

● 宋培彦<sup>1</sup>, 田文波<sup>1</sup>, 冯超慧<sup>1</sup>, 吴柯莹<sup>1,2</sup>

(1. 天津师范大学管理学院, 天津 300387; 2. 吉林大学商学与管理学院, 吉林 长春 130012)

## 融合选择性注意衰减模型的信息简报自动生成方法研究<sup>\*</sup> ——以 UNESCO 科技报告为例

**摘要:** [目的/意义] 简报是重要的情报产品, 摘要和关键词汇集是简报的核心组成部分, 从用户认知角度研究简报生成机制、形成认知与计算的有机结合, 有助于更好地满足用户对专业知识的需求, 提高情报供给服务能力。[方法/过程] 以认知科学中的“选择性注意衰减”理论模型为基础, 将信息简报的生成作为人类信息加工的模拟过程, 首先, 以选择性注意中的“衰减器”模型为理论支撑, 从摘要、关键词、简报三个层级进行一体化设计; 其次, 采用 KeyBERT 和 Transformer 算法, 对联合国教科文组织 UNESCO 科技报告进行关键词抽取和摘要生成, 形成具有一定实用价值的简报生成技术; 最后, 采用信息熵和 ROUGE 值对生成结果进行评价。[结果/结论] 实验结果通过信息熵、ROUGE-2、ROUGE-L 等值进行对比评价, 信息熵值均在 6~8 之间, ROUGE 值满足了摘要可用性水平, 表明选择性注意衰减模型不仅能够提高摘要效果, 覆盖文本的核心信息, 同时该方法自动生成摘要结果较为符合人的认知习惯, 有助于形成可计算、可解释的信息简报生成与知识服务模式。

**关键词:** 选择性注意; 自动摘要; 科技报告; 信息简报; 知识组织

**DOI:** 10.16353/j.cnki.1000-7490.2023.04.007

**引用格式:** 宋培彦, 田文波, 冯超慧, 吴柯莹. 融合选择性注意衰减模型的信息简报自动生成方法研究——以 Unesco 科技报告为例 [J]. 情报理论与实践, 2023, 46 (4): 53-58, 7.

### A Method of Automatic Information Briefing Generation Based on Selective Attention Decay Model: Taking the UNESCO Reports for Example

**Abstract:** [Purpose/significance] Information brief is an important intelligence product in practice, in which abstract and keywords are the core components. Based on cognition computing model, it is beneficial for users to obtain professional knowledge and intelligent services. [Method/process] Based on the “selective attention decay theory” in cognitive science, the generation of information briefs was suggested as the simulation process of human information processing. In this paper, the feasibility of “decay model” in selective attention was analysed firstly in theory, and then the integrated architecture was constructed for keywords extraction, automatic abstracts and briefs generation synergistically. Then, KeyBERT and Transformer algorithms were adapted to extract keywords and generate abstracts from UNESCO’s scientific and technical reports. Finally the entropy and ROUGE indexes were validated to evaluate the quality of generated briefs. [Result/conclusion] The experiment indicates that Information Entropy, ROUGE-2, ROUGE-L are acceptable. The information entropy is 6~8, and the ROUGE index reached the level of usability, which indicates that the selective attention decay model can not only cover the core information of the texts effectively, but also conform to cognitive habits of human. It is a novel potential mechanism for computable and interpretable information briefing generation.

**Keywords:** selective attention; automatic summarization; scientific and technical reports; information briefings; knowledge organization

## 0 引言

通过对科技报告等海量文献信息进行自动标注和摘

要, 并以专题简报形式推送给有关机构或用户, 有利于帮助用户把握前沿热点、及时掌握动态、实现科学决策, 是情报学研究和业务工作的重要方面。信息简报主要涵盖标题、摘要和关键词 3 个部分, 如何模拟人类信息加工机制和过程, 从文本中自动实现关键词提取和文本摘要, 对认知科学和情报技术都具有重要的科学价值。同时, 联合国

<sup>\*</sup> 本文为全国科技名词委 2020 年科研项目“国际组织术语库集成方法研究”的成果, 项目编号: YB20200011。

教科文组织（UNESCO）等机构发布的各类专业报告权威性高、更新速度快，单纯依靠人工编制简报准确性较高但速度偏慢，难以适应大数据环境下的科技情报工作要求，从实践上也需要研制信息简报自动生成技术，提高情报服务能力。因此，开展科技信息简报生成具有重要研究价值和现实意义。

