形容词或副词。作为朴素的做法, 我们保留句子中的名词和动词, 作为句子所讨论事件方面的表示, 即  $\hat{s}_{ij} = \langle \hat{w}_{ij}^1, \dots, \hat{w}_{ij}^m \rangle$  ( $i=1, 2$ ), 其中  $\hat{w}_{ij}^k$  对应于一个词  $k$ , 其权重为

$$\hat{w}_{ij}^k = \begin{cases} tf_{ij}^k \cdot isf^k, & \text{若词 } k \text{ 为名词或动词,} \\ 0, & \text{其他.} \end{cases}$$

据此, 句子  $s_{1i}, s_{2j}$  之间的对比程度可以通过  $\hat{s}_{1i}, \hat{s}_{2j}$  之间的相似度计算。本文采用斯坦福大学的词性标注工具<sup>①</sup>处理语料, 但其他词性标注工具同样适用。

图  $G_{12}$  的每个顶点对应于  $S_1$  和  $S_2$  中的一个句子。若  $s_{1i}$  和  $s_{2j}$  之间的对比程度大于 0, 则  $s_{1i}$  和  $s_{2j}$  之间存在一条边, 其权重为两者的对比程度。我们可以用邻接矩阵  $\mathbf{W} = [W_{ij}]_{|S_1| \times |S_2|}$  表示  $G_{12}$ , 其中  $W_{ij}$  定义如下:

$$W_{ij} = \frac{\hat{s}_{1i} \cdot \hat{s}_{2j}}{|\hat{s}_{1i}| \times |\hat{s}_{2j}|}.$$

### 3.3 句子重要性计算

一个句子的重要性由其对比性和代表性组成。对比性可通过与另一话题中句子的对比程度和它们的重要性计算, 而代表性可由与本话题中句子的相似程度和它们的重要性计算。令  $s_{1i}$  的重要性为  $u_i, s_{2j}$  的重要性为  $v_j$ , 则有

$$u_i = \lambda \cdot \sum_j \frac{W_{ij} \cdot v_j}{\sum_p W_{pj}} + (1 - \lambda) \cdot \sum_k \frac{W_{ik} \cdot u_k}{\sum_p W_{pk}},$$

$$v_j = \lambda \cdot \sum_i \frac{W_{ij} \cdot u_i}{\sum_p W_{ip}} + (1 - \lambda) \cdot \sum_k \frac{V_{kj} \cdot v_k}{\sum_p V_{kp}},$$

其中,  $\lambda \in [0, 1]$  调节对比性和代表性对句子重要性的贡献比重。本文中  $\lambda$  设为 0.5。  $u, v$  可通过迭代的方法计算。算法首先为  $u, v$  随机赋值, 再根据上述公式更新计算  $u^{(n+1)}, v^{(n+1)}$ , 直至结果收敛。

### 3.4 句子选择

基于句子间关系图和句子重要性值, 我们采用

① <http://nlp.stanford.edu/software/tagger.shtml>

贪心算法，综合考虑句子重要性和新颖性，计算句子的排序值。

1) 令  $A = \emptyset$ ,  $B = S_1 \cup S_2$ , 每个句子的排序值为其重要性值，即  $\text{Rank}(s_{1i}) = u_i$ ,  $\text{Rank}(s_{2j}) = v_j$ 。

$$\text{Rank}(s_{pq}) = \begin{cases} \text{Rank}(s_{pq}) - \omega \cdot U_{j,q} \cdot u_j / \sum_k U_{j,k}, & \text{若 } i = 1 \wedge p = 1, \\ \text{Rank}(s_{pq}) - \omega \cdot V_{j,q} \cdot u_j / \sum_k V_{j,k}, & \text{若 } i = 2 \wedge p = 2, \\ \text{Rank}(s_{pq}) + \theta \cdot W_{j,q} \cdot u_j / \sum_k W_{j,k}, & \text{若 } i = 1 \wedge p = 2, \\ \text{Rank}(s_{pq}) + \theta \cdot W_{q,j} \cdot v_j / \sum_k W_{k,j}, & \text{其他,} \end{cases}$$

其中,  $\omega \geq 0$  为话题内信息冗余惩罚值, 控制摘要中信息的冗余程度。  $\theta \geq 0$  为话题间对比提升值, 确保形成对比的句子可以同时出现在摘要之中, 提升摘要的对比性。本文中, 我们取  $\omega = 2$ ,  $\theta = 1$ 。

