

这是作者的文章原稿。该记录版本发表在《国际社会研究方法论杂志》上,可在线获取:<https://doi.org/10.1080/13645579.2019.1576317>。请引用 Jacobs, T. 和 Tschötschel, R. (2019)。 “主题模型满足话语分析:定性方法的定量工具。”国际社会研究方法论杂志

1

2

3

主题模型与话语分析相结合:一种定量工具
定性方法
研究文章

4

5

关键词:话语分析、主题建模、文本分析、霸权、数字
人文科学、计算社会科学

6

托马斯·雅各布斯 (通讯作者)

7

根特大学欧盟研究中心,比利时根特

8

欧盟研究中心 (CEUS), Korte Meer 1, 9000 Gent, België

9

10

11

Thomas Jacobs 是根特大学欧盟研究中心的博士生。他的10个主要研究兴趣包括欧盟贸易政策、国际政治经济学和政治传播。托马斯的大部分著作集中在政治实践中的话语、意义12制造、意义和策略,背离了后结构主义的观点,更具体地说是拉克劳和莫菲的话语理论传统。

12

13

14

在进入学术界之前,他曾在布鲁塞尔的15家传播和公共事务公司担任顾问。

16

罗宾·切尔切尔

17

18

阿姆斯特丹大学阿姆斯特丹传播研究学院,
荷兰

19

20

阿姆斯特丹传播研究学院 (ASCoR), Postbus 15793, 1001 NG
荷兰阿姆斯特丹

21

Robin Tschötschel 是阿姆斯特丹传播学院的博士生

22

阿姆斯特丹大学研究系主任,并在同一所大学的政治、心理学、法律和经济学院担任教学职位。他的研究重点是气候变化和经济增长的公共辩论的结构和影响,并利用了一系列混合方法,其中许多方法借鉴了计算社会科学工具包。

28 主题模型满足话语分析:定量工具

29 定性方法

三十 定量文本分析工具已成为越来越流行的方法
31 各种类型话语分析的操作化。然而,他们的
三十二 应用程序通常仍然相当简单和肤浅,并且无法利用
33 数字时代为话语分析提供了最充分的资源。
34 本文探讨了一种更复杂、更
35 高级文本分析工具,已经在其他领域频繁使用
三十八 文本分析方法,特别是主题建模。我们认为主题
37 建模有望在话语分析传统上
三十九 挣扎,例如规模化、重复性和系统化,这些都超出了
39 更简单的频率和搭配计数的贡献。同时,它
40 并不违背认识论前提和方法论精神
41 我们将展示更激进的话语理论。最后,我们提出
42 两个小案例研究展示主题建模如何使用
43 适当的参数 可以直接增强我们的能力
44 系统地调查和解释大量文本中的话语。

四十八 关键词:话语分析、主题建模、文本分析、语料库语言学、
46 方法论、霸权

47 简介

48 本文认为,主题建模是一种用于大规模文本挖掘的方法

49 个语料库,可以解决部分困扰话语分析的方法论困境,

50 研究文本意义建构的主要理论框架之一

51 演讲.1 话语分析旨在了解思想和现实如何在社会和社会中发挥作用。

¹ 作者非常感谢 Julie Birkholz、Anke Wonneberger、Christophe Verbruggen 和 Jan Orbie 以及对早期草稿的评论的匿名评审本文的作者。我们也非常感谢研讨会的参与者
“定性分析的定量工具:计算社会科学和话语的结合
由我们 Julie Birkholz、Anton Törnberg 和 Petter Törnberg 在

52 话语建构,但它可以在丰富的理论框架内实现的洞见
53 个属于其旗下的框架经常受到实际障碍的限制,
54 话语实证研究。而许多其他流行的定性文本形式
55 分析,例如内容分析,取得了令人印象深刻的方法论
56 意义建构、话语分析过程细致但缺乏理论化
57 遭遇相反的问题:某些类型的话语分析被认为
58 存在着成熟方法论上的缺陷 (Howarth & Torring, 2005, 25, 316-
59 22), 而整个领域被认为迫切需要更系统、更
60 严格的操作化 (Antaki 等, 2003)。我们认为主题建模可以
61 帮助话语分析克服一些实际障碍,并且
62 认为它可以有助于实现更加严谨的方法论,
63 意义建构研究的系统性。
64 使用语料库和数据提供的最重要的方法论优势
65 大规模文本分析工具众所周知:它们减少了所谓的研究人员偏见,
66 允许对语言使用的更细粒度和微妙方面进行精确研究,
67 促进方法论三角测量,并使系统化、大规模化成为可能
68 分析,以及重复和增量变化的研究 (Baker, 2006, 10-14)。这
69 自动文本处理工具在话语分析方面的潜力直接源于
70 这些优势,因为它们都位于“手工”话语分析
71 基于仔细阅读的结果有一定的硬性限制 (Antaki 等, 2003)。那不是为了
72 贬低对文本的仔细手工研究,这永远是

在艾伦举行的“计算社会科学中的社会挑战欧洲研讨会”
2017 年 11 月,图灵研究所在大英图书馆 (伦敦)。
这篇文章将激发和热情的评论、想法、建议和见解
从未实现过。

73 话语分析。但人脑只能吸收有限的文本,检测到特定的内容

74 微妙和细微的程度,并注意语言使用在一定范围内的演变。

75 在每一个领域,自动化工具都帮助我们超越这些限制,创造新的

76 见解和新颖形式的话语分析成为可能。计算机辅助语料库

77 换言之,分析并没有改变话语分析的本质,而是打破了

78 并不断突破其能力的界限。

79 然而,尽管语料库具有相当大的附加值,但其实际使用

80 话语分析在复杂程度上还不是很先进。大多数话语

研究大型文本语料库的 81 项分析使用相当简单的字数统计工具,

82 搭配和索引。更复杂的模型和算法,如主题

83 建模最近才进入话语分析师的考虑范围,

84 并在一定程度上 (Levy & Franklin,2014;Tornberg & Tornberg,2016a;2016b;

85 蒙克斯加德和德曼,2016 年; Jaworska 和 Nanda,2016 年是少数几个例子

86 明确使用主题建模进行话语分析)。这是值得注意的,因为

87 主题建模自 2003 年就已经出现。为了捍卫话语分析

88 社区,然而,主题建模和话语研究之间的无知是

89 相互。专门从事主题建模的学者和计算机科学家很少

90 个例外,对深入了解他们的

91 算法模型语言使用和意义构建理论主题建模

92 隐含地假设。² 因此,本文试图阻止

² 尽管有一些例外 (例如 DiMaggio、Nag 和 Blei,2013 年),但对于主题建模所隐含的语言模型在方法文献。最流行的想法包括像袋子这样的隐喻单词模型以及文档由主题组合组成的概念讨论 (例如 Mohr 和 Bogdanov,2013 年)。

93 主题模型与话语分析之间的相互忽视

94个社区。

95 主题建模本质上是一种方法,而话语分析则是一种

96 在本文中主要将其视为一个理论框架,我们的论点将

97主要采取解释前者如何帮助后者实现其目标的形式

98个研究目标。我们的核心目标是展示主题建模如何扩展

99话语分析在经验上能够取得什么成果,并消除一些理论、

100 个反对两者之间交叉授粉的方法论和实际反对意见

101 传统。此外,我们同样认为主题模型的用户可以受益于

102参与话语理论,因为它们帮助他们解释结果并

103 阐明他们对语言意义构建的理解。这样,

104我们寻求扩大对文本作为一个单元的数字文本分析的普遍理解

105分析,而不是作为一个单位的意义。在这双重努力中,重点将是

106兼容性、相互增值和理论契合度。

107 至于本文的结构,我们首先详细概述这两种方法,

108概述了它们的本体论和认识论前提,以及

109 特征。这些特征可以作为驳斥一系列

110 反对使用主题建模进行话语分析的理论反对意见

111第二节。我们认为主题建模的前提实际上非常适合

112与大多数话语理论所采取的本体论和认识论立场相一致。

113第三部分就主题建模如何扩展

114话语分析可以看到并争论:我们将解释为什么主题建模是

115 特别适合研究霸权问题;它有助于验证;

116 它所达到的系统性水平有助于我们追踪语言的变化和连续性

117 使用。本文的第四部分也是最后一部分包含两个实际示例

118 主题建模在话语分析中的操作化。案例研究

119 是主题建模能够使切实有效的

120 篇对话语分析的贡献,表明主题的操作化

121 建模在实践层面上可以相当简单。

122 什么是话语分析和主题模型?