信息简报本质上是人类对信息再加工的过程，因而信息简报自动生成应该建立在人类信息加工规律基础上。从认知科学角度看，人在阅读文本时，大脑往往是在特定时间特定环境下获取有限的键信息，因此通常会将有限的处理能力分配在焦点信息上，即选择性注意机制<sup>[1]</sup>由此实现从大量信息中快速而且精准筛选出高价值信息、过滤低价值信息，以“省力原则”达到高效的信息处理，这是人类在长期进化中形成的一种认知能力。情报学界也通过眼动实验等证实了人类信息行为也符合这一原则<sup>[2]</sup>。因此，本文以选择性注意模型为理论基础，设计了信息简报生成模型和技术流程，采用 Transformer 计算框架研究信息简报生成技术，最后以 UNESCO 发布的报告为例进行实验验证，为大数据环境下信息简报生成与情报服务提供理论和技术支持。

## 1 相关研究

信息简报的自动生成不仅涉及关键词抽取和文本摘要技术，而且认知科学的相关研究也逐渐增多，将情报学方法与认知理论相结合，对提高信息简报生成质量具有较强的科学性和可行性。

### 1.1 关键词抽取

关键词是信息简报的必要组成部分，能够覆盖和解释文本的核心主题。基于词频的关键词提取简单易行，如高劲松等为了追踪科研热点以及其时间变化趋势，将科研热点进行词频统计并可视化<sup>[3]</sup>；基于图的方法诸如 Textrank 等，它是根据文本词句之间的链接关系进行权重判断，认为一个词句所拥有的有向链接越多就越重要，如张建娥以 Textrank 为基础并融合了其他特征进行关键词提取，提高了结果的召回率<sup>[4]</sup>。基于机器学习的方法是当前关键词抽取的热点，通过在算法的基础上增加语料训练能够提高结果的精准度，如毛立琦等将无监督文本作为数据源，通过交叉熵关键词提取方法的跨领域迁移，完成提取任务<sup>[5]</sup>。综上所述，现有的关键词抽取方法研究相对成熟，但主要是基于概率统计模型，如何使关键词抽取更符合人类的认知特点仍有待探究。

### 1.2 文本摘要

文本摘要分为抽取式摘要和生成式摘要，抽取式摘要基本做法是在原文的基础上，抽取能够表征文本主题的关

键句子形成摘要，重点主要集中在句子重要性的判断、筛选以及排序等<sup>[6]</sup>。章成志等设计了基于细粒度的评论挖掘的书评摘要方法，为图书信息提供多维度、细粒度的评价<sup>[7]</sup>；王晓宇对传统的基于图（Graph-based）的文本图构建成语义图，并在此基础上提出了融合关键词位置信息、概念层级和连接强度等词权重计算方法，进行重要性排序，选择高分节点作为关键词集合，在文本摘要上有一定的提高<sup>[8]</sup>。生成式文本摘要是机器在理解原文的基础上，根据原文核心思想与含义生成摘要。万小军等提出了一种基于模块优化的摘要算法，通过文本摘要解决了主题标注的问题<sup>[9]</sup>。抽取式摘要成熟简单，是情报学领域的主流手段，但易读性和衔接性和生成式摘要有一定差距。而生成式摘要通过深度学习和成熟的语料库，在特定领域内效果显著。总体而言，目前文本摘要方法取得了不错的效果，但文本摘要的过程本质上是模仿人类大脑处理信息的过程，因此大脑深层次认知机制不容忽视。将摘要的过程结合认知学习，能够对文本更深层次的语义关联性和重要性进行判别和过滤，形成更加优质的摘要结果。

### 1.3 认知与计算模型

认知模型与计算技术紧密结合对提高情报技术能力的影响日益显著，将认知计算与文本挖掘相结合，取得了较为理想的效果。曾子明等构建了基于用户注意力机制的 U-BiLSTM 情感分析模型，用于分析情感演化过程，具有较强的可解释性和准确性，F1 值和准确率均有提高<sup>[10]</sup>；周琰等结合了长短时记忆（LSTM）和注意力机制，以“华为 P10”闪存门事件为例，表明基于选择性注意力机制的情感分析能够提高情感分类的成功率，能够较为准确地提取情感特征<sup>[11]</sup>；胡吉明等利用 Text-rank 算法和 CNN-BiLSTM-Attention 集成模型对政策文本进行分类处理，提升了分类的效率和准确度<sup>[12]</sup>。认知模型与计算方法的结合，不仅能够带来更为显著的算法技术效能提升，而且借助认知科学严密的逻辑体系，有助于形成科学意义上的情报学价值阐释，推动情报学从以数据为中心的概率或然性走向以人为中心的情报理性性。

总体而言，简报生成的关键技术是关键词提取和文本摘要，学术界对此已有丰富成果，同时，简报生成本质上是对人类认知信息的模拟过程，认知模型对于简报生成具有重要的指导意义，但当前计算模型与用户认知机理的结合研究仍然偏少，对简报的解释力和计算模式有待改进，因而有必要从认知角度对信息简报的内在机制和计算框架进行审视，以更好地从计算机算法和情报用户需求两个角度提高简报产品的接受度。基于此，本文针对信息简报生成中计算模型与认知机理结合不足的问题，在总结、吸收前人研究取得成果基础上，重点从认知科学“注意力”