4) 若  $B = \emptyset$  或  $A$  的长度(字/词数)达到最大限制, 则输出  $A$  作为摘要。否则转至第 2 步。

## 4 实验

### 4.1 实验设置

由于对比新闻摘要任务的新颖性, 目前还没有公开的评测数据。因此, 我们自行构造了实验数据。我们选择 10 组可比较的新闻话题(5 组英文话题, 5 组中文话题, 如表 2 所示), 利用谷歌新闻搜索引擎<sup>①</sup>, 为每个话题检索 10 篇相关新闻。随后, 为每组话题人工撰写对比摘要。每篇对比摘要包含两个部分, 其中每部分关注于单个话题, 而两部分共同描述了两个话题之间的对比信息。撰写参考摘要时, 撰写者寻找两个话题中共同讨论的相关方面, 再从文中选择合适的句子片段, 进行必要的删减和组合, 使之更为紧凑。若可对比的方面较多, 则将对比内容按照重要程度排序, 选择重要的内容构成摘要。

实验中, 我们采用 ROUGE 作为评价指标。ROUGE 是一个在文档评测领域广泛应用的评测指标, 通过计算机生成的候选摘要和人工撰写的参考摘要之间的重叠程度评价自动摘要的质量<sup>[23]</sup>。在实验中, 我们采用 ROUGE 工具包<sup>②</sup>生成 ROUGE-2 和 ROUGE-SU4 的  $F$  值。此外, 为了评价对比摘要是否与每一个话题相关, 我们还分别计算对比摘要的两部分中的每一部分的 ROUGE 值, 并取两者的平均作为评价指标(记为 M-ROUGE-2 和 M-ROUGE-SU4)。

2) 将  $B$  中的句子按排序值降序排列。

3) 选择  $B$  中具有最高排序值的句子  $s_{ij}$ , 将其移至  $A$  中:  $A \leftarrow A \cup \{s_{ij}\}$ ,  $B \leftarrow B - \{s_{ij}\}$ 。  $B$  中剩余句子的排序值更新如下:

表 2 对比话题数据集

Table 2 Comparable topic pairs in the dataset

编号	话题 1	话题 2
E1	Haiti Earthquake	Chile Earthquake
E2	Chilean Mining Accident	New Zealand Mining Accident
E3	Iraq Withdrawal	Afghanistan Withdrawal
E4	Apple iPad 2	BlackBerry Playbook
E5	2006 FIFA World Cup	2010 FIFA World Cup
C1	汶川地震	玉树地震
C2	2009 年诺贝尔奖	2010 年诺贝尔奖
C3	2009 年奥斯卡奖	2010 年奥斯卡奖
C4	苹果 iPad 2	黑莓 Playbook
C5	2006 年世界杯	2010 年世界杯

### 4.2 实验结果

实验中我们选用了不同的方法作为参照。Manifold/M 方法将对摘要问题看做一个传统的摘要问题, 将两个文档集合混合起来, 采用 Wan<sup>[21]</sup>的方法抽取摘要。Manifold/S 方法则分别对两个文档集合抽取摘要, 每部分摘要长度为  $L/2$ ( $L$  为摘要长度)。CoRank/S 方法采用本文中提出的重要性计算方法, 但在抽取摘要句时不考虑跨集合之间的关系, 而分别从两个集合中抽取  $L/2$  长度的摘要。CoRank/U 方法为本文中提出的对比摘要算法。在英文数据集和中文数据集上的实验结果分别如表 3 和 4 所示。

从表中可以看到, Manifold/M 和 Manifold/S 方法要劣于 CoRank 方法。这是因为它们忽略了话题之间的对比关系, 而简单地选择代表性句子。CoRank/U 方法要优于 CoRank/S 方法, 因为在更新句子排序值时, CoRank/U 方法考虑了不同话题句子之间对比性, 提升具有高对比性的句子的权重, 从

① <http://news.google.com>。

② <http://www.berouge.com>。中文数据评价对工具包进行了修改, 以便识别中文词汇。

而确保结果中可对比的句子同时出现。表 5 和 6 给出了两个实际摘要结果。可以看到,摘要中描述了两个话题间的对比结果,具有较好的效果。

## 5 结论与展望

新闻话题的对比分析在现实中有大量应用需求。本文提出了一种概括可比新闻话题间共同点和

不同点的对比新闻摘要方法。我们利用同一个话题下句子之间的相似性和不同话题中句子之间的对比性,构造句子关系图,采用互相增强的方法计算每个句子的重要程度,抽取合适的句子组成对比摘要。实验结果显示我们的方法优于基准方法,具有良好的效果。

