

ENDOR.COIN 协议

使人工智能预测惠及全人类

由 Endor Ltd.提供支持

2018年2月18日

使人工智能预测惠及全人类

摘要

Endor.coin 通过全民推广人工智能数据分析引领预测分析领域的变革，使其对于所有人而言都具备可及性、可靠性、耐审查性和有用性。

此种变革通过 *Endor.coin* 协议加以实现，提供全球第一个可供商业用户和加密令牌持有者等提出复杂的预测性问题并在数分钟内得到高品质结果的自动化自助预测平台。该平台旨在全民推广数据科学，打破其目前基本仅面向《财富》500 强企业的局面。*Endor.coin* 以麻省理工学院项目团队成员 Alex Pentland 教授和 Yaniv Altshuler 博士创立的新型社会物理学为基础。

- **不断扩大的预测门类：***Endor.coin* 在推出时将包含各类预先确定的令牌相关预测（例如：经预测增加体积、减少波动...的令牌）。这些预测可借助平台专用 *EDR* 令牌进行购买；而后，*Endor.coin* 将扩展其所提供的预测选择，允许用户发送 *预测请求*（RFP）- 推荐将要执行且可供购买的新预测类型。
- **高级用户定制 API：**该技术将为科技精通型用户和专业人士提供一个自助界面，方便其定义任何所需行为模式，使 *Endor.coin* 平台针对此种模式自动生成一种“近似”预测。
- **私人和公共数据自动融合** *Endor.coin* 商业客户（例如：银行、零售商和保险人）可轻松实现专有数据流与平台的融合，通过私人和公共数据融合产生高品质预测结果。*社会物理学*的应用可实现友好的自动数据融合，不需要进行任何清理或数据准备。
- **数据隐私：**由于 *社会物理学*能够开创性地使用在客户端完全加密的数据，数据隐私有保证。
- **由人预测，服务于人：***社会物理学*数据仅处理一次，用户只需支付其所需要的个性化部分，用 1%的成本提供当今科技巨头对应的“99%”的预测能力。

Endor.coin 根植于 *Endor* 技术平台。*Endor* 是一家“高德纳杰出供应商”，最近经世界经济论坛认定为“科技先锋”企业；公司源自麻省理工学院，由 *Innovation Endeavors* 等领先投资者提供资金支持，合作伙伴包括可口可乐、沃尔玛和万事达卡等《财富》500 强企业。

目录

1	价值主张	4
2	序言	6
	2.1 效率低下且陷入困境的市场	6
	2.2 全民推广需要放权和数据分离.....	6
	2.3 技术差距.....	7
	2.4 <i>Endor.coin</i> 协议 - 项目概况	8
3	预测科学的全民推广	10
	3.1 社会物理学 – 源自麻省理工学院的新兴科学	10
	3.2 第 1 阶段: <i>Endor.com</i> - 企业自动预测引擎	13
	3.3 第 2 阶段: <i>Endor.coin</i> 协议 - 大数据科学.....	17
	3.3.1 定义.....	18
	3.3.2 数据提供者.....	18
	3.3.3 预测引擎	19
	3.3.4 预定义预测和预测请求 (RFP)	19
	3.3.5 私人数据分析和自助 API.....	19
	3.4 路线图.....	19
4	可靠、耐审查且责任性	22
	4.1 人工智能预测责任性和真实性.....	22
	4.2 基于分散协议的耐审查性	23
	4.3 隐私保护数据分析: 使用加密数据	24
	4.4 网络效应.....	25
5	启用一个生态系统	28
	5.1 公共数据提供者	28
	5.2 学术研究团体.....	29
	5.3 推动者 – 应用开发人员	29
	5.4 数据主权.....	30
6	令牌实现	31
	6.1 区块链结构	31
	6.2 智能合约.....	32
	6.3 <i>Endor</i> 在发布时的作用	34
	6.4 令牌特权和经济性	35
	6.5 款项的使用	36
7	技术优势和差异化	38
	7.1 源自麻省理工学院大熔炉的科学革命.....	38
	7.2 实际产品, 成熟技术	38

7.3	可用性、对于用户的价值以及对于令牌持有者的价值.....	39
8	团队	40
8.1	团队主要成员.....	40
8.2	顾问	43
附录 A:	<i>所述社会物理学</i>	46
附录 B:	Endor 企业常见应用案例	82
附录 C:	Endor.coin 的预定预测实例	94
附录 D:	知识领域级 API	95
	参考文献	102

第 1 章

价值主张

任何能够快速了解、预测和影响消费者行为的企业都将具备竞争对手无法企及的超强优势。精明的商业领导者通过多种方式影响客户行为，从而提高经营业绩。他们在实施这些方法时需要弄清以下问题：

- 我们最重要的客户是谁？我们该如何获得更多顶级客户？
- 谁有可能尝试这种新发布产品？
- 我们该如何减少对促销的依赖？
- 我们下一家店铺应开在哪里？
- 谁会在下个月将产品 A 换成 B？

为找出这些问题的答案，各组织需要求助于数据科学和预测分析等功能强大的工具。但遗憾的是，这些工具的当前执行效率较低，十分复杂且成本昂贵。

- 需要招募“精英专业人士”：训练有素、昂贵且稀有的科学家和博士学位持有者
- 需要 4-6 次迭代，每次持续数天/数周
- 每出现一种新的业务问题都需要在几周内建立一个新模型
- 产品和行为发生变化时，模型不再适用

采用麻省理工学院最新型 **社会物理学** 技术并基于区块链的 **Endor.coin** 协议是第一个能在几分钟内为任何预测问题提供高品质结果的分散化可靠耐审查行为预测平台，无需进行编码、数据清理或借助博士学位持有者组成的团队。

Endor.coin 项目的关键要素如下所示：

- **技术：**采用麻省理工学院的**社会物理学**技术[1]，趋势预测精度最多可提高 10 倍，可通过完全加密数据实现深度预测（参见我们关于金融市场趋势预测的学术著作[2-6]、相关专利[7-8]和其他评论[9-11]）。
- **创新：**专注于中短期行为模式（数天到数周）自动建模，在任何其他可用技术发现这些信号之前加以感测（参见 Altshuler 博士关于 *FirstMark*“数据推动纽约发展”谈话的见解[12]或金融投资的大规模实时分析[13]）。
- **行业验证：***Endor.coin* 采用 *Endor.com* 所开发的技术，后者源自[14]麻省理工学院，由 Innovation Endeavors 等领先投资者[15]提供资金支持，合作伙伴包括可口可乐[16]、万事达卡[17]和沃尔玛等《财富》500 强企业。参见 *Finnovate 2017*[18]上精选的 *Endor* 产品。
- **奖项及荣誉：***Endor* 是一家**高德纳杰出供应商** [19]，被**世界经济论坛**评为“科技先锋”企业[20]。麻省理工大学项目团队开展的研究还获得其他多项殊荣，包括久负盛名的**美国国防部高级研究计划局网络挑战奖**[21]和**麦肯锡奖**[22]。
- **团队：**全球区块链、数字银行技术和预测分析领域知名专家团队，成员主要来自麻省理工学院计算机科学和人工智能实验室、麻省理工学院媒体实验室和麻省理工学院史隆管理学院。Pentland 教授（联合创始人）系**美国国家工程院院士**，是全球论文被引用次数最多的科学家之一[23]，最近被福布斯评为“全球最具影响力的 7 位数据科学家”[24]。团队迄今为止共发表数百篇科学论文，共获得数十项商业专利，共出版 8 本区块链、机器智能和数据隐私专业著作[25–32]。
- **重塑预测分析：**人工智能和机器学习多年来一直专门供资金雄厚的公司使用，但现在，*Endor.coin* 以极低成本为广大个人和小企业打开了通往卓越技术的大门。所供低成本数据综合分析是深入预测的基础，同时允许数据所有者对自己的数据隐私加以控制。*Endor.coin* 协议融合了新数据源和新预测引擎，形成双网效应—参与者越多，每次预测的成本越低，而预测精度越高。
- **可靠、全分散、责任性、耐审查：***Endor.coin* 协议完全分散，能为预测结果提供完整说明，可防止预测过程中出现任何操纵或偏见。此外，协议的分散性和开放性能够支持任何预测，防止任何单一权限点审查。

第 2 章 序言

2.1 效率低下且陷入困境的市场

世界各地有很多组织收集、持有和仔细维护各种数据。数据科学家和机器学习专家可以处理数据并建立预测模型，而与此同时，世界上还有许多人渴望能够预测未来（从大型公司高级管理人员到中层营销或产品经理，再到绞尽脑汁确定机票购买时机的个人）。为使后者面向自己的问题获得决策服务，所有三种实体必须共存于同一组织内。这意味着 99% 的预测是由/为大型公司利益相关者生成。此外，这一过程就目前来看还十分昂贵，因为数据科学家非常稀少且要价昂贵，况且面向每个项目的预测过程通常还需要数月时间的繁重工作。种种因素为任何对预测感兴趣的人设定了极高的准入门槛（“价格标签”）。

2.2 全民推广需要放权和数据分离

面向“普通人”的人工智能和预测分析全民推广需要开发一种满足以下要求的新范例：

分离：目前除大型企业和研究实验室之外，几乎没有任何商业、学术或非营利组织有能力维持融合 *数据监护*、*数据科学*和 *基于语义问题*的生成的高质量活动。个人、非政府组织和中小型企业通常仅专注于其中一个方面——或者是拥有（或产生）精心维护的数据，或者是雇用强大（且昂贵）的数据科学家，或者是“问题能问到点子上”的专家。因此，为打破瓶颈，使预测不再是《财富》500 强企业的专利，这三个基本要素需要从根本上分离：真正面向普通人的预测协议必须允许（公共或私人）数据提供者能够自由贡献数据，同时确保技术专家能不断贡献人工智能和预测引擎（无缝接入并融合预测协议），所有这一切都允许最终用户轻松利用基于这些数据源并由这些引擎专门为其准备的预测。

责任性：对于一家科技巨头的数据科学部门而言，责任性并非关键特征，因为很显然“所有人都在为同一个老板工作”（例如：负责商业智能和营销预测等方面的相关高阶主管）。但对于数据、智能和计算都是基于特定责任持续租借的平民化平台而言，责任性成为确保“公平竞争”的一种基本要素，可调节所有利益相关者的优势职能追求“长效”，而非鼓励短期收益。

分散：分散性预测框架有两种关键贡献要素，其中一个与工程相关，另一个则围绕耐审查和防偏差。大量实例表明，分散性解决方案往往更加容易升级和扩展。数据源、计算资源和不同类型的预测引擎的增加都极大地受益于分散性解决方案。此外，唯有分散式架构能确保这些预测不会受到通过生成偏见性结果或货币资源仲裁执行的明确审查或隐性审查。此外，高效率的分散还是产生网络效应的关键所在；随着参与人数（数据提供者、预测引擎和预测消费者）的增加，网络效应有助于降低用户平均成本并不断提高预测精度。

2.3 技术差距

遗憾的是，尽管数据科学全民推广的各种优点长期以来已得到验证，相应框架的现实实施一直以来都是一项极具挑战性的任务，主要原因还是技术方面 - 当今主流科学无法支持“一般分散性行为预测”范例。当今存在的各种技术，无论是神经网络（或深度学习）、遗传编程、决策森林还是支持向量机，在任何“实际”机器学习工作开始之前全部需要大量数据清理、处理和理解。这是行业当今面临瓶颈的根源所在：高技能水平数据科学家工资提升，而且人员非常稀缺。附录 A 第 3.1 节详细论述了该主题。

由于目前尚无可 *自动消化任何数据* 从而允许非专业人员和专业人员提出 *任何预测性问题* 的科学突破，行业局限于现有范例，公司不愿意新雇用数据科学家并向其支付 6 位数工资，这严重束缚了行业发展。

2.4 Endor.coin 协议 — 项目概况

为打破上述限制条件，必须创立一种新科学。*Endor.coin* 创始人 Yaniv Altshuler 博士和 Alex “Sandy” Pentland 教授联合提出的 *社会物理学* 是一种数学理论，可对人类群体行为方式进行高效建模。*社会物理学* 理论通过行为数据源中产生的一组数学方程将任何行为数据源自动转换成一组行为集群，不需要对数据语义（或要问的问题）进行清理、预处理或了解。所述的这组行为集群称为“*知识领域*”。

Endor.coin 协议所依据的事实为：行为数据在行为集群正则表述中进行处理时，数据科学传统过程可（最终）分解为基本要素，以分散方式分配至不同执行者。*Endor.coin* 协议基本概况和主要组成部分如下所述：

正则数据表述：提供给 *Endor.coin* 网络的所有数据都会转换为“*知识领域*”正则表述。转换需要借助各种预测引擎（见下文），执行费用由引擎支付。数据经过此种转换之后，从中提取的行为集群可与其他类型数据的集群合并，产生高效的自动预测过程（详细说明参见第 3.1 节和附录 A）。

数据提供者分离：数据转换为“*知识领域*”表述时，数据提供者不再需要积极参与分析的后期阶段，使得数据所有者可以将其（全部或部分）数据与 *Endor.coin* 网络融合，在生态系统中成为自主利益相关者，主要职能是保持其数据质量，控制哪些人可以访问数据的哪一部分，并在财务方面受益于其所提供的预期价值。

预测引擎分离：数据科学家们都知道，数据科学项目约有 90% 的时间都消耗在数据清理和预处理上。由于 *物理学* 革命性地首次实现这些步骤的自动化，各种预测引擎最终可与各种类型的数据源无缝对接。预测引擎提供者在连接 *Endor.coin* 网络方面唯一需要注意的是支持 *Endor.coin* 协议— 被定义为消化数据集的能力（有选择地由引擎定义），并需要提供“*知识领域*”形式的输出（参见附录 D 所述完整规范和 API 代码）。

分散执行：数据提供者利用区块链可捐赠数据（保存在 AWS 中），借助 *Endor.coin* 协议可自由访问。提取行为集群由预测引擎以分散方式实施。最终用户要求 *Endor.coin* 智能合约发布一项特定预测（用于 EDR 费），从而发起询问。每项预测的最佳行为集群由 *Endor.coin* 预测代码（可在项目 GIT 账目上自由访问[33]）进行选择。资金仲裁在数据贡献者当中由智能合约维护，依据是针对该项预测选择的集群。此项仲裁优化了预测的质量和结果的公正性。

数据主权：*Endor.coin* 网络包含的每个数据元素都可标记为“公共”或“私人”数据（相同的数据源可能包含一些标记为“公共数据”的列，而另一些则保持“私人数据”的状态）。所有预测引擎都可使用公共数据源，用于任何未来预测的行为集群。作为回报，其提供者在被协议选中时会获得 EDR 令牌补偿。私人数据元素也可用于各类预测引擎，但要采用加密形式。然而，此种数据源中提取的集群只有在请求进行预测的用户提供数据密钥时才可选作预测数据源。更多详细信息参见第 5.4 节。

责任性和耐审查性：利用区块链可无限保存预测，还可供乐于推断平台信誉、平台所使用的数据或对其进行分析的预测引擎的任何人使用。此外，由于 *Endor.coin* 协议包含一个负责选择用于生成各种预测（以及所用资金的判定）的行为集群的开源预测码 [33]，它会确保不存在任何偏误，仅对精度进行优化。更多详细信息参见第 4.2 和 4.1 节。

全面提升预测效率：最终，*Endor.coin* 协议帮助终端用户以较低成本获得出色的预测效果，主要是得益于过程自动化（无需雇用昂贵的全职数据科学家）以及社会物理学无缝“融合”各类行为数据源的能力。这意味着，即使可口可乐等规模较大的商业客户采用 *Endor.coin* 网络后也能立竿见影 – 通过将其数据标记为“完全私人”，用户可基于专有数据融合（与系统所含公共数据融合）实现深入预测，自行获得该数据并对其进行分析所需付出的选择成本大大降低。此外，这还能提供一种积极的网络效应，加入 *Endor.coin* 网络的用户和数据提供者越多，则成本越低（更多详细信息参见第 4.4 节）。

第 3 章

预测科学的全民推广

3.1 社会物理学 – 源自麻省理工学院的新兴科学

社会物理学是一门革命性新兴科学，利用大数据分析和生物数学定律了解人群行为，使 Endor 能够克服传统机器学习的局限性。此种新型科学源自麻省理工学院 Alex “Sandy” Pentland 教授和 Yaniv Altshuler 博士的研究成果，并由 Endor 借助自身专有技术进一步开发，由此产生的强大引擎可以解释和预测任何人类行为，甚至是迅速变化和发展的行为。

简而言之，社会物理学所依据的前提条件是：所有代表人类活动的事件数据（例如：通话记录、信用卡消费、乘坐出租车和网络活动）都一定包含一个嵌入数据中的人类活动模式构成的特殊组合。人口统计资料所涵盖全部类型的人类数据共有的这些数学不变性可作为新兴行为模式在被其他任何技术发现之前的检测过滤条件。

社会物理学功效说明：试想，一家大型银行的营销部不断致电近期可能需要贷款的客户，部门收益直接源自实际积极响应致电的客户。由于这种持续活动的直接营销成本极高，必须在适当的时间向适当的客户致电：倘若太晚，客户可能已经从他处获得贷款；如果太早，客户可能还未形成实质性的需求。

此种情况下，银行可考虑采用两种工具进行客户预测：银行内部数据科学团队建立的机器学习模型和 Endor 引擎。下面简要描述了每种工具的优点所在：



由机器学习模型确定的客户群包含两种客户：一种客户会积极回应银行营销方案（例如：真肯定），而另一种则不会积极回应（例如：假肯定）。例如，假设真肯定占模型结果的 10%。大量实验表明，Endor 社会物理学引擎还能检测到这 10% 中大部分的客户，其主要有两种差别：（a）Endor 引擎不会报告机器学习模型中大量假肯定；（b）Endor 得出的结果会包含传统模型无法检测的其他许多真肯定情形。由于预测精度/回应更好地彼此结合，销售绩效将大幅改善。



如何实现？检测时间模式：人类现实生活由许多较小的临时事件和变化组成。社会物理学包含人类活动基本动态，因此能更好地发现因所处社会环境最近发生的变化而遵循某种行为模式的小规模人群。因此，社会物理学能够捕捉人类行为数据动态信号，这是因为：倘若没有社会物理学的帮助，此种信号便会缺少任何统计显著性，使传统机器学习和深度学习方法无从分辨。

机器学习和深度学习与社会物理学 - 各种情形下的优劣比较。利用数据科学和大数据分析工具解决业务查询问题时，机器学习和社会物理学都是可行的选择。下表有助于根据问题属性确定适当的工具。

为何选择社会物理学？人们早在上世纪 70 年代便开发了图像处理 and 文本识别等各种数学和统计学机器学习技术解决“静态问题”。此类问题主要为数量较少且相对稳定的“信号”所支配。一个经训练的文本识别模型在处理麻省理工学院一名 2016 级学生的手写文本和分析阿尔伯特·爱因斯坦亲笔信时可实现类似功能。同样，Siri 和谷歌语音识别引擎也都能将肯尼迪《我是柏林人》的著名演讲进行高品质转录。

机器学习更适合：		社会物理学更适合：	原因
数据类型	机械/物理驱动的数据 示例： <ul style="list-style-type: none"> 进行故障预测 监测石油钻井泵控制数据进行故障预测 人脸识别 	人类行为数据： 示例： 分析财务交易，以预测愿意购买付费服务的人群 	人类行为具有不稳定性、不可预测性、杂乱性、复杂性和动态性。从数学角度讲，人类行为由大量“时序”信号主导，每种信号都会影响一小群个人。因此，通过“学习”人类数据建立一致稳定的表述模型难度极大。 Endor 利用社会物理学检测此种时序信号，因此特别适合基于人类的数据。

然而，人类行为有其自身特殊性。在大量“动态信号”支配下，人类行为动态性极高而且极为“琐碎”。经训练后通过信用卡消费数据检测“千禧一代”的传统机器学习模型准确度随时间而迅速下降，需要由技术精湛的专业人员通过不断加入新语义知识进行持续维护。鉴于“千禧一代”行为的频变（和不断变化）趋势，将其定位在数据中不仅需要不断对模型进行再训练，而且需频繁开发新功能来检测这些趋势（即：并非原始数据组成部分的复杂综合行为特性）；语义领域专家与数据专家密切合作才能实现。



“人有生死,国有兴亡,而意志长存。”

机器学习算法自上世纪 70 年代以来变化不大。原本是要用于解决静态问题



“一旦我们接受了自己的极限,就要超越它。”

用作一种基于人类数据的行为预测工具时，社会物理学主要优势如下：

	传统机器学习	深度学习（不含社会物理学）	社会物理学	原因
小数据集	能够分析小数据集，但需借助专业数据科学家，而且是一个耗时极长的过程	每个问题需要使用大量数据	回答人类行为相关问题时所需数据极少。结果自动生成（无需数据科学家参与）	Endor 生成结果时无需借助“大数据”，因为社会物理学已包含人类行为数据的基本动态特性。因此，即使是非常小的数据集也可立即产生准确的预测和可用信号。

3.2 第 1 阶段：Endor.com - 企业自动预测引擎

如以上章节所述，就当今数据科学的发展状态以及机器学习技术的利用率而言，它们仍然是资金雄厚科技巨头的“专利”。这些公司有能为日益稀缺的专业人员提供六位数字年薪。尽管如此，即使对于这些公司而言，数据科学和预测分析还远不是商品。相反，一个旨在实现可靠预测的项目通常组建一个由 2-3 名专业人员组成的团队，每个小组经过数周 4-6 次迭代，在第一个模型建成后通常需要持续维护。因此，按行业标准估算，每个预测项目平均总成本约为 150 万美元。这意味着：

- 大多数公司都无法承受多个项目的资金。
- 为保证投资回报为正值，如今的各种项目必须确保巨额毛利，而且只能关注最核心的业务领域。

	传统机器学习	深度学习（不含社会物理学）	社会物理学	原因
特征与原始数据	需借助技艺精湛的数据科学家和/或领域专家来定义并选择适当的原始数据特征表述	不需要特征，可处理原始数据，但问题类型极其有限（*）	不需要特征，可针对任何类型的人类行为相关预测问题处理各种原始数据。	机器学习需要耗费较长时间将原始数据转换为有意义的特征，这通常是一个手动过程。每个问题和任何新数据类型通常需要从头开始执行这个过程。 尽管深度学习自动涉及特征加工，但仍需要大量数据，而数据要求会随着问题的复杂程度而增加。因此，它仅限于“简单行为”。 另外，深度学习通常局限于“静态问题”，因为深度学习动态分析需要借助大量数据，而普通公司一般无法提供如此大量的数据。 社会物理学可将任何人类行为原始数据自动转换为人类行为集群的规范形式。 Endor 利用该规范表述处理所有数据类型和所有问题（无论数据大小如何），并生成一个统一的人类行为数据集，其利用深度学习功能回答所有预测问题。

