

# 面向查询的自动文本摘要技术研究综述

王凯祥

(中国人民大学信息资源管理学院 北京 100872)

**摘要** 对面向查询的自动文本摘要技术进行系统梳理,分析所用方法的基本思想、优缺点,并总结未来的发展方向。通过分析梳理,总结出了四大类面向查询的自动文本摘要技术:基于图模型的方法、基于机器学习的方法、基于聚类的方法和其他方法。在今后的研究过程中,基于神经网络和多模型融合的方法将成为未来研究的热点,在应用层面上,与实际应用场景相结合的算法研究将成为趋势。

**关键词** 文本摘要,图模型,流排序,主题模型,神经网络

**中图分类号** TP391,G35 **文献标识码** A

## Survey of Query-oriented Automatic Summarization Technology

WANG Kai-xiang

(School of Information Resource Management, Renmin University of China, Beijing 100872, China)

**Abstract** This paper systematically combed the query-oriented automatic summarization technology, analyzed the basic ideas, advantages and disadvantages of the methods used, and summarized the future development direction. By analyzing, four kinds of query-oriented automatic summarization were summarized: the method based on graph model, the method based on machine learning, the method based on clustering and other methods. In the future, the method based on neural network and multi model fusion will become the focus of future research. In the application level, it will become a trend to study the algorithm combining with the actual application scene.

**Keywords** Summarization, Graph model, Manifold ranking, Topic model, Neural network

## 1 引言

随着科学技术发展和进步、信息传播渠道的丰富多样、信息量的暴涨和获取方式的便捷,人们每天都会接收到大量的信息,从这些巨量的信息中找出自己所需的信息,就需要花费大量的时间和精力。自动文本摘要技术的出现,可以帮助人们节省大量的阅读时间,使其在相同的时间内能获取更多的有效信息。随着自动摘要技术的发展成熟,面向查询的摘要技术逐渐兴起。面向查询的自动文本摘要技术是自动摘要技术中的一个特定领域,可以对用户感兴趣的、主动查询的内容进行摘要,更加有针对性地满足用户的信息需求<sup>[1]</sup>。当前国内还较少针对面向查询的摘要技术进行研究,本文通过对国内外相关技术研究现状的系统梳理分析,总结面向查询的自动摘要的方法体系,分析未来的发展方向,为今后的相关研究提供理论参考。

笔者于 2017 年 11 月 20 日使用检索式“SU=(自动文摘+文本摘要+自动文本摘要)and SU=查询”在中国期刊全文数据库进行检索,使用检索式“TS=(summarization AND query)”在 arXiv, Web of Science, ACM, SIAM, ProQuest 进行检索,共检索到中文文献 13 篇、英文文献 291 篇,筛选与所研究题目相关的重要文献后最终得到中文文献 2 篇、英文文献 51 篇。笔者在精读这 53 篇文献的基础上,对面向查询的摘要技术进行了梳理。

## 2 面向查询的自动文本摘要定义及相关概念

文本摘要技术从 1958 Luhn HP<sup>[2]</sup>首次提出使用词频特征作为摘要提取的依据,至今已发展了 60 年。当前文本摘要技术从处理文本结构上分为:结构化和半结构化文本;从文本长度上分为:短文本和长文本;从文档数量上分为:单文档自动摘要和多文档自动摘要<sup>[3]</sup>。从方法上分为:生成式(abstractive)和摘取式(extractive)<sup>[4]</sup>。生成式需要在语义理解的基础上,在词语级别上生成摘要<sup>[5]</sup>。摘取式是通过分析文本统计特征、潜语义特征等,在句子或段落级别上生成摘要。生成式自动文摘在词语层面对句子进行压缩提炼,实现生成摘要的目的。当前主要是使用神经网络,首先对文章进行分词,然后对词语进行 Embedding 处理,通过 CNN 或 RNN 构成的 Encoding-Decoding 神经网络结构等自然语言生成技术<sup>[6-8]</sup>重新生成句子,进而组成摘要。同时,也有部分研究者使用针对特定领域的文章,使用固定模板的方式来生成摘要<sup>[9]</sup>。摘取式自动文本摘要技术的主要思想是:首先通过句子的各种特征给每个句子进行打分,得到句子的权重,然后计算每个句子与其他句子的相似度,根据计算得到的每个句子的得分和与其他句子的相似度进行句子选择和句子排序。选择句子时,选择句子权重最高的,同时满足新选入的句子与选入的句子相似度最低的条件,以保证冗余度最小、信息量最大<sup>[10]</sup>。

