

# 自然语言处理在计算传播学研究中的重要作用

连树国

中国人民财产保险股份有限公司 北京 100022

**摘要:** 信息技术的高速发展,使传播学范式转移现象越发明显,使学科研究对文字数据挖掘技术的依赖性增加,为满足计算传播学的发展需要,本文对自然语言处理技术进行研究,分析将该技术引入计算传播学研究工作中的作用,提出如下结论:在信息全球化大背景下,自然语言处理技术有利于解决学科问题。

**关键词:** 自然语言处理技术;计算传播学;词频分析;语义建模

## The important role of natural language processing in computational communication research

Shuguo Lian

PICC Property and Casualty Co., Ltd. Beijing 100022

**Abstract:** The rapid development of information technology has made the paradigm shift in communication studies increasingly evident, leading to a growing reliance on text data mining techniques in disciplinary research. This satisfies the developmental requirements of computational communication studies. This paper explores natural language processing (NLP) techniques, analyzing their role in integrating these technologies into computational communication research. The following conclusion is drawn: In the context of global information globalization, NLP technology is advantageous in addressing disciplinary issues.

**Keywords:** Natural Language Processing Technology; Computational Communication; Word Frequency Analysis; Semantic Modeling

### 引言:

在物联网迅速发展背景下,我国已经迈入信息化时代,自然语言作为一种保证计算机系统满足人类要求的技术,在信息时代数据处理中,发挥着重要作用。现阶段,在开展计算传播学研究工作中,对自然语言的价值进行分析,不仅可以实现人类与计算机之间的有效交互,保证机器能按照人的要求开展工作,还能为计算传播学研究提供支持。

### 一、自然语言处理技术的概述

计算机技术的出现改变了当前人类的生存方式,人们在日常生活中不仅可以借着计算机享受网络带来的便利,也能感受到物联网给人们提供的安全高效的生活方

式。现阶段,为了满足人们对智能化、现代化社会发展的需要,可以将自然语言处理技术融入到当前的计算传播学研究工作中,确保机器在后续运作过程中不需要人为管理,就可以参照相应控制策略,自行完成机器功能运转操作,从而达到提升智能化控制工作实时性,满足人与机器交互工作要求的目的。

#### 1.应用概述

自然语言处理技术作为一种实现机器与人语言互动的技术方法,可以完成语言、音频等非结构化数据的转化工作,降低机器理解人类语言的难度,保证机器与人能够顺利进行语言互动。在实际应用过程中,自然语言处理技术在实际应用过程中有着自然语言理解与自然语言生成这两大功能,其中,自然语言理解功能可以帮助机器理解人类的语言,而自然语言生成功能可以保证机器与人的顺利沟通。在对非结构内容进行标准化处理时,可以借助机器学习与深度学习这两种方法,完成非结构

---

**作者简介:** 连树国(1978—),男,蒙古族,辽宁,硕士,高级工程师,研究方向:计算机应用,通讯邮箱:lianshuguo@163.com

内容的分词处理、词干提取、词形还原、词性标注、命名实体识别、分块处理工作。将自然语言融入的计算传播学研究中,对机器学习自然语言的处理方法主要包括文本分析法、文本聚集法、关联分析法以及趋势预测法,其中,文本分析法在应用过程中可以通过监督学习的方式,完善模型的分类方法。文本聚集法在应用过程中可以给出特征向量与分类书目,完成样本特征向量相似性的聚类分析工作。关联分析法在应用过程中可以找出特征向量与结果之间的关系。走势预测法则可以对已有数据的时间序列分布条件进行分析,完成数据预测工作<sup>[1]</sup>。

## 2. 统计模型基础

在当前的计算传播研究过程中,常用的自然语言系统模型包括贝叶斯公式条件概率、N元模型、HMM模型等在具体信息分析工作中,结合信息的类型与分析工作量的需要,选择并用合适的模型,可以为分析工作质量与效率的提升提供有力的支持。

### (1) 贝叶斯公式条件概率

在应用贝叶斯公式条件概率判断给出的词串W属于A<sub>1</sub>、A<sub>2</sub>哪一集合的过程中可以先计算概率P(A<sub>1</sub>|W)与P(A<sub>2</sub>|W),并对两个概率的值进行比较,若A<sub>1</sub>的概率值比A<sub>2</sub>的概率值大,则说明W属于A<sub>1</sub>,反之W属于A<sub>2</sub>。