	传统机器学习	深度学习（不含社会物理学）	社会物理学	原因
用户和所需专业知识	机器学习专家，通常需要领域专家协助制作语义特征	深度学习专家	商业用户 你只需要提供“你想要更深入了解的人”的示例。	机器学习需要“学习”大数据集的潜在正常行为或使用先前的领域专业知识。 Endor 已包含人类行为数据的基本动态特性。
数据变化速度	仅限缓慢变化的数据。数据变化还需要领域专家持续干预，以便调整特征。	可以处理动态特性，仅限于缓慢变化的数据（针对人类行为数据的严格限制！）	可以自动方式轻松分析快速变化的数据（无需借助领域专家）。	Endor 引擎特别适合人类行为数据，因此本质上适合具备动态特性的数据。 由于社会物理学是任何人类数据集中嵌入的一组数学不变性，其甚至可以检测到代表极短时间段的信号。换言之，它可以比其他技术提前发现正在产生的变化。
分析范围	特定/有限		范围广/任何有关人类行为数据的问题	对于机器学习而言，每个数据集和问题必须多次重复学习过程，因为自动选择的模型特征需要重新学习。社会物理学是基于不因问题而异的人类行为基本原理。

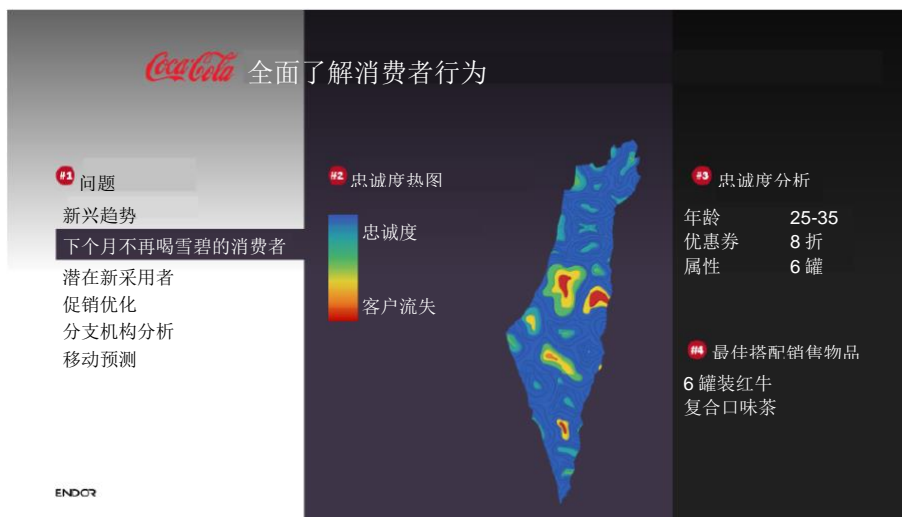
	传统机器学习	深度学习（不含社会物理学）	社会物理学	原因
数据清理	机器学习极易受数据中噪声和间隙的影响。需要进行漫长而昂贵的数据清理；深度学习通常需要将数据仔细转换为深度学习工具可接受的格式		无需进行数据清理	机器学习和深度学习均使用数据驱动的数学模式，以演绎规则、提取信号并生成预测，这需要执行一个精确详尽的数据清理。 另一方面，社会物理学采用外部模式 - 即每个人类行为数据集中嵌入的数学不变性。这可以显著降低数据噪声的影响。 此外，社会物理学将原始数据转换为行为集群，进一步降低数据间隙和噪声的影响（大部分是自动滤出）。
迭代和调整	每项数据调整或问题定义都需要领域专家（商业用户或分析员）和机器学习/深度学习专家相互合作。每次迭代都会耗费数周时间，典型项目通常至少包含 4-6 次迭代。		针对能够修正和修改预测查询的商业用户或分析员设计专用界面。 结果随后根据新定义自动调整。	Endor 引擎特别适合人类行为数据，因此本质上适合具备动态特性的数据。 由于社会物理学是任何人类数据集中嵌入的一组数学不变性，其甚至可以检测到代表极短时间段的信号。换言之，它可以比其他技术提前发现正在产生的变化。

- 公司必须不断进行优先排序，只会将超级昂贵的数据科学工具用于其认为可以（a）获得技术成功+（b）获得较高业务回报的项目。

源自麻省理工学院的 **Endor** 是围绕最新的社会物理学而创建，旨在通过打破当今人们对数据科学的认知方式解决这一问题。借助一支研究者和工程师精英团队在麻省理工学院长达 3 年的研究以及另外 3 年的开发工作，**Endor** 成功开发出全球首个全自动“数据科学服务”引擎，允许各公司上传其所拥有的任何行为数据，并在快速融合（通常只需要几小时）后开始提出关于数据所含任何“对象”（用户、产品、优惠券和位置等）未来行为相关的预测问题。

年成本不到 100 万美元时，公司有机会提出几十个预测问题，在数分钟内快速得到结果。这不仅仅是一种线性改进，更是一种范式转变，因为执行“预测项目”就像“谷歌搜索”那么简单，基本上会对潜在项目进行不必要的优先排序，因为任何决定现在都可以预测为导向。

示例一 - 可口可乐联合研究：Endor 和可口可乐[16]最近联合开展的一项研究表明，社会物理学能够针对各类消费者行为提供精确预测，包括品牌抛弃和忠诚、新产品的采用以及对营销活动的反应等。Endor 的社会物理学引擎将数百万次销售点交易用作原始输入（代表 3 个月期限），检测到近 2000 万相关异常，每项异常代表一个现实单行为组。鉴于每组确切含义未知，行为组用于行为外推。给定样本集（例如：新产品早期采用者）的情况下，系统利用上述多个行为组寻找行为上类似于样本集成员的用户（例如：很快就会体验到新产品的用户）。采用该方法提出 15 个不同的预测问题，每个问题得到一份需求预测报告。为进行验证而保存的样本外数据表明报告准确性较高。



“社会物理学主要是基于大数据的行为分析，并将此种分析推向全新高度。作为 Endor 的合作伙伴，我们倍感幸运。”

Alan Boehme 博士，可口可乐公司首席技术官

示例二 - 推文自动分析：最近开展的一项试验将 1500 万条推文元数据输入 Endor 引擎，作为原始数据进行分析。此外，客户还告知了 50 个已确定是 ISIS 活跃者的推特账号，它们包含在输入的数据中，以验证 Endor 是否能检测到数据中隐藏的其他 74 个账户。借助一台笔记本电脑，Endor 引擎仅用 24 分钟（从原始数据导入系统到输出最终结果）便完成任务，结果表明有 80 个推特账号与所供示例“相似”，其中 45 个（56%）经证明就是 74 个隐藏账号的一部分。值得注意的是，此次测试过程中的误报率极低（35 个误报结果）；因此，客户有能力委托人类专家调查认定的目标。

“这绝对是一项革命性概念，一项真正的技术突破。其所提供的结果令同类工具黯然失色。”

以色列情报部门首席信息官

需使用生态系统使能分散协议：Endor 引擎利用社会物理学和强大的专业团队证明了《财富》500 强企业同样能以少得多的投入博得更多收益。然而，该产品主要是面向大型银行和零售商，年许可费用仍然在 25-120 万美元区间内。这肯定不适用于“长尾”企业，更不要说个人了。

因此，我们明显还需要另一种解决方案；这种方案应能使所有人都以合理价格从社会物理学新技术中获益。因此，该方案必须：

- 具有自主性，并由特定参与者提供的资源组成。
- 产生强大的网络效应，激励参与者不断加入，在成长过程中变得越来越好，越来越便宜。
- 保持公平，确保本质上公正可靠。此项要求不同于其他服务，因为预测还具有所谓的“外部效应”，其产生的原因来自于依靠于它的个人或组织的行动。

基于这一概念，Endor 团队将自豪地展示其革命性进程的下一步 - *Endor.coin* 协议！

3.3 第 2 阶段：Endor.coin 协议 - 大众数据科学

Endor 将基于社会物理学的行为分析在行业内实施成功后便创立 *Endor.coin* 协议，旨在将这种技术推广至长尾企业和专业人员。*Endor.coin* 协议重塑了预测分析技术，将人工智能推广到整个行为预测领域，从而形成一个普遍推广的生态系统。此外，该协议完全分散的特性能够确保可靠性、耐审查性和责任性。行为预测第一次基于合理的费用和安全框架“来到民间”，不再受当今控制着数据和技术的科技巨头操纵。

3.3.1 定义

原始数据：*Endor.coin* 协议支持任何时间相关或交易数据源（例如：呼叫数据记录、ERC20 区块链和应用程序内购买等）。数据提供者必须指明预测依据列。例如，利用 ERC20 可预测并发现感兴趣的未来行为地址或代币本身。可使用推文数据寻找感兴趣的推特 ID 或者感兴趣的位置或散列标签。

处理过的数据：表示已经过行为集群提取并转换为“知识领域”形式的社会物理学典型表述。

知识领域：是指提取自一个或多个原始数据源的一组行为集群。*Endor.coin* 协议可实现任意数量知识领域的拆分或组合。由于不同的行为集群对应不同的预测类型，对于给定的预测，*Endor.coin* 协议会在任意给定时间选择最相关的集群，并生成“知识领域”，用于生成实际预测。

预测：*Endor.coin* 协议支持语句形式为“按行为上类似于 Y 组的概率为 X 组排名”的任何问题。例如，倘若 X 组包含所有 ERC20 代币且 Y 组包含最近量显著增加的代币，预测结果将包含一系列 ERC20 代币，其排序后顶部包含统计上近期最有可能增加量的代币，而底部则包含最不可能呈现此种行为的代币。

3.3.2 数据提供者

Endor.coin 协议支持任何时间相关结构化行为数据，只需通过简单的 API 调用便可实现数据导入，使数据所有者在提供数据的同时还能控制哪些列保持隐私状态、哪些列可供用于公共分析。*Endor.coin* 用户定义的隐私数据仍与公共数据流自动融合，隐私和公共数据融合产生高质量深入预测。可通过自动和友好的方式使用社会物理学“知识领域”（参见附录 D）数据融合，无需进行数据清理或准备。

数据提供者需要为其数据的分析向预测引擎提供者支付 *EDR* 代币，其数据所提供的信息用于预测时可相应获得 *EDR* 代币奖励，从而激励高品质数据流的提供与维护。

3.3.3 预测引擎

Endor.coin 协议基于数据向部分重合行为集群特征空间的投射定义了一种“预测语言”。当前的 *Endor* 引擎将成为插入该网络经发布后立即可用的第一款预测引擎。尽管如此，*Endor.coin* 将促进、协助并资助新预测引擎的开发，旨在建立一套由各类引擎组成的生态系统，提供互补能力，提升性能精度，并增强可靠性。预测引擎数量的不断增加还可保证结果的公正性，因为预测将会基于开源协议从各种预测引擎所提取的一系列集群中自动选择的最相关的集群。

3.3.4 预定义预测和预测请求 (RFP)

Endor.coin 协议发布时包含大量各类预先定义的预测，可利用平台的 *EDR* 代币进行购买。而后，*Endor.coin* 将扩展其所提供的预测选择，允许用户发送 *预测请求* (RFP) - 推荐将要执行且可供购买的新预测类型。

这种在平台支持下逐步扩展预测的方法将利用群体智慧（就像协议用户申请所展现的智慧）优化最新支持的各种预测。

3.3.5 私人数据分析和自助 API

后续版本的 *Endor.coin* 协议将在全“自助”API 中为高级用户提供支持：精通技术的用户和专业数据科学家可使用自助界面轻松加载专有数据，并独立建立新的预测类型。这些预测可定义为私人预测，也可公开共享（在广泛应用时奖励预测开发）。

3.4 路线图

在当今通用软件市场上，我们的方法可与平台即服务或最近提出的区块链即服务等相媲美。*Endor.coin* 代币（或 *EDR*）将用于支持平台交易。*EDR* 作为密钥或软件许可，随着时间的推移，可通过被鼓励扩展应用领域的开发者群体使用更多代币提升性能和规模。此外，我们的管理员将获得一个“仪表盘”，用于分配代币、监测使用情况并根据需要购买更多代币。

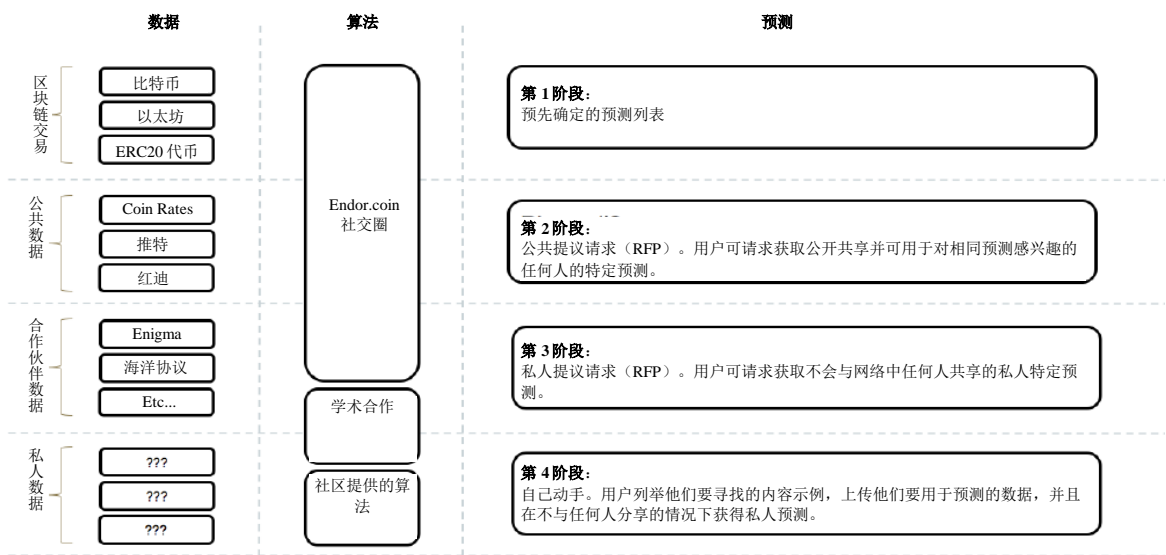
Endor.coin 进化发展的下一步是要通过不断成长的推动者群体有机地扩展我们的生态系统。每个人都可以提供以 **EDR** 支付的预测请求 (**RFP**)，相应的预测请求可由任何推动者处理，并通过智能合约对相应的支付获得奖励，有效扩展我们的预测范围，并由 **Endor.coin** 作为新支持查询嵌入（利用 **EDR** 可进一步供所有人使用）。

在这第二阶段，**Endor.coin** 将演变成一个模块化应用平台，供推动者借助区块链的力量扩展预测领域。为此，**Endor.coin** 为可复用软件组件的作者提供可扩充小额支付流，构件可永久组合和重组，形成不断扩展的查询应用程序库。推动者有权针对每个由其添加到 **Endor.coin** 平台的组件获得一个微使用许可。最终用户安装其所选择的应用程序。要支付的许可是程序所用组件全部微使用许可的总和。**Endor.coin** 通过智能合约向最终用户收取费用，并将款项分配给相关推动者。

随着时间的推移，**Endor.coin** 将发展成为创业编码者和企业的主要开发平台，供其在区块链上建立富数据网络和移动产品，使目前仅面向高技能人员的机器学习得以推广。此外，目前无法满足数据分析算法高效执行所需交易量的区块链将通过不断增长的推动者人群获得这些能力 - 推动者将利用持续扩展查询范围的引擎创建应用程序。

我们希望创建一个可靠的全自动分散式预测架构，确保其对最终用户完全透明。鼓励措施是支付 **EDR** 代币，应用程序基本上是一组插件。可预见一个正向反馈环：创建的应用程序越多，系统中添加的插件越多，可以重新使用的唯一组件越多，最终在整个 **Endor.coin** 生态系统范围内产生相互自强化网络效应。

最终，我们将为高端客户提供一个白金平台，供专业人员和企业提供更复杂的细分化预测请求。为实现上述复杂查询，我们将通过为提供者提供下列 **EDR** 扩展数据存取：推动者请求从实体获取数据时（时间受限），相应的实体提供一个价格（时间受限）。如果推动者同意支付，实体会向以推动者公共密钥加密的数据发送一个链接（此时通过智能合约拨付资金）。推动者和数据提供者均可公布评审结果（供未来积累声誉）。**EDR** 将由数据提供者付款和白金应用程序开发者所供资金之间的套利产生。



第 4 章

可靠、耐审查和责任性

4.1 人工智能预测责任性和真实性

Jeremy Epstein 在其名为《即使不知基于区块链的人工智能为何物也要尝试》[34]的文章中与读者分享了他与自己年轻的女儿和亚马逊人工智能工具 **Alexa** 之间的对话。这次对话围绕“网络中立性”这一主题展开，**Alexa** 在对话过程中分享了大量相关信息。然而，鉴于 Epstein 对主题的轻车熟路，他惊奇地发现，这些信息虽然准确，但都有一定的片面性。这次体验暴露了基于人工智能的未来飞速发展的一些隐患，同时也提醒我们，亚马逊等大型公司在大数据和人工智能方面具有巨大传统优势。

Epstein 语录：

如果将收集、存储和分析尽可能多的数据比作是比赛指标，那么谁已经在竞争中领先了呢？没错，它们是美国的 **FANG**（**F**acebook、**A**pple、**N**etflix 和 **G**oogle）和中国的 **BAT**（**B**aidu、**A**li 和 **T**encent）以及《财富》1000 强和类似的大型跨国企业。

只有它们有途径和财力获取、存储和分析更多数据，并在此基础上建立人工智能模型。此外，只有它们有能力支付 30-50 万美元的起薪以及七到八位数的顶薪。如果你的儿女能成为顶级人工智能科学家，那已经非常了不起了，因为这就意味着不进 **NBA**（美国职业篮球联赛）或 **NFL**（美国职业橄榄球联盟）就能拿到极高的薪水。

所有这些因素形成一种网络效应，使得富人变得更加富有和强大，而创新门槛则变得更高。

不仅创新受到影响，大公司人工智能的封闭特性也意味着社会必须将其信任关到“黑盒子”里。

因此，“预测真实性”完全依赖提供以下特性的基础设施的可用性：

- **责任性：**回想过去，预测提供者可证明其预测的正确性。预测的消费者能基于过去可靠地推出任何预测提供者产生任何类型的准确预测的效率（或缺乏）。换言之，信誉应该无法操纵。
- **真实性：**预测提供者是他们自称的那些人。假冒误导预测消费者应该不太可能。
- **无偏见：**预测提供者之间应是公平竞争，利用市场力量鼓励准确度的提升，而不是通过偏见性预测获得的隐秘收益。
- **准确性：**预测应足够精确，以便向网络提流入资金，在先前三项要求方面关闭正反馈环。*Endor.coin* 协议提供了满足上述所有要求的基础设施：
- ***Endor.coin* 协议责任性和真实性：**区块链使得预测可以无限存储，并可供任何人使用。时效性预测仅发送给消费者时还会以加密公共版本存储，其密钥经过特定时间后才能发布。
- ***Endor.coin* 协议零偏差：**协议为每项预测选择最相关的行为集群（无论是分析引擎还是数据源）。因此，由于协议是一个不受数据提供者和分析引擎提供者控制的独立实体，可从本质上避免偏差。
- **准确度：***Endor.coin* 是基于麻省理工学院的社会物理学技术，经行业大量事实证明可为各类应用情形提供稳定的精确预测。

4.2 基于分散协议的耐审查性

审查是一件棘手的事情。审查表现较明确时较容易发现，因此可利用 IP 代理等“低水平”技术解决方案加以规避。相反，隐式审查则完全是另外一回事。众所周知，谷歌和 Facebook 等企业阻隔了某些类型的搜索查询。这种阻隔往往是出于合理的道德或法律原因，但我们如何才能保证情况总是如此？难道我们能“强迫”谷歌提供一个其已认定的“不能轻松访问”相关网页吗？难道我们能“强迫”Alexa 满足我们提出的一些其认为不合适的请求？答案当然是否定的。

由此而论，预测非常类似于搜索，因为其同样需要经过运营实体或具有管辖权的监管机构的审查。人工智能和预测分析的工作原理表明，它们在潜在审查方面存在三个基础层：

- **数据储存库：**确保数据的完好性、完整性和安全性（输入是否准确可靠，是否能被操纵或盗取？）
- **算法/机器学习引擎：**确保预测不会受到集中管理机构检查，且所有预测请求都会公平执行，服务质量要素与预测主题无关。
- **查询界面：**可靠表述预测查询的输出，有效获取新数据，对与预测主题无关的所支持预测有任何限制条件。

如果一个人完全依赖集中预测源而作出决定，则表示绝对自信地假设上述所有要求都得到满足。对于集中封闭式模型预测方案，如果一个在不了解幕后情形时被要求信赖每一层，则所述置信度难以令人信服（如果不是轻信）。

Endor.coin 协议本质上通过允许执行任何预测以完全分散的可靠方式提供这些主张。数据一旦提供并加载到网络，任何相关预测都可执行，**Endor.coin** 协议会对其进行自动优化。

4.3 隐私保护数据分析：使用加密数据

社会物理学全新科学在数据分析和预测用全民生态系统形成方面具有关键优势：整个数据竖井内可在不影响数据隐私和完整性的同时完成计算，因为社会物理学计算可使用加密数据。特别需要注意的是，**Endor.coin** 分析数据的独特方式使数据提供者能够对在不公开数据本身的情况下以加密方式分析数据的活动进行货币化。与此同时，系统可实现“内隐行为集群”（映射到散列或加密数据空间）的提取，后续用作“相似预测”的依据- 产生不依赖于源数据性质和预测问题语义的准确行为预测。

Endor 目前已采用类似技术与大型银行和金融客户展开合作，允许后者以完全加密的方式加载其数据，完全避免损害数据隐私和完整性。

4.4 网络效应

值得注意的是，*Endor.coin* 协议完全分散化会产生多种网络效应，在应用扩展过程中将其优点体现地淋漓尽致。协议应用增加预计可能产生的主要推动作用如下所示：

用户增加会使平均成本降低：与大型商业客户不同的是，个人和小型企业所需要的预测主要是基于公共数据集。*Endor.coin* 协议使此类数据集能一次性完成分析，所生成的一系列行为集群称为“知识领域”。该数据结构可支持所有基于该数据源的预测用例，需要最终用户支付少量费用对其“个性化要素”进行封闭阻隔 - 即行为集群提取和其感兴趣的具体应用实例之间的差量。

换言之，对于给定的数据集（即比特币区块链）以及具体应用实例，成本计算如下所示：

- **知识领域计算：**约占总成本的 99%，进行一次性计算。此项任务所需资源在该数据流的所有用户之间分摊。
- **预测的个性化要素：**约占总成本的 1%，针对每个用户都要计算（并支付）一次。此外，对于不同用户进行的类似预测，*Endor.coin* 协议会重复利用后续预测过程中首先计算得到的预测见解，在成本不便的情况下进一步提高预测精度。