面向用户查询的自动文摘,指用户在文本集上给定查询语句,摘要系统在此基础上返回与该查询相关的摘要结果。所用到的技术与其他文摘方法类似,但所用方法的侧重点又各不相同。从检索到的 53 篇文献所使用的方法可以看出,当前面向查询的摘要方法主要使用的是机器学习、聚类和图模型的方法。下面将从这几大类方法的角度梳理面向查询的摘要技术。

### 3 面向查询的自动摘要技术

#### 3.1 基于图模型的方法

基于图模型的自动文摘方法主要是通过构建图网络,使用排序算法计算句子权重,常用算法有 PageRank<sup>[11]</sup> 和 HITS 等<sup>[12]</sup>。基于排序的方法是 Mani 等于 1997 年提出的<sup>[13]</sup>,主要思想是把所有句子组成一张有向图或无向图,以句子或词语为图节点,句子之间的相似度为节点之间边的权重,使用图排序的方法计算得到句子节点的权重,根据权重和相似度对句子进行筛选,组成摘要。其主要步骤是预处理、形成文本图、图排序、句子选择、形成摘要。由于在面向查询的自动摘要技术中有查询语句这一条限制,因此在进行图排序时通常会把查询语句所在节点的权值始终置为 1,以得到在查询影响下的句子权重<sup>[14]</sup>。

流排序的算法可以很容易地计算权重在图模型中的传播,所以流排序的算法已经被大量使用在面向查询的图模型方法中<sup>[15]</sup>。当在原有图模型中加入一个查询语句的节点后,使用流排序算法可以计算查询节点的权重在图中传播后各个节点的权重。Cai 等<sup>[16]</sup>在流排序的基础上增加了主题层的排序,首先对文档句子使用聚类算法进行聚类,得到主题层,然后通过流排序算法在句子层和主题层计算查询语句权重的传播结果,得到句子和主题的排序结果。Badrinath 等<sup>[17]</sup>针对单文档的查询摘要提出了一种改进的 LexRank 算法,首先找到包含查询词语的这类直接相关的句子,然后运用 Look-Ahead Strategy,使用随机步长的方法,从直接相关的句子节点处寻找不含查询词但与查询语句间接相关的句子,最后再使用图排序算法计算句子相似度和权重,以减小图大小和运算量。Pandit 等<sup>[14]</sup>使用两个模型对句子进行排序,首先使用 off-line 模型,将段落作为节点,计算节点间的相似度(基于 TFIDF)和节点权重,并根据节点相似度对段落进行分类。在 on-line 模型中输入查询语句时,构建包含查询关键词的子树,计算查询语句与类的相似度以及与节点的相似度,并根据相似度对类和类内节点进行排序。

在进行多文档的自动摘要计算过程中,使用图排序的方法会使图的维度十分巨大,所以在构建图网络时,通过构建多层图模型结构,可以对整个文档集的信息进行分类整合,在充分利用文档信息的基础上又能降低计算的复杂度。Wei 等<sup>[18]</sup>提出了将整个图分成两层,分别是句子层和文档-句子层,在句子排序时,句子层包含了所在文档的相关信息,句子-文档层包含句子和文档的共同信息,从而在句子排序的过程中既方便计算句子之间的相似度,又能使用到文档间和句子间的关联信息。Canhasi 等<sup>[19]</sup>使用了文档、句子、主题的三层网络图模型,通过计算查询语句与句子的相似度,对句子矩阵进行奇异值分解。随后,Canhasi<sup>[20]</sup>基于 pagerank 构建了在句子、查询句、段落、文档、框架 5 个层面的图模型,通过计算