理论角度对信息简报的生成机制进行了理论审视和机制探讨;进而以“选择性注意”模型为基础对信息简报生成的计算框架进行实证研究;最后把关键词抽取和自动摘要计算模型进行一体化设计,为提高信息简报的可解释性和可计算性提供新的支撑。

## 2 总体框架设计

本文根据选择性注意“衰减”模型理论设计了信息简报生成模型,并与 KeyBERT、Transformer 算法结合,为信息简报生成提出了可行的认知计算依据和实现方案,最后从信息熵和 ROUGE (Recall-Oriented Understudy for Gisting Evaluation) 值两个方面进行量化测评和验证。

### 2.1 选择性注意衰减模型与认知负载因素

信息简报生成机制可以从认知科学的“注意”研究中得到有益的启示。“注意”是一种聚焦于特定刺激的能力。大多数情况下,注意的聚焦特性是与“选择性注意”相关联的,即人类能够将注意聚焦或分配在一个特定的位置事物或者信息上。认知心理学家 Broadbent 提出的“过滤器理论”(Filter Theory)认为,人类可以将一些信息过滤掉,允许另一些信息得到更高层次的加工,从而免受“信息过载”的困扰。Treisman 提出的“衰减器”理论是对过滤器模型的改进,未被注意的信息并未被真正过滤,而只是被加工的可能有差异而已,进而通过认知实验表明,物理特性、语言、语义都可以被用户注意到,并用来吸收信息,模型见图 1,通过衰减器,各类信息都有机会被进行加工或者过滤,最终形成记忆。认知负载是指在执行某项认知任务时所需要的认知资源数量。对于熟悉或者简单任务而言,认知负载较低;对陌生或困难任务则需要分配更多的认知资源,其认知复杂相对较高。

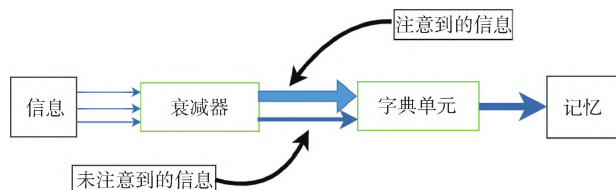


图1 选择性注意衰减模型

Fig. 1 Theoretical diagram of selective attention decay model

选择性注意衰减模型为信息简报的自动生成提供了重要理论依据。表现在 3 个方面。①摘要过程实际上是人类

分配认知资源对信息进行过滤的过程,语言特征和语义是摘要的核心依据,除传统依据词频、位置等外在特征的摘要抽取之外,还应该以语义为中心进行抽取。这对于改进过度依赖统计模型是一个有益的提示。②作为语义载体的词汇和句子对人的认知负载有差异,关键词多数属于低认知负载,而摘要则属于高认知负载,关键词对于摘要属于相容侧干扰项,有助于以较低的认知复杂实现信息的高层加工和特征整合。对信息简报生成而言,应该综合考虑将关键词和摘要一体化设计,提高关键词与摘要语义相容性。③信息摘要往往采用多重任务并行执行,生成多个并行的摘要文本,被注意到的焦点信息固然重要,但重新激活非焦点任务并增强信息仍有可能,可以根据用户的不同需求生成摘要,实现多个维度的摘要生成。

因此,在简报生成过程中,应该对输入文本内容的语义特点进行主动筛选,模拟人类将更多的“注意力”分配给更重要的焦点信息上,将具有相容性的关键词等低负载信息与摘要信息一体化设计。因此本文以选择性注意的衰减器理论为依据,设计了简报生成的基本框架,并采用 BERT 等算法进行语义计算和上机实现。

### 2.2 简报生成框架设计

结合 2.1 节选择性注意力衰减模型理论,本文设计了信息简报生成主要框架,见图 2。

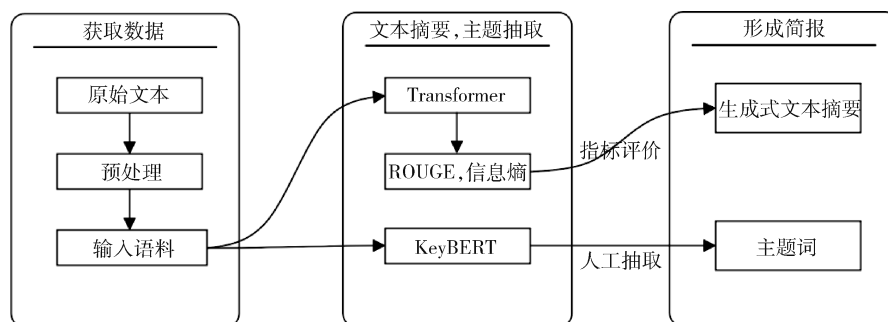


图2 自动简报生成框架

Fig. 2 Automatic brief generation architecture

该框架包括 3 个模块,分别是获取数据、文本摘要和主题抽取、生成简报。

1) 获取数据:文本抽取是生成简报的准备条件,首先根据需求选择信息来源,确定目的文本,剔除图表和公式等信息元素。然后,导入 Jieba 中文分词库对文本进行预处理,过滤掉语气词、标点符号等,形成待加工的语料文本。

