印度议会讨论什么？人民院质询时间的探索性分析

Suman Adhya、Debarshi Kumar S
anyal印度科学培育协会 贾达夫普尔，加尔各答 – 700032，
印度 adhyasuman30@gmail.com、
debarshisanyal@gmail.com

摘要

TCPD-IPD 数据集收录了1999年至2019年间印度议会下议院在质询时间讨论的问题及答复。尽管手动分析如此庞大的数据集十分困难，但现代文本分析工具可以提供一种强大的浏览方式。在本文中，我们对该数据集进行了探索性分析。具体而言，我们呈现了富有洞见的语料库层面统计数据，并对数据集中的三个子集进行了详细分析。在后续分析中，我们重点利用动态主题模型理解主题的时间演化。我们观察到，议会对话确实反映了各个时期的政治和社会经济紧张局势。

关键词： 印度议会、动态主题模型、潜在狄利克雷分配、TCPD-IPD、政治数据

arXiv:2304.00235v1 [cs.CL] 1 Apr 2023

1. 引言

印度议会是印度最高立法机构。其下议院，即人民院的议员由人民直接选举产生，而上议院则由各邦立法议会的议员选举产生的代表组成。尽管议会 proceedings 对政治学家极为有用，但其规模过于庞大，无法通过人工进行分析。因此，有必要使用算法工具对其进行探索。本文分析了 TCPD-IPD 数据集（特里维迪政治数据中心，2019年），其中包含约 29.8 万组英语问答（QA），这些问答于 1999年至 2019年期间，在人民院的质询时间内提出并讨论，涵盖四届人民院任期（第13届：1999年至2004年、第14届：2004年至2009年、第15届：2009年至2014年、第16届：2014年至2019年）。质询时间是议会每个工作日的第一个小时。在质询时间内，人民院的任何议员（简称：MP）都可以向部长提出与政府行政活动相关的任何问题，从而要求政府对其行动负责（Sanyal，2016年；Tripathi 和 Kumar，2021年）。质询时间也是加拿大、澳大利亚和英国等许多其他议会民主国家议会制度不可或缺的一部分（Martin 和 Rozenberg，2014年）。

TCPD-IPD 数据集的技术规格见（Bhogale，2019）。但除（Sen 等，2019）外，该数据集尚未得到探索；该研究旨在针对若干选定主题，判断议员提出的问题是否反映了大众媒体和社交媒体上的趋势。主题建模已被用于分析各国的议会 proceedings，参见例如（Greene 和 Cross，2017；Gkoumas 等，2018；Ishima，2020）。本文通过以下流程研究 TCPD-IPD。首先，基于整个数据集构建静态主题模型，并识别排名靠前的主题。随后，按如下方式选取数据集的一个子集进行进一步分析：（a）构建动态主题模型在该子集上；（b）绘制主题的时间演化以及某个主题中词语的时间演化；（c）分析这些图中各个给定时间点排名靠前的文档。通过这项分析，我们获得了有趣的见解。这证明了我们方法的有效性。我们的具体贡献如下：

1. 我们描述数据集的重要高层统计特征，以及议员参与情况中的偏差（第 2 节）。2. 我们确定整个数据集中的主要主题（第 3 节）。3. 我们通过分别为三个特定部委——财政部、铁路部和卫生与家庭福利部——构建动态主题模型，对与这些部委相关的问答进行更细致的研究（第 4 节）。4. 对于上述每个部委，我们确定了其在某一主题中的概率随时间发生显著变化的词语（第 4 节），以及这些词语所关联的重大事件。因此，我们使用动态主题模型中的词语编排来重构政治史。我们希望从过去获得的见解和经验能够帮助指导未来对重大国家问题的应对。

2. TCPD-IPD 的高层特征

下面列举我们从数据集中获得的一些有趣见解。

2.1. 按部委划分的数据集分布

TCPD-IPD包含与85个不同部委相关的问题。图1展示了排名前十的部委的数据分布（约占整个数据集的50%）。显然，财政部、铁路和卫生与家庭福利部是排名前三的部委。

