

ENDOR.COIN (エンドア・コイン) プ ロトコル

誰もが利用できる人工知能予測を作り出す

提供 : Endor Ltd. (エンドア社)

2018年2月18日

誰でも利用できる人工知能予測を作り出す

要約

Endor.coin (エンドア・コイン) は、人工知能データ分析が誰もが利用しやすく、トラストレスで、検閲耐性があり、全ての人の役立つように民主化し、新たな予測分析を生み出します。

Endor.coin プロトコルは、ビジネスユーザーや一般の仮想通貨トークン所有者でも、複雑な質問に対する高品質の予測結果を数分で得ることができる世界初の自動セルフサービス型予測プラットフォームを提供します。当社は、現在はフォーチュン誌上位500社に独占されているデータ科学の分野の民主化を目的としています。**Endor.coin** は、プロジェクトチームのメンバーである **Alex Pentland** (アレックス・ペントランド) 教授と **Yaniv Altshuler** (ヤニフ・アルトシュラー) 博士によって **MIT** で開発された新たな科学である **社会物理学** に基づいています。

- **予測カタログの絶え間ない拡張**：**Endor.coin** は、事前に定義されたトークンに関する予測によって起動します。(例：ボリュームの増加やボラティリティの減少が予測されるトークンなど) これらの予測は、プラットフォーム専用の **EDR** トークンを使用して購入することができます。**Endor.coin** は、ユーザーが **予測リクエスト (RFP)** を送信すると、実行可能で購入に役立つ新しいタイプの予測を提案する事により、予測の選択肢を広げます。
- **上級ユーザー向け Do-it-yourself (DIY) API**：技術に精通したユーザーや専門家には、セルフサービス・インターフェースが提供され、代わりに **Endor.coin** プラットフォームは「類似した」予測を自動的に生成します。
- **非公開/公開データの自動融合**：**Endor.coin** の商業的顧客（銀行、小売業者、保険会社など）は、独自のデータストリームをプラットフォームに簡単に統合することができます。私的データと公的データが融合した高品質な予測洞察を得ることができます。**社会物理学** データの使用は自動化されており簡単です。クリーニングやデータの準備は必要ありません。
- **データ・プライバシー**：顧客側で完全に暗号化されたデータを使用するという **社会物理学** 独自の性能により保証されています。
- **人々による、人々の為の予測**：使用される **物理学** データは一括処理され、ユーザーは必要な個人化コンポーネントのコストだけを支払います。この為、現在は巨大テクノロジー企業に独占されている予測能力の「**99%**」を、その **1%** のコストで提供することができます。

Endor.coin は、**Endor** の技術的基盤に根ざしています。最近、世界経済フォーラムにテクノロジーのパイオニアとして表彰され、ガートナー社の認定したクールベンダー

(Cool Vendor) である Endor 社は、*Innovation Endeavors* 社などの主要投資家の経済的支援を受けた MIT のスピンオフ企業であり、コカコーラ社、ウォルマート社、マスターカード社などのフォーチュン誌上位 500 位以内の企業とのプロジェクトに取り組んでいます。

目次

1	バリュープロポジション	3
2	序文.....	6
	2.1 非効率で厄介な市場	6
	2.2 民主化には分散化とデータ分離が必要.....	6
	2.3 技術的 ギャップ	7
	2.4 <i>Endor.coin</i> プロトコル — プロジェクト概要	8
3	予測科学の民主化	10
	3.1 社会物理学 – <i>MIT</i> からの新しい科学.....	10
	3.2 第 ¹ 段階： Endor.com – 企業のための自動予測エンジン	13
3.3	第 ² 段階： <i>Endor.coin</i> プロトコル – 大衆のためのデータ科学	17
	3.3.1 定義.....	18
	3.3.2 データ提供者	18
	3.3.3 予測 エンジン	19
	3.3.4 事前に定義された予測と予測リクエスト (RFP)	19
	3.3.5 私的データ分析とセルフサービス API.....	19
	3.4 ロードマップ.....	19
4	トラストレス、検閲耐性、説明責任	22
	4.1 人工知能の説明責任と真正性.....	22
	4.2 分散型プロトコルを通じた検閲耐性.....	23
	4.3 プライバシー保護データ分析：暗号化された データの使用.....	24
	4.4 ネットワーク 効果.....	25
5	エコシステムの有効化	28