这意味着 N 个最终用户的网络管理成本方程应为：

$$\frac{0.99 \times C}{N} + 0.01 \times C$$

C 表示单个商业参与者进行相同预测所用成本。这最终意味着，随着用户数量 N 的增加，预测精度也随之提高，其成本约为大型商业客户所用相同系统成本的 1%。

最后，问题提问者越多，所得结果越令人满意，则成本越低，因为绝大部分成本均为所有活跃用户共同分担。

增加数据提供者数量有助于提高准确度：根据 *Endor.coin* 协议定义，数据提供者需要使用 *EDR* 代币为其数据执行提供支持，根据从数据中提取的行为集群形成“知识领域”。而后，这种初始执行成本由最终用户部分返还数据提供者，或者附带大量利息 - 具体是要根据数据质量及其对各种预测查询的贡献。这种资金仲裁由 *Endor.coin* 协议进行维护，后者可使用各种数据集上由各类分析引擎检测到的所有行为集群，并选择相关性最高的顶部集群。而后，执行代币基于其对最终预测的贡献按比例交付给数据集提供者。

此项机制的结果是，高品质数据的所有者在激励作用下继续支持其数据源（甚至是进一步提升其质量和可用性），而劣质数据的提供者则被完全清除，只是因为其费用未被偿还。此种经济体系利用市场力量自动确保预测协议可用数据源仍保持最高匹配度。因此，新的数据提供者只会提高整体预测精度，却不会造成总体成本的升高。

增加预测引擎的数量有助于提高预测效率：*Endor.coin* 项目第一阶段利用 *Endor* 预测引擎作为行为集群提取的第一个提供者。然而，其他预测引擎预计会适时支持 *Endor.coin* 协议。此类新预测引擎的应用预计会对数据提供者和最终用户产生显著的积极影响：采用不同技术的引擎可基于相同数据源产生不同类型的集群。这意味着，随着支持 *Endor.coin* 协议的预测引擎所用技术种类的增加，请求进行新预测时可供协议选择的集群种类也会增加。这预计主要会有三个方面的影响：

- **精度提高，对新预测的支持增加：**随着各种新集群类型的出现，*Endor.coin* 协议可以更好地选择生成每项所请求预测所用集群。对于现有预测，这有助于提升准确度（因为可以使用基于集群技术的“正交洞悉”）。尽管如此，这也意味着，新预测在该点因次优精度而不受系统支持之前会实现较好的执行成本效益，增加系统所支持的预测种类，因在相邻时间执行的预测之间共享的分析结果进一步提升整体精度。

- **减少每次预测的成本：**系统所支持的预测数量的增加使得为预测而支付资金的最终用户数量增加，从而立即扩大初始数据分析可支付 EDR 代币总数，而后进一步增加数据提供者所得资金量，同时降低最终用户每次预测的成本。
- **新数据源的经济可持续性：**最后，新集群技术的可用性使得某些类型数据源更加经济可用，前提是这些新技术对此类数据源以及可用数据源分析的兼容性更高。此种情况下，这些数据源将迅速成为 *Endor.coin* 协议的可用信息源，向平台支付其初次融合的费用，然后增加可用数据源的数量，并由此产生各种效益。

第 5 章

启用一个生态系统

Endor.coin 项目旨在建立一个综合生态系统，实现数据所有者、开发者、数据科学专业人员、小型企业和个人用户之间的协同合作。通过促使每名参与者都能贡献其资产（包括应添加的新预测相关数据、资金和想法等）- *Endor.coin* 社区将要提供、启用和促进自我延续正反馈业务环。该生态系统主要方面的简要说明参见下文。

5.1 公共数据提供者

Endor.coin 协议将从第一天开始鼓励各方提供注册数据提供者第一阶段提供的公共数据流（经过一个合规与质量保证过程，确保生态系统基于质量最高的数据寻求自我发展），随后由希望丰富可用于分析的数据的任何人员或公司提供，并由 EDR 提供奖励。

潜在数据提供者包括：

- **数据合作伙伴：** Enigma、Twine、Thasos、海洋协议等将其数据流（部分或全部）连接至 *Endor.coin* 平台供分析的项目。
- **社交渠道：** 例如推特、红迪等。
- **数据清理者：** 从各类公开数据源进行数据源的收集、清理、结构化和语义丰富化。
- **区块链协议：** ORBS 或类似协议，方法是建立用于下载并解析数据的节点，并使其可用于分析。

Endor.coin 正在与多家所述潜在提供者进行合作谈判。

5.2 学术研究团体

由于 *Endor.coin* 协议支持综合多个分析引擎进行原始数据群集，公司会利用一些款项鼓励与领域内领先的学术研究团体协作。该活动过程中开发的技术可轻松融合至不断成长的 *Endor.coin* 生态系统，进一步用 *EDR* 代币奖励开发者，前提是其用于提高预测精度。

此项工作将在 *Endor.coin* 科学咨询委员会（参见第 8.2 节详细清单）指导下进行，该委员会由全球数据科学、机器学习和区块链技术领域的业界和学术界领导者组成。

作为依靠麻省理工学院项目创始人开发的社会物理学科学革命发展起来的麻省理工学院分支公司，公司志在创建由 *Endor.coin* 领导的强大而稳固的联盟，成员包括学界领先的研究团队、数据分析和数据集市领域的区块链项目以及数据洞察和预测发展与利用所依托的基础设施。

5.3 推动者 – 应用开发人员

为扩大 *Endor.coin* 可解决问题的范围，领域设置为无限大，且只能由欲构建应用程序者借助社会物理学引擎（“推动者”群体）加以限制。在其推动作用下，*Endor.coin* 将建成一个模块化应用平台，推动者可在平台上借助区块链扩展预测范围。为扩大 *Endor.coin* 可解决问题的范围，领域设置为无限大，且只能由欲构建应用程序者借助引擎加以限制，“推动者”群体 *Endor.coin* 将演化成一个模块化应用平台，推动者可在平台上借助区块链扩展预测范围。*Endor.coin* 使开放式微支付流能够流向可重用软件组件的作者，组件经永久性结合与重组形成高度可定制的实用应用程序不断扩大的库。推动者有权对其添加到框架中的每个组件采用微使用许可。最终用户安装自己选择的应用程序。要支付的许可是该应用程序所使用的组件所有微使用许可的总和。*Endor.coin* 通过智能合约负责向最终用户收取费用，并将款项分配给相关推动者。所有一切都实现自动化并对最终用户透明。激励方式是支付代币。应用程序基本上就是一组插件，可预见一个正反馈环符合以下思路：构建的应用程序越多，系统中添加的插件越多，准备好重新使用的唯一组件越多。这有助于推进整个 *Endor.coin* 生态系统相辅相成的网络效应。

5.4 数据主权

在这个世界上，贪婪的公司为自身利益而消费着我们的隐私；*Endor.coin* 致力于改变游戏规则，将财富回馈给允许他人使用自己数据的人。只需注册 *Endor.coin* 便可根据 *EDR* 代币所开启的各种情景开始执行可展开的过程。会员可以根据自己的数据接收个人见解，或者组成一个团队聚成一个更大的数据池，与小企业合作获得更加个性化的服务，例如：在适当的时间以最优惠的价格深入了解正确的产品。

鉴于人们有权拥有、控制和受益于其生成的数据，*Endor.coin* 向影响小型企业发展的长尾开放了目前仅供资金实力雄厚者使用的分析结果，目前可以在征得数据所有者同意的前提下汇集 *Endor.coin* 会员的数字数据供其使用，所提供的高价值数据分析将为会员带来通过分析得出的高效服务以及货币收益（数据以其他方式使用时）。*Endor.coin* 利用这种基于个人数据独立性的方法为公司和人员释放巨大价值，不会造成两者的价值或信任损失。*Endor.coin* 通过现会员选举会员和品牌参与，在公司和其他组织注册消除数据责任、访问关于客户的优质数据并建立客户关系时以负购置成本支持企业发展。

随着会员人数的持续增加，*Endor.coin* 还将提供一套工具和服务，帮助会员根据他们自己的数据深入了解其生活的方方面面。基于“数据来之于民，用之于民”的宗旨，*Endor.coin* 将个人的固有权利交还给他们，利用其目前已被 Facebook、谷歌、Uber 和许多其他基于数据的企业所打破的生活，推动“平台经济”的发展。通过融合了经证明的企业专业知识、前瞻性法律思维、技术诀窍和消费者价值创造的综合方法，*Endor.coin* 提供了一种独特的商业模式，通过建立双方同意的数据关系解决“数据隐私”问题，通过直接奖励会员使用其数据分析结果使公司产生良好的客户信誉。

为确保过程完整性，我们正在创建独立管理实体 *Endor.coin Trust*，其与 *Endor.coin* 协同合作实现会员利益、价值创造和数据安全的最大化。人们注册 *Endor.coin* 后，其数据由 *Trust* 进行监测，他们会自动成为 *Trust* 成员。*Endor.coin Trust* 以多种方式代表其成员确保和尽可能提高其数据的安全性、隐私性和价值。*Trust* 通过制定并执行个人数据控制、价值实现和隐私标准保护成员数据。

第 6 章

代币实现

6.1 区块链结构

数据存储：区块链并非通用数据库。*Endor.coin* 有一个分散的非链分布式散列表存储器，可通过区块链进行访问，用于存储数据引用而非数据本身。私人数据在传输和存储之前进行加密（使用 AES-256-亚马逊 S3 服务器端加密或亚马逊网络服务 IAM 机制），将访问控制协议编入区块链中。*Endor.coin* 通过非链网络（存储在任何形式的数据库或中央存储器等存储机制中，可导出为结构化格式，最终可上传至 *Endor.coin* 亚马逊网络服务架构）连接现有区块链以及私人 and 公共数据集。在将来发布的版本中，数据层将向外部参与者开放，可通过出售数据获取 *EDR* 代币。数据将会认证并上传至 *Endor.coin* 非链架构，供数据所有者使用，而且如果是标记为“公共”，则可供预测算法认为相关的任何客户使用。价格将由数据提供者决定，并经 *Endor.coin* 协议以动态方式根据对此种预测的需求进行自动调整，鼓励用户为相同数据源分担成本。

消费预测：区块链无法处理大量复杂交易。相同的非链计算网络用于运行（各种预测引擎要求进行的）大量计算。可用结果会在第一时间在整个公共区块链中进行广播，供最终用户使用（使用请求进行预测的用户的密钥进行验证）。与此同时，相同结果用一个不同的临时密钥加密并进行公开广播。该密钥经过预定时间后发布，允许随后对结果进行验证，甚至原先不需要它们的用户也可以对其进行验证。该机制的调整方式形成所述多个临时密钥，成本基于预测的“新鲜度”不断变化。

数据处理：代码执行根据亚马逊网络服务或类似环境（例如：*GOLEM* 项目）中运行的专有超弹性计算层划分为“在区块链上执行”（公共预测，公共预测请求指令，私人预测请求指令）和“*Endor.coin* 基础设施上执行”。由于数据处理和行为集群提取需要复杂而昂贵的执行环境，*Endor.coin* 协议后续版本将开放协议执行层。此种做法能促使一种新型利益相关者加入专门用于强化执行的 *Endor.coin* 生态系统。要分析的数据越大且集群提取算法越复杂（以计算为导向），其运行所需 *EDR* 代币价格便越高。

支付：用户可为每个预测请求支付一定数量的 *EDR*，具体要根据 *Endor.coin* 协议动态定义的预测复杂性、请求时的可用资源和需求以及提出类似问题的用户数量动态确定。预定查询预计会保持相对稳定的价格，而预测请求（*RFP*）在开始时成本相对较高（因为：根据定义，其开始时用户数量较少），在与其相关的用户群体形成时逐步减少。这个群体规模越大，每个人最终支付的金额越小，从而鼓励新预测询问的创建者在网络中的人们之间“传播消息”。

6.2 智能合约

Endor.coin 协议为其最终用户提供多个基本原语：

- *GetPrediction* (*prediction_def*) - 将在第 1 阶段实施
- *PutPredictionReq* (*prediction_def*) - 将于第 2 阶段实施
- *PutData* (*data_def*) - 将于第 2 阶段实施
- *RunCustomPrediction* (*data_alg_prediction_def,price*) - 将于第 3 阶段实施

这些原语允许客户检索 *Endor.coin* 平台上可获取的各种预测的动态成本，随后上传并出售数据，用于预测之目的。原语涵盖了 *Endor.coin* 协议的默认用例，而将来发布的版本将通过支持部署智能合约把更复杂的操作设计在“*Get and Put*”顶部。随着协议向私人数据和自定义私人预测方面发展，基本智能合约（作为第一个公开预测版本）顶部的其他安全层和复杂层将会启用。智能合约使 *Endor.coin* 用户可以编写可花费代币的状态性程序，请求进行预测，用于市场中的数据检索，并验证数据质量证明。用户可以通过向分类账发送交易，在合同中触发函数调用，从而实现与智能合约的互动。智能合约系统经扩展后支持 *Endor.coin* 特定操作（验证证明），支持针对数据上传的合同（后续各阶段将供公共或私人使用），以及通用性更高的智能合约。

getPrediction (prediction_def)：允许用户通过支付 *EDR* 代币检索 *Endor.coin* 网络上存储的预测。客户向检索预测市场订购簿提交一份出价单（将其订单传播到网络上），以启动 **Get** 协议。若匹配预测提供者的询问单，客户会收到针对预测的私人时效链接。收到时，双方签署一份交易单，并将其提交给区块链，用于确认交易成功。预测消费者将支付其要求得到的预测对应的同等价格。预测将可以从 *Endor.coin* 网络门户下载。

putPredictionReq (prediction_def)：允许用户将其他预测类型提交至 *Endor.coin* 平台，实现对这些预测类型的建议。可访问 *Endor.coin* 网络门户中的界面。所有 *Endor.coin* 用户都能看到建议的预测，而且他们可以对其进行排序。排名最高的预测将添加至不断扩大的预测目录。预测请求将通过预测发布者的一个公共地址进行识别，预测添加至目录后，每次其他用户消费预测时，预测发布者都能得到补偿。

putData (data_def)：供数据提供者和合作伙伴使用。数据提供者的数据用于预测时，其会收到 *EDR* 代币奖励，价格由 *Endor.coin* 协议预测算法自动计算。*Endor.coin* 项目第二阶段会提供函数的这种变体。协议第三次发布期间为“自助”模式提供一个 **API**，允许客户上传其专有数据流，将其标记为“私人”，然后与可在平台上访问的公共数据流融合。客户通过向存储市场订购簿提交一份出价单（向区块链提交订单）启动 **Put** 协议，等待数据验证者（例如：预测引擎提供者）提供匹配的询问单。客户需要为通过预测引擎实现的行为集群提取提供资金支持，但随后可以决定最终用户的每次使用对应的价格。

runCustomPrediction (data_alg_prediction_def, price)：协议第三次发布时将执行该方法，支持以“自助”方式形成 *Endor.coin* 目录中未包含的预测。输入内容通过示例或引用公共或专有数据的一个逻辑的描述包含所需行为的描述。确定最高价格，因为请求可由任何可用引擎处理。费用将在对预测发挥作用的不同参与者（例如：预测引擎和数据提供者）之间分摊。

6.3 Endor 在发布时的作用

如前文各章节详细描述的那样，*Endor.coin* 协议实现了各种参与者（例如：数据提供者和分析引擎开发者等）之间的内部通信。从这层意义上讲，协议本身并非预测来源，而是一种使能器和语言，允许“边缘”位置创造价值。类似于各种价值提供者之间实现轻松沟通所需使用的 TCP/IP 协议，*Endor.coin* 协议使数据提供者能够提供其物品，使预测引擎提供其分析服务，推动者提供创造一种协同价值可采用的各种方式。

为此，*Endor.coin* 将在独立法律实体下运行，专注于协议实施，确保周围生态系统的最佳构成，实现增长加速最大化。

然而，为确保快速建立协议的生产级工作流，并允许客户尽快（甚至是在第一天）消费各种准确预测，*Endor.coin* 项目可能会选择要求 *Endor* 扮演多个关键角色。随后在新参与者采用 *Endor.coin* 协议时，*Endor* 的主导地位预计会下降，为由新参与者引入 *Endor.coin* 生态系统的创新和价值让路。

Endor 在 *Endor.coin* 发布后发挥的主要作用如下所述。每项活动都需要专门资助，此种资助将从首次代币发售（ICO）交换 *EDR* 代币过程中收到的款项中划拨。公司目前正就此项合作的确切细节进行谈判。

公共数据提供者：*Endor* 将借助自己的开发团队实施生产级基础设施，从而获取并清理各类数据流，将其上传至 *Endor.coin* 基础设施，作为随时可用于分析的数据。这包括比特币区块链、ERC20 区块链和其他多个专有区块链。此外，这在以后会支持推特和红迪等社交网络内容。*Endor* 将以“成本加成”为基础提供此项服务，并涵盖其人力资源费用。

已处理数据的提供者：*Endor* 将作为支持 *Endor.coin* 协议的第一个分析引擎，从交易性数据流中提取行为集群。*Endor* 将以“成本加成”为基础提供此项服务，并涵盖其亚马逊网络服务费用。

需求提供者：*Endor* 将作为需求提供者，服务希望得到预测的部分现有（和新）企业客户。*Endor* 的客户成功保证人员将使用企业客户支付的美元购买 *EDR* 代币，生成所需预测。

应用与合作开发：为进一步扩展 *Endor.coin* 生态系统并增加基于 *Endor.coin* 的预测的知名度，*Endor* 将积极努力创造基于预测的新业务，其方法是通过合作或形成并公开发布新的预测类型 - 发挥推动者的作用。这需要利用 *Endor* 的业务开发团队及其大型高品质咨询委员会。

6.4 代币特权和经济

如前几章所述，*Endor.coin* 发布的 *EDR* 代币将用作 *Endor.coin* 生态系统所供各类服务的支付机制。部分服务将在项目初期免费提供以促进其成长，*EDR* 代币主要用途如下所述：

购买预定或自助预测：*EDR* 代币预计主要用作预测支付的支付机制。这些预测包括不断扩大的目录中列出的预定预测以及（后续阶段）“自助”预测，可由用户通过专用自助 API 进行

订购。支付将主要涵盖云计算资源（例如：亚马逊网络服务和 *Golem* 等），10%~25%按比例拨付至生成预测所用数据的提供者。预测成本将与生成预测所需云资源的成本挂钩。

预测生成过程的主要组成部分在不同预测之间共享，所形成的成本函数随着用户数量的增加而逐渐减少，时间脉冲与需求突然增加有关：

向数据提供者支付的金额：预测成本的 10%~25%。

数据分析用云资源（亚马逊网络服务或类似资源）：成本约为云成本的 80%~90%，按活跃用户数量分摊（即：预测总成本的 55%~80%除以活跃用户数量 n ）。

个性化预测用云资源（亚马逊网络服务或类似资源）：约占云成本的 10%~20%（即：预测总成本的 10%~15%）。

因此，预测成本定性支付模型为：

$$0.7 \cdot O\left(\frac{C_{PREDICTION}}{n}\right) + 0.3 \cdot C_{PREDICTION}$$

其中， $C_{\text{预测}}$ 表示 Endor.coin 平台推出时的单项预测成本， n 表示平台活跃用户数量。

提交预测请求（RFP）：对目录尚不支持的特定预测感兴趣的用戶可申请利用预测请求机制创建新预测。此种情况下，用戶将支付 *EDR* 代币，其涵盖了技术熟练的 *Endor.coin* 团队成员进行预测创建与优化所需时间成本及其所需要的云资源。从这层意义上讲，预测请求将与云成本和技术熟练的数据工程师的平均市场工资挂钩。

预测服务质量（QOS）偏好要求：购买预测或通过预测请求机制请求获取一项新预测时，用戶可选择购买对预测的收费访问权，收到的时间比发放给非高级购买者的时间提前数小时，或者是在预测请求机制条件下在预定时间内保留对它的独占访问权。此种服务质量要素可通过预测成本乘以预先确定的系数来购买。从这层意义上讲，这种优质服务也会与云资源和市场工资成本挂钩，因为它是预测基础成本的倍增。

为平台提交新数据流：希望为 *Endor.coin* 生态系统提供新数据流的数据提供者需要支付数据初始分析所需云资源及其基于 *Endor.coin* 协议的调整。其目的是为高品质数据源的融合提供积极的激励，以提高预测的整体质量。初始过程费使数据提供者有权使其数据获得 *Endor.coin* 协议支持，作为预测的行为集群来源，在其用于生成预测时还有权获得 *EDR* 代币补偿。

6.5 款项的使用

购买 *EDR* 代币获得的收益将用于实施 *Endor.coin* 协议软件架构，通过营销和战略合作支持其初始成长和采用，为其所需云资源提供资金，并支付所需法务和行政费用。

收益的主要部分将分配给由尖端云和区块链工程师以及世界级数据科学、机器学习和社会物理学专家组成的 *Endor.coin* 研发团队。

Endor.coin 项目和团队一项重要作用是确保尽可能多的主要参与者快速采用 *Endor.coin* 协议。其中一种实现方式就是与学术界开展联合研究活动，从第一天开始就能为 *Endor.coin* 协议和基础架构提供主要研究访问权限。最多 10% 的收益将用于支持此项活动，执行者应为全球领先的研究机构。

最多 30%的收益将用于购买预测引擎或 *ETL* 连接器等相关专有技术，以促进并加速 *Endor.coin* 协议及其周边生态系统的适当引导。

第 7 章

技术优势和差异化

7.1 源自麻省理工学院大熔炉的科学革命

基于麻省理工学院最新社会物理学技术的支持[1]，*Endor.coin* 采用了全球最先进的行为预测技术。本世纪初在麻省理工学院媒体实验室诞生的这项科学突破彻底改变了行为数据分析领域，并且处于技术成就的最前沿，其中包括：赢得著名的美国国防部高级研究计划局网络挑战[21]，推动散户投资者群体的回归[9,13]，并成功预测高效的未知网络攻击[35]。该技术是由一众学术和行业专家开发，他们总共发表了数百篇科技论文，获得数十项专利，并出版了这些方面的若干本著作。

时至今日，这场革命延伸到了公共领域，利用区块链技术使专业人员和非专业人员具备迄今为止仅面向大型零售商、银行和科技巨头的各种能力。社会物理学详细论述参见第 3.1 节和附录 A。

7.2 实际产品，成熟技术

如前文第 3.2 节所述，*Endor* 源自麻省理工学院[14]，并凭借自身力量将社会物理学作为产品实施，为大型银行和零售商提供打破其预测瓶颈的软件即服务解决方案。公司由领先的投资者提供资金支持[15]，并与可口可乐[16]、万事达卡[17]和沃尔玛等《财富》500 强企业开展合作，成功展示了针对各种用例自动生成精确预测的能力。

[27]等方面验证了技术对于区块链和加密货币分析的价值。*Endor* 的产品已经在 *Money-2020* 和 *Finnovate 2017* 等主要活动中亮相[18]。*Endor* 系高德纳杰出供应商[19]，经世界经济论坛认定为“技术先锋”企业[20]。

