面向数据治理的主体注意力趋同分析模型及实证研究

陈荣[1,2]　赵冰燕[1,2]　李建霞[1,2]　严素梅[1,2]
（1. 华东理工大学科技信息研究所，上海 200237；2. 华东理工大学商学院，上海 200237）

摘要：数据治理是数据价值得以充分挖掘与应用的关键，其效能提升需要多主体共同参与。数据治理主体注意力趋同研究主要探究主体关注内容的一致性，对优化数据协同治理体系、保障数据治理的有效性具有重要参考价值。以BERTopic模型和加权关键词主题趋同度计算为基础，构建面向数据治理域的主体注意力趋同分析模型，利用该模型对代表特定领域数据治理主体注意力的数据进行主题建模与趋同度计算，根据主体注意力演化结果剖析主体注意力分配的差异，在此基础上计算主体注意力趋同度，以量化分析相关领域数据治理主体注意力趋同现状。该分析模型为相关决策者了解数据治理主体注意力的变化趋势提供了分析框架与评价维度。以健康医疗领域为例进行实证研究，结果表明：该领域主体在数据管理体系方面的注意力趋同度较高，聚焦于数据管理与数据安全、数据共享等；在数据价值体系方面的注意力趋同度整体较低，主要关注数据质量与服务评价。

关键词：数据治理；主体；注意力趋同；趋同分析；动态演化；健康医疗
中图分类号：G203；G350.7　DOI：10.3772/j.issn.1673-2286.2025.06.002
引文格式：陈荣，赵冰燕，李建霞，等. 面向数据治理的主体注意力趋同分析模型及实证研究[J]. 数字图书馆论坛，2025，21（6）：10-20.

在数字经济加速演进、《“数据要素×”三年行动计划（2024—2026年）》深入推进的背景下，数据成为推动社会经济发展的关键生产要素和国家发展的战略性资源。然而，数据质量缺陷、安全隐患及隐私泄露风险等问题，不仅制约着数据价值的深度挖掘与有效利用，更对个人权益、社会稳定以及国家安全构成潜在威胁，数据治理的重要性愈发明显。作为数据治理客体的数据，其采集、存储、传输等流通过程不仅涉及数据价值挖掘与释放，还与数据风险管控与治理息息相关。当前数据要素流通呈现以下特点：①随着数据规模的增长、技术的发展以及市场需求的多样化，数据要素的跨域流通要求已逐渐突破政府、医疗机构、数据服务商等单一主体能力边界，单一主体的治理受限于技术、人力等资源以及认知局限，已无法满足当前数据要素流通活动的治理需求[1]；②数据要素流通过程呈现出政府机构、企业、公众等多主体权责交织与利益博弈的复杂图景，亟需多主体有效的参与共治。国务院办公厅2021年印发的《要素市场化配置综合改革试点总体方案》、中央网络安全和信息化委员会2021年印发的《“十四五”国家信息化规划》均明确指出构建“共建共治共享”的数据治理新范式的重要性，强调部门与区域、政府、企事业单位与公众等不同层级之间需要提升协同治理能力。2022年印发的《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》（以下简称“数据二十条”）也指出数据治理模式、市场发展秩序的规范需要考虑政

收稿日期：2025-05-17

府、企业、社会多方参与，以协同治理为架构，推行数据要素市场的“数据善治”[2]。多主体协同共治的数据治理模式不仅是相关政策已明确的发展方向，更是优化数据要素流通环境、实现数据价值最大化、提升资源配置效率以及增强治理体系动态适应性的必然要求[3]。

在数据治理多主体协同实践中，主体之间容易产生目标冲突、资源错配以及数据标准、使用规则不统一等情况，导致数据治理协同效能低下。数据治理相关主体对数据治理议题的关注本质上是有限理性下的注意力资源分配过程，主体注意力趋同是实现多主体协同数据治理的认知基础，是突破当前数据治理多主体协同困境的新视角。通过识别数据治理过程中主体注意力分配特征，优化多主体数据协同治理效能，是充分释放数据要素价值、赋能经济社会发展的关键路径，也是落实“数据要素×”行动计划的重要保障。本文面向数据治理域，旨在拓展注意力分配理论在多层次多主体互动关系中的应用，构建包含多主体注意力动态演化与趋同两个部分的分析模型。通过探究政府主导下多主体注意力的动态演化特征及其趋同现状，分析当前数据协同治理面临的困境。

1　相关研究

数据治理指对数据的管理、保护以及优化的过程，旨在提高数据质量，确保数据的准确性、完整性、安全性和可用性，实现广泛的数据共享，并最终实现数据价值最大化[4]。数据治理作为制定战略、政策和标准的顶层框架，指导着具体的数据管理活动。目前，在数据治理研究中，学者大多关注数据规则的制定与落地、数据治理技术的创新与执行。数据管理方面的治理研究聚焦在数据治理操作实施层面，例如基于数据的类型、敏感性和价值选择实施相应的技术解决方案[5]、以数据生命周期为框架实现数据从零碎到集成的治理目标[6]，还有研究从数据管理框架构建[7]、数据质量反馈制度与平台建设[8]、数据质量管控模型建设[9]等方面为数据质量的提升提出参考举措。

学者们也注意到不同主体在数据治理过程中承担的关键作用是不容忽视的。数据治理主体相关研究多围绕政府主体展开，强调政府注意力分配在政策供给与数字权利监管中的重要作用。余丽梅[10]认为政府应当将有限的注意力资源分配到数据治理中，为数据治理提供政策依据和指导；高瑞[11]发现政府在应用数据技术进行数据治理的过程中需要尤其关注政府数字权利的运行、监管机制等，建立并完善有效的监督体系。在不同类型的数据治理中，政府通过发布政策与相关法规对数据进行标准化规范，并作为主导角色进行协调，实现部门与层级之间的协同治理，非政府治理主体则应当在过程中扮演数据来源者、信息传播者、数据监督者等角色[12]。已有研究虽提及多方参与数据治理，但仅把政府定位为制度制定与协同调度的主导者，企业、公众等非政府主体则被认为是技术执行者或监督者，这种主体角色定位忽视了数据治理过程中多主体之间注意力动态调整的变化过程，未能揭示主体注意力分配差异及其对数据治理效率的影响。

