

元宇宙与大数据：时空智能中的数据洞察与价值连接

沈阳, 余梦珑

清华大学新闻与传播学院, 北京 100084

摘要

元宇宙对时空影响的实质是对时空节点进行数据智能化, 并通过虚拟世界对物理世界予以仿真与反馈, 而大数据是在此过程中提升人类对世界洞察力的核心手段。从元宇宙概念演绎与定义逻辑出发, 梳理了元宇宙“小中大全”4个不同层级的概念维度, 并基于对大数据洞察本质的9点思考, 提出了5层数据关联模型。从一元宇宙系统的建立再到多元宇宙系统的连接, 分别探索元宇宙中的数据产生、数据攫取、数据结构分析与数据价值挖掘等问题, 分析在嵌套与融合元宇宙中空间数据、时间数据与国际数据的互联问题, 以期通过数据洞察与价值连接更好地认知世界、描述世界与改造世界。

关键词

元宇宙; 大数据; 人工智能; 数字转型

中图分类号: G206

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2023012

Metaverse and big data: data insight and value connection in spatio-temporal intelligence

SHEN Yang, YU Menglong

School of Journalism and Communication, Tsinghua University, Beijing 100084, China

Abstract

Metaverse realizes the simulation and feedback of the virtual world to the physical world by intellectualizing the data of space-time nodes. Big data is the core means to enhance human insight into the world. Starting from the concept deduction and definition logic of the metaverse, this study sorted out four different levels of conceptual dimensions of the metaverse and proposed a five-level data association model based on the nine-point thinking of big data insight. From the establishment of the metaverse system to the connection of multiple metaverse systems, this study explored the data generation, data collection, data analysis, and data value mining in the metaverse. This study analyzes the connection of space data, time data, and international data in the metaverse, and expects to better understand, describe and transform the world by studying the data insight and value connection in the metaverse.

Key words

metaverse, big data, artificial intelligence, digital transformation

0 引言

元宇宙是高度沉浸且永续发展的三维空间互联网,是人机融生且多感官交互的三元体验互联网,是权力赋予与经济增值的三权价值互联网。元宇宙定义中的3个“三”分别指向三维化、三元化、三权化。其中,三维化是元宇宙空间的基本特征,相较于手机和计算机的二维体验,三维时空将带来高沉浸与多模态的高阶融合交互;三元化强调了元宇宙的交互对象是自然生命、虚拟生命与机器生命三元一体,通过外形共用、交互共情、行为同一、认知共享这4个层次,实现自然人与虚拟人、机器人的多感官交互、时空跳转与数据互联;三权化则是元宇宙中价值的根本体现。Web 1.0“可读”互联网实现了信息的连接,Web 2.0“可写”互联网实现了关系的连接,而元宇宙中Web 3.0的“可拥有”赋予用户身份管理与数据拥有的自主权力。从信息互联到关系互联,元宇宙将实现时空互联、体验互联与价值互联^[1]。在此背景下,围绕元宇宙与大数据的关系展开时空智能中数据洞察与价值连接的研究具有重要意义。

1 元宇宙概念演绎

回顾元宇宙概念的发展历程可以看到,学术界及不同国家对元宇宙的概念在界定时的侧重点都有所不同,在概念膨胀与话语争夺的背后,元宇宙概念的覆盖范围整体包括“小中大全”4个层级。

小概念侧重于元宇宙是移动互联网的升级。Meta公司的CEO扎克伯格(Zuckerberg)提出元宇宙是融合虚拟现实技术,用专属硬件设备打造的具有超强

沉浸感的社交平台,是移动互联网的升级版^[2]。而他所提出的元宇宙是一个时间奇点,在此之后用户将会把大量时间投入具有沉浸感的虚拟世界,则是对元宇宙技术带来的虚实临界转换做出了一种描述。小概念适用于对元宇宙中各种技术突破的表征。