5.1	公開データ・プロバイダー	28
5.2	学術研究グループ	29
5.3	触媒 - アプリケーション 開発者	29
5.4	データ 主権	30
6	トークンの実装	31
6.1	ブロック チェーン構造	31
6.2	スマートコントラクト	32
6.3	開始時の Endor の役割	34
6.4	トークンの特権と トークン経済	35
6.5	収益の使途	36
7	技術的優位性と差別化	38
7.1	MIT フェーンラスからの科学革命	38
7.2	現実の製品。実証済みの 技術	38
7.3	ユーザーに対するユーザビリティと価値、トークン所有者のための価値	39
8	チーム	40
8.1	主要チーム メンバー	40
8.2	顧問	43
	付録 A : 社会物理学の説明	46
	付録 B : Endor の一般的な使用例 - 企業のユースケース	82
	付録 C : Endor.coin の事前定義済み予測の例	94
	付録 D : 知識分野クラス API	95
	参考文献	102

第1章

価値ある提案

消費者の行動を迅速に理解し、予測し、影響を与える能力は、ビジネスを競争よりも不公平なものにする可能性があります。スマートなビジネスリーダーには、顧客の行動に影響を与えてビジネスパフォーマンスを向上させるための多くのアイデアがあります。それらを実行するには、次のような質問に答える必要があります。

- どの企業が当社のトップ顧客ですか？また、トップ顧客をより多くを獲得するには？
- 誰がこの新製品を試してくれそうですか？
- プロモーションへの依存をどのように減らすことができますか？
- 次の店舗はどこにオープンするべきですか？
- 誰が来月から製品 A から B に切り替えますか？

これらの質問に答えるために、組織はデータ科学や予測分析などの強力なツールに目を向けます。残念なことに、これらのツールを実装する為の現在のプロセスは、遅く、苦痛を伴い、高価です。

- 必要な「ユニコーン」：よく訓練された、高価で、稀なデータ科学者と博士号
- 4~6回の反復が必要で、それぞれ数日~数週間かかります
- 新しいビジネス上の質問への回答には、数週間で新しいモデルを構築する必要があります
- 製品と動作が変わると、モデルは崩壊します

MIT の提供する新しい **社会物理学技術**である **Endor.coin** ブロックチェーン・ベース・プロトコルは、数分で予測的質問に高品質の結果を提供する初めての分散型トラストレス