7.3 可用性、对于用户的价值以及对于代币持有者的价值

Endor.coin 主要独特之处在于，*EDR* 代币从发布第 1 天开始便可使用 - 为代币持有者提供预定预测的全面访问权限。另外，代币发布后不久还将选择一组测试用户，他们可以在预先确定的预测之外请求进行其他预测。

而且，认为自己将来可能需要以较低成本获取预测服务的代币所有者将被鼓励购买并持有 *EDR* 代币，因为预计应用量的增加会使预测成本增加，从而推高代币价值，也就意味着对早些时候购买 *EDR* 代币的预测用户的奖励。

第 8 章 团队

8.1 团队主要成员

Yaniv Altshuler 博士

Altshuler 博士现任 *Endor* 首席执行官及麻省理工学院媒体实验室助理研究员，与 Pentland 教授共同创立“社会物理学”- 一种对群体行为进行建模的新科学。该技术为麻省理工学院赢下了著名的美国国防部高级研究计划局挑战[21]，帮助新加坡政府改善了交通堵塞预测能力，并协助数千名金融投资者获得更多财务回报[13]。*Endor* 将此项技术用于精确预测各类用例中的人群行为，高效迎合大型银行和零售商需求[14]。

供职于麻省理工学院并创建 *Endor* 之前，Altshuler 曾担任 IBM 研究人员，并开发出一种用于提高超级计算机性能的新型优化方法。自 2011 年开始，积极参与区块链研究之后，Altshuler 博士发表了 60 多篇科技论文，并申请了 15 项专利。Altshuler 的作品发表在《金融时报》[10]和《哈佛商业评论》[11]等广受欢迎的媒体上。他最近出版的著作包括《社交网络安全与隐私》[26]和《搜索中的集群和网络智能》[25]。

Alex "Sandy" Pentland 教授[科学咨询委员会主席]

作为麻省理工学院媒体实验室创业计划总监以及麻省理工学院连接科学与人类动力学实验室主任，Pentland 教授是全球被引用次数最多的科学家之一[23]，最近被福布斯评为“全球最具影响力的 7 位数据科学家”之一，获此殊荣的还有谷歌创始人和美国首席技术官[24]。他获奖无数，具体包括《哈佛商业评论》麦肯锡奖[22]、美国国防部高级研究计划局互联网 40 周年纪念奖[21]以及他有关隐私的著作所获布兰代斯奖。

他是谷歌、美国电话电报公司、日产和联合国秘书长咨询委员会的创始委员，并作为连续创业者联合创办了十多家公司以及众多社会型企业，包括数据透明化实验室、哈佛-ODI-麻省理工数据应用联盟和数据驱动设计研究院等。他是美国国家工程院院士，同时还担任世界经济论坛领导。他最新出版的著作包括《社会物理学》和《诚实信号》。

Stav Grinshpon

Grinshpon 先生是科技行业资深专家，在思爱普和美国电话电报公司等公司积累了 18 年的产品和管理经验。Grinshpon 是一名世界级网络防御专家，在以色列 8200 技术单位担任技术领导 8 年。Grinshpon 持有 3 项数据分析相关专利，并领导了 *Endor* 的研发

David Shrier 是全球公认的金融创新权威，目前领导着牛津大学在线计划-牛津金融科技与牛津区块链战略以及麻省理工学院未来商务计划。他出版了多部关于金融科技、区块链和网络安全的著作，包括《网络安全新型解决方案》、《前沿金融科技》和《信任：数据》[27,29,32]。

Shrier 现任 **Distilled Analytics** 首席执行官；这是一家源自麻省理工学院研究的机器学习公司，通过行为分析推动金融服务行业的转型。另外，他还担任人工智能驱动的协作技术平台提供商 **Riff Learning** 的董事长。David 是牛津大学赛德商学院的副研究员，还是麻省理工学院媒体实验室的讲师。此外，他还为迪拜政府提供区块链和数字身份识别相关咨询，为中间市场信用流动性提供商 **Millennium Advisors** 提供技术趋势相关咨询，并为基于区块链的数字商品交易公司 **Cleer.digital** 提供战略相关咨询。

David 是世坤投资大学咨询委员会成员；这是一项提供经认证的全免费在线金融工程硕士学位计划。他之前曾建议欧盟委员会以数字技术为核心进行创新商业化，目前是美国金融业监管局金融科技行业委员会委员；该委员会是美国证券行业自律机构，就影响金融服务的最新进展提供咨询。

Mihaela Ulieru 教授

计算智能领域的专家，世界经济论坛区块链拥护者[36]。她对分布式智能系统的研究为区块链治理（作为制度技术推动制造、物流和国土安全领域的变革）奠定了坚实的基础。Ulieru 教授毕业于加州大学伯克利分校，现为世界经济论坛全球议程理事会、新加坡科学与工程研究理事会和加拿大科学技术与创新理事会成员。

Goren Gordon 博士

Gordon 博士现为特拉维夫大学奇特性实验室主任，在这里开发出最先进的计算奇特性模型。作为深度学习和神经网络优化领域的主要专家，Gordon 持有量子物理学博士和神经生物学博士学位。在麻省理工学院媒体实验室与 Cynthia Breazeal 教授合作期间，Gordon 博士充分利用这方面的经验以及另一个医学学位研究了好奇的机器人与好奇的孩子如何进行互动。Gordon 还积极投身科学教育，并开发了“量子电脑游戏”，以游戏的方式向孩子们传授量子物理学相关知识。另外，Gordon 还是量子物理学、脑及跨学科思考方面的一名极受欢迎的讲师。

Arie Matsliah 博士

Matsliah 博士是一名世界级图形分析理论专家，在谷歌、IBM、Intel 和 Lyft 等行业巨头供职 16 年，还担任过总部位于门罗帕克的初创公司 TripActions 的首席架构师。Matsliah 博士共发表过 40 多篇有关算法、复杂性和量子计算基础研究的科技论文。

Shahar Somin-Gavriellov

Somin-Gavriellov 女士是一名统计学习理论专家，持有希伯来大学硕士学位（荣誉学位）。作为一名经验丰富的研究者，Somin-Gavriellov 的研究融合了数据科学的深度理论实现和实际行业经验。她过去还曾是以色列 8200 情报单位一名广受赞誉的分析师。

Edo Eisenberg

金融风险管理专业人士，先前管理着摩根士丹利和巴克莱的零售信用组合。在 NICE 供职 9 年，为美国 10 家顶级银行中的 7 家设计并实施防欺诈解决方案。毕业于杜克大学 MBA 课程。

Lior Regev

Regev 先生是一名经验丰富的软件工程师，他对技术抱有极大热情。作为前智能技术领导者，Regev 在分布式系统、云架构和软件即服务产品方面经验极其丰富。

Liat Yitshaki

Yitshaki 女士是一名公法诉讼专家，持伦敦大学法律与道德文科硕士和法学学士学位（荣誉学位）。Yitshaki 曾供职于麦肯锡公司，并担任英国政府高级法律顾问。

8.2 顾问

Alexander Lipton 教授

Lipton 教授曾担任美洲银行全球定量解决方案常务董事以及美林证券全球量化及信用分析小组常务董事。在此之前，Lipton 教授曾担任城堡投资量化研究主管以及瑞士信贷股权衍生工具主管。Lipton 现为洛桑联邦理工学院金融工程学教授以及麻省理工学院连接科学中心研究员，最近在《科学美国人》上发表了名为《技术如何帮助我们修复破败的金融体系》的文章[37]。

Ron Gross

Ron Gross 是以色列比特币协会创始人兼董事，自 2011 年开始进入比特币和区块链领域，曾担任全球首个 ICO 万事达币（现为 Omni）的执行董事。Gross 是前谷歌员工，还曾担任过 Commerce Science 首席架构师。

Nuria Oliver 博士

Oliver 博士毕业于麻省理工学院，现任沃达丰数据科学研究总监及数据应用联盟首席数据科学家。Oliver 博士此前曾担任西班牙电信研发科学总监以及微软研究员。Oliver 是数据共享领域的世界级专家，将大数据分析作为一种协作形式，私营企业基于该形式为公共利益而共享数据。Oliver 发表了 90 多篇科技论文和著作章节，专利申请 40 余项。

Daniel Tunkelang 博士

Tunkelang 是数据科学领域的世界级专家，曾供职于谷歌和 IBM，现为苹果、Salesforce、Etsy、Yelp 和品趣志等科技巨头的顾问。Tunkelang 此前曾担任领英搜索工程总监，随后担任 Residence 数据科学家，还担任过 Endeca（后被甲骨文收购）首席科学家。Tunkelang 持麻省理工学院硕士学位和卡内基梅隆大学博士学位。

Guy Zyskind

Enigma（受 Enigma Protocol 委托创建保密智能合约）和 Catalyst（一个允许任何人利用成熟的工具和数据启动加密对冲基金的平台）的创始人兼首席执行官。供职于 Enigma 之前，Guy 是麻省理工学院研究生，从事区块链技术的研究与教学工作。Guy 发表过多篇学术论文，最近发表的论文是关于隐私和区块链，包括《Enigma 白皮书》以及《隐私分散化：使用区块链保护个人数据》。Guy 持有麻省理工学院理学硕士学位以及特拉维夫大学电机工程与计算机科学理学学士学位。

Michael Bronstein 教授

Bronstein 教授是一名连续创业者兼领衔研究员，现任哈佛大学研究员及卢加诺大学信息学教授。Bronstein 在主要期刊和会议上发表过 100 多篇文章，持有 20 多项专利，出版了研究专著《数值几何非刚性形状》，还编辑过 4 本著作。Bronstein 是 3 次获得过欧洲研究委员会（ERC）拨款资金的少数几名研究人员之一，还获得过谷歌教员研究奖，曾担任哈佛大学拉德克利夫学院奖学金和鲁道夫·狄塞尔工业奖学金。他是世界经济论坛评选的全球 40 名四十岁以下顶级研究人员之一。

Bronstein 教授积极参与工业应用，是 Novafora（已授权给敦纳广播）联合创始人兼技术副总裁，还是 Invision（后被英特尔收购，Bronstein 目前担任感知运算研究科学家）联合创始人及其 3D 传感技术主要开发者之一。Bronstein 还是 Videocites 的联合创始人兼技术顾问。

潘巍博士

潘博士毕业于麻省理工学院，曾供职于谷歌，是大数据分析、机器学习和复合系统领域的世界级专家。他所发明的“现实对冲”分析方法通过对社交系统的了解和模型以及个人和人群的大数据测量分析金融市场动态和宏观经济。潘博士现为总部位于纽约的创业公司 Thasos Group 的联合创始人兼首席科学家。潘先生之前曾在富达投资担任首席经济师，主要从事系统研究和闪电崩盘分析。

Igor Gonta

Gonta 先生毕业于麻省理工学院工程与计算机专业，现任贝莱德商品与风险管理总监。Gonta 曾担任过 Market Proffit 首席执行官 - 这家公司是一个基于社交媒体会话分析的实时股票人气生成引擎。作为公司创始人的 Gonta 随后将公司出售给一家大型对冲基金。Gonta 职业生涯早期曾担任巴克莱商品销售常务董事以及高盛商品销售副总裁。

Thomas Hardjono

Hardjono 是一名安全专家，专业从事分散式身份验证、区块链和智能合约研究，曾担任威瑞信和麻省理工学院 Kerberos 集团执行董事，共出版过 5 本计算机安全与密码学专业著作[38-42]。

附录 A - 所述社会物理学

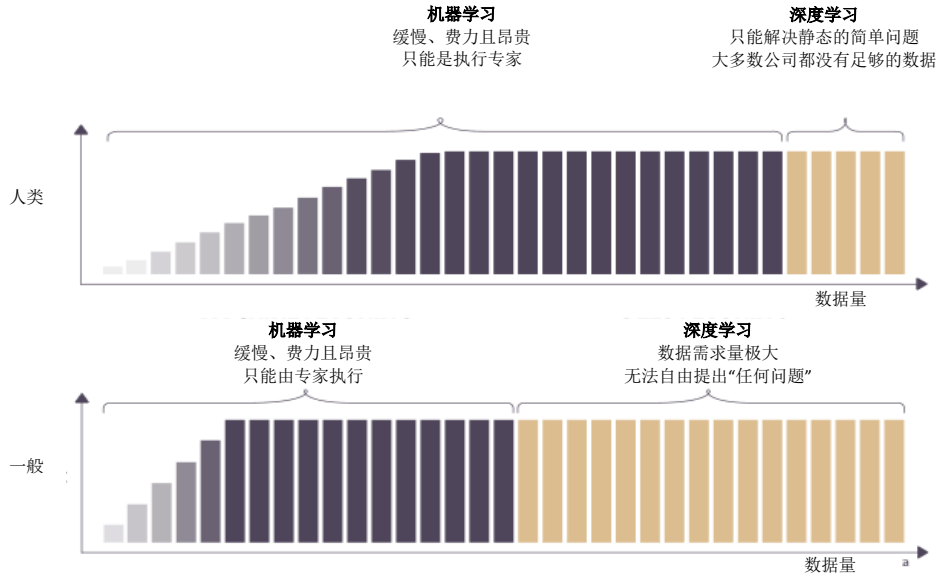
1. 社会物理学原理

说明：本文件旨在综合描述社会物理学、其背后的高级原理以及 Endor 对其进行的技术实施。然而，出于知识产权方面的考虑，Endor 所用社会物理学定律的具体数学公式以及用于检测违反定律情况的机制的具体实施均被刻意省略。

在信息时代，公司从关于其业务经营的大量数据源收集各类数据。数据包含图像和视频、文本和推文、交易和使用记录。然而，大部分数据都来自单一基本来源，即：人。因此，例如：推文和博文由人编写并供人阅读；购买交易和电话呼叫信息传达人类对事物和他人的渴望；关于人们如何与计算机和移动设备互动的使用和应用日志报告。

源自人类行为的数据相当“杂乱”：它具有动态性、复杂性和极高的多变性。根据此类数字数据渠道的记录，人类行为随时间发生显著变化，受复杂的基本社交网络影响，并且通过多模式数据流传送。对于希望分析、理解和预测其客户行为以期改善业务经营的公司而言，这些特征构成重大挑战。

近年来，数据科学家已开始采用“重型”统计方法和机器学习算法尝试处理此种复杂性。这些强大的工具（包括新型“深度学习”技术）收集数据并分析其属性，以便对行为模式进行分类，检测异常情形，并预测未来趋势。尽管如此，这些曾经专门用于图像处理和文本识别等“静态问题”的工具无法轻松应对人类行为数据：学习动态、复杂且多变的数据流难度极大，有时几乎无法实现。

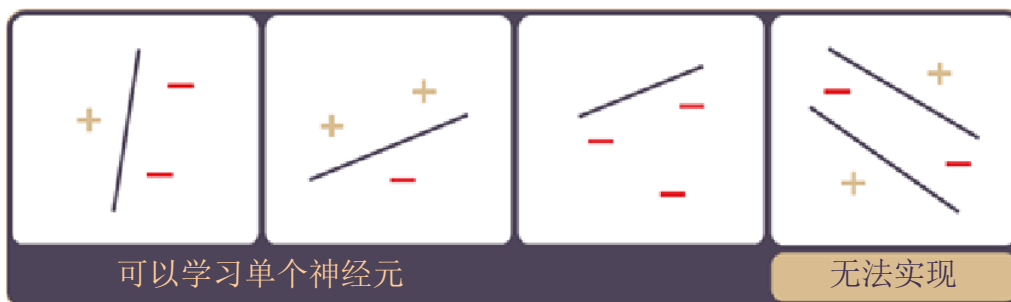


Endor 社会物理学引擎采用完全不同的工作原理。不同于基于输入数据而得出的各种模式，它的运行依据是：所有人类行为数据一定包含一组常用“社交行为定律”，也就是足够多的人在同一空间内作业时出现的各种数学关系式。这些定律支配着人群行为各种统计属性随时间演化的方式，与数据类型、数据创造者的人口统计资料或数据大小无关。**Endor** 将这些定律融入了其数据分析引擎，用于高效提取作为输入资料提供的原始数据所含全部人员的基本社交属性（例如：通话、乘出租车和金融投资等）。

2.1 为什么需要社会物理学？

抽象地讲，学习问题或进行对象分类或预测未来事件的能力需要进行数据分析及其分析算法。所需数据量取决于多个因素：

1. **问题的内部复杂性** - 问题形形色色，而有些问题明显比其他问题严重。问题的“难度”或“复杂性”是指学习算法为成功实现问题学习而必须具备的最小“强度”：强度不足的学习算法无法正确学习问题实例。例如：众所周知，“感知器”（即单个神经元组成的最简单的神经网络）永远无法学习“异或”函数（即布尔“异或”函数），根本原因是，感知器的工作方式（可想象成线性划分输入空间）使得异或函数“过于复杂”，因而无法以此种方式表述；



异或函数示例参见右图（来源：维基百科）

问题基本模型的学习复杂性通常称为问题的 Vapnik-Chervonenkis 维（或 VC 维）。问题学习复杂程度越高，预测算法为实现成功学习所需强度越高，此种算法适当建模需要数据越多。例如，根据定义，具有线性复杂度（是指：其可利用 1 次多项式准确建模）的函数 $y = f(x)$ 所需采样点少于需要更高次多项式以实现准确建模的函数。

算法学习效率

学习算法众多，每种算法都需要不同数量的训练数据和专家领域知识来适当确定模型参数。例如，简单回归需要大量数据和许多问题具体特征，而需要大量数据的深度学习则可自动学习领域特征。基于论述目的，学习过程的整体效率可加以量化，也就是说：学习基本模型所需数据量以及算法自身对适当表述的学习量：

$$\eta_A = \frac{R_A(M)}{D_A(M)}$$

M = 模型位数，是指数据表述或建模所需位数，例如：完美描述模型所需参数个数。就此点而言，**M** 类似于问题的柯尔莫哥洛夫复杂性 - 它是一种已知的数据复杂性理论量度，针对的是可生成给定数据的最短图灵机。模型由问题决定且无法修改，即：如果它是一个简单的问题，则模型会有些位数。

R_A(M) = 模型位数特定表述，是表述基本模型 **M** 时算法 **A** 学习的位数。该数值代表算法的自动特征检测，且与专家必须以手动方式编入算法的领域具体知识量成反比。

D_A(M) = 数据位数，是指算法 **A** 学习模型 **M** 所需位数。

η_A = 算法 **A** 的学习效率，即：算法自动表述学习与保持学习质量所需数据量的比值。高效率算法可借助少量数据学习适当的表述，而低效率算法则需要手动制作特征以及用于对其进行调整的大量数据。

例如，给定 **M** 类问题时，某些算法保持学习质量所需要的数据多于其他算法：

- **逻辑回归**算法通常需要由专家对特征进行手动编码，需要大量数据针对当前具体问题进行调整：

$$R_M(A) \ll 1$$

$$D_M(A) \gg 1$$

$$\eta_A \ll 1$$

- 一次学习算法也需要专家领域特征，但只能使用一些实例微调基本模型并进行预测：

$$R_M(A) \ll 1$$

$$D_M(A) \ll 1$$

$$\eta_A \sim 1$$

- 深度学习算法可自动学习最具信息量的特征，但需要大量数据：

$$R_M(A) \gg 1$$

$$D_M(A) \gg 1$$

$$\eta_A \sim 1$$

- Endor 算法利用社会物理学从较小的数据样本中自动提取相关行为特征：

$$R_M(A) \gg 1$$

$$D_M(A) \ll 1$$

$$\eta_A \gg 1$$

2. 问题变化速度 - 影响产生精确预测所需数据量的另一个因素是问题变化速度的基本模型。某些问题属静态问题，其基本参数不会改变或极少会改变。例如：图像中的人脸几个月都不会改变；面部特征不会那么容易改变。另一方面，大量产生付费服务的基本行为模式可能会随时间改变：或者是在几个月时间里通过社会变化逐渐改变，或者是在几天内因竞争对手成功举行的营销活动而迅速改变。我们将这种动态特性作如下量化：

$$\tau = \frac{\partial T}{\partial M}$$

τ 是问题的持续性或韧性，表示基本模型的变化速度； ∂T 表示模型变化 ∂M 位所需时间。

例如， $\tau = 1$ 天/10 位是指：同 $\tau = 1$ 月/10 位（模型变化速度极慢）相比， $\tau = 1$ 天/10 位是指模型在一天时间内发生显著变化（即：变化速度快）。实际上，动态模型每个阶段都呈现不同的模型，因此需要对模型进行再训练或再学习。

预测分析的运行实施：追逐不断变化的模型

针对给定问题的预测分析可行实现所面临的主要挑战是：如何获得足够的信息以应对建模对象的行为变化。单位时间可累积的信息位数量表示为 I_t 。

极庞大社交网络或搜索引擎（例如谷歌或 Facebook）的运营者通常可在相对较短时间内积累大量信息。然而，希望预测客户行为的大多数公司都不具备此种能力。

此外，即使每天获取大量信息的大型参与者也难以对（a）过于复杂或（b）变化太快或（c）同时存在上述两种情况的问题进行精确建模。

因此，我们可以通过一个简单的方程确定公司执行经营性预测分析解决方案的能力。能够利用所采用学习算法及其不断获取的新数据满足这一原则，从而解决其所要预测的问题的公司能够成功构建用于实现这一目标的运营过程，其中：未能实现该目标的公司不会取得如此这般的成功。