123 借鉴乔根森和菲利普斯对这一传统的概括概述

124 (2002)、Blommaert (2005)、Rogers (2013) 和 Gee (2014) 认为,话语

125 分析本质上是研究沟通和意义建构

126 语境。话语分析是试图描述和理解过程

127 通过它在具体情境中形成、传达和解释意义。

128 通常,这种分析伴随着对这些

129 沟通过程影响着我们日常生活的社会世界

130 批判性话语分析 (CDA)就是一个典型的例子。

131 更具体地说,许多形式的话语分析,如埃塞克斯学派

132 话语理论或德里达解构主义,都归功于后结构主义

133 理解意义建构的生成,将其视为关系开放的

134 练习。该定义的关系部分意味着概念只能成为

重生玩转八零年代第135节

136 现实。开放性部分意味着这些关系不是必要的或预先确定的。

137 是确定的,但有偶然性,非必然性,从根本上说是不完整的。它们只是

138 它们以说话者的发音中所获得的形式存在。最后是“实践”

重生玩转八零年代第139节

140 是由行为者形成、表现和创造的,而不是

141 独立于它们而存在。

142 尽管相信意义最终是开放的并由演员塑造

143 表达这些含义时,所有形式的话语分析都承认,有些含义似乎

144 如此普遍和传统,以至于它们看起来正常和自然。这是

重生玩转八零年代第145节

146 分析,霸权。可以说,霸权意味着一种模式的特权

147 解释模式优于某一特定领域内所有其他可能的解释模式

148 (例如,“负责任的财政政策”通常被理解为债务削减,尽管它

149 可以想象也意味着对富人和中下层阶级征收更多的税

150 较少的)。更简单地说,霸权是指一种占主导地位的、规范化的理解方式

151 这使得一些说话和行为方式变得更加传统,

152 可以接受,而且看起来合乎逻辑。

153 话语分析处理哪种类型的数据和哪类问题取决于

154 使用的话语分析的多样性和风格。例如,CDA 主要着眼于

155 在涉及(隐性)政治的非常具体和有形的互动或声明中

156 维度,而话语理论反映了大规模的思想体系,例如

157 种族主义或新自由主义。然而,从广义上讲,大多数形式的话语分析

重生玩转八零年代第158节

159 部意义构建作品,旨在解构和理解这些观念

160 文本中阐述的内容构成。³

³ 尽管大多数话语分析方法都涉及文本研究,但有些方法强调需要

超越文本和言语所捕捉的内容,并观察实践和行动。

由于显而易见的原因,我们在本文中不讨论这种类型的多模态话语分析。

161 主题建模

162 同时,主题建模是一种旨在降低问题复杂性的方法。

163 通过将每个文本表示为“主题”的组合来构建大型语料库。名称略有不同

164 但具有误导性:主题是文本中重复出现的单词集群,但

165 将这些集群解释为主题、框架、问题或其他潜在概念(例如

166 作为话语)取决于方法论和理论的选择

167 分析师 我们将在下面讨论。虽然主题建模没有内置的

168 人类如何使用语言的模型,以下直观想法有助于理解如何

169 该方法有效。⁴

170 人类有多种语言使用模式来涵盖不同的内容

171 科目。我们沟通的方式数量是不确定的,而且几乎

172 无限的,并不是所有与某个主题相关的词语,也不是所有不同的表达方式

173 谈到它,在各种情况下都会使用。此外,还有很多词可以

174 根据上下文和用法获得不同的含义。利用这个想法,一个

175 一段文本(一份书面文件或一份演讲记录)可以表示为

重生玩转¹⁷⁶

177 最后选择一些与该说话方式相关的单词。主题建模

178 可以理解为这个过程的逆转,其中算法使用观察到的

179 语料库中文本中单词的分布,通常用于推断非排他性聚类

180 共同使用 每一种都代表一种关于特定主题的言语方式。⁵

⁴ 现有文献通常使用更简单但相似的启发式方法来解释该方法(例如。
莫尔和博格丹诺夫,2013)

⁵ 也就是说,如果设置了正确的参数。如果选择的主题太少,主题可能会涵盖
总的主题,一个更抽象的元主题(例如政治,而不是外交政策),或者一个
文本类型。我们将在下文中更详细地阐述这一中心点。

181 实际上,主题建模分析向用户返回三个主要结果(例如
182 个例子,见下文的说明性案例)。第一个结果将所有单词分配为
183 在语料库中每个主题的概率,通过对它们进行排名(使用讨论的启发式
184 上面)根据它们代表语料库中主题的概率(主题-
185 项矩阵)。根据使用的参数,会看到前五到二十个单词
186 作为主题的粗略代表,而主题本质上等同于以下列表
187 “热门词”。此输出是解释主题和研究关系的主要资源
188。第二个输出,即所谓的文档主题矩阵,指定如何
189 每篇课文的189大部分由各个主题组成。该信息可以与
190 个有关文本的上下文数据(作者、日期),以方便演员之间的比较或
191 历时分析。最后,算法会准确概述哪个主题
192 每个文本中的每个单词都已被分配。这有助于分析师掌握
193 每个单词的特定主题含义以及每个主题的上下文含义。
194 由于它完全脱离文本,因此该方法完全与理论无关,并且
195 感应式。因此,主题模型完全可以根据以下功能进行解释:
196 模型的参数及其可操作的更大的理论框架。此功能
197 在各种统计模型和算法程序中共享
198 位学者希望使用主题建模。最常见的模型建立在 Latent 之上
199 狄利克雷分配(LDA),一种由 Blei、Ng 和 Jordan (2003) 开发的方法,但在
200 过去几年,这一模式得到了扩展和完善。其中一个案例
201 本文使用原始模型,有时称为 vanilla LDA,而另一个模型
202 借鉴了结构主题模型,该模型整合了
203 计算机辅助文本处理(Roberts、Stewart & Tingley,2013)。
204 通常,使用主题建模的分析师试图识别一系列主题
205 兴趣,并用它们来定量研究文本语料库,测量