中概念突出了元宇宙作为新技术的集合体,将成为全新的超级平台。微软CEO纳德拉(Nadella)认为元宇宙是全新的应用程序类型与下一代超级平台,在将计算嵌入现实世界与将现实世界嵌入计算的双向构建中,实现数字世界与物理世界的深度融合。事实上,他在2017年《刷新》一书中便谈到了元宇宙(彼时中文翻译为“虚拟空间”),并强调混合现实、AI和量子计算将是微软未来三大发展战略^[3]。中概念强调了元宇宙不仅是互联网的升级,更是全新形态的时空智能场景与社会交互载体。

大概念是笔者团队基于元宇宙特征属性与性能等提出的“三化、三性、三能”。元宇宙中的“三化”是其概念内核,即三维化、三元化、三权化;“三性”是其基本属性,即元宇宙带来的时空拓展性、人机融生性、经济增值性;“三能”则是元宇宙中可实现的应用,包括时空智能、生命智能、合约智能^[4]。

全概念中,元宇宙是将所有互联网技术都包含在内的下一代互联网形态和社会形态,在链接和创造中不断发展和演变,元宇宙将会成为万物互联、虚实共生的世界。

目前,业界对这4种元宇宙概念层级的使用偏向与特征描述通常具有较强的商业路径依赖。为了避免元宇宙概念在使用中的模糊与滥用现象,对其定义的基本逻辑应遵循以下4个基本原则:①区分移动互联网与元宇宙的本质差异;②遵循奥卡姆剃

刀(Ockham's Razor)原则,非必要不增加;③概念能够覆盖当前所有的元宇宙实践;④仅针对主体进行客观描述,不涉及主客体间的互动关系。在明确元宇宙基本概念与定义原则的基础上,可进一步展开元宇宙与大数据研究。

2 大数据洞察论

网络的本质是连接,大数据的本质是洞察,人工智能的本质是进化,区块链的本质是合约。

大数据是人类提升对整个世界洞察力的核心手段,用数据量化来认知世界、描述世界、洞察世界与改造世界。通过数据与语言符号来尽可能准确地描述世界,语言可以理解为通俗化的符号集合,数据主要解决客观描述,而语言主要解决主观共鸣。基于对大数据本质的把握,得出以下9点思考。

①大数据的本质是洞察。衡量人类文明既可以有“能源”指标,也可以有“信息”洞察力指标。这种洞察能力是人类文明进步的标志之一。

②对人类之外的洞察可以理解为无限外洞察,受到数理、物理、地理等规则约束;而对人类自身的洞察是无尽内洞察,受到生理、心理、伦理、事理、法理等限制。

③人类洞察能力从自然态到机械增强态,再到如今的智能辅助态,洞察能力随着技术的进步在日益增强。

④人类若能在数据洞察中知晓关键网络节点,所观察数据的形态越多元、维度越丰富,与之相应的洞察能力也会越强。

⑤人类洞察能力提升规律与摩尔定律类似,在引入合适观测变量情况下,能对人类洞察能力的提升速度进行有效估算。

⑥当人类洞察能力越来越强时,原本

看似同质化的物品将会因洞察力提升带来异质化的全新属性值。

⑦在洞察能力增强的过程中,个体与平台洞察力速度有很大差异。平台方拥有算法、算力与算据的全面优势,洞察能力提升速度远高于个体。

⑧数据造假是数据洞察面临的核心威胁之一。

⑨基于个体洞察合规性,可按照隐私逻辑对数据进行合理分层。

关于第5点中人类洞察能力提升速度的估算可以引入不同的观测变量。如果以人类能观测到的最短时间间隔为例来衡量人类洞察能力,通过整理相关数据可知,21世纪以来,人类每年平均在最短时间间隔洞察方面的提升能力是上一年的1.46倍,当对最短时间间隔的观测达到一定数值后,提升速度有所减慢且提升难度增大。目前,科学家已测量出的最短时间间隔为247仄秒(zeptosecond)^[5]。如果将手机摄像头像素作为另一参考指标,以2000年第一部内置摄像头手机夏普j-sh04的11万像素为基准,到2020年华为P40 Pro系列后置摄像头总像素达到10 200万,这20年的平均增长速率为40.7%。基于观测变量的差异,洞察能力增速可有不同的评估路径,而数据在此过程中起到关键作用。