注意力是意识的聚焦与集中，能够反映个体对环境中的各类信息进行的差别化选择与处理[13]。目前研究主要集中在政府这一单一主体的注意力分配现状、注意力分配差异及影响因素等方面，缺乏对非政府主体注意力的研究。政府注意力相关研究中，政策文本通常被视为政府注意力的数据，对其量化分析是了解政府注意力分配现状的主要途径。Jones等[14]用政策议题内容分析了政府注意力的变化；杨剑等[15]利用政策文本，基于政府注意力选择指向、分配强度分析了政策的潜在主题内容以及演化趋势。注意力差异性分配的比较分析研究限于党派[16]、不同层级与地区的政府注意力[17]，分析的维度涉及时间、地区和内容[18]。政府注意力差异性分配影响因素，包括地方政府本身行政自主权以及事件严重程度[19]、公众舆论[20]、公共资源配置[21]等，Walgrave等[22]、Sevenans[23]认为媒体所关注和反映的社会问题是影响政府注意力突变的主要因素。除了对政府注意力的研究，少数学者对科研人员注意力的分配以及相关选择行为背后的原因展开探究[24]。当前的主体注意力研究多局限于单一类型的主体，且对其差异的分析多停留在静态层面，未能有效拓展至对多主体注意力的动态研究。

要洞察多主体间的注意力协同互动关系，则须将研究视角进一步延伸至主体间注意力分配的一致性，即注意力趋同。主体注意力的趋同指的是主体面对环境中各类信息采取的差别化选择与处理优先级的一致性。主体注意力是否一致与存在关联的研究大多从不同主体对某一关注点的注意力趋同，以及单一主体对不同关注点的注意力趋同等方面展开：Xu等[25]分析省级和地方政府工作报告文本发现，对于安全管理的注意力

存在长期一致性和短期不一致性的特征；樊霞等[26]利用政策文本分析了地方层面与国家层面政策的同构性与差异性；文禹衡等[27]通过量化政策文本和科研论文，探究了数据确权领域政策与科研关注内容的一致性与互动关系；张涛等[28]基于关键词词频研究了开放数据政策与数据安全政策的关联关系；周环等[29]通过关键词频率研究了数据开放政策与隐私保护政策的关联。上述研究通过文本相似度计算探究了政府在不同领域注意力的静态趋同状态，为学者理解主体注意力趋同奠定了基础，但仍缺少对不同主体注意力趋同随时间动态调整的深入研究。

数据要素流通过程中，各参与主体彼此利益交织[30]，不同主体自身属性特征与数据治理目标差异会使其注意力分配产生差异，进而影响数据治理工作的进展。数据治理是释放数据要素价值的必要前提，其效能提升需要以数据治理相关主体行为规律解构与动态适配机制的创新构建为基础，主体对数据治理不同议题关注程度对治理实践的影响不容忽视。既往研究集中在以政府主体为主要研究对象的注意力分配现状、影响因素等方面，涉及的主体层次较为单一，未能回答在数据要素流通的复杂场景中不同类型主体的注意力分配存在什么差异，以及注意力趋同是如何随时间变化的等问题。针对现有研究在多主体注意力动态分析方面的不足，本文聚焦数据治理多主体协同效能优化的目标，面向数据治理域构建多主体注意力趋同分析模型，研究不同主体注意力的分配差异与趋同状况，充分发挥多主体优势，为平衡数据价值挖掘与实现过程中的各方利益提出适配策略，为提高数据治理质量与效率提供实践指引。

2　主体注意力趋同分析模型构建

当前“数据二十条”虽明确提出“构建政府、市场、社会协同推进机制”，但政策落地仍面临主体注意力异步、资源配置失衡等现实问题。本文面向数据治理域构建多主体注意力趋同分析模型，旨在分析多主体在数据治理域的注意力动态演化与趋同现状，构建面向数据治理的主体注意力趋同分析模型（见图1）。模型构建参考了《信息技术服务 治理 第5部分：数据治理规范》（GB/T 34960.5—2018）提出的数据治理域框架，该框架将数据治理域划分为数据管理体系和数据价值体系两个部分，其中：数据管理体系包括数据标准、数据质量、数据安全、元数据管理、数据生存周期；数据价值体系包括数据流通、数据服务和数据洞察[31]。

基于数据治理范畴，多主体注意力趋同研究通过主体注意力演化、注意力趋同度量和趋同现状评价的分析路径，系统解析数据治理相关主体注意力分配现状以及当前治理过程中存在的问题。首先，基于BERTopic模型对相关主体的注意力进行主题建模，了解主体注意力在相应时间跨度内的变化情况；接着，通过关键词词频标准化计算消除文本基数差异对主体注意力趋同度数值计算的影响，量化不同主体的注意力一致性；最后，根据主体注意力动态演化结果以及不同数据治理域的注意力趋同结果，分析主体注意力在不同数据治理子域的变化趋势和趋同现状，对相关主体在数据治理过程中的趋同现状进行综合考量与评价，从而发现当前数据治理多主体协同中存在的问题与挑战，为了解当前数据治理主体协同现状提供可操作的实证工具。

2.1　主体注意力动态演化

通过构建主体注意力演化模型追踪相关主体注意力的变化。基于BERTopic模型对代表主体注意力的数据进行主题建模，并加入时间变量构建主体注意力动态演化模型。BERTopic模型是一种基于深度学习，结合了预训练语言表示模型（如BERT）和聚类技术（如HDBSCAN）优势的主题建模技术，包括文本语义向量嵌入、降维、聚类、序列化、主题特征表示5个步骤。①文本语义向量嵌入：建模的第一步是将原始文本转换为高维语义空间中的向量。模型使用SBERT（Sentence-BERT）或其他预训练语言模型作为嵌入模型，将每一条文本语料转换为固定长度的向量表示。②向量降维：文本向量嵌入后进行降维处理，采用UMAP技术将高维向量降至低维，以便实现后续的聚类与可视化。③聚类分析：模型采用HDBSCAN进行文本聚类，通过调整关键参数控制主题形成的密度阈值以及主题划分的颗粒度。④主题序列化：聚类完成后，模型为每个主题分配唯一编号，并根据聚类中心点的向量位置进行排序，从而实现主题的序列化，为后续主题对齐、演化趋势分析提供索引基础。⑤主题特征提取与表示：模型采用c-TF-IDF方法实现关键词提取。该方法将每个主题簇视为一个独立的类文档，并在此基础上计算词语权重。BERTopic模型支持不同类别词嵌入向量、降维、聚类、主题优化技术的组合，为使用者提