2) 文本摘要、关键词抽取:进行计算之前通过导入 Word2Vec 库对输入语料向量化处理,再通过 Transformer 计算生成新的语料后,利用 ROUGE 值和信息熵进行指标评价,最后进一步处理形成文本摘要;将语料进行向量化



处理后,调用 KeyBERT 算法库进行关键词抽取形成关键词集,再经过过滤和排序处理后形成主题词。

3) 生成简报:标题、摘要和关键词进行合并,并进行一致性检查,生成最终简报。

### 2.3 关键词、摘要一体化生成计算

2.3.1 基于 KeyBERT 方法的关键词抽取 关键词是信息简报的基本组成部分,它是从文本中抽取出来的、能够揭示文本结构和语义中心的词语,可以帮助用户抓住文本重要信息点。KeyBERT 是一种小型且易用的关键词抽取技术,通过对文本进行向量化表示后,再以语义计算为过程导向进行权值计算,使用 BERT 嵌入和简单的余弦相似度创建与文档最贴合的关键词或短语。

本文首先利用 BERT 计算文档的 embedding 值,从而获取文档向量级别的表示。然后针对 n-gram 提取词向量,最后利用余弦相似度来确定与文档最相似的关键词或短语,得到最能描述整篇文档的关键词。在关键词的基础上,结合原文主题内容进一步筛选,对生成式摘要结果实现语义上的内容匹配,为用户提供多维度的摘要参考。

2.3.2 基于 Transformer 模型的摘要句生成 通过注意力模型能够较好地计算句子之间的关系<sup>[13]</sup>。本文采用 Transformer 计算模型,先将输入的文本进行词向量转换,通过 Encoder-Decoder 层进行 Attention Scores 的得分量化计算,识别句子的动态权重关系,最后经 Softmax 计算后输出摘要结果。相比于以往研究抽取式文本摘要或者基于 LSTM 等深度学习摘要,Transformer 模型通过循环多次的自注意力的计算,不但使得其并行计算效率增加,也提高了计算质量<sup>[14]</sup>。

### 2.4 简报质量评价指标

2.4.1 简报信息衡量:信息熵 利用信息熵来判断文本摘要信息含量,可从信息量的维度对生成式文本摘要质量进行量化描述。信息熵是用来衡量系统中信息量大小,从文本信息的角度而言,一段文字的熵值越高,则表明这段文本所含的信息量越高,可以用于验证不同语种表达相同含义下所需文本容量,也可验证特定文本信息含量大小。信息熵的数学公式可以表示为<sup>[15]</sup>:

$$H(x) = \sum_{x \in X} P(x) \log\left(\frac{1}{P(x)}\right) = - \sum_{x \in X} P(x) \log(P(x))$$

式中,  $x$  表示随机变量,它的取值为  $(x_1, x_2, \dots, x_n)$ ;  $P(x_i)$  表示事件  $x_i$  发生的概率,通常情况下信息发生的概率越大,则这个事件所包含的信息越小;  $\sum P(x_i) = 1$ ,即所有随机事件的概率和为 1,引入文本信息量计算中则表示一段文本所包含的随机信息的概率。考虑到文本内容包含语义计算涉及词汇较多,故选取二元模型 (bi-gram)

对文本进行信息熵的计算<sup>[16]</sup>。本实验利用 Python 语言提供的开源代码构建信息熵计算流程,对生成的摘要进行分词处理后再用 Word2Vec 方法进行文本向量化表示,最后通过信息熵公式进行计算并输出结果。

不同认知水平的人群,如学者、儿童等描述同一内容的文字信息熵也存在差异,因此本文通过计算摘要结果的信息熵值,并将该值与中文文本的信息熵标准指标进行对比,辅助验证该生成式简报与人的认知水平是否信息量大致相符。

2.4.2 简报质量评测: ROUGE 值 ROUGE 是评价文本摘要质量的常用指标。它通过参考文本与实验文本之间共现的词语或句子个数,来判断摘要结果的质量,因此计算该指标需要获得由专家手动生成人工摘要作为标准摘要集,与机器生成的摘要作为对比。用公式表示为:

$$ROUGE - N = \frac{\sum_{S \in \text{ReferenceSum}} \sum_{\text{gram}_N \in S} \text{Count}_{\text{match}}(\text{gram}_N)}{\sum_{S \in \text{ReferenceSum}} \sum_{\text{gram}_N \in S} \text{Count}(\text{gram}_N)}$$