各层之间、每层之内的相似度,进一步改善了模型效果。Sakamoto 等<sup>[21]</sup>构建了句子、文档、词语的三层图模型,在词语层使用语义词典计算词语之间的相似度,在句子层通过两个句子之间的最小重复单元定义句子之间的相似度,对词语层和句子层使用 Co-HITS 算法对每个句子进行排序,在排序过程中保持查询语句所在节点的权重始终为 1。

对于多文档图模型的构建,由于句子数量较多,在计算句子之间相似度的过程中复杂度过高,普通的图模型一条边只能连接两个句子,这样会损失掉大量的文档信息,而超图模型不仅可以降低复杂度,而且由于其一条边可以连接多个节点,可以使传统的图模型结构连接超过两个句子,从而可以整合整篇文章的信息。Wang 等<sup>[22]</sup>通过 DBSCAN 的方法进行聚类,构建超图,如果两个句子的余弦相似度超过某个阈值,则添加一条边,如果两个句子在同一个类中,则添加一条边。根据节点之间的相似度和查询语句节点的常量权重,计算句子权重。Silva 等<sup>[23]</sup>在此基础上使用改进的 k 中心聚类方法代替 DBSCAN,根据词语的 TFIDF 值、句子之间的相似度以及与查询语句的相关性,选择最高的 k 个句子作为中心点进行聚类,聚类之后根据句子之间的相似度大小和类之间的相似度构建句子图网络,利用图排序算法对句子进行排序、筛选,形成摘要。Xiong 等<sup>[24]</sup>结合主题模型获得主题分布,使用超图模型获得词与主题、句子与句子的主题分布,通过节点增强和随机游走模型对句子进行排序。Zheng 等<sup>[25]</sup>同样构建了两层超图图结构,但作者使用的是概念层,首先从句子中提取出概念,构建概念与句子、概念与查询的二分图,然后基于这些概念构建句子的超图模型,通过句子所包含的概念的多少和权重大小决定句子得分。

由此可见,使用图模型的方法进行摘要的方法主要可以分为 3 类:构建多层图网络、基础超图模型、其他基于基础图模型的改进。在通过图模型计算句子权重的过程中,普遍使用的是将查询语句所在节点权重置为 1 的计算方法。使用图模型的方法可以使用整个文档集的全局信息,但这必然会导致其复杂度呈指数级上升。

#### 3.2 使用聚类方法的自动文摘方法

该方法主要是利用主题模型或聚类的相关工具,根据计算得到句子的相似度和权重,经过聚类后对句子进行选择 and 排序。聚类前的准备过程主要是计算句子之间的相似度,常用方法有 TF-IDF、语义特征、构建图模型等方法;在聚类之后类内句子的压缩过程中使用的主要方法包括直接选择、使用 pagerank 进行压缩、基于特征选择等方法<sup>[26]</sup>。主题模型的方法加入了语义关联信息,考虑到主题知识信息,可以使含有总结性词语的句子的权重变高,摘要更准确。查询语句的信息主要被用于计算句子相似度、计算类与查询相似度、句子压缩等步骤<sup>[27-28]</sup>。

在根据句子或词语之间的相似度对句子进行聚类时,会加入语义信息和多种特征,以提高相似度计算的准确率和聚类效果。徐晓丹<sup>[29]</sup>使用同义词林进行聚类,得到子主题,根据查询语句与子主题的相似度进行排序,从每个聚类中选择一个得分最高的句子组成摘要。为了提高聚类的准确性, Yang 等<sup>[30]</sup>提取句子的相关性、句子信息量、新颖度等特征加入到聚类模型中。Naveen 等<sup>[31]</sup>改进了聚类方法,使用 PHA-ClusteringGain 与 K-Means 结合的方法进行聚类。