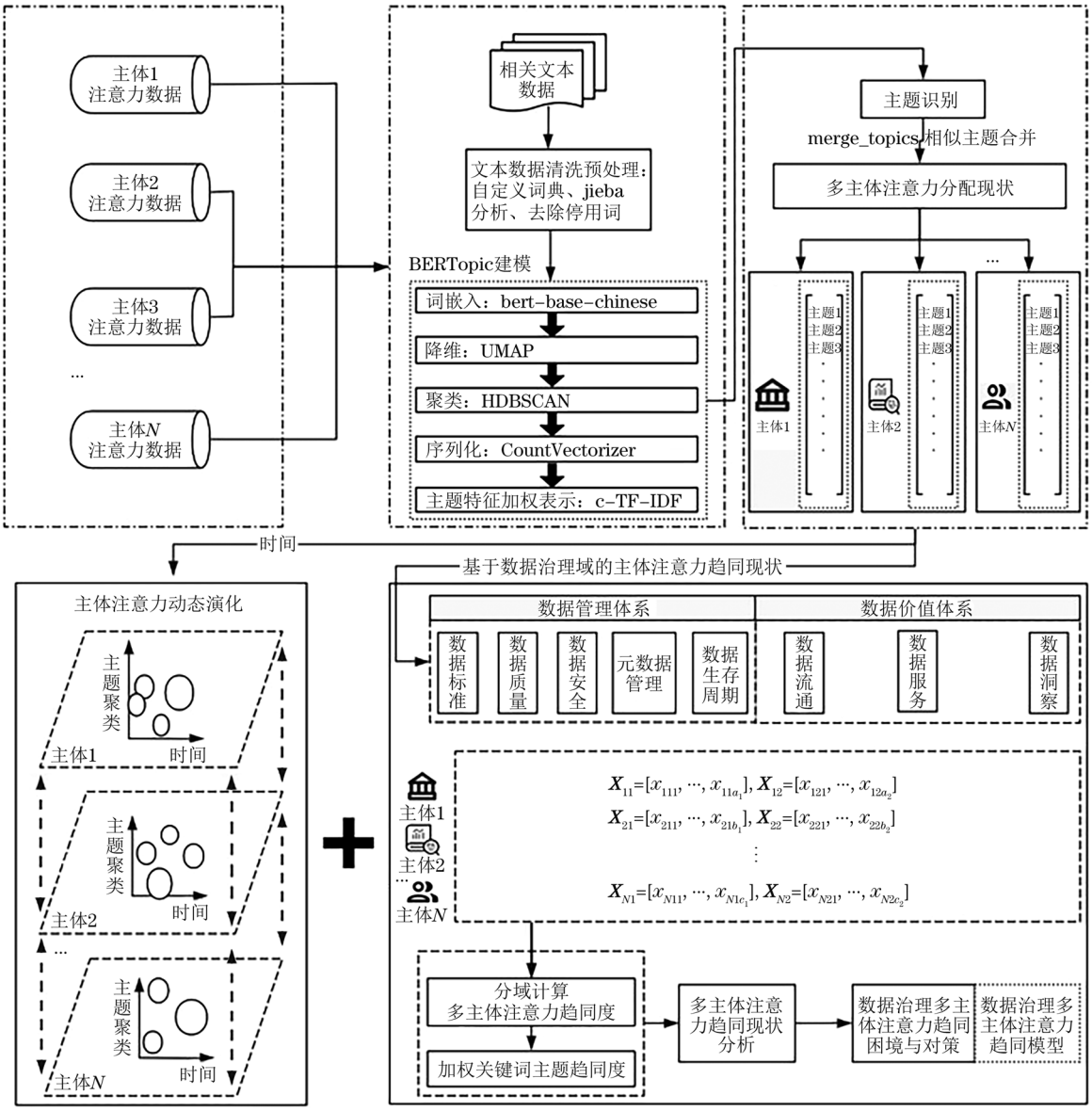

图1　主体注意力趋同分析模型

供了一种较为灵活的主题建模方案。主体注意力动态演化模型以时间为基准，对主体注意力的聚类结果进行可视化展示，演化结果能够反映主体关注内容随时间的变化。通过筛选其中与数据治理高度相关的主题聚类，分析相关主题的演化过程，能够了解主体对数据治理关注内容、关注强度的变化，并结合时间节点分析变化产生的原因。

2.2　主体注意力趋同度计算

主体注意力演化能够揭示主体关注内容随时间的变化，而主体注意力趋同度计算则旨在量化不同主体在数据治理域的注意力一致性水平，在描述性分析的基础上为决策者提供数据治理优化参考依据。主体注意力趋同度的计算中，基于注意力主题建模结果采用加权关键词词频计算的方法，该方法能够反映关键词在文本中的重要性与相关性。参与计算的关键词词频来源于不同的语料，趋同度计算以张涛等[28]基于关键词的主题协同计算方法为基础，用虚拟共现强度指标衡量不同文本语料的关联性。

（1）主体注意力在不同数据治理域的关键词分布。根据不同主体注意力在数据治理域的关键词来分析不同主体注意力关键词分布现状。假设共有N类主体（$N\in\mathbf{Z}$，且$N>1$），按照数据治理域的划分，

将数据管理体系和数据价值体系记为S1和S2，主体m（m为非零正整数）注意力在数据治理域Sn的主题记为$\boldsymbol{X}_{mn}$。

①根据N类主体注意力在数据治理域的关键词x，构建如式（1）所示的主题矩阵。

$$\begin{cases} \boldsymbol{X}_{11}=\left[x_{111},\cdots,x_{11a_1}\right],\boldsymbol{X}_{12}=\left[x_{121},\cdots,x_{12a_2}\right] \\ \boldsymbol{X}_{21}=\left[x_{211},\cdots,x_{21b_1}\right],\boldsymbol{X}_{22}=\left[x_{221},\cdots,x_{22b_2}\right] \\ \vdots \\ \boldsymbol{X}_{N1}=\left[x_{N11},\cdots,x_{N1c_1}\right],\boldsymbol{X}_{N2}=\left[x_{N21},\cdots,x_{N2c_2}\right] \end{cases} \tag{1}$$