検閲耐性行動予測プラットフォームです。コーディング、データクリーニング、専門家チームは不要です。

以下は、*Endor.coin* プロジェクトの主な側面です。

- **技術**：MIT 提供の *社会物理学技術*[1]、トレンド予測のための最高 10 倍の精度を提供するとともに、完全暗号化データから予測洞察を生成する能力（金融市場におけるトレンド予測に関する学術研究[2-6]、関連特許[7、8]、追加レビュー[9-11]参照）。
- **革新**：短期および中期の行動パターン（数日から数週間）の自動モデリングに焦点を当て、現在利用可能な他の技術によって観測できるようになる前にそうした信号を検出します（ファーストマーク社「データドリブン・ニューヨーク」での Altshuler 博士の講演[12]、財政投資の大規模リアルタイム分析[13]参照）。
- **業界の検証**：*Endor.coin* は MIT のスピンオフ企業[14]であり、大手投資家[15]による財務的支援を受けており、コカコーラ社 [16]、マスターカード社 [17]、ウォルマート社などのフォーチュン誌上位 500 社とのプロジェクトに取り組む *Endor.com*（エンドア・ドット・コム）によって開発された技術に基づいています。*Finnovate 2017*[18]で披露された *Endor* の製品をご覧ください。
- **受賞と表彰**：*Endor* 社は、ガートナー社認定クールベンダー [19]であり、*世界経済フォーラム* において「テクノロジーのパイオニア」[20]として表彰されました。MIT のプロジェクトチームが行った研究により、名誉ある *DARPA* ネットワーク・チャレンジ[21]、マッキンゼー賞[22]など数々の賞を受賞しました。
- **チーム**：MIT コンピュータ科学・人工知能研究所、MIT メディア研究所、MIT スローン経営学大学院からブロックチェーン、デジタル・バンキング技術、予測分析の世界的な専門家が集結したチームに率いられています。米国国立科学技術アカデミー会員である Pentland 教授（共同設立者）は、世界で最も賞賛されている科学者の一人であり[23]、最近、*フォーブス*誌によって「世界の最も強力なデータ科学者 7 人」に挙げられました[24]。チームは、数百の科学論文、数十の商用特許、ブロックチェーン、機械知能、データの機密性に特化した 8 冊の書籍を共同で出版しています[25-32]。
- **予測分析の改革**：人工知能と機械学習の能力が裕福な企業に独占されていた年月を経て、*Endor.coin* は個人や中小企業に僅かなコストで優れた技術へのアクセスを提供します。予測的洞察は、データ所有者がデータの機密性を管理しながら、低コストで提供される寄稿データの収集分析に基づいています。*Endor.coin* プロトコルは、新しいデータソースと新しい予測エンジンの統合を可能にし、二重ネットワーク効果を生み出します。参加者が増える程、1 予測当たりのコストは削減され、予測精度は向上します。
- **トラストレス、完全分散化、説明責任、検閲耐性**：*Endor.coin* プロトコルは完全に分散化されており、予測の結果に対する完全な説明責任を提供します。これにより、予測中の操作やバイアスを防止できます。さらに、プロトコルの分散的でオープンな性質は、あらゆる当局からの検閲を防止し、あらゆる予測のサポートを可能にします。

第 2 章 序文

2.1 非効率で厄介な市場

私たちの世界には、データを集め、所有し、慎重に維持する組織がたくさんあります。データを処理し予測モデルを構築できるデータ科学者と機械学習の専門家がいます。未来を予測したいという人もたくさんいます（大企業の高級幹部から大企業の中規模マーケティング・マネージャー、プロダクト・マネージャー、飛行機のチケットを購入するタイミングを決定しようとする人物まで）。現在は、後者が質問の予測を得るためには、**3**つのエンティティが全て同じ組織内に共存する必要があります。これは、予測の**99%**が大企業のステークホルダーによって、またステークホルダーのために生成されていることを意味します。さらに、現在このプロセスは、データ科学者が少なくコストが高額なため、非常に高額のコストがかかります。また、予測生成プロセスでは、**1**つのプロジェクトごとに数カ月のハードワークが必要になることがあります。これは、予測に関心のある人にとっては、非常に高い参加への障壁（および値札）となります。

2.2 民主化には分散化とデータ分離が必要

人工知能と予測分析が民主化され、「一般の人」がアクセスできるようにするためには、次の要件を満たす新しいパラダイムの開発が必要です。

分離： 大企業や研究所は別として、データのキュレーション、データ科学、セマンティック重視の質問の生成を組み合わせた質の高い活動を継続できる商業、学術、非営利団体は殆どありません。個人、NGO、中小企業は、通常、慎重に維持されたデータの所有（または生産）、強力な（そして高額な）データ科学者の雇用、「正しい質問をする」専門家となることといった側面の**1**つにしか焦点を当てることができません。したがって、フォーチュン誌上位**500**社の外で障壁を取り除き、予測にアクセスできるようにするためには、これら**3**つの基本的要素を本質的に分離する必要があります。真に民主化された予測プロトコルは、データ提供者が自由に（公的または私的にマークされた）データを寄稿することができ、技術者はエンドユーザーがこれらのデータソースに基づいた予測を使いやすいように（シームレスに接続してプロトコルと統合した）AIと予測エンジンを提供することができます。