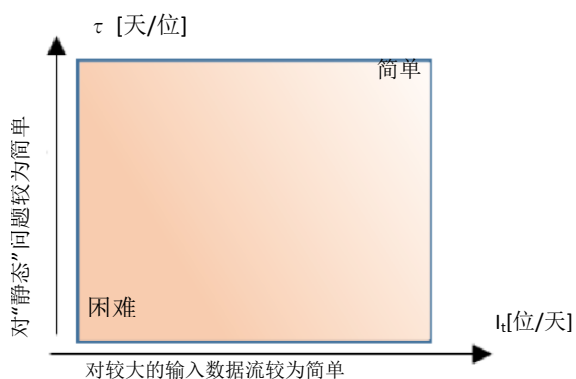
基本运营学习原则：

$$I_t \times \tau \times \eta A > 1$$

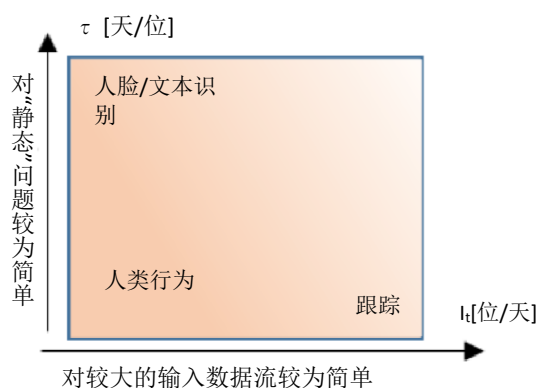
这一原则的实际意义在于，未能构建起可运行预测系统的公司应当：

1. 改善其数据收集带宽，以期每天获取更多相关数据；或者
2. 集中关注更加静态的问题；或者
3. 采用效率更高的学习算法。

这种简单的关系意味着：随着问题日益复杂及其持久性日益降低（如下图所示），每天需要更多信息：



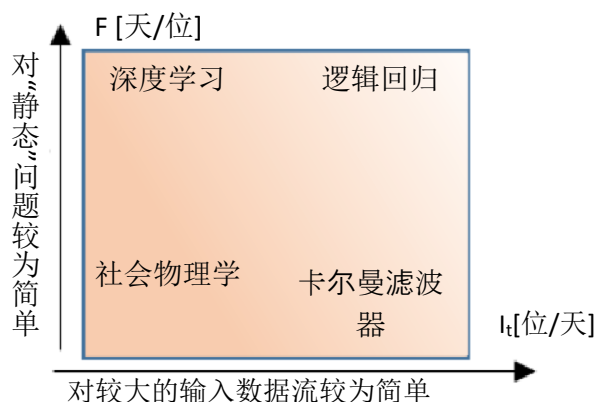
具体分析时一些常见的预测问题如下所示



- **人脸识别：** 该问题变化速度极低（即 $\tau \gg 1$ ），因为人的面部基本不会发生太大变化。因此，即使是 ηA 相对较小的低效率学习算法，其仍然可以非常小，因为此时需要大量信息对问题进行初步学习，但重新训练时不需要大量信息，因为它不会发生太大变化。
- **跟踪操纵移动目标：** 移动目标跟踪时 $\tau \ll 1$ ，因为目标的快速移动（或者在更好的情况下，导航算法为避免检测而引入的其轨迹或动态局部变化中的快速变化）无需使用任何非直接历史数据。然而，这些问题通常对应较大的输入数据流，即 $I_t \sim 1/\tau \gg 1$ ，表示目标位置的动态特性通过输入数据流充分传递至学习算法。因此，可使用相对简单的算法（ $\eta A \sim 1$ ）快速解决问题。

- **人类行为：**人类行为极具动态性 ($\tau \ll 1$)，这意味着其包含变化速度相对较快的元素。此外，尽管想要预测人类行为的公司能够（并且确实已经）获取关于客户的其他信息，它始终是行为变化的部分推导。打个比方讲，这类似于柏拉图的洞穴寓言：传感器感知到的信息穿过一个仅捕捉到它的几个方面的极粗糙透镜。按照我们的形式化模型，这会转化为 $I_t \ll 1/\tau$ ，这又意味着 $I_t \geq 1$ 。为高效预测人类行为，必须采用效率极高的算法 ($\eta A \gg 1$)。

同理，我们可以看到不同的解决方案技术如何以最佳方式应用于每个问题，并用该关系式建模：



该图描述了前述基本运行学习原则 $I_t \times \tau \times \eta A > 1$ 。问题静态性越强，我们得到的相关数据越多，我们与图表右上角的距离越小（预测就会越准确）。越靠近右上角，产生精确预测所需算法效率要求越低。

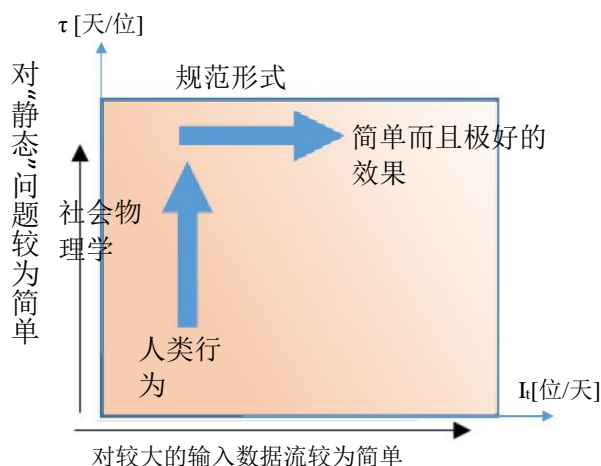
鉴于需要大量信息进行培训的技术不能很好地解决动态性问题，深度学习最适合图像处理和手势识别等基础结构不会快速变化的问题。相比之下，卡尔曼滤波器等可快速处理信息的简单算法可以处理动态性问题，但需要较高的信息吞吐量成功实现预测。逻辑回归等传统机械学习方法在输入训练数据量相对较高（假设问题动态性也相当高）的情况下效率较高。

社会物理学作为永久追逐的解决方案

Endor 从两个正交视角应对针对不断变化的模型不断获取大量数据的需求：

- **将动态问题转换为静态模型：** 社会物理学定律对数据类型和起源不可变且不可知。因此，利用它们将原始数据投射到社会物理学正则表述空间可将初始问题转化为基于静态基本模型的新一类问题实例（得益于社会物理学定律的固有静态特性）。如此一来，初始问题的实际 $\tau \ll 1$ 可转化为极大 $\tau_{SP} \gg 1$ 。
- **根据小数据创建“大数据”：** 如上所述，面对预测客户行为的挑战，大多数公司都会面临一个主要障碍，即：其所处理的相关数据量往往不够充足。动态问题（常见于营销用例）或者涉及引入新元素的问题（例如：预测对新产品的反应，或使用一种新的输入数据）在这方面尤为明显。Endor 将其从客户那里接收到的所有数据流转换成社会物理学规范形式，不受类型、大小和来源影响。Endor 利用该规范形式将所有客户和及其所有查询的所有数据统一并合并为以极高的稳定速度增长（即： $I_t \gg 1$ ）的单个极大数据库。该数据库随后用于在社会物理学规范形式（并非特定查询或客户）实例分析方面训练单一不可变深层学习网络。因此，即使每个客户提供的数据量极其有限，Endor 也能积累大量（规范构成的）数据。

Endor 引擎将难度最大的人类行为预测问题（通过社会物理学）转换为缓慢变化的问题，并（通过规范形式转化）转换为大数据，产生一个“简单高效”的问题后利用深度学习工具加以解决。具体如下图所示：



2.2 社会物理学：概述

上一节简要描述了社会物理学如何利用语义不可知静态数学不变量克服预测人类行为的内在挑战。为更好地理解其可行方式，需考虑各种物理定律，例如：牛顿第二定律或动量守恒定律。任何对象都保持其初始路径，除非受到外力作用。注意，推导这种不可见力的存在性时无需“学习数据”、了解其统计属性或测试以类似方式运行的许多系统。由于物理定律已经给定，任何违反定律的情况都视为异常并且可立即发现，并解释为某种“无形之手”作用的结果。如果发现突然改变方向的物体，则可立即推断出肯定存在某种作用力。如果其变化具有相似性，则它们极有可能都受到了相同的作用力。这种简单的推理必须在理解物理定律的情况下才能实现。

社会物理学的绝对性和缜密性远不及物理定律，但定义却具有相似性。违反社会物理学定律的事物可立即判定为“感兴趣”，前提是数据在现实世界中展示了某些有价值的特性或属性。这不需要学习、基准确定、基线或其他任何数据科学或机器学习工具。违反社会物理学定律的情况可以通过一种十分稳健的方式非常快速地检测到，与生成它的数据类型无关。

2.3 ENDOR：基于社会物理学的支持

2.3.1 数据转换为规范表述

通过提取先前论述的“行为集群”（即：检测到的违反社会物理学不变性的情况）并将其融合成“知识领域”，可将（任何形状或形式的）原始输入数据（只要它源于人类活动）转换为规范形式。此种形式代表了“共同”违反社会物理学定律的人们的群集；换言之，就是相对于特定不变性而言行为“同步变化过高”的人们。这类似于在特定时间以类似方式改变方向的物理对象 - 鉴于造成这些变化的力量不可见，可以较为肯定地推断所有对象都受一种来源的影响。同理，Endor 以行为集群形式进行的数据规范表述包含的人群很有可能受相同“社会力量”的影响，从而具有相同的“现实特征”。任何可使用新原始数据都由客户发送给 Endor（通常是每天或每周一次），允许其他行为集群自动提取。

此种表述方法具有三重益处：

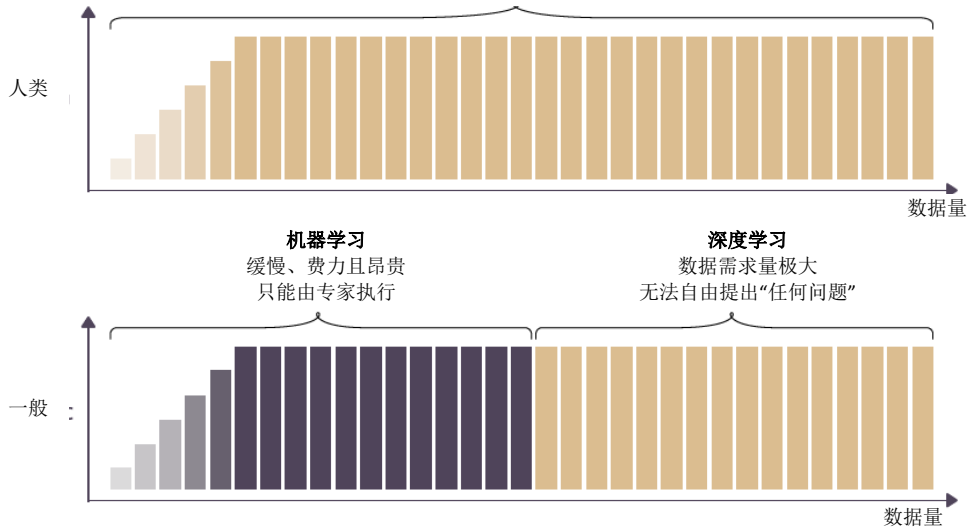
- **自动：**利用社会物理学的数学不变性，因为任何源自人类行为的数据均可自动转换为一系列行为集群（不管是何种输入数据类型），无需进行声明或分析（例如：通话记录、信用卡购物、乘出租车或客户可能拥有的其他任何类型专有数据）。连同对输入数据中极高噪声和间隙的极高适应性，这意味着将未知类型的脏原始数据转换为统一行为集群的过程第一次实现了完全自动（其他信息参见“噪声鲁棒性”章节）。
- **统一：**通过剥离任何领域、人口统计或语义方面的数据，包含行为集群的剩余信息可完美适应下列“查询”阶段。实际上，Endor 可利用这种统一表述在大数据缺失时“形成大数据”，允许过程查询阶段利用无法供数据原所有者使用的深度学习技术！其之所以能够实现，是因为 Endor 深度学习引擎可以使用源自多种数据类型以及大量客户的行为集群 - 全部转换为单一形式。
- **新兴趋势：**简而言之，社会物理学不变性描述了人类群体某些统计属性随时间演化的方式。这种时间相关方面使 Endor 能够轻松发现正在形成的行为变化 - 最近才出现的发展动态，以及大多数情况下没有足够时间生成足够的可观测数据的动态能够利用传统方法以足够高统计学意义检出。利用社会物理学可检测到这些附加信号；除此之外，这往往正是对于各类商业问题较为重要的信号 - 因为它们包含最新趋势相关信息。

“老式”机器学习与预定特征配合使用，可从相对较少的通用数据中提取相关信息。尽管如此，结果在很大程度上取决于所选特征。深度学习独立识别最相关的特征，但需要借助大量数据。每种数据类型和所问问题需要再次找出相关特征，因此需要更多数据。社会物理学可将任何类型的人类行为数据根据其行为转换为人类集群规范形式。此种方式适合少量和大量数据。此外，得益于社会物理学的规范形式，Endor 可吸收所有数据类型和所有问题（无论数据大小），并产生一个巨大的人类行为数据集，可借助深度学习回答任何问题。

ENDOR 社会物理学 → 深度学习

完全自动，可回答任何问题。非专业人士也能使用

社会物理学可将任何类型的人类行为数据转换为规范形式的人类行为集群。得益于社会物理学的规范形式，Endor 可吸收所有数据类型和所有问题（无论数据大小），并产生一个巨大的人类行为数据集，可借助深度学习回答任何问题。



当前预测分析范式（顶部各图）和新范式通过社会物理学（底部各图）实现。

2.3.2 规范表述 (“知识领域”) 查询

由于知识领域包含关于所有用户的测得整体信息，因此可为任何问题提供预测，其简便性和速度相当于简单的数据搜索：对于十亿条记录组成的典型数据，知识领域初次创建通常耗时 1-4 小时。该过程完成后便可使用同一知识领域在数分钟内回答几十个问题。无需借助先验领域知识或提取相关特征。

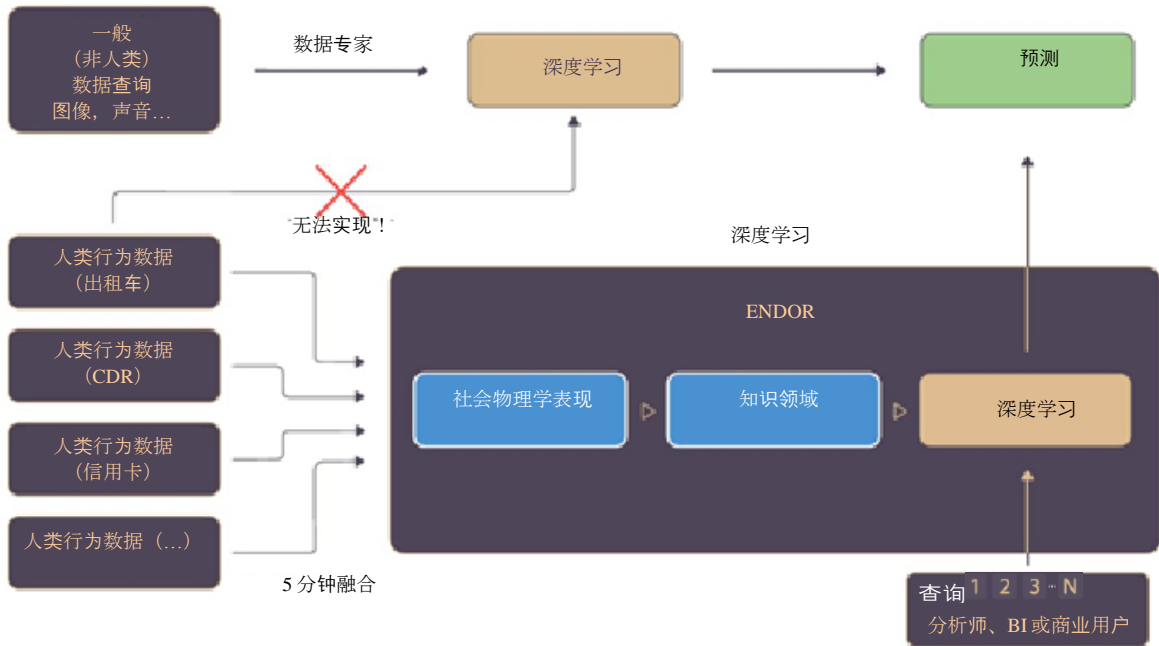
提供任意尺寸的“示例”（明确标示的 ID）便可提交查询。而后，Endor 引擎利用知识领域生成答案：它是一个用户排序表，极有可能“行为上类似于”相关查询。不存在“训练”/“学习”阶段；无需“解释结果”。如果是采用传统方法，则每个具体问题对应的自动过程可能需要数据科学家花费数周到数月时间。

例如，近期极有可能进行炒卖的用户相关查询由一个包含先前炒卖者身份信息的列表进行描述。此外，尝试识别哪些新客户可能转向高级账户的询问将会通过包含近期转用此种高级服务的客户的列表进行描述。然而，两种查询都会使用同一知识领域，无需进行任何类型的再训练。

注意，尽管大部分此类行为预测问题很难通过深度学习加以解决（由于缺少足够的数，而且基本模型频繁变化），但 Endor 利用全部知识领域和询问避开该问题。因此，Endor“创建”其负责提供实际预测的内部深度学习组件所需大数据。这些预测基于特定知识领域，即使它是基于极少量数据。

下图描述了相对于 Endor（下部流）的传统深度学习流（上部流）。对于非人类行为数据（例如：数百万幅图片），倘若接受了数据专家提供的适当训练，深度学习便可以产生高质量预测 - 因为当前深度学习工具专供熟练使用此种工具的工程师。其他方面的人类数据（例如：出租车数据）对应一个基本模型，其频繁变化使其难以通过深度学习建模加以解决（该问题深入论述参见第 2.1 节）。

尽管如此，当被转化为一组行为集群时，出租车数据从其语义结构中提取，并且几乎与从电话记录、信用卡消费或其他任何类型的人类行为数据中获取的行为集群。此外，由于有许多不同的客户提供上述每种类型的数据（即：大量电商平台，每个平台都上传自有购买和网络活跃度数据），深度学习模型训练成为可能。此种情况能够实现，因为模型不会进行原始数据或问题方面的训练，而是进行众多行为集群的培训 - 大量统一表述的数据，还使用静态基本模型进行表征（即：社会物理学定律；它们是数学不变量，因此相较于动态性较高的原始数据行为动态特征具有静态性）。



2.4 数据安全与匿名

2.4.1 Endor 店铺的数据

除自动解决各类行为预测问题之外，Endor 解决方案还提供高水平数据安全并隐匿任何所需数据字段。如前述章节详述内容，预测是利用知识领域完成 - 知识领域是一组无语义行为集群，仅包含大量用户组（每个用户组保证有一个相同社会或行为特征）。显而易见，如果最初原始数据中存在此种信息，则任何敏感或个人信息均该阶段此时不再是系统所用任何数据的组成部分。系统唯一可用的信息就是下例所示用户 ID：

<i>Behavioral_cluster1 =</i>	<i>(ID1, ID17, ID23, ...)</i>
<i>Behavioral_cluster2 =</i>	<i>(ID142, ID4287, ID9711, ...)</i>
...	
<i>Behavioral_cluster748,329 =</i>	<i>(ID5, ID37, ID218, ...)</i>

值得注意的是，客户在数据导入后可在原始数据层散列 ID，从而轻松隐藏该信息。

2.4.2 数据导入 Endor

如之前各节所述，用于从原始输入数据中提取行为集群的方法依赖于显示违反社会物理学不变性的数据模式的用户群的检测。其实现方式是跟踪描绘用户活动同步特性的特定统计特性的动态特征。在实施时，前述内容不需要数据中包含的实际值，而是可通过完全散列替换加以实现。这使得客户在向 Endor 提供完全散列的数据集的同时仍能受益于其出色的预测能力。此外，由于 Endor 是 100%语义不可知，数据字段名称也可散列。

财务记录散列示例如下所示：

散列之前的标头：				
账号	分支	性别	交易类型	目标账号
散列前的数据记录：				
183972	291/30	男	转账	382732
183972	291/30	男	余额查询	不适用
382732	291/30	女	转账	439001
...				
散列之后的标头：				
字段 1	字段 2	字段 3	字段 4	字段 5
散列后的数据记录：				
AjF32sdx	Q2KPbv3A	Wsqp289X	q8Vb3MAs	Je2qx92n
AjF32sdx	Q2KPbv3A	Wsqp289X	q8Vb3MAs	x3PNm78A
Je2qx92n	Q2KPbv3A	m28SbA12	q8Vb3MAs	yL19B4GQ
...				

2.5 社会物理学：数学解释

2.5.1 框架

我们首先介绍了 Endor 的基本原理及其一般数学形式体系，然后提供了两个实施例：（1）面向计算机视觉和（2）基于社交图谱。第一个示例描述了 Endor 与基于传感器数据（与人类无关）配合使用时的缺点，后者描述了社会物理学的概念及其在预测人类行为时的优点。

注：我们在整个论述过程中提供了大量关于社会物理学原理以及 Endor 引擎对其使用方式的数学说明。然而，出于知识产权考虑，此次论述略去了关于社会物理学定律的某些数学方面的详细说明。

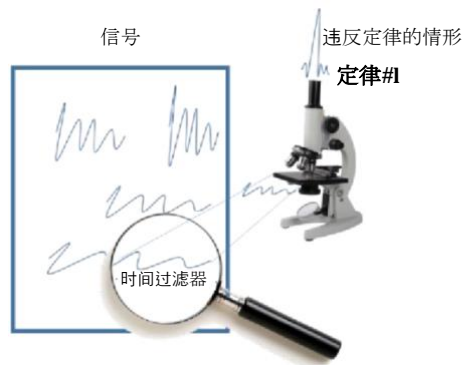
令 $d(x, t)$ 为一个时态数据流，其中 x 表示单个数据点。

令 $L(\cdot)$ 为将原始数据 $d(x, t)$ 转换为定律表述的定律算子：

$$L_{X,T}(d) = \frac{1}{|X| \cdot |T|} \iint_{X \times T} L(d(x,t)) dx dt$$

定律本身是一个公式，使定律算子等于先验常数 C （可以是一个数字或分布类，例如幂次定律等）。常数 C 表示定律所代表的的不变量。

基于说明之目的，我们可以想象一个假想定律，表述白噪声信号随时间的输出变化仅限于一个非常小的阈值。例如，输出尖峰突然变强的信号便违反了该定律。随后便可建立一个过滤信号并检测此类异常输出的机制，用于找出违反该定律的情况。很显然，使用此种机制仅对由所述定律支配的信号（即：白噪声的源头）有意义，因为对于其他信号而言，所述尖峰不能归类为“违反定律的情况”。



另一个示例可以是用于检测辐射吸收材料的异常“组块”的 X 光机。本例中的 X 光机（以及对其进行操作并解释所产生图像的技术人员）作为“违反定律情形检测者”分析 2 维信息流并定位不变性的各种违反情形，确保轰击底片的连续 X 射线束形成均匀吸收的图像（或多或少取决于光片质量和光束连贯性）。

这种不变性（或定律）是一种已知的物理事实，并用于医学用途，方法是将连续 X 射线束人为导向高品质光片，同时穿过第三种物质，以期找出“违反定律的情形”，从而表明该物质中存在吸收 X 射线的元素。例如，骨骼就是这样一种材料，它在此种情况下代表“数据”（即曝光胶片）中检测到的一种“现实现象”，表现为辐射吸收不均，其本质上是违反上述定律的一种情形。



用于执行上述原则的具体的正式数学示例参见第 2.5.2 和 2.5.3 节。

对于社会物理学而言，此种违反定律的情形包含在输入数据中生成一定动态特征一群人，如果假设输入数据源自人类活动并因此符合社会物理学定律，则从统计学角度讲极不可能。

2.5.2 确认某个数据子集是一种违反定律的情形

Endor 引擎执行多个社会物理学定律的定律算子和常数（出于知识产权考虑，其具体细节未进行描述）。本节内容描述了一种违反定律情形的确认。需要注意的是，类似于其他许多计算问题，确认某个特定信号包含一种违反定律情形从根本上不同于这种干扰的检测。本节内容只讨论前者，即：确认“潜在干扰”是否确实违反定律。此处未详细讨论可高速扫描大规模高维数据源并执行即时验证的 Endor 搜索算法。