式中：a、b、c均为正整数，表示不同主体注意力在不同数据治理域的关键词数量。

②代表不同主体注意力的语料$\boldsymbol{Q}$按年份t划分，如式（2）所示。

$$\begin{cases} \boldsymbol{Q}_1=\left[q_{1t_1},\cdots,q_{1t_u}\right] \\ \boldsymbol{Q}_2=\left[q_{2t_1},\cdots,q_{2t_u}\right] \\ \vdots \\ \boldsymbol{Q}_N=\left[q_{Nt_1},\cdots,q_{Nt_u}\right] \end{cases} \tag{2}$$

式中：t_1、t_u代表统计起始年份和终止年份；q代表不同年份不同主体注意力的语料。

③以数据治理域S1、t年为例，统计主体m注意力在该治理域的关键词在各年份相应语料中出现的标准化总频率$P_{m,t}$（某年出现某关键词的主体m语料数量与其语料总数量之比），分别计算相应的关键词标准化词频率$C_{m,t}$。以t=2024为例，具体计算公式如式（3）所示。

$$C_{m,2024}=\frac{P_{m,2024}}{P_{1,2024}+P_{2,2024}+\cdots+P_{N,2024}} \tag{3}$$

（2）主体注意力在数据治理域S1的关键词虚拟共现强度。对于每两类主体注意力的聚类主题，计算数据治理域S1中关键词标准化词频率的乘积与各自标准化词频率和的平方的比值，即不同语料中相关主体注意力关键词虚拟共现强度，如式（4）所示。

$$E_{ij,t}=\frac{P_{i,t}\times P_{j,t}}{\left(P_{i,t}+P_{j,t}\right)^2} \tag{4}$$

式中：i、j取值为1~N，i<j。

（3）主体注意力在数据治理域S1的趋同度。以数据治理域S1为例，主体注意力趋同度记为D_N，对多主体注意力关键词标准化词频率和虚拟共现强度进行加权计算，得到多主体注意力趋同度数值，如式（5）~式（6）所示。

$$a_{12,t}:a_{13,t}:\cdots:a_{ij,t}=\left(P_{1,t}+P_{2,t}\right):\left(P_{1,t}+P_{3,t}\right):\cdots:\left(P_{i,t}+P_{j,t}\right) \tag{5}$$

$$D_{N(t)}=\sum_{i=1}^{N-1}\left(\sum_{j=i+1}^{N}a_{ij,t}\times 2^{C_{i,t}}\times 2^{C_{j,t}}\times E_{ij,t}\right) \tag{6}$$

式中：a_{ij}代表权重系数，根据研究的具体需求和不同主体注意力趋同的相对重要性进行调整。在本文中，权重系数由词频决定，满足$\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}a_{ij,t}=1$。

2.3 主体注意力趋同现状分析评价

基于注意力演化与趋同度计算结果，对相关主体在数据治理域中的注意力趋同现状展开综合分析与评价。注意力趋同度是相关主体在数据治理域的注意力趋同现状的量化表征，该数值是反映数据治理相关主体注意力趋同现状的一个新视角。趋同度数值从量化视角揭示主体注意力对齐或冲突现状，为了解数据治理主体协同现状提供了数据基础。具体而言，趋同度数值越大，则说明数据治理相关主体在特定数据治理子域的目标共识越强，数据治理过程中相关主体能够有效协作、紧密配合，保障数据治理的有效性。相反，趋同度数值越小，则说明数据治理相关主体的注意力趋同度低，多主体数据协同治理可能面临困境与挑战。注意力趋同分析模型通过可视化注意力内容变化、量化不同主体注意力的一致性来解析注意力趋同现状，有利于推动数据治理多主体协同决策从粗略判断到精准调控的更替。

3 实证研究

随着智慧健康医疗向人工智能时代迈进，健康医疗数据数量猛增，数据治理面临诸多难题。据奇安信威胁情报中心监测，2023年国内医疗卫生行业泄露的数据量达到90 252.9万条，约344.7 GB，内容涉及大量个人隐私信息和商业机密[32]。健康医疗领域数据正面临着隐私泄露、未经授权访问、恶意攻击等问题。此外来源多样、数据异构、质量参差不齐等因素使健康医疗数据具有一定的复杂性，阻碍健康医疗数据价值深度挖掘与有效利用。数据治理是健康医疗数据价值得以充分挖掘的前提，是健康医疗领域的关键任务。本文以健康医疗领域数据治理为例，验证上述主体注意力趋同分析理论模型的可操作性。

3.1　主体注意力数据采集

健康医疗领域数据治理涉及的主体包括政府、医疗机构、科研人员、公众、技术提供商等，本文聚焦政府、科研人员、公众。其中，政府制定政策和监管框架，科研人员主要推动数据驱动的医疗研究，公众是数据的提供者也是受益者。三者在健康医疗数据治理中扮演着重要角色，其注意力演化与趋同对保障数据治理的有效性至关重要，3类主体注意力及其数据的选择如下所述。①政府注意力，指中央层面行政机关与立法机关在数据治理议题上的关注焦点。本研究将政府政策文本映射为政府注意力，包含政府对特定议题的关注程度、政策目标、预期效果等信息[33]。②科研人员注意力，指学术界从事健康医疗数据治理研究的科研人员的注意力，将科研人员的产出成果映射为其注意力，包括学术论文、专著和研究报告。与专著和研究报告相比，科研论文更能反映最新动态，关键词和摘要是论文研究主题与内容的高度凝练，能够直接反映研究主题和关注点，提供有关研究内容的丰富语义信息，因此采用学术论文的关键词和摘要文本代表科研人员注意力。③公众注意力，指广义的社会群体在数据治理域的关注内容分布。本研究选取微博和知乎两个主流社交媒体平台上的公众讨论内容作为分析对象。选择这两个平台主要基于平台用户群体与内容特性的互补性：微博平台以实时信息传播和泛社交互动见长，平台讨论氛围轻松活泼，能够反映公众对热点事件和话题的及时关注；知乎平台侧重深度知识分享与讨论，平台讨论氛围相对严肃，能够反映公众对较为复杂的议题的思考。需要说明的是，尽管科研人员本身属于广义公众范畴，但其与普通公众的注意力在专业性、广度与深度上存在差异，因此为了更准确地揭示多主体注意力的动态演化与趋同现状，将科研人员与普通公众的注意力分别视为独立研究对象。