### (2) N元模型

N元模型是自然语言处理过程中一种较为常用的数学模型。在应用这一模型的过程中,若自然语言满足马尔科夫性,那么某一个词语w在某句话r中出现的概率可以用公式 $P(w_i | w_1 w_2 \dots w_{i-1}) = P(w_i | w_{i-n+1} \dots w_{i-1})$ 进行计算,并推导出一个句子概率的计算公式:

$$p(S) = p(w_1 w_2 \dots w_m) \\ = p(w_1) p(w_2 | w_1) \times \dots \times p(w_i | w_{i-n+1} \dots w_{i-1})$$

在实际分析过程中,公式中n值越大,模型的精确度越高,但考虑到模型所用参数与所需训练集也会随n的增大而增大,而训练集不够大,那么会导致大部分分子或连接词搭配在语料库中的出现次数降低,甚至不出现这一情况的出现,必然会导致数据稀疏。知识短缺影响模型在后续处理工作中的性能,因此在应用n源模型开展分析工作时,需要对数据进行平滑处理<sup>[2]</sup>。

### (3) HMM模型

HMM模型是一种极为重要的数学模型,被广泛应用于自然语言处理与语音识别工作中,在当前的计算传播学研究工作中,这一模型也被应用到了其他领域。

## 二、自然语言处理技术的应用功能

在当前的信息技术发展过程中,自然语言处理技术的应用原理在于利用预处理技术完成非结构内容的标准

化处理工作,被标准化处理后的内容可以被积极正确理解并完成特定功能。在当前的计算传播学的研究过程中,自然语言处理技术较为广泛应用功能包括词频分析、情感分析与语义建模。

### 1. 词频分析

词频分析是当前计算传播学中较为常用的功能之一,在当前百度指数、抖音热榜等网站热词排名分析过程中,词频分析技术可以帮助上述网站利用分词技术,对相应数据信息进行预处理,然后借助分布式大数据系统流处理方法完成词语信息的数量统计工作。

### 2. 情感分析

情感分析技术是利用算法参照特定情感类型完成文本分类处理工作的分析方法,较为常见的情感分类包括积极、消极、高兴、悲伤等情绪。在应用情感分析技术的过程中,可以通过情感分类、主观性判断、意见总结、评论有效性分析等方式,完成网络空间中文字信息评价对象与被评价主体、传播话题内容与情感倾向等内容态度的有效分析。在当前的计算传播学研究过程中,情感分析技术主要被应用于口碑分析工作中,在分析过程中不仅可以以评分高低为参照,完成相应信息的评价积极或消极分析工作,还可以通过人工标注、词典匹配等方式,完成参与评价人员的主观评价情感倾向的分析工作,从而为产品优化工作的开展提供参照<sup>[3]</sup>。

### 3. 语义建模

考虑到人类语言极为复杂,在不同的语境下,词语会拥有不同的含义,为提升词语分析工作的准确性可以将语义建模引入到计算传播学研究工作中,利用算法结合语境的方式完成词语背后隐藏含义的分析工作,在当前的计算传播学中,语义建模技术是一种较为成熟的主题解析技术。举例来说,语义建模技术可以被应用于提取论坛网络用户语义、对比用户知识框架与阐述异同等工作中,完成响应信息对受众影响的研究。

## 三、自然语言处理技术的应用价值

### 1. 满足计算传播学研究的诉求

在当前社会发展过程中,自然语言处理技术的应用可以切实满足计算传播学研究在传播范式调整、文字文本研究等方面的诉求。

#### (1) 传播学范式调整

从传播学范式调整的角度看,计算传播学研究的深入使得传播现象研究不局限在功能研究方面,而是开始描述复杂多元传播现象与群体特征,并在研究过程中提炼新研究主题及其背后的意义。同时,计算传播学研究在应用过程中可有效地平衡研究者主观思想和文本研究