我们所验证的定律数学公式始终以下面的形式给出：

$$L_{x,T}(d) = C$$

假设定律有明确表达，局部偏离（表示为 ξ ）可通过偏离测量进行验证，具体如下所示：

$$\xi(\Delta x, \Delta t) = L_{\Delta x, \Delta t}(d) - C$$

其中， Δx 表示 X 的一个子空间，而 Δt 则表示时间窗。因此，每个 X 子空间和任何时间段都可计算该偏离，并生成一个量度表示子空间在给定时间段内违反相关定律的程度。

该量度与预定阈值 $\xi_{\text{阈值}}$ 进行比较可检测到违反定律的子空间：

$$r(\Delta t) = \{ \Delta x : | \xi(\Delta x, \Delta t) | > \xi_{\text{threshold}} \}$$

定律违反情形阈值 $\xi_{\text{阈值}}$ 选择时应确保从定律中延伸出的信号同时出现超出阈值的情况极不可能发生。此种做法可自动验证某个数据子集是否违反某个定律，具有足够高的统计学意义，没有数据本身语义的任何先验知识。

注意，由于信号在时间和空间上都发生变化，不同的时间窗会产生作为违反定律情形被检出的不同子空间。Endor 使用一组预先确定的固定时间窗（源自定律而非数据）： $\Delta t = 1$ 天，7 天，30 天，90 天。

如果数据动态性较高，则较长的时间窗不会产生任何偏差组；如果数据处于静态，则较短的时间窗不会生成任何偏差组。时间窗无论如何都不会产生“无意义组”，因为根据定义，噪声无法产生稳定的违反定律情形（或者按照更加正规的说法，噪声产生足够明显的违反定律情形的概率接近于零，因为这正是选择阈值 $\xi_{\text{阈值}}$ 时所采用的方式）。

知识领域是对于所有相关时间窗的所有定律全部组差的集合：

$$K_{sphere} = \{r(\Delta t) : \forall \Delta t, L\}$$

该知识领域每个数据集计算一次，因为该过程不受询问查询影响，而是受源自原始数据的内部行为结构影响。从抽象的角度来讲，Endor 社会物理学引擎将匿名原始数据“压缩”成行为相关规范表述。

2.5.3 示例

下面提供的两个示例展示了利用定律检测数据中的违反模式。通过说明，我们利用已知数学现象论证此种机制。

例 1：视觉

令 $d(x,t)$ 表示特定时间段 t 内特定像素 x 的颜色。此时可定义同色定律，规定任何子区每种颜色都必须是某种预定颜色 C ：

$$L_{x,t}(d(x,t)) = \frac{1}{|X| |T|} \iint_{X \times T} L(d(x,t)) dx dt$$

该算子利用区域 x 和时间窗口 t 为该输入数据计算平均色。

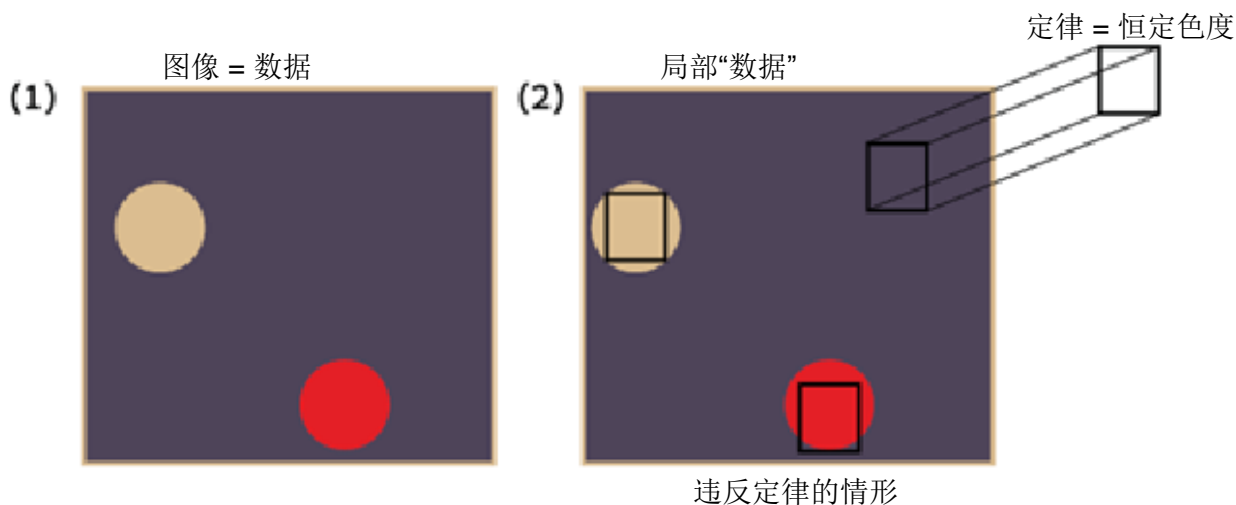
本例通过否定描述了定律的概念：不存在关于视频像素的固有先验已知定律（当然不会假设每个子区都会使用相同的平均色...）。考虑到该低算子和常数 $C=$ 蓝色，此时可将 Δx 定义为 $N \times N$ 像素的方形窗口，并使用单时间段 $\Delta t = 1$ 。

因此， $N \times N$ 窗口与淡蓝色的局部偏差如下所示：

$$\square \square \square \square \xi(\Delta x, \Delta t) = L_{\Delta x, \Delta t}(d) - C$$

在 2 维图像中，其可轻松用于测量色斑与淡蓝色的偏差，将此种偏差与预定阈值进行对比。偏差超出该阈值的色斑将归类为“具有相似特性的像素集群”，表示方框中违反“淡蓝色定律”的所有像素。

我们可以使用略微更加一般化的假法定律对此进行说明：假设图像始终“平滑”（即：它们在较小尺度内是单色，缺少“峰值”或局部极大值/极小值）。该假设条件下，我们可以观察下列“数据”- 包含一个平滑颜色“合法”背景的一幅图像，上有两个色斑（一个为蓝色，另一个为红色）。大量小方形随机区域的采样可轻松定位这两种“违反定律的情形”。



然后可以红色像素形式提出一个问题。该像素将识别为包含在红色违反定律情形中，将红色斑作为结果返回。

请再次注意，这些示例仅用于描述社会物理学的数学与力学特性，而不是暗示其适用于基于简单视觉的应用，因为此类应用可采用传统计算视觉或深度学习方法。

例 2：无尺度网络

本例中， $\mathbf{d}(\mathbf{x}, \mathbf{t})$ 抽象地表示一幅以 \mathbf{x} 为节点的图。定律算子是度分布算子，公式为：

$$\bar{L}(x) = \ln(x) = \begin{cases} 1 & \text{x has degree n} \\ 0 & \text{otherwise} \end{cases}$$

该向量算子为每个节点的读数生成 1。所有图节点算子结果相加为图形度数得到一个基数向量（除以节点数量时等同于度分布）。

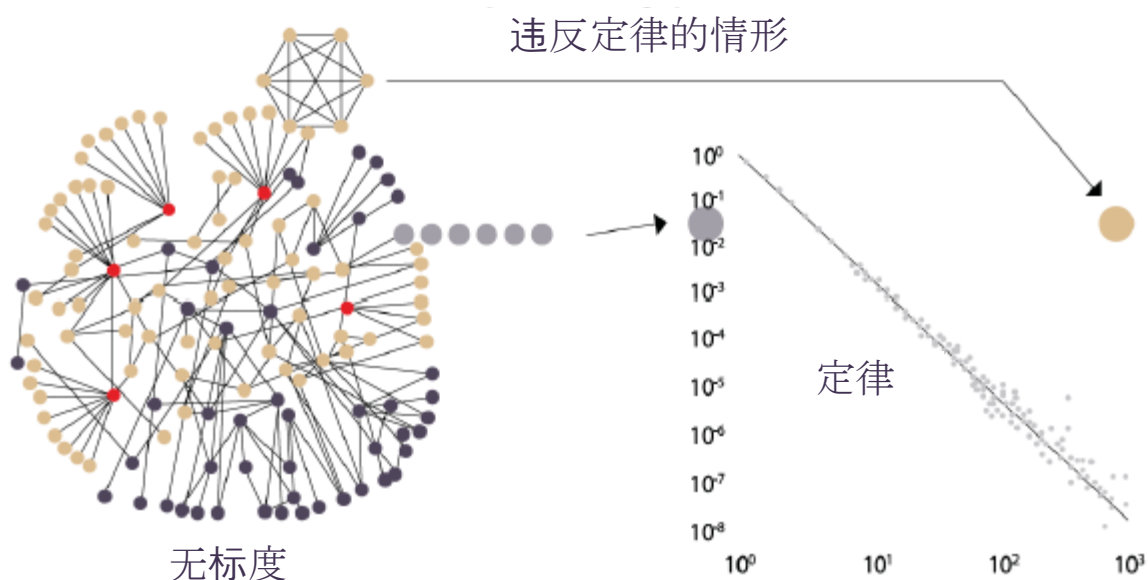
本例中应假设图形是一个无标度网络。因此，假设可使用幂率度分布的定律常数（用于某个归一化常数 α ）：

$$\overline{C} = c_n = \alpha \cdot n^{-\gamma}$$

该定律常数又可表示为：

$$L_{X,T}(d) = \frac{1}{|X| \cdot |T|} \int \int_{X \times T} L_n(d(x,t)) dx dt = C_n$$

该定律意味着整幅图应遵守所有节点度数的幂律分布规则。尽管如此，在许多大面积现实无标度图形中可能会与此种分布存在较大的局部偏差。例如，这可能发生在团（即：完全连接的子图）或链（即：构成没有任何节点有超过 2 个相邻节点的连接树的节点子集）周围。定律的此种偏差或违反定律的情形如下图所示，包含其结构化表述（左）和邻接矩阵（右）：



注意，在给定定律的情况下，通过各种量度可轻松确认此种违反定律的情形，例如：

$$\xi(\Delta x, \Delta t) = \sum_n \frac{|L_{\Delta x, \Delta t; n}(d) - C_n|}{2}$$

□

该偏差等于差值累积平方（而此类量度的另一个示例可以是两种概率分布的 *KL 散度*）。其中， Δx 表示图形所有可能包含的子图。显而易见，在一幅输入图中无法扫描所有可能存在的子图，因为它是名为“*非多项式难解问题*”的一类“高难度问题”。从这层意义上讲，我们必须区分验证违反定律情形（需要了解定律的详细信息）和检测到一种违反定律情形（需要专门针对每种定律开发的一套专有技术）。此种检测技术是 **Endor** 专有技术的组成部分，专为社会物理学数学定律而定制。

再次回到我们的无标度示例，假设我们拥有一种高效率技术能在图形中找出所述局部干扰，它们将会产生一组可以下式表述的子图：

$$r(\Delta t) = \{ \Delta x : | \xi(\Delta x, \Delta t) | > \xi_{threshold} \}$$

而且，该定律暗含的知识领域定义为：

$$K_{sphere} = \{ r(\Delta t) : \forall \Delta t, L \}$$

一旦包含违反社会物理学定律情形的知识领域可供使用，其便可用于针对任何给定的标记范例（定义为一组来自相同领域的对象）检测“类似情形”。在我们的示例中给出了一系列图节点，因此，所有其他图节点可根据其与标记范例共享的集群数量进行评分。另外还可使用不同的评分标准，只要它们仅将检测到的集群以及标签列表作为输入内容，以便产生群体评分形式的输出。我们将此种标准或评分机制称为“评分器”，并在下一节对其进行详细说明。

注意，不同的时间窗会产生不同的知识领域，代表着图节点之间差异非常显著的联系。此外，这些集群还代表着并非由社交媒体或社交网络等外部数据源产生，而是由客户自有内部交易数据源产生的行为联系。因此，**Endor** 可以检测到未在任何可用数据中明确表现的隐性行为集群。

2.5.4 回答问题：简单的“评分器”

知识领域可供使用后，用户便可以通过提供询问提出业务相关问题 - 具有语义意义的标记数据列表（可能会非常小）：

$$y. \in X.$$

本节提供了两个“评分器”示例 - 使用行为集群的函数 + 用于生成排名群体列表作为输出内容的询问。

我们论述的第一个评分器是一个简单的分群聚合评分器，为每个候选者 $\tilde{y} \in X$ 计算下列得分：

$$score_{\tilde{y}} = \sum_{\Delta x \in K_{sphere}} \begin{cases} 1 & \tilde{y} \in \Delta x, y \in \Delta x \\ 0 & \text{otherwise} \end{cases}$$

每个可能存在的“候选者”都根据与标记数据组成部分共生的知识领域中的集群数量进行评分。这种简单的评分器用于计算 X 组成部分与输入的标签列表之间行为相似处的数量。与标记输入列表中的对象共享集群的对象列表表示为：

$$\tilde{y} = \{x \in \Delta x : y \in \Delta x, \Delta x \in K_{sphere}\}$$

例如，对于给定的有色点以及示例 1 中描述的“同色定律”，即一个作为输入内容给出的点周围的带色区域。对于“无标度图形定律”中给定的一个节点，输出内容将会是包含局部违反度分布的该节点的所有子图中的节点。

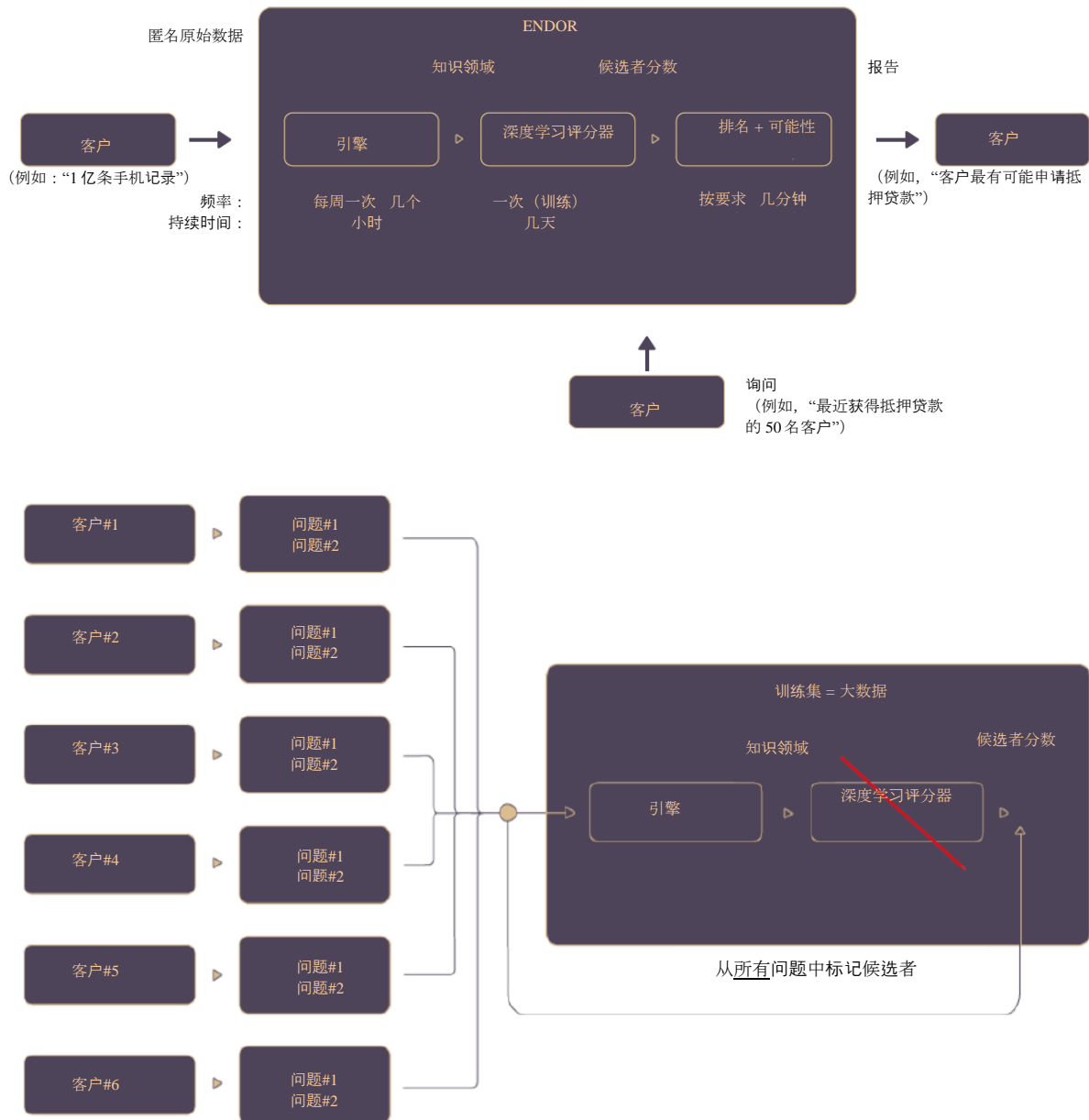
2.5.5 回答问题：深度学习评分器

上一节所述的评分器直观地描述了构成社会物理学知识领域的集群如何针对任何“示例输入”（即：询问）产生高品质“类似预测”（未提前获知询问性质）。这个简单的示例描述了社会物理学规范数据表述的各种优点。然而，为实现精确预测，Endor 根据最先进的深度学习算法开发出一种更加稳健的评分器。

如上文所述，当涉及到人类行为时，深度学习的高效使用需为其训练阶段提供大量数据。利用原始数据难以达到这种效果，因为每个标记询问通常包含少量标签，而每次询问往往只有极少量实例，而每个客户对应的不同询问数量也相对较少。这种动态性较高的且较为复杂的小型标签数据不是非常适合深度学习。Endor 社会物理学引擎通过将来自全部客户的数据集以及所有来自全部询问的实例转换成单一规范形式的较大集合而克服了这种局限性：由人的集群以及对应这些集群的示例组成的知识领域。来自所有这些数据源的所有数据的融合使得 Endor 能够创建足够大的标记训练 - 训练一个深度学习网络为每个人评分，前提是给定了标签和知识领域。该过程是一次性完成，所产生的高品质已训练深度学习模型可高效处理任何基于新数据源和新查询而产生的任何新集群。

注意，这种受过训练的模型不需要定期再训练（对于新查询和新的数据集类型均是如此），因为其已在表述为一系列集群（即社会物理学规范表述形式）的（大量）数据实例上进行了训练。基于此种方法，Endor 融合了社会物理学和深度学习的各种优点：社会物理学基于违反社会物理学定律的情形将匿名原始数据转换成规范形式，而深度学习算法再根据所形成的知识领域和（可能是极少量的）标记数据对候选者评分。

这种流程如下图所示：



2.6 噪声鲁棒性

Endor 预测平台主要优点之一，便是其对于数据间隙和噪声数据富有弹性。传统而言，每个数据分项目的第一步都是“数据清理”，主要任务是对下列不需要的数据段进行检测和必要的调节：

- 数据间隙（即：缺失数据的阶段，全部或部分）
- 原始数据中的胡乱插入部分
- 语义含混，例如：一个可能以多种形式出现在数据中的类别名称
- 数值归一化问题
- 需要进行的数值装箱
- ...

机器学习算法一般对噪声极度敏感，由于数据通常伴有噪声，减轻这些数据问题通常需要高昂的成本和长时间的操作。出现此种情况的原因是，现实世界中数据产生时呈现原始形式（金融交易和电话通话等），以便传统机器学习方法能够对其进行分析，然后转化成一种综合形式，称为“特征”或“特性”。这种集合也会受到少量错误数据值或针对不同用户对应不同数量值的显著影响。

Endor 通过分析原始数据本身克服了这一问题（如前面各章节所述）。此外，Endor 引擎不通过执行数据统计分析寻找可用于预测的模式，而是使用社会物理学定律 - 在数据以外且不受其影响的数学不变性。此种方法具有多种显著优势，主要是来自于下列基本概念：

噪声无法形成“不能随意出现”的数据模式

（后者定义为基于分析显示可存在可忽略概率的数据模式）

为了解此项观察在先前未进行任何数据清理的情况下如何实现任何人类行为数据分析结果的自动提取；提取过程在违反其中一条社会物理学定律的数据中寻找数组对象，也就是能够显示我们能够证明无法在数据中自然产生（利用社会物理学所驱动的数学分析）的数据模式的组。

这意味着：鉴于噪声可以将分析结果从我们的平台上隐去，则根据定义，它（几乎）无法创建一种可作为违反定律情形被检出的数据模式。噪声数据不能违反社会物理学定律，只有人类驱动信号数据才可以。

3. 结果

本节提供的各种用例描述了 Endor 预测系统的使用方法。整个预测过程的详细说明，利用包含大型金融投资平台 7 天价值活动精确回答 4 个不同的预测性问题（包括利用谷歌 Tensor Flow 深度学习平台实现的一项比较分析）；

1. 以全自动方式使用 Endor 应对卡歌网挑战的一个实例。
2. 可口可乐案例研究：Endor 利用销售网点数据，数据在 24 小时内针对大量商业问题提供了多种精确预测。

3.1. 利用超过 7 天的金融活动精确自动地回答四个预测性问题

本节展示了基于 Endor 系统的整个预测过程：

- 描述所用原始数据
- 数据转换为社会物理学表述（即：由一系列行为集群组成的知识领域）
- 四个预测性问题的定义
- 社会物理学形式的询问论证
- 总预测精度
- 对比谷歌 Tensor-Flow 深度学习平台。

3.1.1. 数据：

本例所用数据源自零售金融投资平台，包含投资群体成员的所有投资交易。数据进行匿名处理，并在麻省理工学院内部公开用于研究目的（可按要求分享数据）。

数据集全面汇总如下所示：

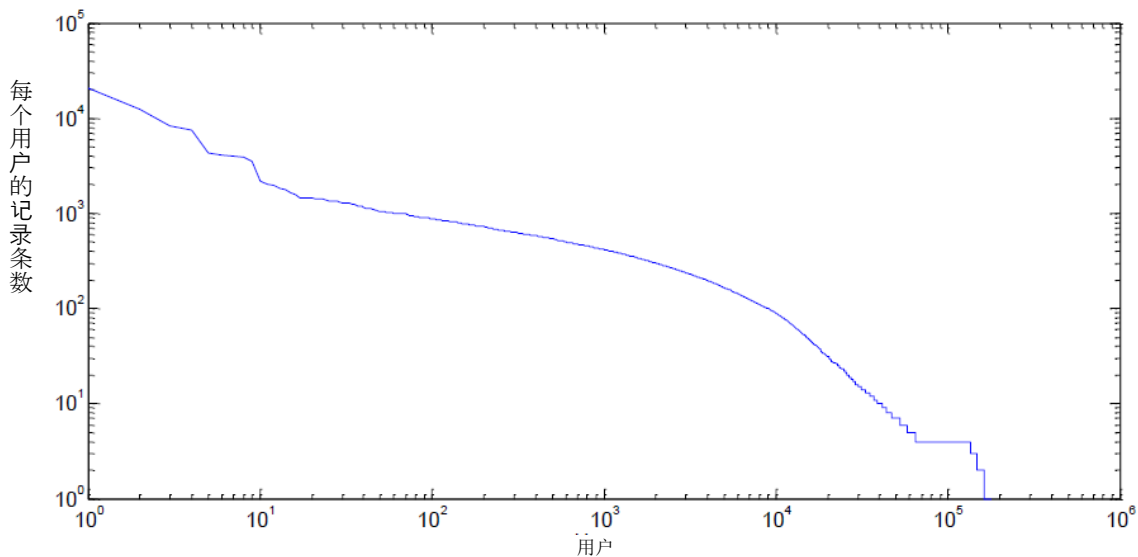
- 7 天数据
- 3,719,023 行
- 178,266 名唯一用户
- 数据未进行语境和语义说明
- 12 个数据字段：