206空间致力于随着时间的推移或不同参与者的特定主题。在这个过程中，

207 主题大多被视为内容或问题显着性的衡量标准（例如 Jacobi、van

208 Atteveldt & Welbers, 2015），或作为问题框架（例如 Boydstun 等人,2013；

209迪马吉奥、纳格和布莱,2013）。然而,正如我们所讨论的,话语分析更关注

重生玩转八零年代第210节

211 患病率。那么,我们必须回答的第一个问题是,是否以及在什么情况下

212什么情况下话题可以包含话语而不是内容？

213主题建模与话语分析的兼容性

214 而上面关于主题建模如何拆解和表示的描述

215 对于熟悉话语分析观点的学者来说,文本听起来可能已经很有希望

重生玩转八零年代第216节

217主题建模的弱点确实值得将其用作话语分析的工具。

218在这方面,我们遵循两条论证思路。在元理论层面,我们发现

219 主题建模的假设与视图之间有很好的匹配

220话语是一种关系性的、开放的意义建构实践。从认识论的角度

221 在水平上,我们认为主题建模背后的方法论思想 它是如何的

222旨在生成关于语料库中的文本和单词的知识 符合

223 进行话语分析的分析过程。

224 我们为强调话语的理论兼容性而做出的巨大努力

225 分析和主题建模可能看起来像是一个相当哲学的练习,但我们

226坚信事实并非如此。由于许多形式的话语分析都坚持这样的观点：

227意义完全是象征性的,仅在语言和实践中产生,外部

228对于话语分析研究来说,验证在认识论上通常是不可能的,因为它们

229否认话语必然对应于外部现实。由于没有

230 外部验证的可能性,如果话语分析要实现内部有效性,那么内部有效性至关重要

231 避免在分析实践中陷入“任何事情都可能发生”的陷阱 (Antaki 等人, 2003 年)。这是

232 通过证明, 虽然分析所依据的假设是

重生玩转八零年代 第233节

234 至关重要, 以方法论上令人信服且正确的方式应用于当前的案例

235 (Marttila, 2015, 105-114)。我们在下面几页的论述旨在使这一点

236 与主题相关的话语分析的内部效度论证类型

237 种建模方法。

238 元理论拟合

239 正如所讨论的, 大多数形式的话语分析都会考虑单词的含义

240 保持关系和开放。这意味着意义是从单词所在的上下文中产生的

241 被使用, 并且它不是单词本身的固有特征。主题建模

我们认为, 242 与这种语言和意义观非常吻合。由于话题是

243 原始语料库中使用的所有单词的概率分布, 每个单词

244 在每个主题中都有主要人物, 其含义因主题而异。这是分析师的

[第 245 章] 第 245 章

246 与其他主题相关。同样, 单词的含义也是特定于主题的, 并且基于

247 出现在其突出的主题中的其他单词。这几点

248 确保主题建模方法与话语分析的假设相一致

重生玩转八零年代 第249节

250 (1) 主题模型明确地模拟了“多义性”(参见 DiMaggio, Nag 和 Blei,

251 2013), 单词可以根据不同的含义获得多种含义的概念

252 它们所处的上下文。事实上, 主题建模的作用可以概括为

253 追踪语料库中每个单词的上下文多样性

254 独立于一个词在其他领域或最常见的

255 假设。这意味着主题建模与开放性理念相同

256 话语分析所固有的意义。

257 (2) 主题本身通过 i) 它们建立的关系获得其意义

258 其中包含的单词之间的关系,ii)出现在

259 这些主题之间建立多个主题,以及 iii)通过频繁的合作

260 与其他主题一起出现。同样,单词通过链接获得意义

261 多个主题中的其他词。因此,主题建模共享的想法

262 话语分析所固有的意义的相关性。

263 在介绍主题建模时,我们提到主题可以被解释

264 框架、主题等等,但强调最合适的解释

265 取决于方法的使用方式 换句话说,取决于分析师的选择。尽管

266 关于主题状态以及主题模型实际代表什么的模糊性

267 可能被视为一个缺点,本文提出相反的观点,声称它给出了

268 主题建模具有显著的方法论多价性。我们认为主题建模

269 模型的构建应考虑特定的研究目标,而不是

270 统计优化,因为我们相信主题模型告诉我们并显示的内容

271 我们,在很大程度上取决于我们试图通过以下方式回答的研究问题:

272 模型和所分析的数据。因此,模型的参数应该是

第273章 被选中是为了促进这些研究问题的最佳答案,而不是

274 比实现最大统计拟合和显著性。⁶ 而不是调整研究

275 个问题,以便可以通过主题建模来回答,主题模型应该是

276 以回答研究问题的方式构建和解释。

⁶ 事实上,该技术的统计局限性导致存在多个“局部模式”,需要
由分析师进行调查和比较(参见 Roberts, Stewart, & Tingley, 2016)。

第277章 对一个主题所代表的内容的各种解释是紧密相连的
 278 事实上,主题的数量通常是由分析师预先选择的⁷。这
 第279章 最后一点引起了关于如何选择“正确”或“正确”的很大争议。
 280 个“自然”主题数(Arun 等人,2010;Wallach 等人,2009;Zavitsanos 等人,
 281 2008)。我们认为,这一争议不能通过量化措施来解决。
 第282章 仅统计主题质量;选择最终取决于分析师想要如何
 第283章 解释主题。虽然其中一些统计措施仍然有用(对于使
 284 候选模型的预选),我们强调定性解释的作用和
 285 在选择主题数量时,应考虑研究设计的要求。无论
 286 参数怎么微调,有些选择总是主观的
 287 由研究员制作。反射性地、有意识地处理主观选择是最好的
 288 分析师可以实现的目标,本文旨在为实现这一目标提供蓝图。
 289 使用主题模型作为话语分析的方法。
 290 建立主题建模作为一种适合话语方式的方法
 第291章 分析想要将意义建构作为一种开放的、关系性的实践来研究,其关键在于
 292 现在设计主题模型,以便他们可以追踪话语。我们的假设是,这
 293 如果语料库在主题和风格上足够连贯,并且
 294 主题总数足够大。在这种情况下,大多数主题都会
 第295章 不再列出语料库中涵盖的各种主题或主题,而是改为
 296 包含语言使用方面更细致入微的方面。无论哪一种
 297 分析师在更总体层面上青睐的更高级别实体(主题、框架、
 298 叙事等),通过增加主题数量或主题和风格

⁷ 在一些统计方法中,分析师选择影响结果的其他参数
 以类似的方式进行建模过程。

299 语料库的连贯性,这些可以分解为包含组合的主题

300 个词,可以解释为各种话语单位,通过这些

301 更高级别的实体被构建和组成。

302 这个分解过程将从较少数量的主题开始,主题越多

303 语料库的连贯性。如果一个语料库只包含来自单一类型的文本,并且讨论

304 特定主题集,高级实体将会减少,因此

305 分解将从较少的主题数开始。主题数在

306 个语料库仅包含贸易政策演讲 (如第一个案例研究中)与

307 随机收集的新闻文章、观点文章和文章中的主题数量

308 广告涉及各种问题。因此,主题和问题的分解

309 将主题转换为包含语言使用片段的主题将在前者更早开始

310 比后来的语料库中的要多,如果我们逐渐增加主题的数量。

311 简而言之,我们认为,通过使用大量主题,通过关注

312 一个明确的元主题 (如贸易政策或国民经济),以及

313 使用仅包含单一类型文本 (演讲、报纸文章)的语料库,

314 主题建模成为话语分析人员的有用工具。这个假设是

第315章 被人欺负了

第316章 应用于文献 (Tornberg & Tornberg, 2016a, 6-7; 2016b; Munksgaard &

317 Demant, 2016) ,但我们支持这一主张的最重要论据是理论上的。

318 至关重要的是,我们可以阐明 “主题”和 “主题”

319 通过引起人们对主题中事实的关注,将主题分解为 “话语”主题

320 建模中,文档不被分配到一个主题,而是被视为

321 主题数量 (Grimmer & Stewart, 2013) 。如何解释主题取决于

322 所选主题的数量,因为该数量会影响主题的 “粒度”

323 分解 (参见 Maier 等人, 2018 年) 。这背后的逻辑很简单:如果

第324段 由于整体文档的增加,归属于单个文档的主题也随之增加
主题模型中的主题数量为325,这显然并没有增加数量

第326段 文档中讨论的主题或问题。相反,涵盖每个主题的主题数量
327文件中提到的主题不断增加,其中一个主题
328个主题特征,每个特征都包含该主题的不同方面,不同的
329种表示它的方法,以及谈论它的不同方式。

330 因此,通过增加

第331段 语料库中的主题总数表明该文档来自主题集合
332 变成代表这些主题的语言使用模式的集合,每个模式
333 成为主题的一部分。⁸ 换句话说,更高级别的实体主题建模识别
334 在语料库中,例如主题、框架或叙述(当出现的数量
335 主题很小),可以分解为构成较小级别的实体(这
336(当主题数量较多时出现)通过增加主题总数。