式中,  $N$  表示 N-gram,代表分母为参考文本中 N-gram 个数和分子为摘要结果中 N-gram 的个数; ReferenceSum 表示参考摘要。进行评价时通常以 ROUGE-1, ROUGE-2 为指标。

除以上两个指标外,为增强评价的科学性,本文还参考了最长公共子序列计算方式,即 ROUGE-L,其计算公式为:

$$R_{\text{LCS}} = \frac{\text{LCS}(C, S)}{\text{len}(S)}$$

$$P_{\text{LCS}} = \frac{\text{LCS}(C, S)}{\text{len}(C)}$$

$$F_{\text{LCS}} = \frac{(1 + \beta^2) R_{\text{LCS}} P_{\text{LCS}}}{R_{\text{LCS}} + \beta^2 P_{\text{LCS}}}$$

式中,  $C$ 、 $S$  分别表示参考文本和实验文本;  $\text{LCS}(C, S)$  表示文本  $C$  与  $S$  之间最长公共子序列长度;  $\text{len}(S)$  和  $\text{len}(C)$  分别表示两个文本的长度;  $R_{\text{LCS}}$  表示召回率,  $P_{\text{LCS}}$  表示准确率,  $F_{\text{LCS}}$  即是 ROUGE-L。  $\beta$  则表示一个非常大的数,因此经过推导公式后发现,  $F_{\text{LCS}}$  几乎等于  $R_{\text{LCS}}$ 。本文生成的摘要主要为单文档摘要,故采用 ROUGE-2 和 ROUGE-L 作为评价指标。

## 3 实证研究

### 3.1 数据采集

本实验数据选自 UNESCO 教育方面的报告,按照 2.1 节流程图所示,利用 Python 中 selenium 库选取了 UNESCO 中 10 篇具有代表性的报告,进行初步人工预处理后储存在数据库中。然后,对数据库中的文本数据进行分词,去除停用词等操作后完成预处理。

3.2 关键词与文本摘要自动生成

按照 2.2 节提出的总体框架进行实验验证。首先,将预处理后的文本导入,通过 Jieba 分词库进行分词处理,再利用 Word2Vec 转化成向量表示便于计算。然后,依据选择性注意衰减模型理论,利用开源工具搭建的 Transformer<sup>[17]</sup> 及 KeyBERT 算法框架,将采集数据预处理为机器可识别的词向量后,通过已有框架进行计算,分别生成文本摘要和关键词集。

在此基础上,本文建立了可量化的信息熵和 ROUGE 值,对文本摘要结果进行量化评价;在关键词集的基础上,抽取围绕文本主题叙述的关键词作为线索,并按照文本顺序对关键词进行排序,最终形成规范化的信息简报。部分实验结果见表 1。

基于本文提出的选择性注意衰减理论的计算框架,将

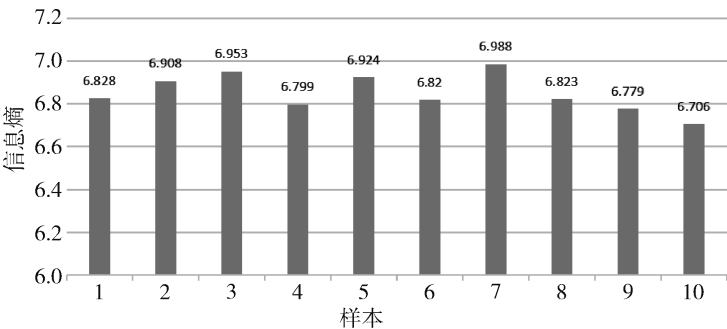


图 3 摘要质量测评:信息熵

Fig. 3 Quality evaluation: information entropy

文本摘要与关键词抽取深入到了语义理解层次,再重新组织文本生成完整的信息简报<sup>[18]</sup>,使得摘要结果比较流畅自然,表明该方法更符合人的认知习惯,具有较强解释力。