字段名称	类型	#唯一值
时间 1	时间	501573 唯一
时间 2	时间	4 唯一
用户 ID	INT32 类别	178266 唯一
记录 ID	INT32 类别	1053574 唯一
特性 1	INT32 类别	24 唯一
特性 2	INT32 类别	7 唯一
特性 3	INT32 类别	134 唯一
特性 4	INT32 类别	77527 唯一
特性 5	INT32 类别	10 唯一
特性 6	INT32 类别	27 唯一
特性 7	INT16 类别	9 唯一
特性 8	双数字	3772 唯一

数据集的下列各重要事项应加以注意

- 数据提供时为原始未聚合形式，且包含用户级事件。
- 数据未进行语境和语义说明。
- 未进行数据清理。数据包含噪声、间隙、重复记录等。
- 数据包含一个极不均匀的用户记录分布（78%的用户不超过 10 次交易，但有 4000 名用户超过 200 项记录。中位数为每个用户 4 条记录）：

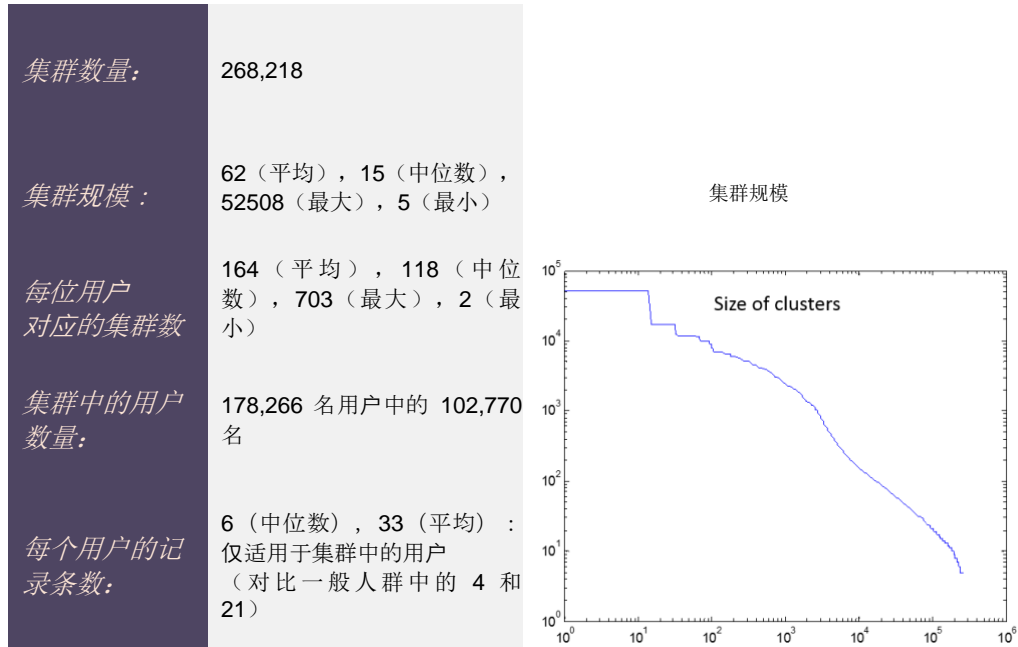
用户平均记录条数双对数表述



3.1.2. 自动集群提取

数据进行第一次分析时，Endor 系统检测并提取“行为集群”- 数据动态特征违反社会物理学数学不变性的用户群。这些集群是基于所有数据列，但仅限于最后 7 天 - 因为数据是作为输入资料提供给系统的。

来自系统检测到的数据集的行为集群作如下汇总：



3.1.3. 预测查询

定义的预测查询如下所示：

- **新用户变成“大户”**：过去 2 周，加入的用户在随后 90 天内至少产生 500 美元的可用资金
- **降低活跃度**：上一周仍活跃的用户在随后 30 天内的活跃度会降低 50%（但不会出现炒卖，仍会继续交易）
- **“大户”炒卖**：上周活跃且当前仍活跃的“大户”（根据最近 90 天的活跃度），后续 30 天内变得不活跃
- **将在 Apple 中进行第一次交易**：从未投资苹果股票但会在未来 30 天内第一次购买的用户

可以看出，上述所有问题都涉及延伸超过 7 天数据集的数据 - 过去的的数据（用于产生“搜索群体”或“示例”）和将来的数据（用于确认预测）。为避免为系统提供任何超出 7 天期限的信息，询问采用包含用户 ID 值的列表。例如：

- **查询名称**：新用户变成“大户”
- **搜索群体**：一组包含过去 2 周内（数据结束前）加入用户的用户 ID。
- **“示例”**：一组包含“大户”的用户 ID（即：数据结束前 90 天内产生超过 500 美元可用资金的用户）。
- **待寻找目标列表**：一个用户 ID 列表，它是“搜索群体”的子集，包含数据结束后 90 天内产生 500 美元以上可用资金的用户。该列表仅用于确认目的，且并不提供给系统。

3.1.4. 询问的知识领域表现

需要再次注意的是，搜索查询的定义与行为集群的提取以及知识领域的生成（独立于查询定义）完全正交。因此，在系统检测到的集群中分析查询的体现十分有意义：尽管存在以下情况，集群是否包含与查询定义有关的信息：

- 集群以全自动方式提取，不使用关于数据的语义信息，并且
- 集群提取后再定义查询，而且不影响这个过程。

此项分析的实现方式是，测量包含极高浓度“样本”的集群数量；换言之，也就是寻找包含“示例数量远超统计预期”的集群。大量此类集群（假如数量远大于同一群体随机取样时收到的数量）证明该过程能够以全自动方式提取价值较高的相关语义分析结果。

下表展示了此项观察，对比包含一定量“样本”的行为集群数量，同时比较包含相同数量样本的“随机集群”数量。随机集群是指从客户群体中随机取样的 N 组用户，N 等于系统检测到的行为集群数量，随机取样群组的大小等于系统检测到的集群大小。此种做法用于验证行为集群所含信息，因为它们所包含的“目标用户”高度一致集群数量远远大于随机样本（注意，它们是在“问题”定义之前由系统检出）。

所要求“样本”数量单位为“基线”。即：用于“降低活跃度”查询（基线约为 11%）的“X5 基线”是指与“样本”有 55%重合度的集群（即：其有 55%的成员也出现在“先前降低活跃度的用户”列表中。）

包含许多目标客户的集群		#随机集群	#行为集群
降低活跃度	来自基线的 X2	0	98
	来自基线的 X5	0	11
“大户”中的炒卖	来自基线的 X2	212	1678
	来自基线的 X5	67	525
	来自基线的 X10	21	114
从未购买	来自基线的 X2	3962	60044
	来自基线的 X10	1332	25542
	来自基线的 X20	415	9090
新大户	来自基线的 X2	38	1898
	来自基线的 X5	0	65

3.1.5. 预测结果

下表列出了四项查询的预测精度

- 基线：随机人群样本中请求目标客户的平均部分，代表一种随机猜测的准确度。
- 候选者：搜索人群的规模。例如，在“将成为大户的新用户”中，候选者数量是指新用户数量。
- 前 100 名：预测报告前 100 名成员中被请求的目标客户的部分（前 250 名和前 500 名同理）。

	基线	前 100 名	前 250 名	前 500 名
将成为“大户”的新用户 过去 2 周，加入的用户在随后 90 天内至少产生 500 美元的可用资金	7.5% (2270 名候选者中的 170 名)	37%	28.8%	21%
降低活跃度 当前活跃用户当中在接下来的 30 天内活跃度会降低 50% (但不会炒卖) 的用户	11.4% (2233 名候选者中的 255 名)	21%	23%	20.2%
“大户”中的炒卖 当前活跃“大户”，上周较为活跃，接下来的 30 天内会变得不活跃？	1.66% (4141 名候选者中的 69 名)	10%	10.8%	9.2%
将第一次在 Apple 中交易 从未购买苹果股票，但会在未来 30 天内首次购买的用户。	0.5% (161382 名候选者中的 839 名)	14%	12%	10%

根据当前和预期情况，精度会随着预测表的深入逐渐降低。

3.1.6. 与 Tensor-Flow 的对比

本节内容对 Endor 系统和谷歌 Tensor Flow 得出的预测结果进行了对比。需要注意的是，同其他任何深度学习库一样，在处理类似于所述数据的数据时都会面临一些困难。

- 每名用户记录数量极不均匀的分布需要对数据进行一定程度的经典化，进而需要：
 - 由对数据科学至少有一定了解的个人进行手工作业。
 - 对数据语义有一定的了解；这需要耗费一些时间，还需要能够接近数据所有者或提供者
- 阳性和阴性样本极不均匀的分布会导致单分类出现结果过适，因此需要进行一定程度的重大调整。这需要深度学习专家参与（不同于业务、产品或营销专家可使用的 Endor 系统，在机器学习或数据科学方面无先决条件）。

我们要求一名深度学习专家花 2 周时间根据 Tensor Flow 制定一套解决方案，其应具有足够的专业知识进行数据处理。创建的解决方案使用下列辅助技术：

- 将数据序列调整为每位客户 200 条记录，并用中性记录为记录少于 200 条的用户补足流量。
- 创建 200 个训练集，每个包含 1,000 名客户（50%已知为正标签，50%未知），然后用这些训练集训练模型。
- 序列分类（带 128 个长短时记忆的回归神经网络）与 2 个输出神经元（正，负）配合使用，总结果是两者分数之差。

下表把使用这些技术所得结果（红色）与 Endor 预测结果（蓝色）进行了对比：

	基线	前 100 名	前 250 名	前 500 名
将成为“大户”的新用户 过去 2 周，加入的用户在随后 90 天内至少产生 500 美元的可用资金	7.5% (2270 名 候选者中的 170 名) (2135 个示例)	37% 21%	28.8% 27.2%	21% 19.6%
降低活跃度 当前活跃用户当中在接下来的 30 天内活跃度会降低 50%（但不会炒卖）的用户	11.4% (2233 名 候选者中的 255 名) (366 个示例)	21% 8%	23% 18.8%	20.2% 19.4%
“大户”中的炒卖 当前活跃“大户”，上周较为活跃，接下来的 30 天内会变得不活跃？	1.66% (4141 名候选者 中的 69 名) (21156 个示例)	10% 11%	10.8% 12.4%	9.2% 8.4%
将第一次在 Apple 中交易 从未购买苹果股票，但会在未来 30 天内首次购买的用户。	0.5% (161382 名候选者 中的 839 名)	14% 1%	12% 0.8%	10% 1%

观察结果：

- Endor 在 4 次询问中有 3 次优于 Tensor Flow，而且第 4 次获得了相同精度。
- 随着任务“越来越困难”，Endor 的优势越来越明显 - 主要是关注前 100 名而非前 500 名。
- “动态性较低的询问”（变成大户，炒卖，降低活跃度降低）和“谁将第一次在 Apple 中交易”询问之间存在明显区别，它们 (a) 动态性更强，而且 (b) 基线极低，因此对于后者，Endor 准确度是另外一种方法的 10 倍！
- 如前文所述，此处显示的 Tensor Flow 结果是由深度学习专家进行 2 个星期的手动改进获得，而 Endor 结果则是 100% 全自动获取。

3.2. 3 小时内自动攻克卡歌网挑战

在另一个示例中，我们使用卡歌网比赛（称为“获取有价值的购物者挑战”）中得到的公开数据测试了 Endor 系统。数据包含近 3 亿条销售网点记录，涉及成千上万名顾客，而挑战最终预测，得到某些促销优惠券的用户将成为回头客户。

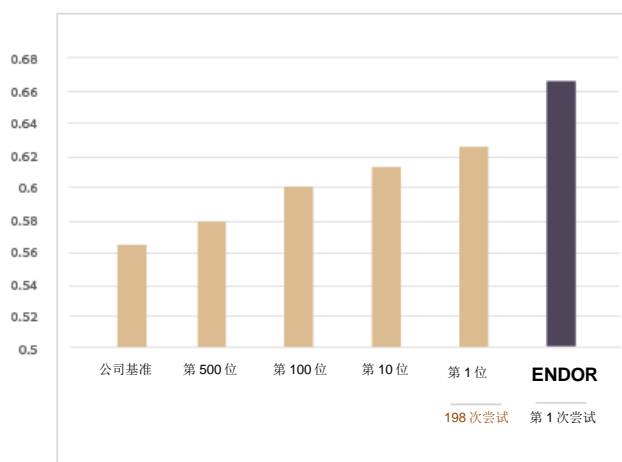
初始挑战涵盖 952 个团队并持续了 3 个月。

Endor 设备在为挑战参与者提供的原始数据上全自动运行，得到的预测结果优于最初拔得头筹的团队。



- 952 个参与比赛的数据科学团队
- 3 个月的挑战

ENDOR :
几乎无点击，第 1 名



附录 B : Endor 企业常见应用案例

如果你有自己的个人甲骨文数据库，会是怎样一种情景？如果能够使用先前只由谷歌和 Facebook 使用的强大数据引擎，那会是怎样一种情况？如果你能使用迄今为止最强大的 Endor 引擎获得最可靠的答案，那会是怎样一种情景？为通过高级机器学习分散化而使区块链预测能够供所有人使用，*Endor.coin* 赋予你这种目前为止仅为技术巨头所用的能力，而它们还要借助内部专业数据专家团队才能实现。

Endor 企业客户所申请的常见预测用例作如下汇总。它按 Endor 目前服务的下列主要领域对用例进行分类：

- 零售银行
- 保险公司
- 零售和电子商务

然而，由于使用了社会物理学技术，新用例可轻松获得支持 - 平均需要 Endor 销售工程师几小时的服务时间。



ENDOR



应用实例：零售银行

ENDOR

常见用例



销售更多

- 购买偏好
- 交叉销售
- 向上销售
- 出手阔绰者
- 新服务 - 早期采用者

问题示例

- 购买偏好：过去 3 个月里，哪些合格的活跃客户会在下周接到电话时决定贷款 X 美元？谁会在没有信用卡的情况下贷款？短期贷款？长期贷款？
- 交叉销售：对于过去 6 个月内使用 X 产品的顾客，如果下个月提供 Y 促销，那么谁会开始使用 Z 产品
- 出手阔绰者：对于最近 30 天内加入的新客户，谁可能会在未来 3 个月内每月花费 X 美元以上？
- 新服务：对于近 3 个月一直保持活跃的客户，谁会注册 2 周以前推出的一项新服务？

常见用例

销售更多



其他示例

- 对于近 90 天以来所有保持活跃的信用卡客户，谁会在获得第二张卡时再增加 X 美元的花费？□
- 对于近 90 天内不活跃信用卡客户，如果我们将 A 产品换成 B 产品，哪些人会开始使用信用卡？□
- 对于近 6 个月来所有保持活跃的客户，谁会在未来 3 个月内出现欠款？□
- 对于近 6 个月来全部保持活跃的客户，下个月谁会愿意获得零利率？□
- 对于所有活跃的黄金和白金账户，下个月谁会愿意为一张卡缴纳申请费？□
- 对于近 6 个月来产生利息的客户，谁会在未来 3 个月内清偿其欠款？□
- 对于在我们旅游网站上购买机票的人，谁会购买酒店夜间服务？□
- 白金和黄金组合中哪些客户会在未来 2 个月内在高成本类别中挽回积分？□
- 对于白金卡高价值客户，谁会在下个月致电取消该卡？
- 对于近 12 个月来一直保持活跃的客户，未来 2 个月内谁有可能在所有产品中花费 Y 美元以上？
- 对于近 6 个月来一直保持活跃的 Y 产品客户，未来 2 个月内谁有可能对 Y 产品花费 Z 美元以上？□
- 对于去年未持有任何信用卡的客户，如果下星期联系，谁会在下个月办理一张新信用卡？

常见用例



销售效率

- 促销效率
- 跨渠道粘性
- 客户忠诚度

问题示例

- 促销效率：对于活跃客户，谁会在收到文本消息/电话/Y 促销后注册高级服务 X？
- 数字卡：对于至少在所列网站上购物 Y 次的客户，谁会在下个月使用回报服务进行购物？
- 对于至少进行过一次免税购物且最近 X 个月未在国外购物的客户，谁会在未来 3 个月内使用信用卡进行海外购物？
- 对于所有活跃客户，谁会在未来 30 天内飞往纽约或英国？
- 对于最近 90 天内刚加入的客户，谁还没有将卡激活？谁会在收到 10 美元/20 美元奖金后将卡激活？
- 对于所有近 6 个月内的活跃持卡者，谁会在下个月停止使用我们的卡？

ENDOR

常见用例

销售效率



其他示例

- 对于未打开月度消费信用促销电子邮件的客户，如果我们下周(1)向他们致电或(2)发送短信，谁会在未来 3 个星期内办一张新卡？
- 对于已有 3.000-24.000 点积分的已婚客户，如果我们明天向其发送电子邮件，谁会在未来 4 天内用积分兑换香水？
- 对于在餐馆中使用信用额度 100 万美元以上的信用卡并住在 X 区的客户，如果本周向其发送电子邮件，谁会在下周将 X 产品的消费额提高 Y%以上？
- 如果我们下周向客户提供加油特惠，哪些客户下个月会用信用卡加油？
- 对于近 6 个月内重新商议未偿清债务的客户，如果明天发送一条短信提醒他们付清账户欠款，谁会在下个月出现违约情况？
- 对于至少 2 个月未使用过信用卡的客户，如果下周向其发送加油特惠活动信息，谁会再次开始使用手上的信用卡？

ENDOR

常见用例



数字转换

- 提高在线活跃度
- 从分支过渡到在线
- 重新定位目标网站访问者
- 首次购买者，多次购买者，长期购买者
- 出手阔绰者

问题示例

- 对于已经 2 个月处于不活跃状态的在线客户，如果向其致电/给予 X 奖励，谁会在未来 7 天内恢复活跃状态？
- 对于所有采用传统交易方式（分支，呼叫中心）的客户，如果下周向其提供 X 促销，谁会开始使用数字交易方式？
- 对于最近 7 天的新增的在线用户，谁的交易总额可能会在前 3 个月内达到 X 美元以上？
- 对于过去 3 个月内通过数字渠道注册购买新产品的所有客户，如果向其提供 X 促销，谁会在下个月内注册购买第二种产品？
- 如果我们重新将其设为目标，近 3 个月内的哪些网站访问者可能会注册我们的网上银行服务？

ENDOR

常见用例

数字转换



其他示例

- 对于上个月注册的新在线用户，谁会在未来 3 个月内继续使用在线服务？有哪些人不会继续？
- 对于所有网上银行客户，谁会在下个月停止使用网上银行？
- 对于上个月新注册的所有在线用户，如果下周给予 X 奖励，谁将成为经常性在线用户？
- 对于所有在线客户，谁会使用 X 功能而不使用 Y？如果向其致电，谁会在下个月使用 Y？
- 对于过去 3 个月内注册购买新数字产品（X，Y，Z）的所有客户，如果向其提供 X 促销，谁会在下个月内注册购买第二种产品？
- 对于过去 3 个月内通过数字渠道获得产品的所有客户，谁会在 1 个月内离开？
- 对于近 3 个月内所有网站访问者（网络跟踪器 ID），谁有可能在被重新作为目标时注册我们的网上银行？
- 对于只有一种银行产品的客户，谁有可能在下个月购买第二种产品？
- 对于从未访问我们网站的客户，如果我们明天向其发送电子邮件，下周谁会访问我们网站？



ENDOR

应用实例：保险

ENDOR

常见用例
你每天如何进行预测？作出决定？提升绩效？

你需要回答以下问题



销售

- 谁会增加一个附加承保范围 X？
- 如果下个星期与其联系，谁有可能向上销售/交叉销售？
- 对于取消预定生效日期的客户，谁会重新购买？
- 谁会购买一周前推出的新产品服务？ □
- 谁会在下周申请一项新计划？ □
- 对于新客户，谁会成为出手阔绰者/高级消费者？ □□



营销

- 谁会对促销作出回应？
- 谁会在开始 90 天内终止合同？
- 谁会使用我们的移动服务（应用）？
- 谁会在 2 个月内前往纽约和墨西哥？
- 谁有可能推荐一位朋友？



数字

- 谁会将活跃度提升到 X 美元以上？
- 谁会使用新数字服务？
- 哪些不活跃者会恢复活跃状态？
- 谁会从分支转变到网上购买？
- 谁会对促销作出回应？
- 如果被重新设定为目标，哪些网站使用者会发生转变？

常见用例



销售更多

- 购买倾向
- 交叉销售
- 向上销售
- 出手阔绰者
- 新服务-早期使用者

问题示例

- 购买倾向：对于所有潜在客户，谁有可能会增加 X 保单？对于所有现有客户，谁有可能会扩大附加承保范围？
- 交叉销售：对于 X 计划的客户，如果向其提供 Z 促销，谁可能会购买 Y 产品/服务？
- 出手阔绰者：对于最近 30 天内加入的新客户，谁可能会在未来 3 个月内每月花费 X 美元以上？
- 新服务：对于近 3 个月一直保持活跃的客户，谁会注册 2 周以前推出的一项新服务？

常见用例



数字化

- 提高在线活跃度
- 从分支过渡到在线
- 重新定位目标网站访问者
- 首次购买者，多次购买者，长期购买者
- 出手阔绰者

问题示例

- 对于已经 2 个月处于不活跃状态的在线客户，如果向其致电/给予 X 奖励，谁会在未来 7 天内恢复活跃状态？
- 对于所有采用离线交易方式（分支，呼叫中心）的客户，如果下周向其提供 X 促销，谁会开始使用数字交易方式？
- 对于最近 7 天的新增的在线用户，谁的支付总可能会在前 3 个月内达到 X 美元以上？
- 对于过去 3 个月内通过数字渠道注册购买新产品的所有客户，如果向其提供 X 促销，谁会在下个月内注册购买第二种产品？
- 如果被重新设定为目标，近 3 个月的哪些网站访问者有可能会扩大在线保险范围？

ENDOR

常见用例



销售效率

- 促销效率
- 新产品
- 客户忠诚度和推荐
- 针对性促销

问题示例

- 促销效率：对于活跃客户，谁会在收到文本消息/电话/Y 促销后注册高级服务 X？
- 针对性促销：对于所有活跃客户，谁会在未来 30 天内前往目的地 X？
- 推荐：对于所有活跃客户，如果给予 X 奖励，谁有可能会向朋友推荐？
- 新产品：对于所有活跃客户，如果下周进行接触，谁有可能会使用新数字服务？