采集代表3类主体注意力的数据，限制时间为2015年1月—2024年11月，搜集并经过数据清洗预处理，得到2 321个相关政策数据条目、8 079个相关论文数据条目、3 898个有效公众讨论数据条目。

3.2　主体注意力动态演化结果

利用BERTopic模型对代表政府、科研人员以及公众注意力的文本数据进行主题建模，在此基础上加入时间变量，分别得到主题动态演化结果，如图2~图4所示。

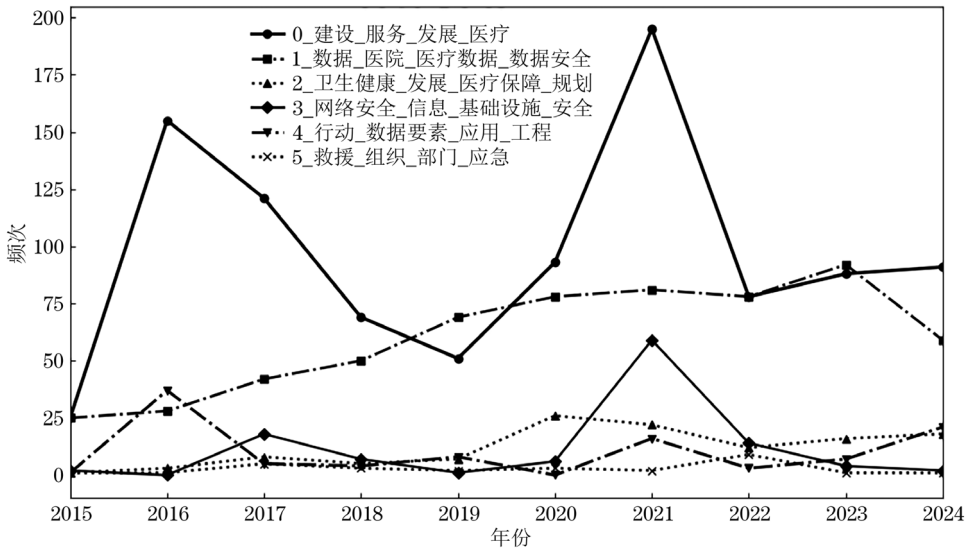

图2　政府注意力动态演化

由图2可以看出，政府注意力短期波动明显，关注内容集中在医疗数据的应用与服务、网络信息与数据安全、医疗保障与规划、基础设施建设等方面，波动主要出现在2016年与2021年。2016年国务院办公厅发布《国务院办公厅关于促进和规范健康医疗大数据应用发展的指导意见》，中共中央、国务院印发《“健康中国2030”规划纲要》，从建设权威互联互通的健康信息平台、培育健康医疗数据应用新业态等方面对健康医疗