表 1 部分生成式文本摘要  
Tab. 1 Partial generated text summarization

序号	结果
文本 1: 摘要	联合国教科文组织发起了一项全球范围的“快乐学校倡议”。学校应该成为支持社会凝聚力的场所,创造跨越差异的社区,也应该通过快乐和参与来培养学生对学习的终生热爱,而不是通过将学习成绩置于一切之上来阻止学生学习,从而损害个人的幸福。“幸福学校”倡议正在走向全球,联合国教科文组织正在向全世界发起,它的目的是帮助全球的学生和教育工作者。它将成为全世界学校和教育工作者的一个全球模式,这是一个国际学校教育的国际模式,以帮助整个大陆和国家的年轻人和教育工作者
文本 1: 关键词	联合国教科文组织,全世界,年轻人,幸福危机,生产力,压力,难民,学校,提高,学生,凝聚力,平等
文本 2: 摘要	2017 年,在全球范围内,超过 2/3 的国家中,学习工程、制造和建筑或信息和通信技术 (ICT) 的人中只有不到 1/4 是女性。国际教育协会指南针指出,这可能会导致更少的表现出色的女孩进入 STEM 高等教育领域教育系列简报。在几乎所有的教育系统中 (87%),男孩比女孩更经常回答他们想从事涉及数学的工作,男孩和女孩在数学领域工作的愿望与他们对自己在该学科的能力的信心密切相关。这表明,解决女孩对科学和数学的信心问题应该继续成为政策制定者的关注点。该报告发表在 IEA 的 TIMSS 2019 年数据特刊上,样本为 25 万名学生,显示 8 年级时希望从事数学或科学相关职业的男孩多于女孩
文本 2: 关键词	联合国教科文组织,国家,科学研究,统计学,1/4,女孩,取得成功,成就,分析,信心,不同
文本 3: 摘要	一项对西班牙巴塞罗那近 3000 名儿童的队列研究发现,在调整了社会经济地位后,那些暴露在西班牙高污染水平下的儿童,其认知发展的增长低于污染较少的学校的同龄人的发展增长。在东南亚,如果一个孩子的温度高于平均水平 2 个标准差,预计他将减少 1.5 年的学业;学年平均温度上升 1.8℃,学习成绩就会下降 1%,超过 32.2℃ 的 6 天也是如此。假设最佳温度低于 22℃,教室里的温度从 30℃ 降到 20℃ 将使考试成绩平均提高 20%。污染的空气会大大降低认知能力,也许是不可逆转的,因为怀疑有神经毒性。在以色列,即使暂时暴露在灰尘中,也会降低学生的考试成绩和中学后的教育程度
文本 3: 关键词	高温,高污染,平均气温,造成,考试成绩,中学,年限,降低,认知,发展

通过将结果进一步处理计算文本摘要的信息熵,对 10 篇文献的摘要信息含量进行评价。计算结果如图 3 所示,横坐标轴代表了 10 篇摘要结果,纵坐标代表了信息熵,柱状图顶部显示了每个摘要结果的信息熵值。

图 3 表明,所抽取样本信息熵平均值达到 6.852;同时为验证摘要质量,本文与中文小说文本的信息熵进行对比发现,基于 2-gram 的中文小说文本信息熵范围介于 4 ~ 8 之间,与本文的信息熵基本一致,表明该方法生成的摘

要结果信息熵达到中文文本信息熵标准,信息量符合人的认知水平。

本实验采用 ROUGE 值进一步验证摘要结果的质量。验证召回率需要合适的参考摘要,为了使研究结论更具有客观性,本文将 10 篇文本语料人工撰写摘要,作为计算 ROUGE 值的参考答案,进而通过计算人工生成参考摘要与实验得到摘要之间的词文本共现率计算出 ROUGE 值<sup>[19]</sup>,从图 4 所示,由于 ROUGE-2 运算时考虑了两个连