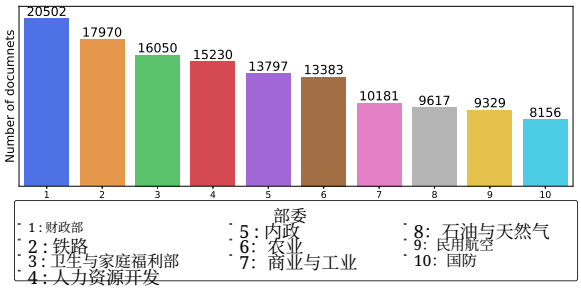


图1：按部委划分的数据分布。

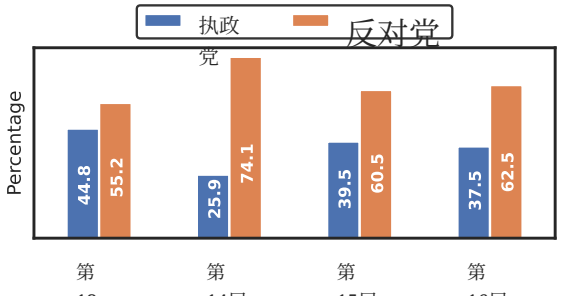


图 2：执政联盟与反对党的参与情况。

2.2. 按任期划分的数据分布

在总计298,292个问题中，人民院各届任期内提出的问题数量如下：第13届：73,531个；第14届：66,371个；第15届：79,401个；第16届：78,989个。我们发现，在该数据集涵盖的这段时间里，人民院每年开会的次数始终超过50次，只有2004年（48次）和2008年（46次）例外。这与2004年（9,398个问题）和2008年（9,851个问题）提出的问题数量最低相吻合。2004年的三次会期中有两次，以及2008年的所有会期，都属于第14届人民院。尽管许多新闻报告都提到了开会次数较少这一情况（Anuja，2020年11月25日），但我们没有发现有报告提及其对质询时间的影响，而这一影响正是我们在上文明确指出的。

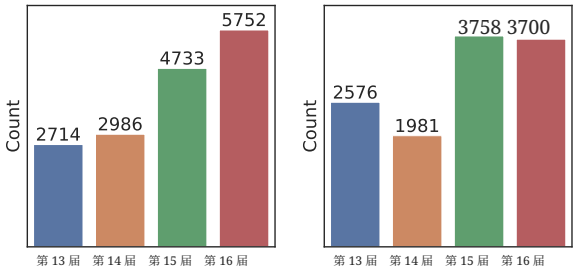
2.3. 议员的参与

与此前各届人民院任期不同，在第16届任期内，执政联盟提出的问题数量超过了反对党。当时，执政联盟在议院中占有历史性的65%席位。由于按照规则，执政联盟的数值实力高于反对党，我们对数据进行标准化，以了解双方的参与情况假设双方代表人数相等。我们认为，这样可以更公平地比较双方的参与情况。设 R_n 和 O_n 分别为执政联盟和反对党的议员人数，R_q 和 O_q 分别为执政联盟和反对党提出的问题数量。那么 $R_{pp} = \left(R_q \times \frac{O_n}{R_n+O_n}\right), O_{pp} = \left(O_q \times \frac{R_n}{R_n+O_n}\right)$。如图2所示，在转换后的空间中，反对党提出的问题仍多于执政联盟，这符合健全民主的预期。

2.4. 参与中的性别与种姓偏见

在该数据集涵盖的四届人民院任期内，男性提出了91.6%的问题，女性提出了8.4%的问题。同一时期，男女议员的平均比例为8.3:1。因此，性别比例失衡与问题的分布情况相关。值得注意的是，旨在为女性保留人民院三分之一席位的《妇女保留席位法案》自2010年起一直悬而未决（Marwah，2019），因而使这种偏差得以延续——就种姓分布而言，80.6%的问题由普通种姓议员提出，其余则来自保留席位

类别。请注意，人民院 24.03% 的席位为保留类别预留，其余席位属于普通种姓。图 3 展示了议会中提出的有关性别和种姓问题的问题数量；附录 6.5 列出了我们用于此项分析的关键词。



(a) 与性别相关。　　(b) 与种姓相关。

图 3：每届人民院中与性别 和种姓 相关的讨论。

3. 面向 TCPD-IPD 的主题模型

我们使用主题建模来从主题角度考察 TCPD-IPD 数据集。研究人员发现，采用吉布斯采样进行推断的潜在狄利克雷分配（LDA）（Blei 等，2003）所产生的主题往往优于现代基于变分推断的主题模型；例如，参见 Blei 等（2017）和 Lisena 等（2020）。因此，我们选择使用 LDA，而非神经主题模型。我们对整个 TCPD-IPD 语料库进行了预处理，使用 LDA 提取了 50 个主题，并为其手动标注（见附录 6.1.-6.3.）。我们过滤掉了一些噪声较多、异质性较强的主题，并从其余主题中选出排名前十的主题绘制于图 4。显然，经济增长和科学在邦级和国家层面都受到高度关注。