在使用聚类模型时,也会结合其他模型,以提升摘要效果。

### (1)与主题模型的结合

Luo等<sup>[32]</sup>使用PLSA(Probabilistic Latent Semantic Analysis)和PHITS(Probabilistic Hyperlink-Induced Topic Search),从相关性、覆盖度、新颖度3个方面构建了PRCN的面向查询的主题概率模型框架。在相关性方面,通过主题模型计算得到句子之间的相似度,以及句子与查询语句之间的相似度;在覆盖度方面,通过对主题的排序和句子的选择,保证摘要的结果对各个主题内容的覆盖率达到最大;在新颖度方面,通过MMR(Maximal Marginal Relevance)的方式结合贪心算法来保证内容的新颖度。Li等<sup>[33]</sup>结合聚类方法和图模型算法,首先根据LDA得到的主题分布和句子分布构建图模型,并将查询语句所在节点的权重设为常量,然后在图模型中不断迭代计算节点权重。Yang<sup>[34]</sup>基于hLDA(hierarchical Latent Dirichlet Allocation model),结合n-gram模型,提出了一种考虑上下文关系的主题模型,该方法不仅考虑到了不同层次之间的主题分布,而且考虑了相同层次中的词语相关性。

### (2)与图模型的结合

Sondur<sup>[35]</sup>首先根据句子的词频、位置、线索词、句子长度等特征对句子打分,得到句子权重;然后通过流排序的方法得到句子与查询的相似度权重。作者认为一个聚类结果中包含的与查询相关的句子越多,该聚类的权重就越高。通过前两步得到的结果,计算得到聚类权重的大小,最后从每个聚类中选出权重高的句子作为摘要。Shafieibavan等<sup>[36]</sup>同样使用聚类和图模型的结构提出了一种生成式的摘要方法,使用ADW比较句子的相似度,构建无向图,对与查询语句的相似度小于阈值的边进行剪枝操作;然后使用Chinese Whispers对剪枝后的网络进行聚类,在每个类中构建词语级别的MSC(Multi-Sentence Compression)网络,使用pagerank算法得到词语权重;最后根据词语的词性、语法规则、权重选择词语组成摘要。Sun等<sup>[37]</sup>针对聚类之后的每个类内句子的压缩过程,提出了一种新的方法,对每个聚类采用多句子压缩的方法,首先构建句子图和词语图,找到k条最短路径,然后使用定义好的子模块函数选出最优的路径,路径所通过的节点句子即组成摘要,其中子模块函数是根据句子的主题覆盖度、与查询的相关性、摘要的多样性等原则设计的。

基于聚类的方法,是根据文本单元的特征进行聚类的。通过改进聚类方法、增加聚类特征、提高相似度计算准确度,可以增强聚类效果;同时,与其他模型相结合,能提高该方法的扩展性、灵活性、准确性<sup>[37]</sup>。

## 3.3 基于机器学习方法的自动文摘方法

基于机器学习方法的自动摘要方法主要是使用有监督学习的方法,通过分析提取特征项,加入模型进行训练,将问题转化为二分类问题,所以进行训练时就需要大规模地标记语料库。基于机器学习的自动摘要主要包括以下几个步骤:特征选取、算法选择、模型训练、文摘提取和模型评测等。特征选择主要是提取文本的统计特征和语法语义特征,其中统计特征主要有:词频、句子位置、句子长度、包含关键词的多少及权重;语法语义特征主要有:句法结构、修辞、实体词关联度等<sup>[38]</sup>。

Li等<sup>[39]</sup>使用贝叶斯主题模型进行有监督训练,在提取的特征中加入句子与查询语句的相似度,但由于特征数量较少,

效果提升并不显著。Ouyang等<sup>[40]</sup>通过人工摘要计算得到每个句子的重要度,并以此为标签,然后使用SVR的有监督学习的方法进行训练。Shen等<sup>[10]</sup>同样使用人工摘要计算得到句子标签,作者使用基于图模型的方法得到文档句子与人工摘要句子的相似度,并将其作为标签,然后通过提取句子位置、长度、与查询句子的相似度等特征,使用SVM进行训练;同时作者还改进了损失函数,添加了冗余的后处理,使得结果更加准确。