数据与洞察力间相辅相成的关系,在第6点思考中得以体现。事实上当人类洞察能力逐步增强、对事物观察能力日益敏锐时,那些原本看似没有明显区别的物件之间,也会由于数据颗粒度更加细腻,呈现出本质差异。技术与数据一方面能让人类在洞察同质化物品时发现其内在的异质差异,另一方面异质信息又会构成全新的数据集,辅助洞察能力进一步增强。

个体与平台增速的差异在第7点思考中可被通俗地解释为大数据“杀熟”,其内在逻辑是平台方的数据洞察能力增长速度

远快于个体。个体在算力、算据与算法方面都处于劣势地位，增速慢于平台方。平台方的规模效应与多模态数据带来的数据合围效应结合，更加能够从多个维度认知世界。因此需要相应的法律法规与行业准则来约束与规范平台方行为，避免平台方将数据洞察优势用于不正当竞争，损害个体利益。第8点思考对数据风险与伦理问题做出了警示，数据洞察的最大挑战不仅是人类洞察能力不足，还包括在数据洞察过程中各利益攸关方出于利益动机进行数据造假，从而使人类洞察世界与真实世界产生背离与失真。因此，无论是在平台社会还是元宇宙空间中，都需要警惕数据造假带来的洞察与认知偏差。

遵循第9点思考的逻辑可对数据进行隐私层级划分。在大数据生态圈逐渐成熟的演化过程中，数据本身的开采深度与应用层次也在不断深化。数据层次可分为5层，如图1所示。表一层是通用公开大数据，表二层是需要登录才能获取的垂直领域大数

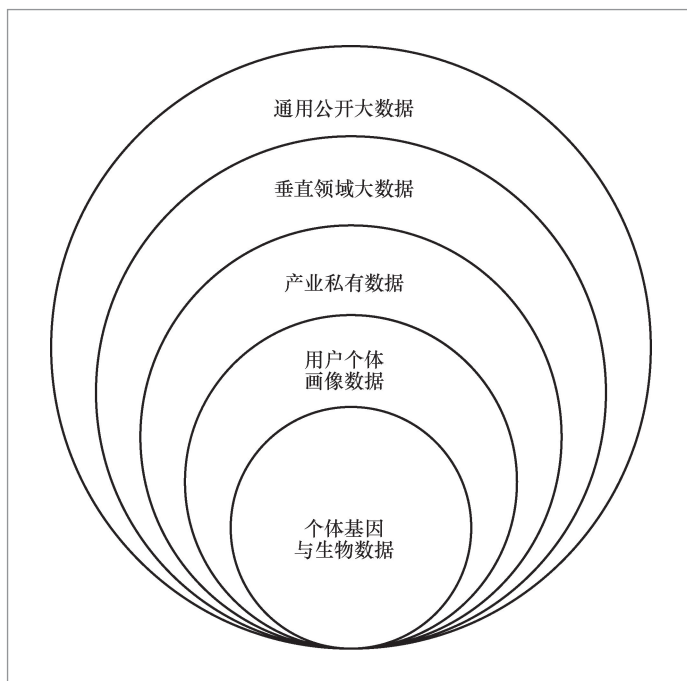


图1 5层数据关联模型

据，里一层是产业私有数据，里二层是用户个体画像数据，最里层是个体基因与生物数据。其中，表层数据是机器可以进行公开采集的数据，且数据应用企业可以通过技术手段进行整合；而里层数据则是需要个人授权才能合法获取与分析的数据。

5层数据关联模型由表及里、从通用到垂直、从企业机构到个体用户、从全局图景到基因画像，不断延伸数据触角。在强化公有数据采集整合基础设施的同时，需要引导更多产业领域私有数据的聚合，通过构建多层次数据闭环与通用化、一站式数据平台，不断驱动行业领域的应用创新。

3 一元宇宙中的大数据

一元宇宙即单个元宇宙系统中的大数据问题，涉及元空间中数据产生、数据攫取、数据结构分析与数据价值挖掘等维度。元宇宙对时空影响的实质是对每个时空点进行数据智能化，并赋予其数据、信息、知识、符码、智慧、连接，但随着密度提升、强度（频度）提升，也伴随着时空信息与连接过载问题。未来会出现时空信息增加的摩尔定律，即每隔一段时间每个时空点的数据、信息、知识、智慧等都大幅度增加。