4. 按部委分析

我们选取了三个部委——财政部、铁路和卫生与家庭福利部——进行进一步分析。

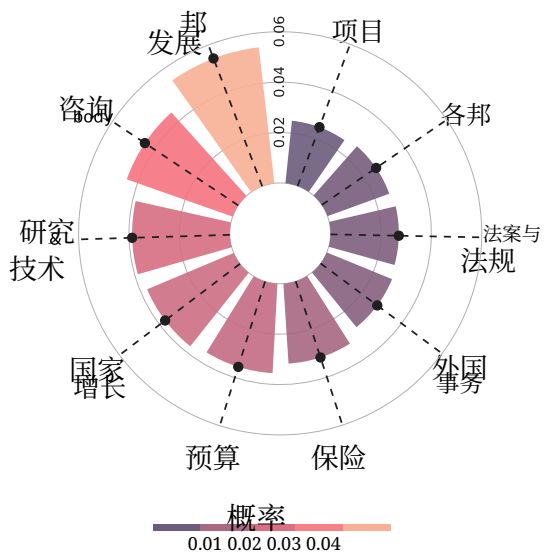


图 4：印度议会问答会议中讨论的热门主题。

我们将这三个部委的主题数量均设为 20。为了对主题的时间变化进行建模，我们必须在经典的动态主题模型（（Blei 和 Lafferty，2006年））与更现代的动态嵌入主题模型（D-ETM）（（Dieng 等，2019年））之间做出选择。前者本质上是针对序列数据对 LDA 的改编（下文称为 LDAseq），后者则由 David Blei 及其合作者开发。我们的实验发现，LDAseq 生成的主题效果更好（见附录 6.4）。Dieng 等，2019年的研究也报告了类似的观察结果。

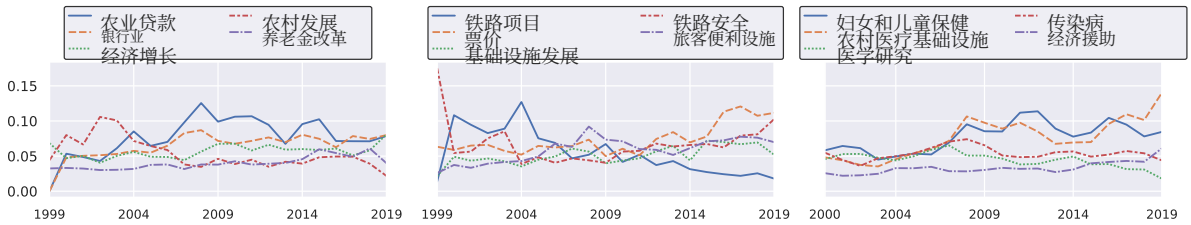
利用 LDAseq，我们获得了如图 5 所示的各部委五个主题概率的时间演化。在财政部，2008 年和 2014 年“农业贷款”主题的峰值与当时宣布的债务豁免有关，不过农民贷款始终是一个重要主题。对“农村发展”的关注逐渐减少，而“银行业务”、“经济增长”（当国内生产总值变化出乎意料地高或低时，相关提问会增多）和“养老金计划”等其他主题则保持得相对稳定。在铁路部门，21世纪初出现了许多新的政府项目及相关的议会问答，但关注点逐渐转向“基础设施发展”和“乘客便利设施”，相关问题主要围绕不断增加的私人参与展开。人民院上一届任期内宣布了许多铁路安全项目，这一点体现在“铁路安全”主题出现频率的增加上。2014 年客运和货运票价大幅上涨，引发了议会的激烈讨论。在卫生与家庭福利部，国家卫生使命于 2005 年启动后，关于妇女和儿童保健以及农村医疗基础设施的讨论有所增加。该十年初期对医学研究的关注促成了许多一流医疗机构在全国各地的建立，但人们的兴趣逐渐减弱。

主题中的罕见词语通常更具信息量，能够捕捉具体事件或问题。因此，在接下来的小节中，我们从各部委选取若干具有代表性的主题，展示所选主题中罕见词语概率随时间的演化。在每张图中，我们还用议会质询时提出的示例问题，对一个主导词语进行了注释。