第337章 我们的主张显然是,在某些情况下,可以解释这些较小级别的
338 借助话语分析将实体视为话语元素。

339 如果确实如此,我们可以追溯语料库中的各种话语是如何
340 建造,以及它们出现的地点和时间。从第一个案例可以看出
341 研究中,单个话语常常是由多个话语元素组成的
第342段 主题。这意味着通过研究主题之间的关系(无论是在
第343段 定量共现和定性联系),我们可以揭示如何

⁸ 虽然部分依赖于所使用的源材料和语料库的预处理,但它是
一个主题中包含的一些语言模式几乎没有解释性,这是很常见的
价值。随着主题总数的增加,这个数字会明显增加,但随着
这些“无意义”的主题在小 k 模型中通常被忽略,没有理由不这样做
在大 k 模型中略过它们。

344 个话语是由较小的话语元素组装和配置的。相似地，

第345章 如果我们研究谁何时使用话语准主题，我们可以揭示以下模式：

第346章 特定群体在特定时间点使用的言论。

第347章 值得注意的是，主题模型不会自动进行讨论

第348章 主题数量增加时的分析；较大主题模型的主题确实

349 根据定义不包含话语元素。我们只是认为它们所包含的内容

350 可以解释为话语元素，如果我们理解单词之间的关系

351 它们通过话语分析的视角来揭示。因此，增加主题的数量

352 不一定将主题话题分解为话语话题。在某些情况下，它

第353章

354 需要分析性解释来理解它们并梳理出话语

355 个元素。因此，主题建模不能完成话语分析员的

第356章 为她或他工作，这只是一个促进他努力的工具。

第357章 认识论契合

第358章 除了符合语言使用和意义构建的理念之外

第359章 话语分析遵守并包含实施的实际可能性

360 这个想法，主题建模作为一种方法也允许主观解释的空间

361 分析师，这同样是话语分析的核心要素。

362 作为一种无监督方法，主题建模本质上是一种归纳方法

363 到语料库分析。这与监督技术相反，在监督技术中，分析师预先

364 定义类别或量表，训练算法来准确地重现它们，然后

365 以演绎方式将评分/分类扩展到整个语料库（Grimmer &

第366章 Stewart,2013 年)。⁹ 主题建模仅仅代表了

367 语料库,对用于分析的文本之外的任何内容一无所知。

368 因此,分析师的任务是解释和理解主题

369 模型向他或她展示了语义关系和意义构建过程

语料库中有370 个工作。在解释模型结果时,分析师可以而且应该

第371章

372 语料库的来源。因此,分析师的主观输入继续发挥

373 正如话语分析中所保证的那样,发挥着至关重要的作用。人们可以说而不是做

374 分析工作本身,算法为分析师提供了一个浓缩或

375 对大型语料库进行转换,然后分析师发布分析结果本身。

376 该算法通过将某些词放在

377 不同的主题,并且某些单词相互链接以形成更大的单元

378 意思。但分析师的工作就是解释一个词的不同含义

重生玩转八零年代第379节

380 我们可以通过说明该方法如何概述来使这个想法更加具体

重生玩转八零年代第381节

382 话语分析。例如,如果一个术语仅归属于一个特定主题(即其

383 其他主题中的概率低得可以忽略不计),该主题可以说包含霸权

第384章 该术语在语料库中的解释:该主题中包含的其他单词

第385章

386 由于不存在其他解释,该语料库占据主导地位并标准化。在一个

第387章 主题模型中,“利润”一词只出现在一个主题中,周围是

⁹ 一些更高级的主题建模允许将算法用作

监督工具也是如此(McAuliffe & Blei,2008)。这些版本显然被排除在外
从我们在这里提出的论点来看。

388个词,如“贪婪”、“剥削”、“老板”、“资本家”、“牟取暴利”,一目了然

389 霸权主义对盈利的解释是负面的、反

390 资本主义的。

第391段 如果一个概念在与某个主题相关的许多主题中重复出现,则可以使用类似的逻辑。

392 某个问题:在这种情况下,这个概念大概是霸权主义的共同构成部分。

393 话语,只要其含义在不同的上下文中保持稳定

394包含在不同的主题中。如果“增长”一词再次出现在五个不同的主题中,

第395段 主题分别为财政审慎、社会福利、政府目标、

396健全的经济政策和古典经济学,每次都具有类似的积极

397内涵,它可能是霸权经济观的重要组成部分

398 在语料库中明确表达。

399 如果它出现在不同的主题中会导致不同的含义

400归因于一个概念,然而,我们可能正在目睹一场关于其概念的斗争

401 解释。如果“增长”在气候变化和

第402段 不平等,但在消费者福利和商业话题上却具有积极含义

403 健康,语料库很可能包含关于如何表示该术语的争论。

404 我们在这里提出的关于认识论和元

405 主题建模和话语分析之间的理论契合也隐含地包含

406 我们认为主题建模可以从明确参与中受益的原因

407 话语理论。自动文本分析工具总是包含隐含的和

408语言和意义生成模型必然是不完善的

409语言作品(Grimmer & Stewart,2013,3-4)。话语分析理论

410意义建构帮助我们阐明我们如何看待这种必然不完美的

411模型看起来像主题建模,让我们反思如何将其与主题建模相协调

412 理论上严谨的实证文本分析。

413 主题模型与话语分析相结合的附加价值

414 我们已经预示了使用主题建模的一些原因

415 话语分析可能是可取的,当我们建立了理论基础来做

因此。在下文中,我们将这些建议更加明确,并提供更

417 有力的论据,说明如何使用主题建模突破界限

418 话语分析可以实现实证研究。使用大型语料库的广泛好处

第419章和简单的软件工具已经被深入讨论过 (Baker,2006;

420 Kennedy,2014 年)。本节重新探讨了其中一些主题,但特别关注

421 为什么主题建模特别有利于话语分析师愿意

422 参与其中。它提出了至少三个维度,其中是这样的:研究

423 霸权,语境中的语言研究,以及验证和系统化。

424 话题建模与霸权研究

425 话语分析从主题中受益的最具创新性的方式

426 建模,在后者中促进了一种研究文本霸权的新方法。

427 话语分析通常着眼于相当小的数据体,因为其局限性

它采用的手册、仔细阅读方法。结合其感应

429 方法,这意味着话语霸权的研究往往被迫集中于

霸权崩溃或建立的430个时刻,最能了解其霸权

自然 (Wood & Kroger, 2000, 34;Gee, 2014, 37-38;Jorgensen & Phillips, 2002, 138-

432 174)。归纳逻辑赋予以下观察更大的示范价值:

433 打破或创造一种模式,而不是确认它的模式。如果你什么都不知道

434 关于天鹅,发现一群十只白天鹅可以给你很多信息。

第435章你看到的接下来的二十只白天鹅并不会增加你对天鹅的了解,但是

436 发现一个黑色的斑点。

437 同样,观察一些明显霸权的
438 解释被毫无问题地再现,并不能很好地说明霸权主义
439 逻辑是有效的;然而,它被提出、拒绝或质疑的例子却远远
440 更多信息。例如,在经济领域,学者们研究了
441 当前自由主义金融和资本思想霸权的特征
442 回顾自由主义解释战胜
443 18 世纪和19世纪的替代概念。他们也非常重视
444 当代历史上的稀缺时刻,当这种偶然状态
445 短暂演绎再现其霸权短暂顺利再现
446 由于外部错位或主动电阻而出现故障 (De Goede,2005;
447 吉布森-格雷厄姆,2006) 。