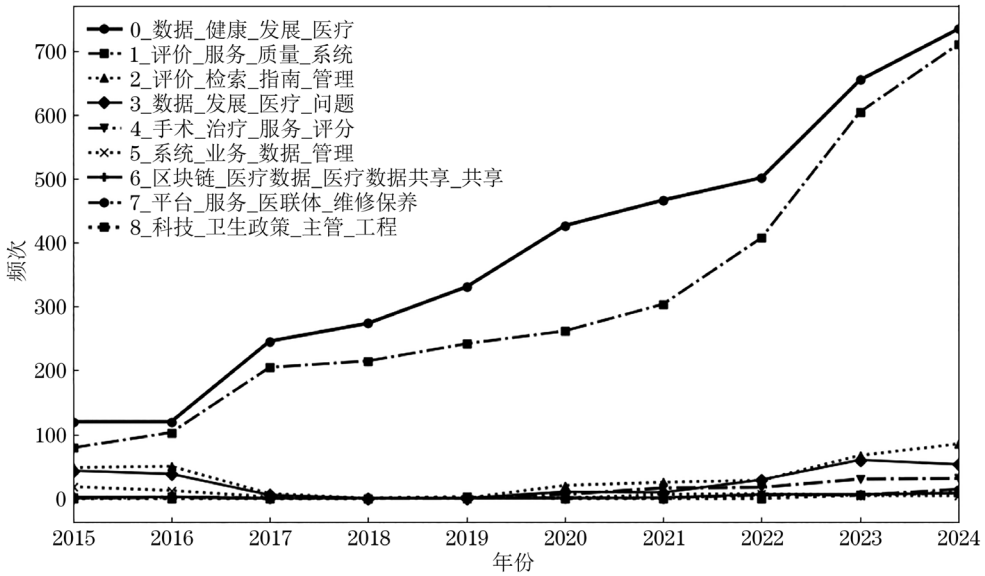

图3　科研人员注意力动态演化

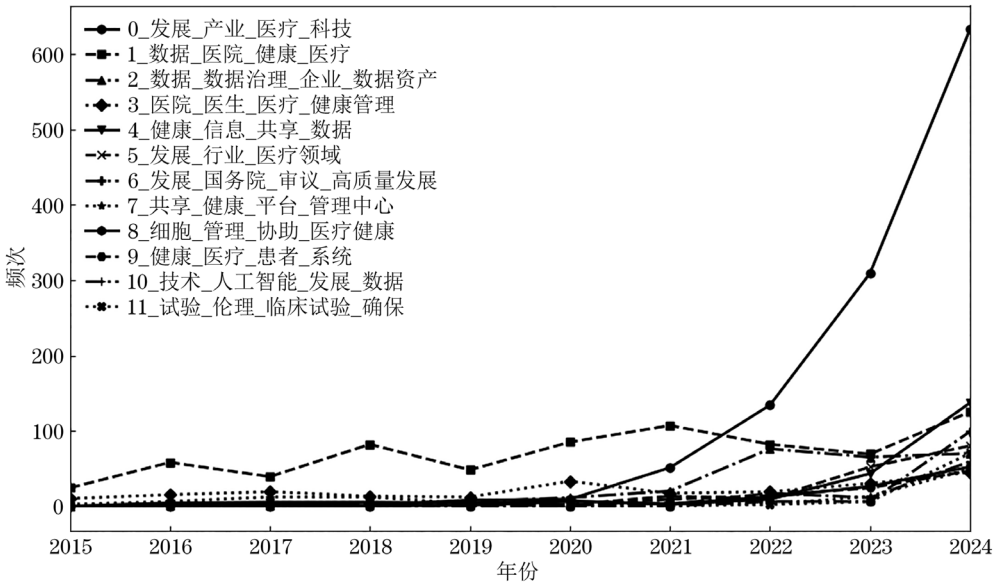

图4　公众注意力动态演化

数据的应用与规范、服务与创新作了整体部署。2021年国家标准《信息安全技术 健康医疗数据安全指南》正式实施，为保护健康医疗数据安全提供了可采取的措施。政府对健康医疗数据安全、网络安全的关注度相对稳定，《医疗卫生机构网络安全管理办法》《数据安全法》等的颁布体现了政府在确保医疗数据安全、有效、合规利用等方面制定了保障措施。

由图3可以看出，科研人员对健康医疗数据治理的关注内容集中在医疗数据共享、隐私保护、医疗信息化建设、医疗数据的应用等方面，涉及区块链、人工智能技术。与政府注意力变化不同，科研人员对健康医疗数据安全与应用的关注表现在更细粒度的信息检索系统的评价与管理、标准化区块链在医疗数据共享中的应用、健康医疗数据平台的维护等方面。另外，科研人员也致力于医疗数据共享方案、数据质量、信息化系统、服务质量的评价研究，通过科学的方法对医疗数据管理、应用以及服务进行评估。

由图4可以看出，公众对健康医疗数据治理领域的关注议题分布较为广泛，注意力呈现多元化分布特征，涉及医疗健康政策法规、医疗企业动态与科技创新、新兴技术在医疗领域的应用、医疗资源优化配置等方面。与政府、科研人员关注内容的差异之处在于，公众关

注内容还包括民生领域的改革与发展等现实问题，如养老服务体系的数字化建设、互联网医疗资源的优化配置。

3.3　主体注意力趋同度结果

主体注意力趋同度计算结果如表1所示，其中：E_{12}、E_{13}、E_{23}分别表示政府与科研人员、政府与公众、科研人员与公众注意力关键词的虚拟共现强度；D_{S1}、D_{S2}分别表示数据管理体系、数据价值体系的注意力趋同度。根据趋同度结果绘制主体注意力趋同度随时间的变化曲线，以揭示趋同度动态演化的阶段性特征（见图5）。政府、科研人员以及公众3类主体在数据管理体系与数据价值体系中的注意力整体趋同水平有待提升，趋同度数值在0.38上下浮动（取值范围为0~0.50）。具体而言，主体注意力在2016年（D_{S1}=0.389 6，D_{S2}=0.291 6）和2021年（D_{S1}=0.378 4，D_{S2}=0.346 7）存在明显差异，在数据管理体系上呈现的共识度（均值0.382 7）高于数据价值体系（均值0.364 9）。这一差异表明，相较于如何在数据价值挖掘与实现过程中实现收益分配、产权界定等更具挑战性的议题，各方对于数据管理的基础性、操作性规则（如质量、安全、标准）形成了相对一致的理解，反映出当前多主体在数据治理认知结构上形成“重管理、轻价值”的非对称格局。

表1　注意力趋同度

年份	数据管理体系				数据价值体系			
	E_{12}	E_{13}	E_{23}	D_{S1}	E_{12}	E_{13}	E_{23}	D_{S2}
2015	0.222 4	0.181 3	0.236 5	0.381 1	0.215 4	0.172 8	0.236 6	0.379 5
2016	0.246 8	0.247 4	0.250 0	0.389 6	0.214 5	0.231 4	0.246 6	0.291 6
2017	0.249 9	0.250 0	0.249 9	0.378 5	0.234 8	0.236 1	0.250 0	0.363 2
2018	0.236 0	0.226 8	0.248 6	0.396 6	0.216 8	0.190 4	0.244 3	0.381 5
2019	0.247 7	0.230 4	0.241 1	0.376 6	0.249 5	0.249 8	0.248 6	0.385 3
2020	0.249 5	0.220 0	0.226 6	0.384 3	0.247 9	0.241 7	0.232 0	0.395 9
2021	0.249 8	0.250 0	0.249 7	0.378 4	0.248 6	0.226 5	0.235 8	0.346 7
2022	0.245 9	0.221 5	0.238 1	0.396 8	0.222 4	0.212 5	0.249 0	0.385 6
2023	0.248 9	0.245 2	0.239 8	0.394 5	0.169 7	0.152 4	0.248 0	0.374 1
2024	0.250 0	0.229 0	0.229 2	0.350 8	0.229 9	0.249 6	0.234 8	0.345 5

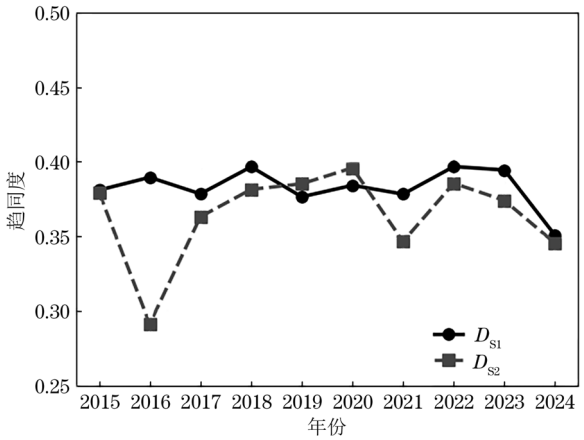

图5　注意力趋同度可视化结果

3.4　主体注意力趋同分析

3.4.1　数据治理域维度

（1）数据管理体系。3类主体注意力趋同度的变化与政策发布节奏基本一致，表现出政策导向的特点。根据趋同度计算结果，3类主体注意力趋同度整体处于相对稳定且缓慢增长的趋势。2015—2017年，政府发布政策以数据标准化、信息平台建设为核心，科研人员聚焦数据采集方法与医疗信息系统建设，公众围绕医院信息系统体验进行反馈，3类主体关注的内容开始出现交集。2018年以后，“健康中国2030”“互联网+医疗健康”等政策落地实施，3类主体关注内容中医疗信息平台建设、医疗数据融合等占比较大。