説明責任： 巨大テクノロジー企業のデータ科学部門では、説明責任は「誰もが同じボスのために働いている」という事実によって自動的に提供されるため、重要な機能として必須ではありません（例：ビジネスインテリジェンス、マーケティング予測などを担当する経営幹部）。データ、インテリジェンス、計算が常時臨機応変に借用されている民主的なプラットフォームでは、説明責任が不可欠になり、「フェアプレー」を保証するためには、短期的な収益を促進するのではなく、長期的にすべてのステークホルダーの設計クライテリアを調整することが重要です。

分散化： 分散型予測フレームワークには、エンジニアリング関連のものと、検閲耐性とバイアス防止を中心としたものがあります。多くの例で示されているように、分散化ソリューションは、拡大と拡張が容易な傾向があります。より多くのデータソース、計算リソース、さまざまなタイプの予測エンジンを追加することで、分散型ソリューションのメリットが大きくなります。加えて、分散型アーキテクチャは、これらの予測が、偏りのある結果の生成または貨幣資産の調停によって実行される明示的または暗黙的な検閲を受けないことを保証する唯一のものであります。さらに、参加者数（データ提供者、予測エンジン、予測消費者）の増加に伴い、予測精度を絶えず向上させながら、ユーザーあたりのコストを押し下げる新興ネットワーク効果の鍵となるのは、効率的な分散化です。

2.3 技術的ギャップ

データ科学の民主化のメリットは以前から明らかになっていますが、残念ながら、現実にはこのようなフレームワークを実行することは難しいことです。主な理由は技術的なものです。今日の一般的な科学は単純に「ジェネリック分散型行動予測」のパラダイムをサポートできていません……。ニューラルネットワーク（またはディープラーニング）、遺伝的プログラミング、決定フォレスト、**SVM** など今日存在するテクノロジーは全て、「現実の」機械学習作業を開始する前に大量のデータサニタイジング、処理、理解を必要とします。これが、業界が現在直面しているボトルネックである熟練したデータ科学者の給与の上昇、人手不足の原因です。このトピックに関する詳細はセクション 3.1 と付録 A にあります。

あらゆるデータを自動的にダイジェスト化でき、専門家でも、専門家でなくても予測的質問をすることができる科学的な突破口がなければ、業界は既存のパラダイムに拘束され、新たなデータ科学者を採用するペースが大きくボトルネックとなり、数十万ドルの給与を支払っていました。

2.4 Endor.coin プロトコル — プロジェクトの概要

これらの限界を超越するためには、新しい科学を開発しなければなりません。Endor.coin 創設者の Yaniv Altshuler 博士と Alex 「サンディ」 Pentland 教授によって開発された社会物理学は、人間の群衆の行動を効率的にモデル化する数学理論です。行動データソースに表される一連の数学的方程式を通じて、社会物理学理論は、任意の行動データソースを行動クラスターのセットに自動的に変換することを可能にします。これには、データのセマンティクス（または質問）のクリーニング、事前処理、または理解を必要としません。この行動クラスターの集合体は、「知識球体」として知られています。

Endor.coin プロトコルは、行動データが行動クラスターの標準的な表現で処理されているときに、データ科学の伝統的なプロセスは（最終的に）それを基本的なコンポーネントに分解することができ、分散化された方法で、それぞれを異なる実行者に割り当てます。以下は、Endor.coin プロトコルの基本的概要と主要コンポーネントです：

正統的データ表現： Endor.coin ネットワークに寄稿された全てのデータは、「知識球体」の正統的表現に変形します。これは、さまざまな予測エンジン（下記参照）によって実行でき、実行コストはエンジンによって支払われます。データがこの変換を経ると、そこから抽出されたさまざまな動作クラスターを他のタイプのデータのクラスターと一緒にバンドルすることができ、効率的かつ自動的な予測プロセスが得られます（詳細はセクション 3.1 および付録 A を参照）。