打造元宇宙首先需要建立三维化空间（加时间维度是四维），而“数据升维”则是搭建三维化空间的基础，即人类需要在“0”与“1”的二进制代码中完成三维建模，并通过终端设备进行三维模型展示与交互。事实上，完备元宇宙的操作系统（operating system, OS）从底层设计出发到技术路径实践，应当是完全三维化的操作系统，并直接产生三维数据。但目前操作系统是基于二维信息处理设计的，因此在利用二维操作系统建构三维空间的过程

中,需要通过数据升维来丰富元宇宙中的视觉感知与体验感知。

三维元空间中的主要交互对象是虚拟人。目前为了节约算力,在二维数据中普遍采用伪三维虚拟人用于交互,但从长远来看,三维虚拟人是趋势所在。从人与空间角度看,元宇宙底层大数据必然是海量三维数据。因此,元宇宙对数据计算的要求远高于二维交互空间。各类AIGC(AI-generated content),需要在二维基础上逐渐升级成三维AIGM(AI-generated metaverse)。三维空间也正在从静态化建设向动态化空间发展。元宇宙空间中的数据结构将呈现出动态参数化特征,空间、人物、货币皆是如此。

从搭建框架来看,首先需要考虑数理部分,即该元宇宙空间的数理规则是什么?是 $1+1=2$?还是 $1+1\neq 2$?数理是元宇宙运行与演替的绝对规则。其次是元宇宙中的物理定律。在不同元宇宙星球设定中,可以选择性地解放部分物理束缚,并给定不同重力加速度,从而使空间中的地理地貌特征发生改变,虚拟人在其中的行为也随之遵循其特定物理规定值。例如《荒野大镖客2》中的游戏世界,玩家可以在不同星球空间中有差异化体验。在元宇宙中,可以通过大数据对经典力学进行模拟,但这个过程中的真正难点在于从微观量子力学一直模拟到宏观经典力学。单个立方厘米铜就有 10^{22} 个铜原子,这是目前计算机难以承受的巨大算力。通过数据实现单个立方厘米从微观力学到宏观力学的贯通性模拟,需要花费10年,乃至更长时间。

当完成初步空间模拟后,元宇宙需要进一步考虑对人类社会的模拟,即如何利用大数据将人类模拟出来。思路一是将智能体的自我生存设为最高准则,让其自我演化;思路二是对人类在社交媒体账号中的性格进行提取、映射,将其作为初始性

格,并不断博弈演化。

从元宇宙经济系统来看,当参与博弈的主体同时包括虚拟人与自然人时,为了避免平台主导经济系统的通货膨胀,需要引入区块链的智能合约。在传统经济社会中,不受制约的滥发货币会造成货币贬值。而在元宇宙中,为了避免货币滥发与贬值,设计者们引入了货币通缩机制。在智能合约的规定下,整个元宇宙的货币受到严格的发行限制,如可以在启动阶段设置永不增发,也可以每隔一段时间规定好货币生成减半。在元宇宙中,大数据不可避免地与区块链结合,使其具备合约性。当在元宇宙中建立人货场三者后,元宇宙运营也由此开始。

在元宇宙运行过程当中,攫取用户数据的方式与传统移动互联网存在较大差异。在移动互联网中,人们主要通过手指与手机的交互产生数据,搜索引擎采集搜索数据,社交网络采集社交数据,电商平台采集购物数据,部分互联网公司单日采集的交互数据可达百万条。而在元宇宙中采集数据的主要器官已经从手上转移到眼睛。从身份验证开始,移动互联网通常将手机验证码作为验证方式,元宇宙中更加适切的则是通过智能眼镜进行虹膜识别验证。元宇宙中多模态数据场包括追踪眼球关注焦点变化、脸部表情变化、手势姿态变化、体态移动变化。元宇宙为了能识别用户在空间中的移动,需要能够获取上下前后左右6个方位的移动数据,6DoF(degree of freedom)数据结合了平移和旋转^[6],不仅包含用户在X、Y、Z三轴上平行移动的数据,也包含其在三轴上旋转移动的数据。不难发现,元宇宙大数据与移动互联网大数据的一个重要区别在于,移动互联网中数据是以手指的交互数据为核心,而元宇宙的数据是以眼睛的交互数据为核心。前者涉及的触觉研究与设计较多,而后者还