常见用例



长期购买者

- 续保
- 终止
- 回头客

问题示例

- 终止：对于 X 保险的所有新客户，谁可能会在前 90 天内终止合同？
- 回头客：对于要在生效之日前取消的所有客户，谁会重新签约？
- 续保：对于近 60 天签有新保单的所有客户，谁会在生效之日前取消？

ENDOR

ENDOR

应用实例：零售和电子商务

ENDOR

常见用例

你每天如何进行预测？决定？提升绩效？

你需要回答以下问题



销售

- 下个月谁会购买 X？
- 谁有可能购买一周前推出的产品？
- 谁会转而购买高级产品？
- 谁会成为出手阔绰者？
- 哪些 X 顾客会购买 Y？



营销

- 谁会对促销作出回应？
- 需要维持哪些活动？
- 哪些客户会进行网购？
- 谁会使用我们的移动应用？
- 谁会将活跃度提升到 X 美元以上？



电子商务

- 哪些网络跟踪器 ID 会转化？
- 谁会对 X 促销作出回应？
- 谁会成为首次购买者？
- 谁会提升活跃度？
- 如果给予 X 促销，哪些非活跃客户会成为活跃客户？



店铺/产品

- 哪些店铺的 Y 产品销售额会达到 X 美元？
- 哪些店铺会表现不佳？
- 哪些产品下月销售额会增加 20% 以上？



战略

- 预警系统
- 何处会出现客户流失？
- X 产品在何处会获得助推力？

常见用例



销售更多

- 购买倾向
- 交叉销售
- 向上销售
- 未来的出手阔绰者
- 新产品 - 早期采用者

问题示例

- 对于过去 3 个月一直保持活跃的客户，未来 30 天内谁有可能购买 352553？
- 对于过去 30 天购买产品 219 的所有客户，未来 90 天内谁有可能购买产品 890？
- 对于近 3 个月来一直保持活跃的 32236 产品客户，未来 7 天内谁有可能会升级到高级产品？
- 对于近 6 个月一直保持活跃的客户，谁可能会购买我们 3 个月前推出的新产品？
- 对于近 30 天内所有的新客户，谁的总购买金额会在前 3 个月内达到 1 万美元以上？

ENDOR

常见用例



销售效率

- 更合理的细分
- 促销效率
- 跨渠道粘性
- 客户忠诚度

问题示例

- 对于近 6 个月来支出额处于前 10% 的客户，谁会在 2 个月内将支出额至少减少 20%？
- 对于近 6 个月来支出额处于后 20% 的客户，谁会在 3 个月内将活跃度至少提高 30%？
- 所有顾客当中，如果下周给予 20% 的赠券折扣，谁会购买产品 6347？
- 所有顾客当中，如果下周向其发送文本消息，谁会购买产品 6347？
- 对于近 90 天内进行网购的所有客户，谁有可能会在下个月使用移动应用程序？

ENDOR

常见用例



提高活跃度

- 高效的**用户获取**（重新设定目标）
- **首次购买者**
- **多次购买者**
- **提高活跃度**

问题示例

- 对于所有网站访问者（网络跟踪器），如果下周通过谷歌重新设为目标，则谁有可能会成为客户？
- 对于所有网站访问者（网络跟踪器），谁前**3个月**的购买总额会超过**1万美元**？
- 对于最近**6个月**当中状态不活跃的所有客户，谁会在下个月成为**FTB**？
- 对于最近**7天**当中第一次下购物订单的所有客户，谁会在**60天**内下**2个**新订单？

其他...



产品/店铺

- 哪些服务下个月增长量会超过**20%**？
- 哪些服务在未来**3个月**内减少量可能会超过**50%**？
- 哪些店铺未来**3个月**内**X类**产品的销量会超过**50%**？



经营

- 谁明天会就**X**联系呼叫中心/技术支持？
- 谁明天会在分支**X**内进行交易？



策略

- **X品牌**下个月何处会出现客户流失（**顶风**）？
- **Y品牌**何处会获得助力？

不存在“预定菜单”

任何商业用户询问业务增长需要什么

附录 C - Endor.coin 的预定预测实例

Endor.coin 的宏伟蓝图主要面向区块链所影响的保险、银行、电子商务或医疗保健行业；首先，我们为加密货币分析结果提供了一个史无前例的预测平台，支持加密持有者寻找潜在交易客户。经过区块链证明的分散式基础架构允许任何人在任何数据源上测试任何假设，无需透露其实际交易策略。*Endor.coin* 可帮助用户在任何其他现有技术之前预见未来。

下面是 *Endor.coin* 平台推出后所支持的预定区块链预测的一个实例。根据群体要求，新用例会逐步增加，仍然是基于预测请求（*RFP*）。

加密货币地址预测：这些预测会得到一个预定地址列表（即：“上个月至少有一次外向交易的比特币地址”），并根据它们最近一段时间与预定行为的相似度进行排序（即：“上周至少收到 0.1 个比特币的地址”）。所产生的列表顶部将包含最类似于预定行为的地址（因此将来在统计学上更可能会显示相同行为），而结果列表底部的地址与预定行为相似度最小。

- **活跃地址：**上个月至少活跃一次的所有地址中最类似于最近交易次数显著增加的地址。
- **交易活跃地址：**上个月至少活跃一次的所有地址中最类似于上个月交易 10 个比特币的地址。
- **变成不活跃地址：**上个月至少活跃一次的所有地址中最类似于最近交易量剧减 50% 的地址。

代币预测：这些预测得到一个预定代币列表（即：“上个月交易额至少为 1000 万美元的代币”），并根据其与最近一段时间预定行为的行为相似度进行排序（即：“平均月交易量增加 2 倍的代币”）。所产生的列表顶部将包含最类似于预定行为的代币（因此将来在统计学上更可能会显示相同行为），而与预定行为相似度最低的代币则置于列表底部。

- **可获利代币：**最近一个月交易额至少 100 万美元的所有代币中最类似于上月价格相对于比特币增长超过 50%的代币。
- **不可获利代币：**最近一个月交易额至少 100 万美元的所有代币中最类似于上月价格相对于比特币下跌超过 50%的代币。
- **不稳定代币：**上月交易额至少 500 万美元的所有代币中最类似于上个月 10 种最不稳定代币的代币。
- **稳定代币：**上月交易额至少 500 万美元的所有代币中最类似于上个月 10 种最稳定代币的代币。
- **交易量增加的代币：**上月交易额至少 2000 万美元的所有代币中最类似于上个月交易量翻番代币的代币。
- **交易量减少的代币：**上月交易额至少 2000 万美元的所有代币中最类似于上个月交易量减少 50%代币的代币。

附录 D - 知识领域级 API 访问

本附录包含“知识领域”数据结构（*Endor.coin* 协议的基本构件）的规范，完整描述了向其提供通路的类别，用于实现未来分析引擎，与 *Endor.coin* 网络连接。

使用说明： 集群对象包含 3 个实体：

- 稀疏矩阵 M ，维度为（|可搜索对象| × |行为集群|），代表可搜索对象和行为集群之间的连通性， $M_{i,j}$ = 指标的 1 iff 可搜索对象 $i \in$ 集群 j 。可搜索对象是指代币、钱包地址、位置、电话号码或数据中包含的任何其他类型的对象，用作预测依据。
- 阵列 A_M 将每个可搜索对象 SO 映射到稀疏矩阵 M 中的一个指标上。
- 包含其他各种集群属性的数据框 D_M ，由每个预测引擎定义和计算。例如，此种属性可以是集群大小、内部连通性与外部连通性之比以及预测引擎用于生成该集群的内部模块等。

如要提取和构建供使用的集群对象，则需提供下列文件：

- 构建稀疏矩阵 M ：
newFormat_unified_blocks_connectivity_matlFile.spmat
newFormat_unified_blocks_connectivity_matlFile.spmat
newFormat_unified_blocks_connectivity_matlFile.spmat
- 构建映射阵列 A_M ：
newFormat_TranslationTabletmp_numbers.spmathlp
- 构建集群属性数据框 D_M ：
newFormat_unified_blocks_blk_data.mat

上述文件会置于具体路径中，由下列“集群-提取器”引用为：
<cluster_files_path>

使用示例：将相关的 5 个集群文件放到<cluster_files_path>中之后，集群可使用下列代码提取：

```
>>> extractor = clustersExtractor(cluster_files_path)
>>> pop_to_clusters_map, clusters_props, translation_pop = extractor.retrieveClusters()
```

本例中，“pop_to_cluster_map”指向稀疏矩阵 M 中。“clusters_props”指向包含集群属性的数据框 D_M 。“translation_pop”指向阵列将每个可搜索对象映射至其在稀疏矩阵 A_M 中的指标。

集群提取器类别：下面是知识领域 API 的完整描述：

```
import os
from cStringIO import StringIO
import struct
import tempfile
import numpy as np
import re
import scipy.io as scio
import pandas as pd
from tempfile import NamedTemporaryFile
from scipy.sparse import csr_matrix

class clustersExtractor(object):
    def __init__(self, cluster_files_path):
        self.__path = cluster_files_path
        self.__mat_init_name = 'blk_data'
        self.__mat_fields = {
            'src': 2,
            'type': 1,
            'blk_type': 1,
            'field': 1,
            'fieldby': 1,
            'N': 1,
            'WN': 1,
            'deg': 1,
            'thrs': 2,
            'SUB_CLUSTERS_FILE': 1,
            'percInternal': 1
        }
    def get_path(self):
        return self.__path
    def retrieveClusters(self):
        pop_to_clusters_map = self.__build_pop_to_clusters_map()
        clusters_props = self.__build_mat_props_df()
```

```

        translation_pop = self._get_translation_pop()
        return pop_to_clusters_map, clusters_props, translation_pop

def _build_pop_to_clusters_map(self):
    dims_for_mat = self._get_dimensions_for_matrix()
    indices = self._get_ir_list()
    indptr = self._get_je_list()
    pop_cluster_map = self._build_pop_clust_matrix(dims_for_mat,
                                                    indices, indptr)

    return pop_cluster_map

def _get_dimensions_for_matrix(self):
    dims_file_name = "MtFile.spmat"
    all_files = self._get_files_in_path(dims_file_name)
    full_path = all_files[0]
    a = self._open(full_path)
    f = StringIO(a.read())
    mat_sizes = struct.unpack('<2IQ', f.read(16))
    return {'n_rows': mat_sizes[0], 'n_cols': mat_sizes[1],
            'nnz': mat_sizes[2]}

def _get_ir_list(self):
    ir_file_name = "IrFile.spmat"
    all_files = self._get_files_in_path(ir_file_name)

    full_path = all_files[0]
    remote_file = self._open(full_path)
    data = remote_file.read(10 * 1024 * 1024)
    local_temp_path = os.path.join(tempfile.mkdtemp(),
                                    ir_file_name)

    with open(local_temp_path, 'w') as f:
        while data != '':
            f.write(data)
            data = remote_file.read(10 * 1024 * 1024)

    ir = np.fromfile(local_temp_path, dtype=np.int32)
    os.unlink(local_temp_path)
    return ir

def _get_je_list(self):
    je_file_name = "JeFile.spmat"

    all_files = self._get_files_in_path(je_file_name)

    full_path = all_files[0]
    data = self._open(full_path)
    local_temp_path = os.path.join(tempfile.mkdtemp(), je_file_name)
    with open(local_temp_path, 'w') as f:
        f.write(data.read())

```

```

        je = np.fromfile(local_temp_path, dtype=np.int64)
        os.unlink(local_temp_path)
        return je

def __build_pop_clust_matrix(self, mat_dims, indices, indptr):

    nrows = mat_dims['n_rows']
    ncols = mat_dims['n_cols']
    nnz = mat_dims['nnz']

    data = np.ones(nnz)
    try:
        mat = csr_matrix((data, indices, indptr),
                        shape=(ncols, nrows))
    except Exception as e:
        msg = """Couldn't build population to
        cluster match due to: %s, aborting.""" % str(e)

        raise ValueError(msg)
    return mat

def __build_mat_props_df(self):
    mat_names_list = self.__get_mat_files_names()

    if len(mat_names_list) > 1:
        prop_names = 'temp_'
        mat_files = [scio.loadmat(StringIO(self.__open(mat_name).read()))
                    for mat_name in mat_names_list]
        props_df = self.__build_multiple_clusters_props(mat_files, prop_names)
    else:
        prop_names = 'blk_data'
        mat_file = scio.loadmat(StringIO(self.__open(mat_names_list[0]).read()))
        props_df = self.__build_single_clusters_props(mat_file, prop_names)
    return props_df

def __build_multiple_clusters_props(self, mat_files, prop_names):

    all_props_df = pd.DataFrame()
    for mat_file in mat_files:
        df = self.__build_single_clusters_props(mat_file, prop_names)
        all_props_df = all_props_df.append(df)
    all_props_df.index = range(len(all_props_df.index))
    return all_props_df

def __build_single_clusters_props(self, mat_file, prop_names):
    try:
        clusters_props = mat_file[prop_names]
    except Exception:
        msg = """Field %s doesn't exist in mat file,
        but expected. Cannot continue""" % prop_names
        raise ValueError(msg)

```

```

cluster_props = {}
for prop, counts in self.__mat_fields.iteritems():
    try:
        mat_values = clusters_props[prop][0][0]
    except (KeyError, IndexError):
        ValueError("""Field %s doesn't exist in mat file,
please remove it from config file.""" % prop)
        break
    if counts == 1:
        cluster_props[prop] = mat_values.flatten()
    else:
        for i in np.arange(counts):
            cluster_props[prop + '_' + str(i)] = mat_values[:, i]
cluster_prop_df = pd.DataFrame(cluster_props)
# noinspection PyUnresolvedReferences
cluster_prop_df.index.names = ['cluster']
wn_sizes = cluster_prop_df['WN'].astype(float)
n_sizes = cluster_prop_df['N'].astype(float)
cluster_prop_df['W_pcentg'] = wn_sizes / n_sizes
del mat_file
return cluster_prop_df

def __get_translation_pop(self):
    spmat_help_name = ".spmathlp"
    all_files = self.__get_files_in_path(spmat_help_name)

    full_path = all_files[0]

    f_spmatlp_data = self.__open(full_path).read()
    local_temp_path = os.path.join(tempfile.mkdtemp(), spmat_help_name)
    with open(local_temp_path, 'w') as f:
        f.write(f_spmatlp_data)
    with open(local_temp_path, 'rb') as f_spmatlp:
        num = struct.unpack('<Q', f_spmatlp.read(8))
        # noinspection PyTypeChecker
        ids = np.fromfile(f_spmatlp, dtype=np.double)

        if num[0] != len(ids):
            msg = """translating ids went wrong. Found %d ids,
where expected %d ids, aborting""" % (len(ids), num[0])
            self.__logger.error(msg)
            raise ValueError(msg)

    os.unlink(local_temp_path)

    return ids

def __get_files_in_path(self, name):
    all_files_in_dir = list(self.__list_dir())

```

```

    relevant_files = [file_name for file_name in
                       all_files_in_dir if name in file_name]
    return relevant_files

def __open(self, path):
    real_path = os.path.expanduser(path)
    if not os.path.isfile(real_path):
        raise LookupError("{} does not exist".format(real_path))

    return open(real_path, 'rb')

def __get_mat_files_names(self):
    mat_names_list = self.__get_files_in_path(self.__mat_init_name)
    mat_names_list.sort(key=lambda x: int(re.search(r'\d+', x).group()))
    return mat_names_list

def __list_dir(self):
    return [os.path.join(self.__path, f) for f in
            os.listdir(os.path.expanduser(self.__path))]

```

参考文献

- [1] 维基百科 - 社会物理学 (2017)。
URL https://en.wikipedia.org/wiki/Social_physics
- [2] W. Pan, Y. Altshuler, A. Pentland; 解码金融交易网络中的社会影响和群体智慧: 隐私、安全、风险和信任 (PASSAT); 2012 年国际社会计算会议 (SocialCom), 电气与电子工程师协会, 2012, pp. 203-209。
- [3] Y.-Y. Liu, J. C. Nacher, T. Ochiai, M. Martino, Y. Altshuler; 在线金融交易前景理论; 公共科学图书馆期刊, 9(10) (2014) e109458。
- [4] Y. Altshuler, W. Pan, A. Pentland; 基于社会扩散模型的趋势预测; arXiv.org, 2011。
- [5] P. M. Krafft, J. Zheng, W. Pan, N. Della Penna, Y. Altshuler, E. Shmueli, J. B. Tenenbaum, A. Pentland, 分布式贝叶斯推理中的人类集体智慧; arXiv 预印本; arXiv: 1608.01987。
- [6] Y. Altshuler, A. S. Pentland, G. Gordon; 社会计算、行为-文化建模和预测中的社会行为偏差和知识管理优化; 施普林格, 2015, pp. 258-263。
- [7] Y. Altshuler, A. Pentland; 最佳性能网络调节用方法和装置; 美国专利 8,914,505 (2014 年 12 月 16 日)。
URL <https://www.google.com/patents/US8914505>
- [8] W. Pan, Y. Altshuler, A. Pentland, N. Aharony; 网络中行为预测与修正的方法和装置; 美国专利 9,098,798 (2015 年 8 月 4 日)。
URL <https://www.google.com/patents/US9098798>
- [9] 调整社交网络, 获得群体智慧 (麻省理工学院媒体实验室网站) (2017 年)。
URL <https://www.media.mit.edu/research/highlights/tuning-social-networks-gain-wisdom-crowd>
- [10] 市场洞察: 重视推特对市场的影响 (《金融时报》) (2017 年)。
URL http://web.media.mit.edu/~yanival/Markets_Insight.htm
- [11] 回音室之外 (《哈佛商业评论》) (2017 年)。
URL <https://hbr.org/2013/11/beyond-the-echo-chamber>
- [12] 预测分析再思考 (FirstMark's Data Driven) (2017)。
URL <http://firstmarkcap.com/insights/rethinking-predictive-analytics/>
- [13] 麻省理工学院用 100 万美元验证社交媒体是否能为投资者赚钱 (2013 年)。
URL <https://tinyurl.com/MIT-1M-USD>
- [14] Endor 发明“谷歌预测分析” (2017 年)。
URL <http://news.mit.edu/2017/endor-inventing-google-predictive-analytics-1220>
- [15] Endor 领导投资者 - 创新的力量 (2014 年)。
URL <http://www.innovationendeavors.com>

- [16] A. Boehme, Y. Altshuler ; 利用社会物理学预测消费者行为 ; 网络科学 (NetSci) , 2017 年。
- [17] 万事达卡携 5 家新创业公司加入创业路径加速计划 (2016 年) 。
URL <https://tinyurl.com/MasterCard-Endor>
- [18] Endor - Finnovate Fall 2017 (2017 年) 。
URL <https://www.youtube.com/watch?v=69rUQlog-qA>
- [19] Endor - 高德纳杰出供应商 (2017 年) 。
URL <https://www.gartner.com/doc/3727117>
- [20] Endor - 经世界经济论坛认定的技术先锋 (2017 年) 。
URL <http://widgets.weforum.org/techpioneers-2017/>
- [21] 美国国防部高级研究计划局网络挑战 (2011 年) 。
URL <http://archive.darpa.mil/networkchallenge/>
- [22] 2012 年麦肯锡奖获奖者 (2012 年) 。
URL <https://hbr.org/2013/04/the-2012-mckinsey-award-winners>
- [23] 谷歌学者 - Alex “Sandy” Pentland 教授 (2017 年) 。
URL <https://scholar.google.com/citations?user=P4nfoKYAAAAJ&hl=en>
- [24] Tim Oreilly : 全球最具影响力的 7 位数据科学家 (2017 年) 。
URL <http://www.forbes.com/pictures/lmm45emkh/6-alex-sandy-pentland-professor-mit/>
- [25] Y. Altshuler, A. Pentland, A. M. Bruckstein ; 搜索中的集群和网络智能 ; 施普林格 , 2017 年。
- [26] Y. Altshuler, Y. Elovici, A. B. Cremers, N. Aharony, A. Pentland ; 社交网络安全与隐私 ; 施普林格科学商业媒体 , 2012 年。
- [27] H. Shrobe, D. L. Shrier, A. Pentland ; 网络安全新型解决方案 ; 麻省理工学院出版社 , 2018 年。
- [28] J. Clippinger, D. Bollier ; 从比特币到火人节及其他 : 数字化社会的身份验证与自主性探索 ; ID3 and Off The Common Books , 2014 年。
- [29] T. Hardjono, D. Shrier, A. Pentland ; 信任::数据 : 一种新型身份识别与数据共享框架 , 2016 年。
- [30] A. Pentland, T. Heibeck ; 诚实信号 : 它们如何塑造我们的世界 ; 麻省理工学院出版社 , 2010 年。
- [31] A. Pentland ; 社会物理学 : 好主意如何扩散 - 新科学之教训 ; 企鹅出版社 , 2014 年。
- [32] D. Shrier, A. Pentland ; 前沿金融科技 : 未来商业从区块链和数字银行到预测市场等的转变 ; Visionary Future 出版社 , 2016 年。
- [33] Endor.coin 协议 GIT (2017 年) 。
URL <https://github.com/orgs/EndorCoin>
- [34] 我们为什么需要基于区块链的人工智能 - 即使你甚至都没听说过 (2017 年) 。
URL <https://tinyurl.com/Blockchain-Blockchain-based-AI>
- [35] Y. Altshuler, N. Aharony, A. Pentland, Y. Elovici, M. Cebrian ; 偷取现实 : 当罪犯变成数据科学家 (或相反) ; 智能系统 , 电气与电子工程师协会 26 (6) (2011 年) 22-30 。 doi : 10.1109/MIS.2011.78 。
- [36] M. Ulieru ; 区块链是什么 ? 它如何真正地改变这个世界 ? 世界经济论坛。
- [37] 技术如何帮助我们修复破败的金融体系 (2017 年) 。
URL <https://tinyurl.com/technology-fixing-our-financia>

- [38] J. Pieprzyk, T. Hardjono, J. Seberry ; 计算机安全基本原理 ; 施普林格科学商业媒体 , 2013 年。
- [39] T. Hardjono, L. R. Dondeti; 组播与群组安全; 阿泰克, 2003 年。
- [40] T. Hardjono, L. R. Dondeti;
无线局域网与城域网安全 (阿泰克计算机安全) ; Artech House, , Inc. , 2005 年。
- [41] J. Seberry, T. Hardjono; 关于印尼语和马来语密码分析, 1989 年。
- [42] S. G. Ong, J. Seberry, T. Hardjono, A. D. F. Academy.关于普通话 (拼音) 密码分析 , 1991 年。

免责声明: 该白皮书仅供讨论之用。
Endor.coin 不保证该白皮书所得结论的准确性。

版权所有 © 2018 Endor.coin。