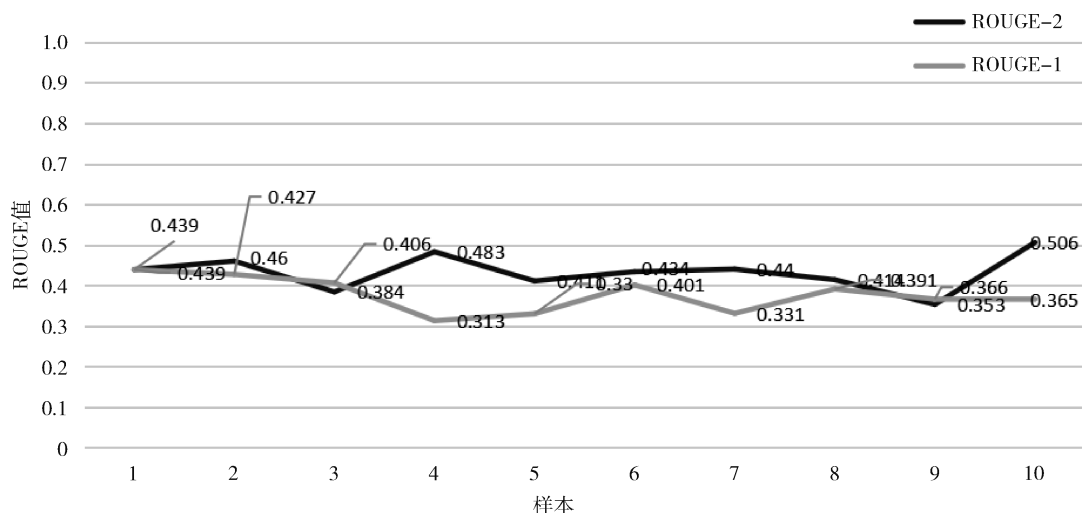


图4 摘要质量测评: ROUGE 值  
Fig. 4 Quality evaluation: ROUGE value

续词向量,而 ROUG-L 是依赖多个连续词向量,ROUGE-2 普遍大于 ROUGE-L。

本文所抽取样本在 ROUGE-2、ROUGE-L 两个指标平均值分别为 0.432 和 0.367,与现有文献公开的结果基本持平<sup>[20]</sup>,但生成式摘要的可读性更高。其原因在于通过词、句子和篇章的语义计算,实现对不同认知负载信息的一体化加工,有助于生成简洁、自然的简报,可读性和可解释性有所提高,由此也说明了选择性注意力机制对简报生成方法具有重要指导意义。

### 3.3 简报生成

简报包括标题、关键词和摘要三部分。通过摘要生成和关键词抽取两个维度的信息处理,能够较为充分描述文本信息。关键词与文本摘要作为简报产品的重要组成部分,相辅相成,在内容和主题把握上有着较强的联系:摘要关键词的扩展深化部分,关键词是摘要的核心信息,二者具有较高的一致性<sup>[21]</sup>。文本摘要和关键词并行计算、相互参照,通过若干主题词的描述可以一定程度上对文本要点进行挖掘,共同构成信息简报。

## 4 结论

本文将认知科学中的选择注意力衰减理论和信息简报生成方法相互结合,设计了融合关键词抽取和生成式摘要的信息简报生成框架,以 UNESCO 数据为例,采用 Transformer 和 KeyBERT 方法进行了实证研究。本文的创新性和实践意义主要表现在:从理论角度,本文设计了基于认知科学“选择性注意衰减模型”的简报生成模型和框架,能够有效地解释简报生成机理,形成可计算的简报生成方法,促进认知科学与计算模型的有机结合;从应用角度,本文提出了关键词与摘要一体化生成框架,并采用真实数

据进行实验,简报信息熵和 ROUGE 值测度表明其信息量和召回率指标不低于主流摘要方法,而且能够提高关键词、摘要和简报的语义关联性,具有更好的可读性和可解释性,从而可以较好地满足用户的情报

需求,这对于实现情报监测、决策支持等工作实践有一定的参考价值。今后,对选择性注意的影响因素和机制进一步探究,将认知模型与情报技术进一步结合,持续提高信息简报可解释性与语义关联性,是需要继续研究的方向。□

### 参考文献

- [1] 章成志,胡少虎,张颖怡.通用语料的眼动数据对微博关键词抽取的性能提升探究[J].情报学报,2021,40(4):375-386.
- [2] 李开容,胡博,陈雨嘉,刘博,常明.奖赏预期对选择性注意的影响——基于客体和空间的注意的区别[J].心理科学,2022,45(1):133-141.
- [3] 高劲松,黄梅,付家炜.基于年度高频关键词的研究热点提取与可视化研究[J].现代情报,2020,40(12):130-139.
- [4] 张建娥.基于多特征融合的中文文本关键词提取方法[J].情报理论与实践,2013,36(10):105-108.
- [5] 毛立琦,石拓,吴林,马涛.基于领域自适应的无监督文本关键词提取模型——以“人工智能风险”领域文本为例[J].情报理论与实践,2022,45(3):182-187.
- [6] 贾君枝,王醒.基于微数据的语义标注应用研究[J].情报理论与实践,2016,39(2):54-58.
- [7] 章成志,童甜甜,周清清.基于细粒度评论挖掘的书评自动摘要研究[J].情报学报,2021,40(2):163-172.
- [8] 王晓宇,王芳.基于语义文本图的论文摘要关键词抽取算法[J].情报学报,2021,40(8):854-868.
- [9] WAN Xiaojun, WANG Tianming. Automatic labeling of topic models using text summaries [C] // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Berlin, Volume 1: Long Papers). 2016: 2297-2305.