需要多种光学原理。如何在交互中欺骗、迎合,乃至满足视觉系统,也将是元宇宙算法研究中需要重点关注的问题。

依据元宇宙开放世界的属性,元宇宙将会产生更多自定义数据,用户提交的各种内容,加上AIGC的流行,使元宇宙数据类型更加多样化。基于元宇宙定义,比较好的数据模型是从底层设计便考虑元宇宙数据资产确权问题,让每个不可分的非同质化数字资产都各有其所,从基础设施层面就内嵌Web 3.0的内在运行机理。

从用户角度来看,需要关注多感知交互数据的产生,即“如何有效地平衡能耗与功能”。可行思路是将不同计算设备分散开,智能手表识别手部姿态,智能耳机实现麦克风与音响功能,智能眼镜实现虹膜识别,以此降低多功能交互需求的设备重量。把元宇宙装备当成整体设备的一环,而不是过度突出元宇宙装备,是未来5~10年的一个可行策略,这也决定了多感知系统的多设备基层需求。此外,还需要集成虚拟人与机器人数据。通过外形数据共用、交互数据共通、行为数据共一、认知数据共享这4个维度实现虚拟人和机器人在数据层面的互联互通,并实现自然人与虚拟人、机器人的三元一体。

在元宇宙中流行的内容将是移动互联网信息流的升级,即时空数据流。3类容易广泛普及的时空数据流预测是:美颜时空流,类似抖音升级版;社交数据流,类似微信升级版;时空数据定位,类似搜索引擎升级版。如何对元宇宙中受欢迎的时空数据流进行有效的信息价值提取也是数据处理系统研发时需要考量的问题。

关于元宇宙中的数据处理,一方面,元宇宙中大数据维度将更加丰富且数据量更为庞大,这就对处理元宇宙数据的算力提出了更严苛的要求;另一方面,元宇宙大数据中关于人的数据有明显增多,特别是未

来脑机接口技术的成熟将会带来更多生物数据与空间数据的融合处理需求。此外,人类对于虚拟全息人的数据需求也在明显增加,虚拟全息人既要从外表模仿人类,也要尽可能从内部模拟出人类细节。

总结起来,元宇宙中的大数据特征包括:①数据更海量且维度提升;②数据更人化且直观化;③数据引入智能合约具备价值属性;④从微观到宏观的数据模拟将带来巨量新需求;⑤多模态、多感官交互将成为新常态,并带来新数据形态。元宇宙在5层数据关联模型中应贯穿这5项基本特征,时空智能中大数据可应用层次多元且价值丰富。

4 多元宇宙中的大数据

元宇宙的多元宇宙分为两类情况:一类是元宇宙对元宇宙的嵌套,类似电影《盗梦空间》中梦境对梦境的嵌套;另一类则是元宇宙与元宇宙的融合、联通、集成。

从多元宇宙空间数据来看,从一元宇宙、多元宇宙到跨元宇宙,再到超元宇宙,元宇宙涉及的大数据处理愈加复杂。需要通过联邦计算(federated computing)^[7]等方式,保证数据安全与隐私,并实现跨机构数据的有效流通。多元宇宙中大数据处理的复杂性不仅是数据连通的问题,差异化经济系统与价值观的内嵌也使不同元空间之间深度互联的难度加大。例如在“红楼梦元宇宙”与“西游记元宇宙”中进行物品交换的同时,要保证两个不同元宇宙时空价值体系的稳定,则需要在数据融合的过程中考虑区块链如何有效进行智能合约设计。

从多元宇宙时间数据来看,数据结构与算法差异会对人类的时间感知产生影

响。元宇宙构成了一种全新的时间存在形式,虚拟世界的时间与现实世界的时间既存在同构性、孪生性,也存在异质性、原生性。基于多重时间线的并构与拼接构建了新型数字虫洞,可随时“跳转”“回溯”与“联结”,在数据与算法驱动空间中潜移默化的影响用户对时空流速度的感知体验。特别是在考虑经济计价与货币利率等情况下,需要在数据系统中格外关注不同元宇宙中的规定流速差异与实际流速差异。