由于有监督机器学习的方法需要大量的标记数据,而人工标注的成本高,同时不同的标记员的标记结果可能会相差很大,导致标记数据含有大量噪声。Azar等<sup>[41]</sup>在训练过程中通过添加噪声改进效果,将查询语句和文章句子一起使用TFIDF编码后加入随机噪声,放入自编码器进行训练。Feigenblat等<sup>[42]</sup>提出了一种无监督的训练方法,为了实现无监督的训练,作者通过提取相关性、多样性、长度、位置、显著性5种特征,以及迭代优化目标函数,计算句子权重,取得了当前的最好结果。

Valizadeh等<sup>[43]</sup>使用多模型融合的方法改进摘要效果,在特征提取方面,抽取了更加多样和丰富的特征;在模型训练过程中,对于给出的多份人工摘要,考虑到每个人的认知和行为偏好各不相同,作者在使用标记语料时保留了每个人的摘要特点,给每份人工摘要建立一个模型,在测试集上生成摘要时,先经过多个模型给出摘要后,再对这些产生的摘要进行主谓分析和MMR去除冗余,从而形成最终的摘要。

与此同时,使用神经网络方法的生成式摘要也逐渐发展起来。Liu等<sup>[44]</sup>认为深度学习的各个隐藏层可以表示文本的复杂结构,作者基于受限玻尔兹曼机,将网络结构分成3个部分,即内容过滤、结构重组和摘要生成。内容过滤时主要过滤非关键词,提取重要词汇,选择候选句。结构重组时对网络结构进行剪枝操作,优化网络结构。最后,通过动态规划的方法从候选句中选出摘要。Rush<sup>[6]</sup>提出一种在RNN,Encoding-Decoding结构的基础上加入Attention机制的生成式摘要方法。该方法首先对分词后的词语进行Embedding处理,将Embedding之后的词语依次输入到编码器中,将整篇文章压缩成了一个向量,再经过解码器,将该向量解码成词语,解码生成的词语就组成了我们所需的摘要,其中编码器和解码器的结构均使用RNN结构。但由于该结构是把整篇文章压缩成一个向量,对于文章这类长文本,在压缩过程中会产生大量的信息损失,所以作者在解码过程中加入了Attention机制,该模型可以在解码器的解码过程中只关注与要解码词语语义关系较近的词语,从而提高解码的准确度。Nema等<sup>[45]</sup>和Johan将此方法用于面向查询的自动文摘中<sup>[7]</sup>。

基于机器学习的方法由于使用了大量的标记语料,可以提取丰富的特征用于模型训练,但同时其仍有许多不足之处:需要大量的人工标记语料,成本较高;由人工摘要转为标签时会产生信息损耗<sup>[46]</sup>;不同人对文章集的理解不尽相同,各个人工摘要的内容侧重点也不同。深度学习的方法给这些问题的解决带来了契机,虽然评测结果较好,但句子不通顺、指代不明等问题仍比较普遍<sup>[7]</sup>。

## 3.4 其他方法

除了基于图模型、聚类、机器学习的方法,还有一些其他方法应用于面向查询的自动摘要中。

从文本特征处理提取的角度, Song 等<sup>[47]</sup>提出了一种增加复合查询的方法。当给定一个查询时,通过复合查询的方法获得与源查询相关的一些查询语句,这样可以获得更多、更全面的关于查询的信息,然后再通过句子的特质性、信息量、冗余度等特征进行句子选择,组成摘要。Wang 等<sup>[48]</sup>为了提高句子选择的准确性,通过多种子模块对句子进行打分筛选,包括计算句子权重时的词频特征、查询相关特征、语义特征、查询独立的特征、覆盖度特征(主题覆盖度、命名实体覆盖度、内容覆盖度)以及新颖性特征(主题多样性、词汇多样性、语义多样性)等,经过这些模块的计算后,综合筛选句子。Satyananda 等<sup>[49]</sup>在基于统计特征的方法上加入了语义特征,在计算词语之间的相似度时,根据词语之间的词频、 $n$ 元组共现次数、词语间相对次序,同时结合语义词典计算相似度,最后再根据相似度计算得分并排序。