客观性的要求，在信息技术飞速发展的背景下计算传播学研究使得假设外碎片化、小概率、大容量事件信息融入了学科研究内容当中，提升了传播要素研究的深入性。

## (2) 文字文本研究

现阶段，计算传播学研究范式对文字文本研究相关工具有着极为迫切的需要。具体来说，由于图像数据一般为矩阵格式的rgb像素文件的集合，在处理图像数据的过程中，需要综合应用图像处理、计算机视觉等学科知识，这一情况的出现使得图像数据的研究工作难度大幅度提升。面对上述情况，文字文本因具备在网络空间中有着获取难度低、数据处理所需背景知识较少、数据计算量偏低等优势，成为当前新闻传播研究的主要对象，也是计算传播学研究中获取数据挖掘分析的主要内容。在当前的计算传播学研究过程中，网络数据获取分析是研究工作的重要环节之一。然而在信息化时代背景下，受网络数据来源缺乏管制、数据总量呈指数增长等因素的影响，计算传播学研究工作面临了巨大的挑战。对此，为了尽可能提升研究工作的质量与效率，在当前的计算传播学研究过程中构建一套行之有效的计算及文字语义识别方法，并将其作为网络空间语义文本的分析工具成了提升研究工作数据获取能力、处理效率、分析能力的重要举措。在此背景下，自然语言处理技术受到了研究人员的关注，并且将这一工具应用到当前的计算传播学研究工作过程中，可以切实满足研究人员对于个性化、定制化数据获取、组织、处理的需要<sup>[4]</sup>。

### 2. 为计算传播学研究创造机遇

自然语言处理技术的不断发展为计算传播学研究带来了新的发展方向，也为传统计算传播学研究中问题的解决提供了新的思路。

首先，自然语言阅读理解技术的快速发展，推动了计算传播学文本研究朝着精细化、定制化的方向发展。具体来说，自然语言阅读理解算法在应用过程中可以使机器按照语境完成语义理解工作。实现语言要点的总结、归纳与提炼，切实解决语言多样性、歧义性、鲁棒性与知识依赖性。需要注意的是，尽管自然语言阅读理解技术在发展的知识依赖的机器阅读理解技术后，其技术能力得到了显著的提升，但在实际的文本分析过程中，检测无法回答的问题、合理答案的分区等工作仍存在一定的困难，这一情况的存在，降低了计算传播学研究工作的质量水平。

其次，在当前的计算传播学研究工作中，研究人员往往会先提出相应模型，然后利用统计学等学科方式完成信息关联性验证工作，然而受文字文本数据中往往包含大量难以察觉隐含变量，且这些变量具有小样本、离

散、高维等特点的影响，应用上述方法开展文本关联验证工作，必然会导致模型设计验证难度有所提高。为了解决这一问题，可以在分析过程中应用无监督学习的自然语言处理技术，在没有提前给出具体模型的情况下，完成文本数据特征向量的分析工作，及时找出文本信息中包含的隐含变量，然后以此为基础开展模型设计工作，以便达到弥补假设模型先天不足的目的。

再次，在当前的计算传播学研究过程中，自然语言处理技术的应用可以进一步提升研究工作的客观性，具体来说，为缩短计算传播学与客观科学间的距离，计算传播学研究人员想找寻更为客观的传播对象研究方式作为自身的研究重点。近年来，随着自然语言处理技术的不断优化升级，将其应用到计算传播学研究中，成了提升研究结果准确性与客观性的重要举措。举例来说，在实际的研究工作中，研究人员可以将调查问卷与访谈替换成信息技术挖掘下的数据集与验证集，通过对海量数据进行分析，并弱化个体间异质性的方式，提升数据挖掘工作的可靠性与准确性。

最后，随着自然语言处理技术的不断优化升级，传播模式分析与伦理探讨研究越发深入。具体来说，自然语言处理技术在实际应用过程中，不仅具备语言理解功能，还具备自然语言生成功能，并且这一功能可以被应用到数据新闻的生成工作中。近年来，在自然语言处理技术不断发展的背景下，这一技术不仅切实满足了当前网络空间对于积极生成文字文本的需求，也为计算传播学研究对象、模式与伦理研究工作的开展提供了新的议题。

## 四、结论

总而言之，计算社会学可以通过收集分析网络空间中与人类行为相关的数据的方式，在消除异质性与噪声因素的同时，实现复杂社会系统的有效分析，在找出系统中社会现象的同时，丰富人们的社会认知。现阶段，开展自然语言处理在计算传播学中应用价值的研究工作，可以为计算社会学的发展提供助力。

### 参考文献：

- [1]王敏, 张子柯. 计算传播学的仿真研究范式: 优势、挑战与发展[J]. 新闻界, 2022(10): 64-74.
- [2]公宣迪. 数智时代传播范式、理论与实践的变迁与重构——以计算传播学的研究视角为例[J]. 青年记者, 2022(12): 28-30.
- [3]石豪. 自然语言处理在计算传播学研究中的应用、挑战与机遇[J]. 传播与版权, 2021(04): 55-58.
- [4]张伦. 计算传播学范式对传播效果研究的机遇与挑战[J]. 新闻与写作, 2020(05): 19-25.