从多元宇宙与国际数据联通的层面来看,不同国家的元宇宙道路可能存在科技树(technology tree)^[8]分叉现象,即在元宇宙发展过程中选择不同升级路线,进而走向不同发展结果。例如美国强调民主自由中的个人权利实现,《头号玩家》《失控玩家》等元宇宙电影中凸显的个人英雄主义将其体现得尤为明显。而中国则强调用户责任与自由之间的平衡,集体主义与共识的达成,并警惕沉迷与成瘾的风险。与之相应地,中美元宇宙底层数据逻辑必然不同。因此,在基于不同国家而形成的多元宇宙中进行数据对接与转换时,需要进行数据规则翻译与公约制定,从而更好地实现国际数据的联通与安全保障。

5 结束语

元宇宙装备在未来会逐渐成为人类器官的延伸。通过AR眼镜、脑机接口等技术,将实现比移动互联网时代更加深入与高频的交互模式。在这个过程中,不仅需要关注元宇宙为新一轮大数据的产生与数据价值的挖掘提供的新想象空间与产业路径,也需要关注其蕴含的数据安全与风险问题。研究者可进一步展开理论与实践探索,以期化弊为利,让元宇宙与大数据真正造福于社会。

参考文献:

- [1] 清华大学新闻与传播学院元宇宙文化实验室. 元宇宙发展研究报告3.0[R]. 2022. Metaverse Culture Laboratory, School of Journalism and Communication, Tsinghua University. Research Report on Metaverse Development 3.0[R]. 2022.
- [2] ZUCKERBERG M. Founders letter[Z]. 2021.
- [3] 萨提亚·纳德拉. 刷新: 重新发现商业与未来[M]. 陈召强, 杨洋译, 译. 北京: 中信出版社, 2018. SATYA N. Hit refresh: the quest to rediscover microsoft's soul and imagine a better future for everyone[M]. Translated by CHEN Z Q, YANG Y Y. Beijing: CITIC Press, 2018.
- [4] 沈阳. 元宇宙的三化、三性和三能[J]. 传媒, 2022(14): 21-22. SHEN Y. The concept, attribute and application of the metaverse[J]. Media, 2022(14): 21-22.
- [5] GRUNDMANN S, TRABERT D, FEHRE K, et al. Zeptosecond birth time delay in molecular photoionization[J]. Science, 2020, 370(6514): 339-341.
- [6] JEONG J B, LEE S, RYU E S. Rethinking fatigue-aware 6DoF video streaming: focusing on MPEG immersive video[C]// Proceedings of 2022 International Conference on Information Networking. Piscataway: IEEE Press, 2022: 304-309.
- [7] ABREHA H G, HAYAJNEH M, SERHANI M A. Federated learning in edge computing: a systematic survey[J]. Sensors (Basel, Switzerland), 2022, 22(2): 450.
- [8] 查尔斯·辛格, E·J·霍姆亚德, A·R·霍尔. 技术史[M]. 王前, 孙希忠, 译. 上海: 上海科技教育出版社, 2004. CHARLES S, HOLMYARD E J, HALL A R. A history of technology[M]. Translated by WANG Q, SUN X Z. Shanghai: Shanghai Science and Technology Education Press, 2004.

作者简介



沈阳(1974-),男,清华大学新闻与传播学院教授、博士生导师,清华大学新闻与传播学院元宇宙文化实验室主任,清华大学新闻与传播学院新媒体研究中心执行主任,主要研究方向为AI和大数据、新媒体、元宇宙、网络舆论、虚拟人和机器人。现为近10个部委专委会专家以及多个二级学会副会长或秘书长、中宣部“文化名家”暨“四个一批”人才入选者、教育部新世纪人才计划入选者。



余梦琰(1992-),女,博士,清华大学新闻与传播学院博士后、助理研究员,主要研究方向为元宇宙、科学传播、新媒体。

收稿日期: 2022-12-27

通信作者: 沈阳, 124739259@qq.com