从算法改进的角度, Sowmya 等<sup>[50]</sup>将摘要问题转化为约束问题,使用动态规划的方法进行解决。首先基于句子与主题的相似度、词项 TFIDF、位置、内容覆盖度等特征,构建句子图,使用动态规划的方法寻找在规定长度下权重值最大、冗余度最小即通过节点的权值综合最大、通过边的权重最小约束下的最优解。

其他一些方法主要是从文本特征的处理和提取角度进行改进,或者将一些新的方法应用于自动摘要中,这些方法的提出和改进丰富了文本摘要领域的技术方法,同时对其他模型方法的优化具有一定的参考价值。

**结束语** 当前面向查询的自动文本摘要的研究已有较多的成果,但目前国内的相关研究还比较少。从每年的摘要效果上来看,效果最优的方法呈现不断迭代更新的趋势,从 2013 年的图模型、2016 年的超图模型,到 2017 年的机器学习、深度学习,新模型方法的提出对摘要效果的提升十分显著,在原有方法的基础上,研究者们也不断进行改进以期提升效果。

笔者认为面向查询的自动摘要技术在今后主要有 3 个发展趋势:1)基于深度学习的摘要技术研究将成为热点。当前深度学习的各项技术飞速发展,标记语料不断增加,给其广泛应用提供了基础条件。基于深度学习的方法可以更加全面地把握全文信息,更好地表示语义信息,在自动摘要领域的应用将不断完善,效果也会逐渐提升。2)多模型融合的方式,将为自动摘要的发展提供更多思路。通过多模型融合,充分发挥各个模型的优势,提升摘要的整体效果。3)在应用层面,与真实场景相结合的技术改进方法,也将成为研究的趋势<sup>[51]</sup>。面向查询的自动文本摘要,主要是为了满足用户的个性化需求,所以结合网站、微博等搜索引擎的结果进行自动摘要,会有更大的实用价值。

## 参 考 文 献

- [1] CHEN Z M, YI J, ZHAO Y. Automatic summarization based on user's query expansion[J]. *Application Research of Computers*, 2011, 28(6): 2188-2190.
- [2] LUHN H P. The automatic creation of literature abstracts[M]. IBM Corp., 1958.
- [3] BHASKAR P, BANDYOPADHYAY S. A Query Focused Multi Document Automatic Summarization[C]// The 24th Pacific Asia Conference on Language, Information and Computation. Tohoku University, Sendai, 2010.
- [4] 康世泽, 马宏, 黄瑞阳. 一种基于神经网络模型句子排序方法[J]. *中文信息学报*, 2016, 30(5): 195-202.
- [5] SUN R, WANG Z, REN Y, et al. Query-Biased Multi-document Abstractive Summarization via Submodular Maximization Using Event Guidance[M]. Springer International Publishing, 2016.
- [6] RUSH A M, CHOPRA S, WESTON J. A Neural Attention Model for Abstractive Sentence Summarization[J]. arXiv:1509.00685, 2015.
- [7] YING W, XIAO X, LI S, et al. Improving Query-Focused Summarization with CNN-Based Similarity[J]. *Acta Scientiarum Naturalium University Pekinensis*, 2017, 53(2): 197-203.
- [8] WADHVANI R, KUMAR R, GYANCHANDANI M, et al. RIN-Sum: A System for Query-Specific Multi-Document Extractive Summarization[J]. *International Journal of Advanced Computer Science & Applications*, 2017, 8(3): 106-112.
- [9] QUMSIYEH R, NG Y. Web Search Using Summarization on Clustered Web Documents Retrieved by User Queries[C]// 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). 2015: 401-404.
- [10] SHEN C, LI T. Learning to Rank for Query-Focused Multi-document Summarization[C]// IEEE International Conference on Data Mining. 2012.
- [11] 何新宇. 基于中心词耦合度和 PageRank 的文本自动摘录算法[J]. *中山大学研究生学刊(社会科学版)*, 2013(3): 43-50.
- [12] MIHALCEA R. Graph-based ranking algorithms for sentence extraction, applied to text summarization[C]// ACL 2004 on Interactive Poster and Demonstration Sessions. 2004.
- [13] MANI I, BLOEDORN E. Multi-document summarization by graph search and matching[C]// Fourteenth National Conference on Artificial Intelligence and Ninth Conference on Innovative Applications of Artificial Intelligence. 1997.
- [14] PANDIT S R, POTEY M A. A Query Specific Graph Based Approach to Multi-document Text Summarization: Simultaneous Cluster and Sentence Ranking[C]// International Conference on Machine Intelligence and Research Advancement. 2014.
- [15] LEI K, ZENG Y F. A Novel Biased Diversity Ranking Model for Query-Oriented Multi-Document Summarization[J]. *Applied Mechanics & Materials*, 2013, 380-384: 2811-2816.
- [16] CAI X, LI W. Mutually Reinforced Manifold-Ranking Based Relevance Propagation Model for Query-Focused Multi-Document Summarization[J]. *IEEE Transactions on Audio Speech & Language Processing*, 2012, 20(5): 1597-1607.
- [17] BADRINATH R, VENKATASUBRAMANIYAN S, MADHAVAN C E V. Improving Query Focused Summarization Using Look-Ahead Strategy[C]// European Conference on Advances in Information Retrieval. Springer-Verlag, 2011: 641-652.
- [18] WEI F, LI W, HE Y. Document-Aware Graph Models for Query-oriented Multi-document Summarization[M]. Springer Berlin Heidelberg, 2011.
- [19] CANHASI E, KONONENKO I. Weighted archetypal analysis of the multi-element graph for query-focused multi-document summarization[J]. *Expert Systems with Applications*, 2014, 41(2): 535-543.
- [20] CANHASI E. Query Focused Multi-document Summarization