（2）数据价值体系。3类主体围绕数据在医疗服务中的应用，关注内容包括医疗服务平台建设、服务创新优化等，趋同度的提升与技术落地应用、服务可感知性增强密切相关。3类主体注意力趋同度波动较显著。2015—2017年，主体关注内容存在较大差异：政府关注医疗大数据发展潜力等；科研人员关注医疗服务创新与管理评估，公众则关注就医体验与医疗资源的公平等。2018—2020年，主体关注内容逐渐转向智慧医疗、

远程医疗。2021年后，3类主体关注内容中基层医疗数字化、健康医疗服务等占比较大，主体对健康医疗数据价值的认知与应用层面的融合趋势日益增强。

3.4.2 主体维度

（1）政府注意力。2016年，政府注意力主要集中在医疗数据安全防护（如隐私保护技术）、应用场景拓展（如移动医疗与大数据应用）和系统设计（如医疗信息化架构）；2021年，主要关注数据治理体系化（如主数据管理系统构建）、应急响应能力（如新冠数据实时监测）与服务模式创新（如线上医疗）。总体而言，政府注意力重心从技术工具开发向制度技术融合转移，强调治理框架的完整性与敏捷性。

（2）科研人员注意力。2016年，科研人员注意力聚焦在医疗信息化基础（如电子病历标准化）、服务模式优化（如分级诊疗）与资源配置效率；2021年，注意力聚焦数据共享机制（如交换平台建设）、质量评估体系（如诊疗效果量化指标），研究呈现技术落地应用到数据价值释放的纵深趋势。科研人员注意力的变化说明数据要素在临床决策与公共卫生中的赋能作用不断被强化。

（3）公众注意力。2016年，公众关注宏观政策导向（如数据治理法规）、技术风险担忧（如数据泄露）与行业规范诉求（如互联网医疗监管）；2021年，注意力转向智慧医疗成果体验（如健康码应用）、疫情防控参与（如行程数据上报）与信息安全反馈（如隐私保护实践），体现了公众从被动接受到主动参与的行为转变。公众注意力从风险感知向权益诉求转移，凸显了当前时代背景下数据治理与个体生活的强关联性。

4 结语

本文构建了面向数据治理的主体注意力趋同分析模型，以主体注意力为切入点，为数据治理主体注意力趋同现状的分析与评价提供了新的分析视角。实证研究以健康医疗领域数据治理为例，研究结果表明，政府、科研人员以及公众3类主体在健康医疗领域数据管理体系和数据价值体系中的注意力分布存在一定的差异。数据管理体系中，整体注意力趋同性较好，原因在于要解决健康医疗数据治理领域数据管理流程不畅、数据使用权限限制等问题，需要从原始数据的录入、元数据标准的建立、统一的主数据管理等方面入手[34]，因此主体注意力整体偏向数据管理体系。尽管数据管理体系中整体注意力趋同水平较为稳定，但公众参与仍显不足，原因在于相关政策文本、科研内容的技术认知门槛高、专业化信息占比大，公众难以根据相关内容展开广泛且有效的讨论。数据价值体系中主体关注内容容易受技术发展的影响，缺乏持续性机制支撑，注意力趋同度呈现波动的特点。

针对上述情况，提出以下建议：①构建公众友好型数据传播与应用机制，通过数字素养提升工程以及多渠道科普，增强公众对数据治理相关议题的认知与参与感；②尝试建立政策与科研之间的协调机制，在政策制定过程中结合相关研究的仿真反馈，提升政策的前瞻性与实操性；③推进不同领域数据价值转化路径的透明化与协同化，逐步实现以共识为基础、以价值为导向的数据治理格局。

未来可以将所提出的模型应用于不同领域主体数据治理协同研究，在进一步优化注意力主题表征的基础上，通过多样的实证研究收集和分析数据，检验分析模型的有效性与实用性。在此框架的基础上，结合数据治理域与数据治理全过程，构建面向数据治理的多主体协同评价体系，以系统地衡量数据治理相关主体注意力分配的合理性与协同治理的有效性。

参考文献

[1] 黄科满，许多，杜小勇. 数据社会化视角下的数据流通安全治理：五位一体框架[J]. 大数据，2024，10（6）：5-15.

[2] 新华社. 中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见[EB/OL]. [2024-09-16]. https://www.gov.cn/zhengce/2022-12/19/content_5732695.htm.

[3] 林镇阳，陈荣源，郭明军，等. 多元主体协同治理的数据要素价值生态体系研究[J]. 技术经济，2024，43（11）：1-13.

[4] 吴信东，董丙冰，堵新政，等. 数据治理技术[J]. 软件学报，2019，30（9）：2830-2856.

[5] TALLON P P，SHORT J E，HARKINS M. The evolution of information governance at Intel[J]. MIS Quarterly Executive，2013，12（4）：189-198.

[6] MULYANA R，RUSU L，PERJONS E. IT governance mechanisms influence on digital transformation：a systematic literature review[C]//Twenty-Seventh Americas' Conference

on Information Systems（AMCIS），Digital Innovation and Entrepreneurship，Virtual Conference，2021：9-13.

[7] SHAH S I H，PERISTERAS V，MAGNISALIS I. DaLiF：a data lifecycle framework for data-driven governments[J]. Journal of Big Data，2021，8（1）：89.

[8] 周林兴，徐承来，宋大成. 智慧城市视域下政府数据质量优化反馈机制研究[J]. 情报杂志，2021，40（7）：146-156.

[9] 丰佰恒，佟泽华，韩春花，等. 科研大数据质量管控模型仿真研究[J]. 情报理论与实践，2021，44（9）：33-42.

[10] 余丽梅. 地方政府推进数字政府建设的注意力分配与变迁研究[J]. 图书情报工作，2025，69（3）：121-134.

[11] 高瑞. 政府数字权力监督反思与体系建构[J]. 领导科学，2025（2）：134-143.