448 虽然这种方法在认识论上是有道理的,但它并不完全符合
449 话语分析如何从根本上理解霸权
450 本体层次。研究十九世纪资本主义全球化的胜利
451 关注其替代方案以及资本主义在危机等疲软时刻后的复苏
2008 年的452确实告诉了我们很多关于它的特征,就像第一只白天鹅和
453 罕见的黑天鹅就是这样。但霸权包含着正常化的概念,
454 标准化。它是指对一个想法毫无疑问地接受为常识
455 这不是天生的,也不是不可改变的。因此,在达成共识的时刻
重生玩转八零年代第456节
457 明显没有问题且毫无疑问的重复。只看特殊但
458 个关于崩溃或制度的信息性实例意味着我们以一种相当
459 间接派生方式:我们假设它的存在,然后寻找它的根源或它的
460 瞬间崩溃。

461 寻找形成规律和常态的模式、惯例和逻辑

第462章 我们日常生活的语法是一种更真实地理解霸权的方法

463话语分析 (Glynos & Howarth, 2007)。然而,由于这些常规和正常的情况

464只包含较少的独特信息 (它们是白天鹅 10 到 30), 它们较少

第465章 具有指导意义。这是一个精读语篇分析的问题,对于实际应用来说

第466章理由只着眼于少数案例,因此存在以一概而论的风险

467信息量过于有限。

468 这就是主题建模的作用所在,因为它为我们提供了一种解决方法

第469章 这个问题。它使我们能够用以下几个非常有见地的案例来补充

第470章 霸权的无数正常的、平凡的、个人无益的

第471章 可以毫无障碍地再现。后一种类型的数据可能教育意义不大,但它们

472 更加丰富,通过主题建模我们可以概述大量

473全面地看待它们 (继续天鹅的比喻,我们可以看看成千上万

474千只天鹅)。因此,由于主题模型可以帮助我们检测什么是

475在文本语料库中不断重复出现 (或不断缺失但被假定出现), 它们

第476章

477标准化及其随时间的微妙转变和调整。

478 虽然其他更简单的定量工具可以实现类似的过程,但它们

第479章 需要研究人员对研究的性质和内容做出一些假设

第480章 霸权话语。例如,关键词频率分析只有在以下情况下才有效:

481知道驱动霸权解释的关键词。主题建模,

[第 482章]

483或先验假设,什么可能被视为正常。这有助于我们发现

484 个例程和规范化逻辑,否则我们可能无法发现,确切地说

485因为我们在多大程度上将它们视为理所当然并认为它们是理所当然的。

486 主题建模和语境中语言使用的研究

487 主题建模还有助于文本中单词的研究

第488章 背景。首先,大多数主题建模工具不仅仅为分析师提供

第489章 哪些文档中存在哪些主题 (文档-主题矩阵)的概述

[第 490 章]

491 语料库中的文本。这创建了一个快速实用的程序来在

492 主题模型作为语料库和文档中语言使用的聚合

第493章

第494章

第495章 等,2003,13-16)。

第496章 其次,主题建模同样可以帮助我们避免相反的问题,过度

第497章 分析时过多关注特殊的背景细节 (Antaki 等人,

498 2003)。在这方面至关重要是主题建模将每个单词分配给一个

499 话题。因此,我们不能轻易地跳回文本上下文,文本

500 上下文本身也是量化的。这有助于采用系统的方法来研究

501 文本上下文,因为可以整体跟踪哪些主题在文章中占主导地位

502 以关键词、主题或话语为特征的文本。因此,通过主题建模,

503 文本语境研究也可以量化和系统化。这有助于

504 分析师避免从一项具体陈述中得出仓促的结论,并让他或

第505章 她轻松地概述了术语、主题或话语所处的各种上下文

506 已使用。

507 请注意,这种可能性比简单文本具有重要优势

508 分析工具可以量化所选关键词的 (共同)出现,但不能

第509章 量化出现在其周围的单词 这意味着该术语的上下文

510 个正在分析的数据仍需要人工读取、解释、统计和分析,

511 造成总结性分析不足或局部过度解释的风险。正如它

第512章 512

513单词和话语出现的上下文,主题建模是一种有价值的

重生玩转八零年代 第514节

515语境在意义建构中的重要性。

516话语分析中的主题建模、验证和系统化

517 第三,主题建模满足了可复制性和

第518章话语分析的系统化。这两个概念中的第一个可能是

第519章

520话语分析。但即使在解释主义框架内,也需要

521证明一个人的语境解释确实代表了

重生玩转八零年代 第522节

523无论是有意还是无意 (Johnston,2002;Louw,Todd & Pattamawan,2014;

524贝克和莱文,2015;毛特纳,2015)。事实上,有人建议,话语

事实上,分析很容易犯使用其数据进行预预测的错误。

现有观点 (Rogers,2013,74;Antaki 等,2003,19-21;27-30)。

主题建模在这方面有两个方面的帮助。首先,显然

528 创建量化选项。数据的代表性和

解释的意义和可靠性可以通过统计来证明。第二,

530或许更重要的是,考虑到话语分析的精神,主题建模

531 有助于定性验证我们的解释是否有意义,即使

532正在处理一个非常大的语料库。正如我们所讨论的,将主题解释为

533有意义的单位仍然是分析师的任务。这种解释借鉴了

534阅读文本、主题知识以及个人观点和经验。

第535章

536 个主题词矩阵以及属于该主题的单词的具体出现率

537 在整个语料库的文本中,分析人员还可以验证他或她的

538 对主题的解释表面化,用他或她对这个主题的解释来诠释

539 当他或她遇到文本中分配给它的单词时。这样,分析师

第540章 不服气的人

第541章 话题模型中的单词和话题的出现

所分析文本中有542 个具体表述。

543 此外,系统性主题建模允许检测

重生玩转八零年代第544节 这样,将

545 关注语言使用细节是话语分析的特征,可以

546 的练习次数是精读的几倍。这使得

547 主题模型是分析话语演变的合适工具

548 并逐步变化,例如 (Jaworska & Nanda,2016) 。如果时间跨度

549 变得太长或变化太微妙,如果一个人

550 完全依赖于仔细阅读。同样,主题建模为比较提供了空间

551 话语研究。它有助于研究差异、相似之处、

552 以及不同政党政治家言论的变化。

553 个 案例研究

554 到目前为止,我们已经在相对抽象的层面上讨论了主题建模如何

555 促进话语分析,讨论其本体论、认识论和

556 方法论契合。在以下两个案例研究中,我们旨在展示这种

557 组合在实践中发挥作用,使之前的一些见解

558 个 部分更加具体、切实。

559 首先,主题建模和主题建模的元理论和认识论契合。

560 话语分析正在展出,因为这两个案例清楚地展示了大主题的主题

561 个模型包含词汇集合,分析师可以主观地和反射性地
562 解释为话语元素。第一个主要是在质量上,第二个
563 一个混合了定性和定量解释的案例。此外,两个案例研究
564 还提到了主题模型如何促进话语霸权的研究,以及
第565章 第二个案例进一步努力展示其对于语言使用研究的效用
566 在上下文中和内部验证中。然而,必须强调的是,这些
567 个案例研究绝不是成熟的、独立的分析。他们不呈现
568 自给自足的实证研究或结果,而只是试图说明一些
569 上面提出的抽象方法论论证。
570 我们的第一个案例涉及国际贸易政治演讲的话语
571 通过建立干预措施主题模型,在欧洲议会制定政策
572 1999 年至 2016 年期间的议会全体会议。¹⁰该案特别关注
573 我们如何能够将话题中包含的话语元素解释为霸权和
574 标准化。第二个案例研究分析了奥地利报纸文章,
575 反过来,批判性地质疑了支持增长的立场在经济中占据霸权的观点。
576 公共经济话语。在这里,使用话语模式的分解
第577章 不服气