- Based on Five-Layered Graph and Universal Paraphrastic Embeddings[C]// Computer Science On-line Conference. Springer, Cham, 2017:220-228.
- [21] SAKAMOTO K, SHIBUKI H, MORI T, et al. Fusion of heterogeneous information in graph-based ranking for query-biased summarization[C]// BSG@SIGIR. 20015:19-22.
- [22] WANG W, WEI F, LI W, et al. HyperSum: hypergraph based semi-supervised sentence ranking for query-oriented summarization[C]// ACM Conference on Information and Knowledge Management. ACM, 2009: 1855-1858.
- [23] D SILVA S, JOSHI N, RAO S, et al. Improved Algorithms for Document Classification & Query-based Multi-Document Summarization[J]. International Journal of Engineering and Technology, 2012, 3(4): 404-409.
- [24] XIONG S, JI D. Query-focused multi-document summarization using hypergraph-based ranking[J]. Information Processing & Management, 2016, 52(4): 670-681.
- [25] ZHENG H T, GUO J M, JIANG Y, et al. Query-Focused Multi-document Summarization Based on Concept Importance[C]// Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer International Publishing, 2016: 443-453.
- [26] LI C G. Query-Focused Multi-Document Summarization Based on Dominant Sets Cluster[J]. Computer Systems & Applications, 2014, 3(1): 50-55.
- [27] SURESHRAO D Y, SUBHASH S B, DASHORE P. Analysis of Query Dependent Summarization Using Clustering Techniques [J]. International Journal of Computer Technology & Electronics Engineering, 2012, 2(1): 213-217.
- [28] PATIL G V, MRUNAL B. Query Dependant Single Document Summarization using Partitional Clustering; K-Means Clustering Approach[J]. International Journal of Computer Science Engineering & Technolo, 2011, 1(5): 191.
- [29] 徐晓丹. 基于子主题和用户查询的多文档摘要系统[J]. 计算机系统应用, 2011, 20(3): 112-115.
- [30] YANG L, CAI X. Semi-Supervised Co-Clustering for Query-Oriented Theme-based Summarization[J]. Research Journal of Applied Sciences Engineering & Technology, 2012, 4(18): 3410-3414.
- [31] NAVEEN G K R, NEDUNGADI P. Query-based Multi-Document Summarization by Clustering of Documents[C]// International Conference on Interdisciplinary Advances in Applied Computing. 2014.
- [32] LUO W, ZHUANG F, HE Q, et al. Exploiting relevance, coverage, and novelty for query-focused multi-document summarization[J]. Knowledge-Based Systems, 2013, 46(1): 33-42.
- [33] LI Y, LI S. Query-focused multi-document summarization: Combining a topic model with graph-based semi-supervised learning [C]// The 25th International Conference on Computational Linguistics; Technical Papers. 2014: 1197-1207.
- [34] YANG G. A Novel Contextual Topic Model for Query-Focused Multi-document Summarization[C]// IEEE International Conference on TOOLS with Artificial Intelligence. 2014.
- [35] SONDUR P P B S. Multi-Document Text Summarization using Mutual Reinforcement and Relevance Propagation Models Added with Query and Features Profile[J]. International Journal of Advanced Computer Research, 2013, 3(3): 59.