由于新闻话题和比较语义的复杂性,新闻对比

表 3 英文数据集上的评测结果

Table 3 Evaluation results on the English dataset

模型	ROUGE-2	ROUGE-SU4	M-ROUGE-2	M-ROUGE-SU4
Manifold/M	0.197	0.244	0.148	0.186
Manifold/S	0.204	0.238	0.166	0.193
CoRank/S	0.219	0.263	0.172	0.213
CoRank/U	0.251	0.292	0.182	0.219

表 4 中文数据集上的评测结果

Table 4 Evaluation results on the Chinese dataset

模型	ROUGE-2	ROUGE-SU4	M-ROUGE-2	M-ROUGE-SU4
Manifold/M	0.133	0.154	0.112	0.130
Manifold/S	0.128	0.154	0.108	0.131
CoRank/S	0.140	0.162	0.115	0.137
CoRank/U	0.157	0.179	0.142	0.161

表 5 “2006 年世界杯”和“2010 年世界杯”的对比摘要结果

Table 5 Comparative summary result of “World Cup 2006” vs. “World Cup 2010”

话题	World Cup 2006	World Cup 2010
摘要	Italy claimed a fourth world title in a penalty shoot-out victory over France after the two sides finished a goal apiece following extra-time in Berlin's Olympic Stadium on Sunday.	Spain have won the 2010 FIFA World Cup South Africa final, defeating Netherlands 1-0 with a wonderful goal from Andres Iniesta deep into extra-time.
	France captain Zinedine Zidane, sent off for head-butting Marco Materazzi late in Sunday's World Cup final loss to Italy, won the Golden Ball award for the tournament's best player.	Uruguay star striker Diego Forlan won the Golden Ball Award as he was named the best player of the tournament at the FIFA World Cup 2010 in South Africa.
	But has this World Cup been a success for football or just for the hosts, Germany?	SOUTH Africa and its people have indeed benefited from the on-going FIFA 2010 FIFA World Cup.
	An average of 52, 500 fans packed into the 12 stadiums for the 64 matches, making the tournament the second most-attended in the past 24 years after the 1994 tournament in the U.S. About 32 billion people tuned into the games, according to FIFA.	A global TV audience of more than 700 million watched Sunday's World Cup final, according to the tournament's organizers.

表 6 “汶川地震”和“玉树地震”的对比摘要结果

Table 6 Comparative summary result of “Wenchuan earthquake” vs. “Yushu earthquake”

话题	汶川地震	玉树地震
摘要	北京时间 5 月 12 日 14 时 28 分,位于北纬 31 度、东经 103.4 度的四川省汶川县发生里氏 7.8 级地震,造成重大人员伤亡和重大财产损失。	4 月 14 日 7 时 49 分,青海省玉树藏族自治州玉树县发生 7.1 级地震,给当地人民群众生命财产造成严重损失。另据中国地震网消息,在 7 时 49 分的 7.1 级地震前,还连续发生数次 4 级以上地震。
	地震发生后,9000 多名解放军和武警官兵紧急调动赶赴地震灾区参加救灾行动。中国地震局启动了一级预案,由 184 人组成的国家地震灾害紧急救援队和 40 人组成的国家地震灾害现场工作队已抵达成都,正连夜乘坐军用卡车奔赴汶川地震灾区。汶川县干部群众正奋力抗震抢险。	地震发生后,根据国务院的指示,中国地震局下达了派遣地震灾害紧急救援队伍的命令,启动地震应急救援协作联动机制,命令国家地震灾害紧急救援队和青海、西藏、甘肃、陕西、宁夏等地地震灾害紧急救援队和现场应急工作队队伍赶赴地震灾区开展抢险救援工作。
	国家民政部会同财政部向四川地震灾区紧急下拨 2 亿元中央自然灾害生活补助应急资金,帮助地震灾区群众解决生活困难。	玉树地震发生后,不少单位和个人纷纷伸出援手,为地震灾区群众捐款捐物。