[12] 陆莉，邵瑞华. 公共安全数据治理体系：交叉学科视角下的要素范畴与逻辑框架解析[J]. 情报资料工作，2022，43（1）：32-40.

[13] LAUWEREYNS J. The anatomy of bias：how neural circuits weigh the options[M]. Cambridge：The MIT Press，2010.

[14] JONES B D，BAUMGARTNER F R. The politics of attention：how government prioritizes problems[M]. Chicago：University of Chicago Press，2005.

[15] 杨剑，石孝宇，郭正茂，等. 新时代我国青少年体育政策注意力配置及演进：基于LDA主题模型的政策文本分析[J]. 体育学研究，2024，38（3）：77-88.

[16] HEMPHILL L，RUSSELL A，SCHÖPKE-GONZALEZ A M. What drives U.S. congressional members' policy attention on Twitter? [J]. Policy &Internet，2021，13（2）：233-256.

[17] 胡冲，刘振宇. 注意力视角下的政策扩散：选择方向与分配强度：以长江经济带“水十条”为例[J]. 行政与法，2020（2）：10-20.

[18] 李燕凌，苏健. 地方政府建设数字乡村的注意力分配差异与政策逻辑：基于435份地方政策文本的量化分析[J]. 中国农村观察，2024（1）：127-145.

[19] CHEN W，ZHANG H B，HU Y W. How local governments balance multiple agendas in small-scale outbreaks responses：a multi-case comparative study from four major cities in China[J]. Cities，2024，148：104900.

[20] JENNINGS W，JOHN P. The dynamics of political attention：public opinion and the Queen's Speech in the United Kingdom[J]. American Journal of Political Science，2009，53（4）：838-854.

[21] 文宏，赵晓伟. 政府公共服务注意力配置与公共财政资源的投入方向选择：基于中部六省政府工作报告（2007—2012年）的文本分析[J]. 软科学，2015，29（6）：5-9.

[22] WALGRAVE S，BOYDSTUN A E，VLIEGENTHART R，et al. The nonlinear effect of information on political attention：media storms and U.S. congressional hearings[J]. Political Communication，2017，34（4）：548-570.

[23] SEVENANS J. How mass media attract political elites' attention[J]. European Journal of Political Research，2018，57（1）：153-170.

[24] 章熙春，莫秋平，朱绍棠. 基于注意力分配视角的科研人员行为选择研究[J]. 华南理工大学学报（社会科学版），2022，24（3）：74-83.

[25] XU J，LU L D，WEI J C. Hierarchical difference in attention allocation of local governments：explaining change and stability in safety management[J]. Safety Science，2022，152：105789.

[26] 樊霞，李传举，张巧玲，等. “同构”还是“差异”：我国专精特新政策注意力分布比较研究[J]. 科学学与科学技术管理，2024，45（5）：24-40.

[27] 文禹衡，付张祎. 基于计量分析的我国数据确权政策与科研协同研究[J]. 现代情报，2022，42（10）：58-70.

[28] 张涛，马海群. 基于政策文本计算的开放数据与数据安全政策协同研究[J]. 情报理论与实践，2020，43（6）：149-155，141.

[29] 周环，幸强国，唐泳. 基于政策文本计算的数据开放与隐私保护政策协同度研究[J]. 图书馆论坛，2021，41（11）：118-127.

[30] 陈丽娜. “三权分置”数据产权制度基本问题阐析[J]. 重庆邮电大学学报（社会科学版），2024，36（2）：88-100.

[31] 国家市场监督管理总局，国家标准化管理委员会. 信息技术服务 治理 第5部分：数据治理规范：GB/T 34960.5—2018[S]. 北京：中国标准出版社，2018.

[32] 新京报. 奇安信：2023年国内医疗行业泄露数据逾9亿条[EB/OL]. [2024-09-16]. https://baijiahao.baidu.com/s id=1799033114609880080&wfr=spider&for=pc.

[33] 林西，董瑜，郑新曼. 情感分析视角下政府注意力配置量化研究：以国务院政府工作报告为例[J]. 图书情报工作，2024，68（11）：136-147.

[34] 阮彤，邱加辉，张知行，等. 医疗数据治理：构建高质量医疗大数据智能分析数据基础[J]. 大数据，2019，5（1）：12-24.

作者简介

陈荣，女，硕士，副研究馆员，研究方向：专利信息、信息分析、信息检索、竞争情报，E-mail：rchen@ecust.edu.cn。
赵冰燕，女，硕士研究生，研究方向：信息计量、数据分析。
李建霞，女，学士，副研究馆员，研究方向：绩效评价、科学数据。
严素梅，女，硕士，副研究馆员，研究方向：文献计量、创新评价、竞争情报。

Model for Analyzing Convergence of Attention Among Data Governance Entities and Empirical Research

CHEN Rong[1,2] ZHAO BingYan[1,2] LI JianXia[1,2] YAN SuMei[1,2]
(1. Institute of Science and Technology Information, East China University of Science and Technology, Shanghai 200237, P. R. China; 2. School of Business, East China University of Science and Technology, Shanghai 200237, P. R. China)

Abstract: Data governance is the key to fully unlocking and leveraging the value of data, and its effectiveness enhancement requires the collaborative participation of multiple entities. Research on the convergence of attention among data governance entities primarily explores the consistency of their focus areas, offering significant reference value for optimizing collaborative data governance systems and ensuring the effectiveness of data governance. Based on the BERTopic model and weighted keyword topic convergence calculation, this study constructs an attention convergence analysis model tailored to the data governance domain. This model is used to perform topic modeling and convergence calculation on data representing the attention of data governance entities in a specific field. By analyzing the differences in attention allocation based on the evolution of entity attention, the model calculates the convergence degree of entity attention, thereby quantitatively assessing the current state of attention convergence among data governance entities in the relevant field. This analysis model provides decision-makers with an analytical framework and evaluation dimensions to understand the trends in the attention of data governance entities. Taking the healthcare sector as an example, the study indicates that entities in this sector exhibit a high degree of attention convergence in data management systems, focusing on data management, data security, and data sharing. However, attention convergence degree in data value systems is generally low, with primary focus on data quality and service evaluation.

Keywords: Data Governance; Entity; Attention Convergence; Convergence Analysis; Dynamic Evolution; Health Care

（责任编辑：王玮）