4.1. 财政部

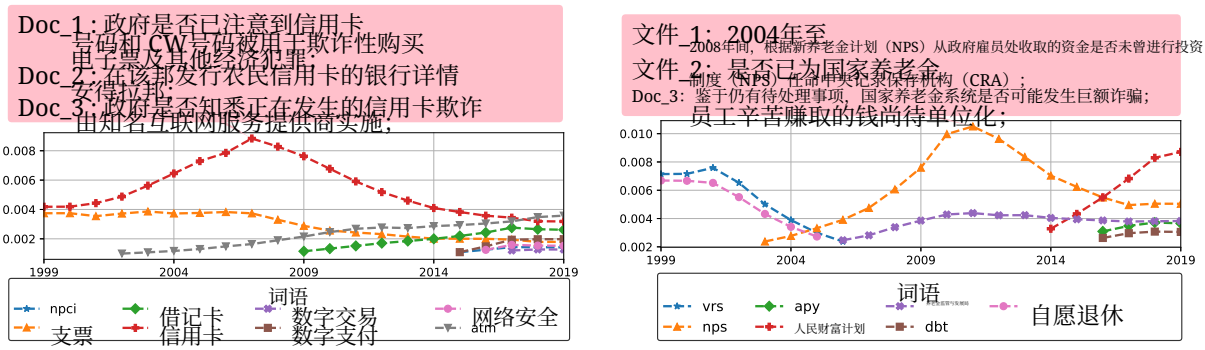
我们选取了两个主题，「银行业」和「养老金改革」，并在图 6 中绘制了其中若干所选词元的概率随时间的变化。在图 6a 所示的「银行业」主题中，我们发现「信用卡」一词在 2007 年达到峰值。分析显示，这段时间的大多数问题都与印度日益增多的信用卡欺诈，以及部分银行信用卡利率突然上升有关；这些情况也恰逢大众媒体的相关报道。信用卡相关讨论的另一个方面是农民信用卡（KCC），这是一种面向农民的低利率信用卡，1998 年推出，并于 2004 年得到显著改进。关于借记卡的讨论稳步增加，这与印度借记卡使用率不断上升相吻合；借记卡可避免陷入信用卡债务陷阱。随着废钞政策实施，以及政府更加重视端到端数字交易（而非现金交易），「自动取款机」「数字交易」和「网络安全」等词语日益受到关注，而「支票」等较传统支付方式的受关注程度则有所下降。

图6b展示了养老金改革这一主题。2003年末，印度政府宣布废除当时针对所有新入职员工实行的政府资助的养老金制度，并规定这些员工将纳入由新成立的临时养老金基金监管与发展局（PFRDA）管理的国家养老金制度（NPS）。国家养老金制度（NPS）使参保人能够为退休后收入进行有计划的储蓄。2009年，国家养老金制度（NPS）的覆盖范围扩大至所有印度公民。作为一项重大变革，国家养老金制度（NPS）引发了诸多疑问，并在2011年前后达到高峰；议员希望了解该计划的详情，包括其表现、实施挑战、向非正规部门的扩展情况，甚至其安全保障。21世纪初，政府在国有化银行推行自愿退休计划（VRS），以减轻公共财政的负担。2000—2001年，来自各个组织的数千名员工接受了VRS。鉴于该国失业率居高不下，VRS引发了民众的广泛恐慌，也促使议会提出尖锐质询，要求政府说明其对劳动力的未来规划。其他显眼的词语“apy”“dbt”和“jan dhan”指的是近期的金融普惠计划。



(a) 财政部 (b) 铁路 (c) 卫生与家庭福利

图5：使用 LDAseq 对 TCPD-IPD 的3个子集中的主题进行时间演化分析。



(a) 主题：银行业 (b) 主题：养老金改革

图6：在 TCPD-IPD 的财政部子集上运行 LDAseq 得到的选定主题。

4.2. 铁路

在这里，我们仅重点介绍一个主题，即图7所展示的“基础设施发展”。Ob-

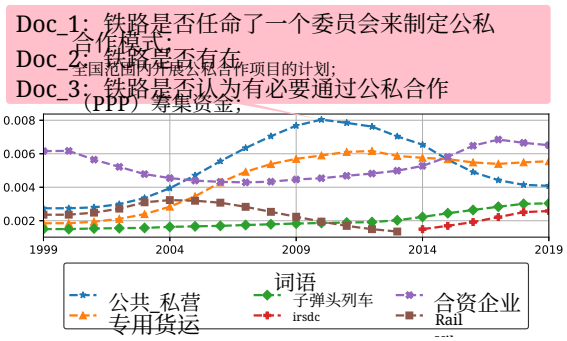


图7：铁路中的主题基础设施发展 。

“公私”一词在2010年前后达到使用高峰，当时铁路部推出了一种新的公私合营模式，以实现印度铁路现代化。事实上，人民院一直以来都对公私合作模式存在疑问和担忧，因为经济私有化可能导致票价上涨、失业和劳动非正规化（Makhija，2006年；Reddy，2019年）。“合资企业”也是许多讨论中的相关术语。早期（21世纪初）提及这一术语时，讨论重点是印度铁路与其他公共部门公司的合资企业；而近期（自2015年以来）的关注重点则是私营企业参与者日益重要的作用。事实上，议员之间也出现过类似的交流，讨论民航部对航空公司实行私有化的问题。“rail vikas”和“irsdc”等术语指的是印度铁路拥有并负责铁路维护工作的公司。货运量增长促使人们在2005年提出建设专用货运走廊（DFC）的构想，并引发了许多与其成本、进展和扩建相关的问题（“专用货运”）。关于引入高速列车的讨论由来已久，但在政府于2009年12月提交愿景文件、孟买—艾哈迈达巴德高速铁路走廊于2017年开工建设后，相关讨论开始升温。