(下转第7页)



风险。

5) 评估所有信息基础设施和设备的管理状况。

6) 识别启动业务流程再造 (BPR) 的合适时机, 利用信息和信息技术去设计更好的可操作的业务流程。

7) 识别进一步提高知识型工作的绩效的机会, 重点关注新的“群件”工具。

8) 评价所有重要的 IT 项目, 弄清楚对变革管理是否有足够的重视, 并通过定义组织的愿景、规划变革和规划宣传, 加上标准化的监督和控制程序来强化项目管理和变革管理。

9) 建立培训和职业发展规划, 以培养更多能真正从事信息资源管理的专业人才。

10) 将所有这些具有创新性的项目和工作综合到一个规划中, 以保证信息资源能够部署到最需要的地方, 将组织的信息基础设施从原来内省式、不灵活的庞然大物改造成外向型的、灵巧的组织耳目。

## 7 结论

信息资源管理作为一种管理活动和管理学的一个重要分支, 它的历史发展和实际内涵可以简单地归纳为: 一种新的资源观和管理哲学。两类主要管理对象: 信息资源, 人类及其代理的相关管理活动和行为。三大来源: 政府文

书管理、计算机数据处理、工商领域的数据处理管理。四个主要应用领域: 政府、企业、非营利性机构、网络。五大目标: 合规、效率、公平、安全、价值实现。六个主要管理维度: 理论、政策法规、技术、经济、组织、文化。七个管理层次: 技术管理、内容管理、项目管理、IT 服务管理、信息化投资管理、战略信息管理、知识管理。八个管理结构要素: 首席信息官、规划、安全、技术集成、咨询委员会、企业模型、信息集成、数据管理。十个管理优化步骤。□

## 参考文献

- [1] 郭沫若. 古代文字之辩证的发展 [J]. 考古学报, 1972 (1).
- [2] LEWIS B R, SNYDER C A, RAINIER JR R K. An empirical assessment of the information resource management construct [J]. Journal of Management Information Systems, 1995, 12 (1): 199-223.
- [3] WILSON T D. Information management [M] //International Encyclopedia of Information and Library Science. London: Routledge, 1997: 187-197.
- [4] COLLIN N. Information management in crisis [J]. Computer Audit Update, 1995 (February): 11.
- 作者简介: 赖茂生, 男, 教授, 博士生导师。  
录用日期: 2023-03-08
- (上接第 58 页)
- [10] 曾子明, 孙晶晶. 基于用户注意力的突发公共卫生事件舆情情感演化研究——以新冠肺炎疫情为例 [J]. 情报科学, 2021, 39 (9): 11-17.
- [11] 周琰, 刘越, 蔡俊. 基于注意力机制的微博情感分析 [J]. 情报理论与实践, 2018, 41 (3): 89-94.
- [12] 胡吉明, 付文麟, 钱玮, 田沛霖. 融合主题模型和注意力机制的政策文本分类模型 [J]. 情报理论与实践, 2021, 44 (7): 159-165.
- [13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [EB/OL]. [2022-11-20]. <https://arxiv.org/pdf/1706.03762.pdf>.
- [14] EL-KASSAS W S, SALAMA C R, RAFAA A A, et al. Automatic text summarization: a comprehensive survey [J]. Expert Systems with Applications, 2021, 165 (6): 113-129.
- [15] 马峥. 基于“反事实”思想测度学术期刊对知识系统信息熵变化的贡献 [J]. 情报学报, 2022, 41 (7): 745-761.
- [16] 程慧平, 程玉清. 基于 AHP 与信息熵的个人云存储安全风险评估 [J]. 情报科学, 2018, 36 (7): 145-151.
- [17] NIU Zhaoyang, ZHONG Guoqiang. A review on the attention mechanism of deep learning [J]. Neurocomputing, 2021, 452 (8): 48-62.
- [18] 赵旸, 张智雄, 刘欢, 李婕. 基金项目摘要的语步识别系统设计与实现 [J]. 情报理论与实践, 2022, 45 (8): 162-168.
- [19] PUDUPPULLY R, DONG L, LAPATA M. Data-to-text generation with content selection and planning [C] //Proceedings of the AAAI Conference on Artificial Intelligence. 2019: 6908-6915.
- [20] WAN X, LUO F, SUN X, et al. Cross-language document summarization via extraction and ranking of multiple summaries [J]. Knowledge and Information Systems, 2019, 58 (2): 481-499.
- [21] 唐晓波, 翟夏普. 基于混合机器学习模型的多文档自动摘要 [J]. 情报理论与实践, 2019, 42 (2): 145-150.
- 作者简介: 宋培彦 (ORCID: 0000-0003-1055-2717), 男, 博士, 教授, 硕士生导师。研究方向: 知识组织, 自然语言处理, 科技大数据。田文波 (ORCID: 0000-0002-4251-2763, 通信作者), 硕士生。研究方向: 知识组织与知识聚合。冯超慧 (ORCID: 0000-0001-6503-4111), 硕士生。研究方向: 情报学。吴柯莹 (ORCID: 0000-0003-3013-6657), 硕士生。研究方向: 情报学。
- 作者贡献声明: 宋培彦, 论文理论框架、研究思路和方案设计, 论文撰写与修改完善。田文波, 论文算法与实验设计与上机实现, 论文撰写与修改。冯超慧, 数据分析、论文修改等。吴柯莹, 论文实验改进与数据分析。
- 录用日期: 2022-12-12