- [36] SHAFIEIBAVANI E, EBRAHIMI M, WONG R, et al. A Query-Based Summarization Service from Multiple News Sources [C]// IEEE International Conference on Services Computing. 2016.
- [37] LI J, LI S. Query-focused Multi-document Summarization; Combining a Novel Topic Model with Graph-based Semi-supervised Learning[J]. arXiv: 1212. 2036, 2012.
- [38] KATIYAR S, BORGOHAIN S. Lexical Similarity Based Query Focused Summarization Using Artificial Immune Systems[M]. Springer International Publishing, 2015.
- [39] LI J, LI S. A Novel Feature-based Bayesian Model for Query Focused Multi-document Summarization[J]. Transactions of the Association for Computational Linguistics, 2014, 1: 89-98.
- [40] OUYANG Y, LI W, LI S, et al. Applying regression models to query-focused multi-document summarization [J]. Information Processing & Management An International Journal, 2011, 47(2): 227-237.
- [41] AZAR M Y, SIRTS K, ALIOD D M, et al. Query-Based Single Document Summarization Using an Ensemble Noisy Auto-Encoder [C]// Australasian Language Technology Association Workshop. 2015.
- [42] FEIGENBLAT G, ROITMAN H, BONI O, et al. Unsupervised Query-Focused Multi-Document Summarization using the Cross Entropy Method [C]// The International ACM SIGIR Conference. 2017.
- [43] VALIZADEH M, BRAZDIL P. Exploring actor-object relationships for query-focused multi-document summarization[J]. Soft Computing, 2014, 19(11): 1-13.
- [44] LIU Y, ZHONG S H, LI W. Query-Oriented Multi-Document Summarization via Unsupervised Deep Learning[J]. Expert Systems with Applications, 2012, 2(21): 1699-1705.
- [45] NEMA P, KHAPRA M, LAHA A, et al. Diversity driven Attention Model for Query-based Abstractive Summarization[J]. arXiv: 1704. 08300, 2017.
- [46] 森田一. Enhanced Models for Query-Oriented Extractive Summarization[J]. Journal of the Japanese Society for Artificial Intelligence, 2015, 30(1): 125.
- [47] SONG W, YU Q, XU Z, et al. Multi-aspect query summarization by composite query[C]// International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2012: 325-334.
- [48] WANG L, RAGHAVAN H, CARDIE C, et al. Query-Focused Opinion Summarization for User-Generated Content[J]. arXiv: 1606. 05702, 2016.
- [49] M R V V, Y P K S, SATYANANDA REDDY C. A Hybrid Method for Query based Automatic Summarization System[J]. International Journal of Computer Applications, 2014, 68(6): 39-43.
- [50] SOWMYA K S. Query-oriented Unsupervised Multi-document Summarization on Big Data[C]// International Conference on Computing Communication and Networking Technologies. 2016.
- [51] SHAFIEIBAVANI E, EBRAHIMI M, WONG R, et al. A Query-Based Summarization Service from Multiple News Sources [C]// IEEE International Conference on Services Computing. 2016.