4.3. 卫生与家庭福利部

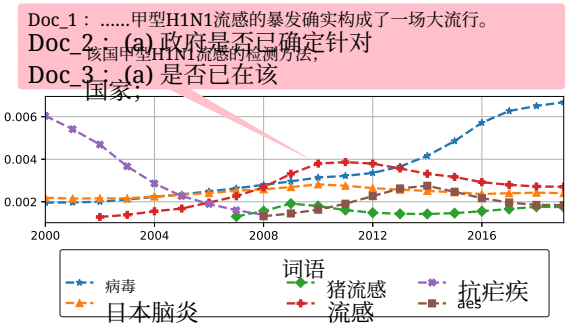


图8：卫生与家庭福利部中的主题 传染病 。

此处，我们选择了“传染病”这一主题。图8展示了该主题中词语概率的演变，

显然，“病毒”一词受到的关注日益增加，因为印度屡次遭受猪流感病毒（如H1N1）的侵袭，包括2009年猪流感演变为大流行，以及2014—2015年（Kshatriya等人，2018）。人民院议员询问了病例数量、测试的实施、政府干预、疫苗接种行动、世界卫生组织的作用，以及开发国产疫苗的努力。疟疾也是印度的一种主要疾病，但人民院的一份答复告诉我们，疟疾死亡人数逐年稳步减少，而议会对疟疾的关注度持续下降也印证了这一点。2008—2014年，印度有数千人死于急性脑炎综合征（AES）（Ghosh和Basu，2016）；2005—2011年，也有数千人死于日本脑炎（Adhya等人，2013）。在这些不幸事件发生期间，人民院就疾病展开了热烈讨论。

5. 结论与未来工作

我们确定了 TCPD-IPD 的显著特征，并展示了数据中三个重要子集的主题时间演化。未来，我们将尝试自动检测主题变化点，用触发事件（例如 VRS 的宣布）为其添加注释，自动总结特定时间包含特定主题或词语的主要文档，并推动针对议会讨论的最敏感主题影响（例如 VRS 对中年员工的影响）开展调查性报道或研究。我们希望本研究有助于提出更具探究性的议会质询，并制定更完善的国家政策。

6. 附录

6.1. 主题建模的数据预处理

我们删除了数据集中的标点符号，随后将词语转换为小写并进行词形还原。我们还删除了停用词，并筛除文档频率低于 0.001 或高于 0.95 的剩余词语。我们只保留至少由 3 个字符组成的词语。最后，我们删除了其中不足 3 个词语的文档。预处理后，我们创建了二元语法，以便更好地捕捉词共现统计。

6.2. 主题模型的配置

我们使用以下超参数来运行 LDA、LDAseq 和 D-ETM。

1. **LDA** （Blei 等，2003年）：我们使用 Gensim 对 LDA 模型的实施。为确保结果可复现，我们使用固定的随机种子，即将随机状态设置为 $= 2021$。我们将 passes 设置为 20，并对其余超参数使用默认值，具体为：

每个参数。

2.**LDAseq** （Blei 和 Lafferty，2006年）：我们使用 Gensim 实施的版本。我们将随机邦值设置为 2021，并将迭代次数设置为

20. 对于其余超参数，我们使用默认值。

3. **D-ETM** （Dieng 等，2019）：我们使用原始实施[1]，批量大小为 $= 64$，训练轮数为 $= 100$，其余超参数保持默认值。

6.3. TCPD-IPD 中的主题

我们使用 LDA 从整个 TCPD-IPD 数据集中提取了 50 个主题，因为当主题数量以 25 为步长从 25 变化到 200 时，该模型取得了最高的一致性得分（见图 9）。主题及其人工标注见表 1。