¹⁰ 我们使用 MALLET 构建了一个标准的 LDA 主题模型,包含 120 个主题,构建了超过 3,500 个主题
每 20 次迭代进行一次超参数优化,并进行 40 次磨合,
我们与其他具有较高和较低主题计数的模型进行了验证。它基于
语料库,包含 11.744 条经过预处理的讨论国际贸易政策的演讲
1999 年至 2016 年间的欧洲议会全体会议,我们对其进行了词形还原
并从中删除了停用词。演讲摘自
欧洲数据库使用 SPARQL 关键字查询特定于 121 个术语和短语
国际贸易政策。

第578章 重生之人

579霸权。

第580章 案例1:欧洲议会的贸易政治

581 在欧洲议会案例研究中,我们正在寻找现存的

582贸易政策制定中的霸权。政策领域的霸权实践可以

583被视为构成任何政治家在行使职权时必须遵循的正常、适当的规则

第584章在这个政策领域内行动(参见Glynos和Howarth,2007年关于社会逻辑的内容)

第585章更详细的讨论正是在这种常态化、社会化的基础上

586 本质上,我们可以着手开发一种启发式方法来研究霸权,因为它对

587期望这样的规范化会在语言使用中留下痕迹。主要的经验

588 我们预计任何形式的霸权语言使用都会表现出的特征是,

589 定义,它持续地表现出来,无论说话者是谁,他的意识形态如何,

590 具体(子)问题或时间。因此,我们寻找代表相当

第591章 语料库中每个演讲的稳定份额。¹¹我们发现出现率的主题是

第592章 在所有演讲中都相对稳定,无论何时或由谁发表,

593然后被定性地解释为话语(片段)。这给我们留下了几个

594个相关话语,我们在这里分析其中两个,分别由两个和

¹¹ 每次演讲中每个主题所占份额的变异系数(CV)被用作衡量标准这种稳定性。CV较低的主题在所有文本中具有较低的标准差与每篇文章的平均份额(以及在语料库中的份额)进行比较。更简单地说,CV值较高的主题通过以相对较高的比例出现在语料库中来增加其份额在相对较少的演讲中,所有其他演讲。CV较低的主题通过代表相对稳定的主题来获得份额每次演讲的份额,在任何方向上都没有大量显著的异常值。后者显然是这里感兴趣的主体。

三个不同的主题。

[此处插入表1]

主题 15 和 20 被解释为将贸易确立为一种围绕

围绕合作与伙伴关系。系统地将国际贸易与条款联系起来

比如“关系”、“伙伴”、“合作”、“关系”和“伙伴关系”,这些主题

代表贸易实践涉及合作、联系。术语

“协议”、“好处”、“支持”、“重要”、“重要”、“必要”和

“促进”进一步为这种伙伴关系注入了积极的情绪。贸易作为

合作伙伴之间的关系被认为对相关各方有利,因此,

必然会有所追求。事实上,所有欧洲议会议员都借鉴了

这些主题中所包含的话语表明,他们都认为贸易

两国关系应得到进一步推动、支持和发展。贸易关系是

被认为重要,甚至必不可少。当然,议员们对什么

那些贸易关系应该是什么样子,或者关于贸易的积极潜力应该如何

换句话说,仍然有可能讨论不想要的负面影响

特定贸易政策的后果,或辩论商业政策是什么

实现贸易关系的内在福祉所必需的,但更加

基本层面上,贸易似乎总是被呈现为一种固有的

积极的。

第二组主题 (55、90、99)共同包含组织论述,

阐明贸易和贸易关系始终融入更广泛的全球范围的理念

体系。诸如“秩序”、“体系”、“世界”、“全球”、“组织”、“多边”等术语

和“框架”就暗示了这种趋势,也是相对稳定的方式

参考了 WTO 的第 620 条。贸易关系不仅仅是相互之间孤立的联系

621 合作伙伴,他们被话语性地构建为构成一个更大的整体,一个全球性的

622 交易系统。这三个主题中的其他术语,如“机会”、“利益”和

623 “繁荣”,暗示了之前合作论述的一些延续,

624 确立了贸易是一种互利伙伴关系。同样,在

625 包含合作话语的某些方面,我们也可以找到

626 组织,通过诸如“秩序”和“框架”等术语。合作伙伴维护并

第627章合作的好处在一个大的整体上发挥出来。

628 这种有组织的话语给出了贸易由一种

629 逻辑、有序的特征。贸易关系网络有一个结构,一个

第630章组织的一致性,但这种结构不会自动出现。它需要

第631章被人欺负了

632 “实施”和“改革”。贸易关系的结构性并不是

本质上,这些术语在话语中的存在表明,政治干预是

634 实现这一目标所需的条件。因此,贸易必然涉及政策制定,因为贸易关系

第635章需要建立系统。同样,干预和政治行动的类型

636 政治家们希望看到整个议会的意见可能存在很大差异,但

第637章

638 贸易关系似乎是理所当然的,与政治意识形态或国籍无关。

639 案例 2:奥地利公共增长辩论

640 在第二个案例研究中,我们感兴趣的是新闻媒体如何理解

641 经济增长和重大经济危机。经济增长是一个突出的概念

642 在政治和学术研究方面都做出了许多学术贡献,并受到公众的关注

第643章致力于其原因和后果。但有些令人惊讶的是,研究

第644章公众对经济增长的理解和态度是相当罕见的。这

645稀疏的文献提出“增长霸权”(例如 Schmelzer,2016)

重生玩转八零年代第646节

647论点,例如强调环境问题,或将经济增长与

648 不平等加剧和其他社会问题。本研究将增长霸权

第649章假设要在一个特殊案例中得到检验:奥地利媒体。

650 我们分析了有关经济增长的报纸文章¹²,和

第651章剖析有关该主题的公共言论并调查他们如何

第652章随着时间的推移而演变。首先,我们确定了与经济增长相关的主题并进行了研究

第653章随着时间的推移,它们的显着性。接下来,我们确定了这些主题中的一些论述,并且

第654章研究了他们如何将其表达为一个概念。然后我们分析了之间的相关性

第655章 主题13, 调查这些话语的不同元素通常如何

656文章内合并。这使我们能够分析霸权和非霸权

657报纸上有关成长的深入论述。我们的期望是

658篇带有明确或隐含支持增长立场的论述占据了主导地位,

重生玩转八零年代第659节

¹² 选取主要报纸 (《新闻报》、《标准报》、《皇冠报》
Zeitung,Kurier 和 Kleine Zeitung) ,2006 年 9 月至年底出版
2016年8月,且至少包含以下关键词之一:“经济增长” ,
“不平等”、“可持续性”、“就业”和“失业” 相关关键词
关于经济增长的争论。该语料库总共包含 52,593 篇文章。

⁺³ 在本例中,我们使用了结构主题模型 (Roberts, Stewart, & Tingley 2013),
允许并建模主题相关性,从而实现这种类型的查询。预处理后
(停用词删除、词干提取和删除提及次数少于 15 次的单词) ,我们运行
具有不同参数的多个模型。此处给出的结果基于
模型包含 200 个主题,并与具有相同或更低主题的其他模型进行了验证
主题很重要。

660

[此处插入图1]

第662章

我们发现,毫不奇怪,经济危机在

2008-2011 年期间 (图 1 左上图)。正如预期的那样,该主题占据了

与危机相关的话语以负面的术语来描述经济衰退。例如,

关键词“戏剧性”、“严重”、“迷失”和“大规模”描述了该主题

负面情绪和合法的直接促增长政策干预。这

相关 (校正系数 = 0.28)复苏主题明确地将严重的危机与

“复苏”以“强劲”“增长”为标志。定性地考察一些文章

在这两个主题以及与危机主题相关的其他主题上得分很高的人 (即

乐观和预测)证实了这一解释。这与我们的

猜想公共话语强调促进经济增长作为公众

良好的和理想的政策结果。因此,特别是在

危机相关的新闻报道中,经济增长乏力被视为一个主要问题,揭示了

有利于增长的立场。

第676章

[此处插入表2]