摘要仍有大量问题有待研究。未来我们将引入更多的语义信息, 提高对比摘要质量。利用维基百科等语义资源, 计算概念之间的相关程度, 可以帮助挖掘更多的比较点, 同时也有助于计算句子与话题之间的语义联系。我们将进一步研究文本间对比程度的度量方法, 通过区分比较点和比较对象, 更准确地计算文本之间的对比性。我们将结合机器学习的方法, 自动调整摘要模型中的参数。此外, 我们将进一步扩展对比摘要任务, 包括研究多话题之间的对比、区分话题之间的共性和差异、研究跨语言新闻话题之间的对比等。

### 参考文献

- [1] Kennedy C. Comparatives: semantics. Concise Encyclopedia of Philosophy of Language and Linguistics. Oxford: Elsevier, 2009: 68–71
- [2] 马建忠. 马氏文通. 上海: 商务印书馆, 1898
- [3] Jindal N, Liu B. Identifying comparative sentences in text documents // Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2006: 244–251
- [4] Jindal N, LIU B. Mining comparative sentences and relations // Proceeding of the 21st National Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2006: 1331–1336
- [5] Kim H D, Zhai C. Generating comparative summaries of contradictory opinions in text // Proceedings of the 18th ACM conference on Information and knowledge management. New York: ACM, 2009: 385–394
- [6] Paul M J, Zhai C, Girju R. Summarizing contrastive viewpoints in opinionated text // Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2010: 66–76
- [7] Zhai C, Velivelli A, Yu B. A cross-collection mixture model for comparative text mining // Proceeding of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2004: 743–748
- [8] Wang D, Zhu S, Li T, et al. Comparative document summarization via discriminative sentence selection // Proceeding of the 18th ACM conference on Information and knowledge management. New York: ACM, 2009: 1963–1966
- [9] Liu B, Hu M, Cheng J. Opinion observer: analyzing and comparing opinions on the web // Proceeding of the 14th international conference on World Wide Web. New York: ACM, 2005: 342–351
- [10] Witte R, Bergler S. Next-generation summarization: contrastive, focused, and update summaries // International Conference on Recent Advances in Natural Language Processing (RANLP). Borovets, 2007. <http://rene-witte.net/next-generation-ummar-ation>
- [11] Lerman K, Mcdonald R. Contrastive summarization: an experiment with consumer reviews // Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2009: 113–116
- [12] Wan X, Jia H, Huang S, et al. Summarizing the differences in multilingual news // Proceedings of the 34th Annual ACM SIGIR Conference. New York: ACM, 2011: 735–744
- [13] Mani I. Automatic summarization. Amsterdam: John Benjamins Pub Co, 2001
- [14] Dang H T, Owczarzak K. Overview of the TAC 2008 update summarization task // Proceedings of the 1st Text Analysis Conference. Gaithersburg: National Institute of Standards and Technology, 2008: 12–28
- [15] Hovy E, Lin C Y. Automated text summarization and the SUMMARIST System // Proceedings of the TIPSTER Text Program Phase III. Stoudsburg: Association for Computational Linguistics, 1998: 197–214
- [16] Rau L F, Jacobs P S, Zernik U. Information extraction and text summarization using linguistic knowledge acquisition. Information Processing and Management, 1989, 25(4): 419–428
- [17] Nomoto T, Matsumoto Y. A new approach to unsupervised text summarization // Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval. New York: ACM, 2001: 26–34
- [18] Shen D, Sun J T, Li H, et al. Document summarization using conditional random fields // Proceedings of the 20th international joint conference on Artificial intelligence. San Francisco: Morgan Kaufmann Publishers Inc, 2007: 2862–2867

- [19] Gunes E, Radev D R. LexPageRank: prestige in multi-document text summarization // Proceedings of EMNLP 2004. Stroudsburg: Association for Computational Linguistics, 2004: 365–371
- [20] Mihalcea R. Graph-based ranking algorithms for sentence extraction, applied to text summarization // Proceedings of the ACL 2004 on Interactive poster and demonstration sessions. Stroudsburg: Association for Computational Linguistics, 2004: 170–173
- [21] Wan X, Yang J, Xiao J. Manifold-ranking based topic-focused multi-document summarization // Proceedings of the 20th international joint conference on Artificial intelligence. San Francisco: Morgan Kaufmann Publishers Inc, 2007: 2903–2908
- [22] Wan X, Yang J. Multi-document summarization using cluster-based link analysis // Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. New York: ACM, 2008: 299–306
- [23] Lin C Y, Hovy E. Automatic evaluation of summaries using N-gram co-occurrence statistics // Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology. Stroudsburg: Association for Computational Linguistics, 2003: 71–78