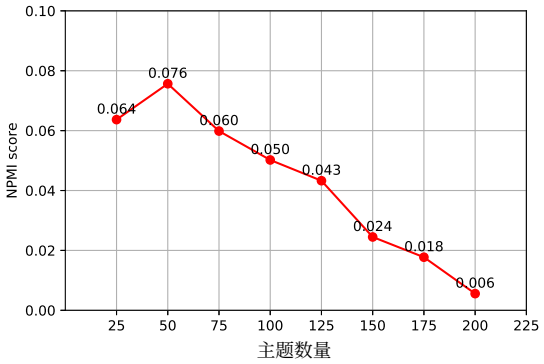


图 9：LDA 在整个 TCPD-IPD 数据集上针对不同主题数量得出的 NPMI 值。

请注意，我们无法运行 LDAseq，因为它耗时过长，即使运行了 72 小时仍未结束。我们也无法运行 D-ETM，因为它出现了后验坍塌问题；即使更改超参数，也无法解决这一问题。

6.4. TCPD-IPD 子集中的主题

我们对三个部委（财政部、铁路部以及卫生与家庭福利部）的数据子集开展了动态主题建模试点研究。我们将主题数量设为 20。我们按年份将每个数据子集划分为若干切片，并将每个切片按 8:1:1 划分为训练集、验证集和测试集，然后在这些数据上运行 LDAseq 和 D-ETM。随后，我们对各模型在所有时间切片上的结果取平均，计算其一致性（NPMI）、多样性和主题质量（$= NPMI \times Diversity$）。表 2 显示，LDAseq 表现最佳，因此我们选择它从各部委的数据集中提取主题。正文图 5 中展示的主题，是根据其排名前五的词语人工标注的。表3展示了这些主题的构成。

6.5. 性别和种姓相关讨论的关键词提取

为了生成分析性别与种姓相关讨论所需的关键词，我们采用了 SkipGram，

[1]https://github.com/adjidieng/DETM

主题中的前五个词	人工标签
国家的、外国、协议、国际、签署	外交事务
标准、质量、安全、使用、污染	污染
城市、城市、住房、德里、建设	城市发展
安全保障、邦、警察、内政、事务	法律与秩序
道路、高速公路、国家的、建设、邦	道路建设
银行、贷款、RBI、信贷、账户	银行业
委员会、邦、审查、建议、报告	咨询机构
邦、安得拉邦、北方邦、马哈拉施特拉邦、泰米尔纳德邦	各邦
基金、千万、拨款、年、邦	预算
天然气、石油、牛奶、生产、天然气	石油业
航空、机场、国防、民用航空、航空公司	民用航空
邮政、雇员、中央、办公室、官员	中央雇员
出口、进口、产品、贸易、纺织品	市场概况
项目、完成、工作、成本、制裁	项目
案件、法院、高等法院、个人、残疾	司法系统
太阳能、可再生能源、能源、系统、电力	可再生能源
法案、邦、提供、区段、规则	法案与法规
服务、信息、提供、电信、网络	信息技术
支付、税收、纳税、收入、金额	*Tax*
电力、工厂、产能、供应、邦	电力厂
体育、肥料、有限公司、限制、公司	化肥厂
女士、行动、投诉、案件、报告	申诉
水、河流、邦、资源、饮用水	水资源
教育、学校、大学、学生、人力资源	教育领域人力资源
计划、邦、发展、提供、实施	国家发展
煤炭、矿山、生产、矿物、采矿	煤炭开采
农村、县、地区、村庄、功能性	农村问题
德里、孟买、城市、古吉拉特邦、金奈	大都市区
健康、邦、家庭福利、药物、医疗	卫生服务
答复、人民院、回复、声明、提交	议会
农民、农业、作物、农业、生产	农业
中国、孟加拉国、岛屿、尼泊尔、灾害	自然灾害
增加、年、率、国家、减少	国家增长
钢铁、港口、连通性、产能、主要	港口航运
部落、日程、少数群体、奖学金、部落	部落事务
食品、价格、消费者、邦、粮食	食品价格
部门、私营、政策、公共、投资	私营部门
女人、儿童、就业、工人、劳动	妇女与儿童劳动
非政府组织、比哈尔邦、社会、组织、组织	*NGO*
旅游业、文化、地点、游客、发展	旅游业
工业、发展、基础设施、单位、计划	工业
企业、问题、指南、申请、保险	保险
研究、技术、中心、培训、国家的	研究与技术
铁路、列车、车站、乘客、铁路	铁路
土地、森林、地区、环境森林、邦	野生动物保护
邦、提案、设定、恰蒂斯加尔邦、提议	杂项
车辆、采购、网站、授予、合同	杂项
年、明细、邦、过去三年、数量	杂项
由于、满意、损失、原因、影响	其他
对此、反应、乔杜里、马诺吉、真实	其他

表 1：整个 TCPD-IPD 数据集中的主题。

数据集	连贯性 LDAseq	D-ETM	多样性 LDAseq	D-ETM	主题质量 LDAseq	D-ETM
财政部	0.088	0.078	0.652	0.496	0.057	0.039
铁路	0.129	0.094	0.686	0.558	0.088	0.052
健康	0.103	0.096	0.617	0.571	0.064	0.055