第678章

然而,尽管这种对经济增长的总积极态度

在 2008-2011 年期间表现强劲,但在后来几年逐渐消退。公众

辩论变得更加平衡,为批评经济的言论提供了空间

增长与全球经济“模式”的总体情况。这段话的主要内容是

被人欺负了

社会正义、不平等和民主。批评部分可以从

685 个关键词 使用“资本主义”或“新自由主义”等术语已经表明了

686 批判立场 但仔细阅读一些文章样本后,我们发现批评更多

687 深刻的。为了说明这一点,在中右翼报纸《Die》上发表的一篇文章

688 Presse 指出:“因此,增长放缓是生存的必然结果。但它要求

689 不同的经济、生活方式、文明和社会条件的变化”。最多

690 文章并不像例子那样强烈地站在批评者一边,但通常提出

691 从平衡的角度进行批评。

692 回到图 1,展示主题显著性随时间的变化,我们看到 2011 年之后

693 个有利于增长的话题和那些更具批判性和反思性的话题在

694 随着时间的推移语料库。我们将此解释为公众讨论的一个迹象

695 经济增长目前并不像乍看起来那么霸权

696 至少在奥地利是这样。这个论点很大程度上取决于我们如何使用主题模型

697 在这种情况下:我们从语料库检查转向定性研究、解释和

698 验证,到(说明性)量化。

第699章 最后的话

700 在本文中,我们认为主题建模可以为

701 话语分析。我们讨论了将话语分析与

第702章 和主题建模,讨论了它们的理论兼容性,假设

第703章

704 通过两个例子说明这种组合的可行性,说明必要的

第705章 有条理的分析步骤。我们想在这个结论中强调的是,

重生玩转八零年代 第706节 并非所有话语

707 分析可以而且应该使用主题模型进行。虽然主题建模具有

708 将文本解构为其话语元素的潜力,无论这是否适用于

第709章 具体案例取决于熟悉理论、方法、

第710章 最终,主题建模并不能转换话语

711 分析为精确或定量科学;而不是解决所有挑战,

712方法改变了一些需要提出的关键问题。

713 需要重新表述的第一个问题涉及

714 代表性。手工话语分析经常面临工作的批评

715数据有限,不适合对整个学科领域的论述提出主张

716 实践。关于解释的范围和适用性的主张通常是相当

717模糊(即“分析的许多文章”、“很少发现的功能”)。因此,读者

第718章 当分析师声称他们的材料是

719 足够丰富,可以进行概括。系统化主题建模

720引入话语分析帮助分析人员通过促进赢得读者的信任

721 更大的语料库,允许他们展示整个语料库(而不仅仅是样本

722或插图),并通过透明化它所代表的

723 个单独的主题(从而揭示了他们的论点的范围

724不适用)。尽管如此,系统化的改善并没有使信任问题

725 消失,它只是改变了它。因为分析师选择了用于

726 进一步分析潜在的无限数量的替代方案,挑剔的读者现在

727必须相信语料库的模型确实具有代表性,并且他们可以

728挑战分析师通过展示替代模型来验证这一说法。

729 其次,主题建模改变了我们对解释的看法。

第730章 传统的话语分析通常会在文本中引用说明性的引文来显示

731 数据和分析师工作之间的关系。读者是否接受

732 分析师对数据的解释取决于他或她是否相信

733 分析师没有挑三拣四,也没有被确认偏差误导。为了验证这一点

第734章 不是这样的,验证分析,挑剔的读者可以积极寻找

735 个反例。主题建模简化了这个过程,因为它允许

[第 736 章]

737 因此,应适用对该主题的特殊解释。读者现在可以批判性地

738 透明地评估特定词语组合的解释,并且

第 739 章 详尽无遗,而不是必须相信所提供的片段总结了

第 740 章 分析师的工作很好。不过,如上所述,读者不得不问,类似的

741 可以使用替代模型规范得出解释和结论。

第 742 章 最后,使用主题模型将

第 743 章 一项具体研究的结果以及有关霸权和权力的更大主张。

744 手工话语分析通常通过观察关键节点来研究霸权

745 和激烈的话语斗争,声称现存的曲目是霸权的。

第 746 章 我们认为,主题建模允许话语分析将焦点转向日常生活

第 747 章 正常和常规。在本阅读中,主题始终出现在

第 748 章 因此,整个语料库可以构成霸权曲目的表征。但有一个话题是

749 仅仅是词语的集合,其意义取决于它与其他

750 剧目正在发挥作用。例如,我们能否相信,在一个主题中捕捉到的话语

751 不会自我反省吗?如果是这样的话,同样的话也可以用在

第 752 章 在不同的地方和不同的时间,但它们的意义却不一样。

第 753 章

754 使用特定主题的各种上下文,以及其含义是否

其中捕获的 755 确实始终很稳定。

756 这些观点表明,主题建模推动了方法论

757 话语分析的界限,而不是完全废除它们。然而,尽管这些

758 转型也带来了新的限制,但我们相信潜在的收益是值得的

759 尝试使用主题建模,我们鼓励研究人员在

766 话语分析传统来探索主题模型如何丰富他们的技艺。

761 同样,我们认为,经常从事主题建模工作的研究人员会

762 批判性地反思他们的解释实践,并从参与中获益

重生玩转八零年代第763节这种相互参与可以

第764章 为以新方式解决旧方法论问题开辟途径,他们可能

765 甚至引发了一些迄今为止尚未引起关注的全新争论。

第766章 参考文献

767 Antaki, C., Billig, M., Edwards, D. 和 Potter, J. (2003)。话语分析意味着做

768 分析:对六个分析缺陷的批评。在线话语分析

第769章 1(1)、1-39。

770 Arun, R., Suresh, V., Madhavan, CEV 和 Narasimha Murthy, MN (2010)。

771 用潜在狄利克雷分配找到自然数量的主题:一些

第772章 观察。亚太知识发现与数据会议

773 矿业。柏林:施普林格。

第774章 贝克, P. 和莱文, E. (2015)。挑选合适的樱桃?基于语料库的比较

775 以及关于男性气质的新闻文章的定性分析。话语与

第776章 通信, 9 (2), 221-236。

777 Baker, P. (2006)。在话语分析中使用语料库。伦敦: A&C Black。

第778章 Blei, D., Ng, A., & Jordan, M. (2003)。潜在狄利克雷分配。机器学习杂志

779 学习研究, 3, 993-1022。

780 布洛马特, J. (2005)。话语:批判性的介绍。剑桥: 剑桥

第781章 大学出版社。

782 Boydston, A. (2013)。制作新闻:政治、媒体和议程设置。

783 芝加哥:芝加哥大学出版社。

784 De Goede, M. (2005) 美德、财富和信仰:金融谱系。明尼阿波利斯:

第785章 明尼苏达大学出版社。

786 DiMaggio, P., Nag, M., & Blei, D. (2013)。利用主题建模之间的亲和力

787 以及文化的社会学视角:应用于报纸报道

788 美国政府艺术资助。诗学, 41(6), 570-606。

789 吉, J. (2014)。话语分析简介:理论与方法。伦敦:

790 劳特利奇。

- 791 Gibson-Graham, JK (2006). 后资本主义政治。明尼阿波利斯:明尼苏达大学
明尼苏达出版社。
- 793 Glynos, J 和 Howarth, D. 社会和政治理论中的批判性解释逻辑。
- 794 伦敦:劳特利奇。
- 795 Grimmer, J., & Stewart, B. (2013). 文本作为数据:自动化文本的前景与陷阱
政治文本的内容分析方法。政治分析,21(3),267-297。
- 797 豪沃斯 D 和托芬 J. (2005 年)。欧洲政治中的话语理论:身份、
政策与治理。伦敦:Springer。
- 799 Jacobi, C., van Atteveldt, W. 和 Welbers, K. (2015)。大数据定量分析
使用主题模型来分析大量新闻文本。数字新闻,4(1),
89-106。
- 802 Jaworska, S., & Nanda, A. (2016)。通过说好话来做好事:主题建模
企业社会责任的辅助话语研究。应用
语言学,14,1-28。
- 805 约翰斯顿 H. (2002 年)。框架和话语分析中的验证和证明。B。
克兰德曼斯和 S.斯塔根堡。社会运动研究方法 (62-91)。
明尼阿波利斯:明尼苏达大学出版社。
- 808 Jørgensen, M 和 Phillips, L. 《话语分析的理论与方法》。伦敦:Sage。
- 809 Kennedy, G. (2014). 语料库语言学导论。伦敦:劳特利奇。
- 810 Levy, K., & Franklin, M. (2014). 驾驶监管:使用主题模型来检验
美国卡车运输业的政治争论。社会科学计算机
评论,32(2), 182-194。
- 813 Louw, S., Watson, T. 和 Pattamawan, J. (2014 年)。采摘成熟的樱桃:提取物
定性研究中的选择。国际会议论文集:
应用语言学研究 (155-169)。曼谷:蒙固国王
吞武里科技大学。
- 817 Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A. 等人。
(2018)。LDA 主题建模在传播研究中的应用:迈向
有效和可靠的方法。沟通方法与措施,12 (2-
3), 93-118。
- 822 玛蒂拉, T. (2015)。后基础话语分析:从政治差异到
实证研究。柏林:施普林格。

第823章 莫特纳,G. (2015) 。制衡。语料库语言学如何做出贡献

824 CDA. R. Wodak 和 M. Meyer 批判性话语研究方法 (154-179) 。

825 千橡市:鼠尾草。

826 Mohr, JW 和 Bogdanov, P. (2013) 。简介主题模型 :它们是什么以及为什么它们很重要。诗学 ,41(6) ,545–569。

第827章

828 Munksgaard, R., & Demant, J. (2016) 。混合政治与犯罪 犯罪的盛行和加密货币市场政治话语的衰落。国际药物

829 政策 ,35,77-83。

830

831 Roberts, M., Stewart, B., & Tingley, D. (2013) 。结构主题模型及其应用社会科学。在 NIPS 2013 主题模型研讨会上发表的论文:

第832章

833 计算、应用和评估。太浩湖。

第834章 罗杰斯 (R) (2013) 。批判性话语分析。 A. Trainor 和 E. Graue, 审阅

835 社会科学中的定性研究 (66-81) 。伦敦:劳特利奇。

第836章 施梅尔泽,M. (2016) 。增长霸权。剑桥:剑桥大学

第837章

838 Törnberg, A 和 Törnberg, P. (2016a) 结合 CDA 和主题建模 :分析网上伊斯兰恐惧症和反女权主义之间的话语联系

第839章

840 论坛。话语与社会 ,27(4) ,401-422。

841 Törnberg, A 和 Törnberg, P. (2016b) 社交媒体话语中的穆斯林 :结合主题建模和批判性话语分析。话语、语境与媒体 ,13,

第843章

844 Wallach, H., Murray, I., Salakhutdinov, R., & Mimno, D. (2009) 。评估方法对于主题模型。第26届国际年会论文集

845 机器学习 (1105-1112) 。纽约:计算机协会

846 机械。

847

848 伍德 L. 和克罗格 R. (2000) 。进行话语分析:研究行动的方法在谈话和文本中。千橡市:Sage。

850

第851章 表格和数字

852表 1. 关于欧盟贸易政策的霸权话语。

ID	名称	15 关联性	简历	特色词语
			1,934	贸易关系 合作伙伴 协议 经济 合作 贸易 重要 联盟 地区 利益 欧洲 投资 发展 政治 国家 关系 经济 政党 伙伴关系
20		关联性	1,373	政策发展战略 欧洲经济目标 支持 工会 国际报告 社会要素 促进可持续合作 重要性 秩序 方法框架
55		系统性 1,341		秩序体系 保障措施 重要成员 有效规则 委员会监管 国家信息 制定实施 保障 时间 提案机关 提供立法
90		系统性 2,229		贸易世界经济增长市场全球经济机会欧洲自由国家工作重要繁荣开放总 统全球化创造利益壁垒
99		系统性 2,904		贸易世界组织国际世贸组织体系规则多边公平全球发展国家改革框架贸 易利益全球化水平发达国家发展中

第853章

854

第855章

856

857表 2.奥地利学派的增长论述。

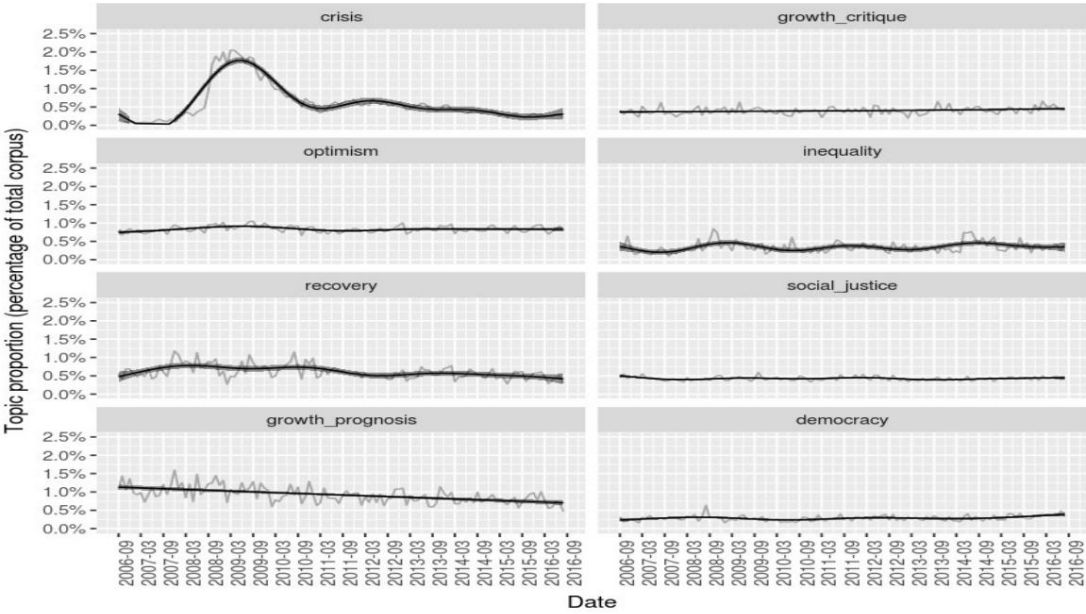
858

ID 名称	描述	特征词
1 不等式	不平等和资本主义（批判）	国家、财富、不平等、经济、资本、福利国家、新自由主义、国家、市场、金钱、再分配、schulmeister、私人、私人、金融市场、经济政策
72 社会正义	社会正义和公民权	社会、公民、主张、中心、责任、贡献、根本、战略、方法、塑造、弱、主张、依赖、独立、访问
81 成长批判	对增长和经济的批评	社会、经济、世界、福利、全球化、经济、资本主义、增长、市场经济、进步、更多、模型、变化、资源积极、显著、当前、期望、发展、后果、尽管、消极、仍然、尽管、强劲、形势、稳定、乐观
87 乐观	谨慎乐观	
94 预测 经济分析与预测	恢复	今年、预期、预测、经济增长、微薄、下一个、预期、之后、上升、预计、今年、预测、经济、下沉、专家
115 恢复	经济的 全球范围内的复苏	增长、强劲、复苏、全球经济、全球、经济、繁荣、转型国家、经合组织、疲弱、投资、国民经济、衰退、全球、缓慢
126 危机	危机， 原因和 结果	危机，经济危机，金融危机，经济，后果，失业，戏剧性，严重，迷失，时间，深，经济衰退，硬，很多，大规模

859

860

86. 图 1。随时间变化的主题显著性。



第862章