表 2：主题质量分析。

它是基于神经网络的 Word2Vec 算法中用于生成词嵌入的模型之一（Mikolov 等，2013）。我们在整个 TCPD-IPD 数据集上运行 SkipGram，并根据生成的词向量的余弦相似度，找出关键词“性别”和“种姓”各自最接近的二十个词。这些词列于表 4。随后，我们统计了包含这些词的文档数量。表中以斜体显示的词会引入许多无关文档，因此我们未将其纳入统计。

6.6. 某些术语的解释

正文中脱离印度语境后难以理解的词语解释如下。

1. ‘印度国家支付公司’：印度国家支付公司，由印度储备银行在

Year	主题中的前五个词	人工标签
	财政部	
1999	比率，增长，百分比，增加，年	经济的增长
2009	比率、百分比、增长、增加、年	
2019	增长、百分比、比率、经济、部门	
1999	银行、RBI、储备、问题、指南	银行业
2009	银行、RBI、问题、指南、储备	
2019	bank, rbi, 欺诈，问题，交易	
1999	项目、邦、世界、发展、银行	农村发展
2009	项目、邦、发展、基础设施、农村	
2019	项目，发展，邦，基础设施，基金	
1999	bank, loan, 信贷，NABARD，邦	农业 *loan*
2009	银行，贷款，信贷，农民，年	
2019	loan, bank, 农民，信贷，计划	
1999	计划，雇员，养老金，利息，存款	养老金改革
2009	计划，养老金，基金，存款，利息	
2019	计划，账户，养老金，邦，总理	
	印度铁道部	
1999	工作，完成，进展，项目，线路	铁路项目
2009	工作、完成、区段、线路、轨距转换	
2019	工作、区段、完成、线路、轨距转换	
1999	货运、交通、乘客、货物、车票	票价
2009	车票、乘客、货运、增加、计划	
2019	乘客、车票、票价、货运、列车	
1999	系统、安全、轨道、委员会、报告	铁路安全
2009	系统、轨道、安全、委员会、信号	
2019	系统、轨道、安全、列车、委员会	
1999	项目、公司、铁路、发展、企业	基础设施发展
2009	发展、项目、走廊、铁路、识别	
2019	发展、项目、走廊、铁路、高速	
1999	车站、设施、提供、站台、提供	乘客便利设施
2009	车站、设施、提供、站台、工作	
2019	车站、提供、设施、站台、计划	
	卫生与家庭福利部	
2000	研究、研究所、研究、委员会、开发	医疗研究
2009	研究、研究、临床试验、研究所、委员会	
2019	研究、医疗、研究、临床试验、委员会	
2000	儿童、人口、计划、国家的、生殖	女人与儿童健康护理
2009	儿童、项目、国家的、护理、女人	
2019	儿童、护理、女人、国家的、提供	
2000	疾病，疟疾，控制，case, 死亡	传染病
2009	疾病、控制、案件、报告、疟疾	
2019	患者、治疗、提供、免费、计划	
2000	中心，care, area, 服务，农村	农村医疗基础设施
2009	患者、治疗、提供、免费、医院	
2019	国家的、公共、医疗保健、提供、包括	
2000	患者、治疗、提供、医院、免费	经济援助
2009	项目、千万、基金、支出、成本	
2019	项目、基金、千万、成本、完成	

表3：TCPD-IPD 数据集中各部委的时间主题。

Word	前20个邻近词
性别	**性别**，**性别平等**，**性别差异**，**女性**、**女人**，**性别敏感度**，**性别意识提升**，**性别差距**，**性别平等**，**女孩**，**性别敏感**，**女性识字率**，**性别**，**男女**，**差异**，**儿童性别**，**儿童**，**女童**，**识字率**，**性别选择性**
种姓	**种姓**，**部落**，**scs**，**obcs**，**obc**，日程，**种姓部落**，**表列部落**，达利特，**社会正义**，**scs sts**，**表列种姓**，*Vijay Sampla*，赋权拿破仑，归属，*Subbulakshmi Jagadeesan*，**暴行**，**少数群体**，*Pal Gurjar*，赋权 *Smt*

表4：使用 Word2Vec 得出的每个关键词的前20个邻近词。

财政部下设立，旨在推动印度的数字支付和结算系统。2. ‘NABARD’：农业和农村发展国家银行，在财政部领导下运营。它监管向农村社会提供金融援助的机构。3. 农民信用卡（农民信用卡）：农民信用卡。4. ‘apy’：阿塔尔养老金计划。“计划”意为计划。

5. ‘dbt’：直接福利转移。

6. ‘jan_dhan’：总理人民财富计划。译为“总理人民财富计划”。7. ‘rail vikas’：铁路发展公司由印度铁道部所有，负责建设铁路基础设施。8. ‘irsdc’：印度铁路车站发展公司。9. ‘scs’、‘obc’、‘obcs’、‘scs sts’、‘表列部落’、‘表列种姓’：这些词语指代印度历史上处于不利地位的社群。“表列种姓”“表列部落”和“其他落后阶层”分别缩写为 SC、ST 和 OBC。

7. 书目参考文献

Adhya, D.、Dutta, K. 和 Basu, A.（2013）。印度的日本脑炎：国家首都区发生疫情的风险。*International Health*，5(3)：166–168。Anuja, G. V.（2020年11月25日）。议会今年的开会次数可能创历史新低。*Mint*。Bhogale, S.（2019）。TPCD-IPD：TCPD印度议会代码本（质询时间）。特里维迪政治数据中心，阿育王大学。Blei, D. M. 和 Lafferty, J. D.（2006）。动态主题模型。载于第*23*届国际机器学习会议论文集，第113–120页。Blei, D. M.、Ng, A. Y. 和 Jordan, M. I.（2003）。潜在狄利克雷分配。*Journal ofMachine LearningResearch*，3：993–1022。Blei, D. M.、Kucukelbir, A. 和 McAuliffe, J. D.（2017）。变分推断：面向统计学家的综述。*Journal of the American Statistical Association*，112(518)：859–877。Dieng, A. B.、Ruiz, F. J. R. 和 Blei, D. M.（2019）。动态嵌入主题模型。*CoRR*，abs/1907.05545。Ghosh, S. 和 Basu, A.（2016）。印度的急性脑炎综合征：不断变化的形势。神经科学年鉴，23(3)：131。Gkoumas, D.、Pontiki, M.、Papanikolaou, K. 和 Papageorgiou, H.（2018）。探索希腊议会全体会议的政治议程。载于*LREC 2018*研讨会论文集——*ParlaCLARIN*：创建和使用议会语料库。Greene, D. 和 Cross, J. P.（2017）。使用动态主题建模方法探索欧洲议会的政治议程。*Political Analysis*，25(1)：77–94。Ishima, H.（2020）。选举改革如何改变立法演说：来自澳大利亚维多利亚州议会1992–2017年的证据。*ElectoralStudies*，67：102192。Kshatriya, R.、Khara, N.、Ganjiwale, J.、Lote, S.、Patel, S. 和 Paliwal, R.（2018）。从印度H1N1（猪流感）疫情中汲取的经验：基于流行病学和临床特征的结局预测因素。*Journal ofFamily Medicineand PrimaryCare*，7(6)：1506。Lisena, P.、Harrando, I.、Kandakji, O. 和 Troncy, R.（2020）。ToModAPI：用于训练、使用和比较主题模型的主题建模API。载于第二届*NLP*开源软件研讨会（*NLP-OSS*）论文集，第132–140页。Makhija, A. K.（2006）。印度私有化。*Economic and Political Weekly*，第1947–1951页。Martin, S. 和 Rozenberg, O.（2014）。议会提问的作用与功能。Routledge。Marwah, V.（2019）。性别、种姓与印度女性主义：妇女保留席位法案这一案例。载于印度妇女与性别研究，第151–163页。Routledge India。Mikolov, T.、Sutskever, I.、Chen, K.、Corrado, G. S. 和 Dean, J.（2013）。词语与短语的分布式表示及其组合性。神经信息处理系统进展，26。Reddy, V. Y.（2019）。印度私有化。发展中国家的私有化，第178–198页。Sanyal, K.（2016）。监管监管机构：议会的作用。*EconomicandPolitical Weekly*，51(13)：16–19。Sen, A.、Ghatak, D.、Kumar, K.、Khanuja, G.、Bansal, D.、Gupta, M.、Rekha, K.、Bhogale, S.、Trivedi, P. 和 Seth, A.（2019）。利用大众媒体、社交媒体和议会质询时间数据研究印度经济政策话语。载于第*2*届*ACM SIGCAS*计算与可持续社会会议论文集，第234–247页。Tripathi, V. 和 Kumar, R.（2021）。大流行时期的议会：定位反对党。*EconomicandPolitical Weekly*，56(33)：16–19。

8. 语言资源参考文献

特里维迪政治数据中心，阿育王大学。（2019年）。*TPCD-IPD*：*TCPD*印度议会数据集（质询时间）*1.0*。