データ提供者の分離： データが正統的な「知識球体」表現に変形されるにつれて、データ提供者は分析の後の段階に積極的に参加する必要がなくなります。これにより、データ所有者は、（完全または一部の）データを Endor.coin ネットワークに統合し、データの品質の維持、一部にアクセスする者による制御、将来の価値からの財政的利益に焦点を当てたエコシステムの自立的ステークホルダーとしての役割を果たすことができます。

予測エンジンの分離： データ科学者の中で知られている秘密は、データ科学プロジェクトに費やされた時間の約 90% がデータサニタリーと事前処理に費やされていることです。革新的な側面のために、社会物理学は、まず自動的に段階を進み、さまざまな予測エンジンが最終的に様々なタイプのデータソースにシームレスに接続できるようになります。Endor.coin ネットワークに接続する予測エンジンのプロバイダに必要なものは、（エンジンによって定義されたように選択的に）データセットをダイジェスト化する機能として定義され、「知識球体」の形式で出力を提供する Endor.coin プロトコルをサポートすることだけです（付録 D の完全な仕様と API コードを参照）。

分散化実行： ブロックチェーンを使用。データ提供者は、（AWSに保存された）データを寄稿すると、**Endor.coin** プロトコルを使用できるようになります。行動クラスタの抽出は、様々な予測エンジンによって分散化された方法で行われます。**Endor.coin** スマートコントラクトに特定の予測（**EDR** 手数料）の発行をリクエストすることによって、エンドユーザーは質問することができます。各予測について、最良の行動クラスタは、**Endor.coin** 予測コード（プロジェクトの **GIT** アカウント [33]で自由にアクセス可能）によって選択されます。資金調停は、その予測のために選択されたクラスタに従って、データ寄稿者と予測エンジン間のスマートコントラクトによって処理されます。この調停によって、予測の品質が最適化され、偏りのない結果が提供されます。

データ主権： **Endor.coin** ネットワークに寄稿されたすべてのデータ要素には、「公開」または「非公開」のフラグを付けることができます（同じデータソースには、「公開」とマークされた列と「非公開」の列が混在している可能性があります）。公開データソースは、すべての予測エンジンからアクセス可能で、今後の予測のための行動クラスタのソースとなります。その代わりに、プロトコルにより選ばれた時に、提供者は **EDR** トークンを受け取ります。公開データ要素は、暗号化された方法にも関わらず、様々な予測エンジンによって利用することができます。しかし、そのようなソースから抽出されたクラスタは、これらの予測をリクエストしたユーザーがデータ・キーを提供する場合にのみ、予測ソースとして選択することができます。詳細は、セクション 5.4 をご参照ください。

説明責任と検閲耐性： ブロックチェーンを使用すると、予測は無期限に保存され、プラットフォームの評判、それに使用されたデータ、またはそれを分析した予測エンジンの推定に関心を持つ人は誰でも利用できます。さらに、**Endor.coin** プロトコルには、各予測の生成に使用される行動クラスタ（およびそれに対して支払われる資金の調停）を選択することを担当するオープンソースの予測コード[33]が含まれています。予測は偏りがなく、精度の最適化が保証されています。詳細はセクション 4.2 と 4.1 をご参照ください。

全ての人の為の予測効率： 最終的には、**Endor.coin** プロトコルにより、エンドユーザーは低コストで優れた予測を得ることができます。これは、プロセスの自動化（高額なフルタイムのデータ科学者を雇用する必要性を減らす）に基づいており、さまざまな種類の行動データソースをシームレスに融合することができる社会物理学の能力です。これは、コカコーラ社などの大規模な商業顧客であっても、**Endor.coin** ネットワークへの移行からすぐに利益を得ることができることを意味します。データを完全に非公開にすることで、独自のデータをシステムに提供された公開データと融合させ、このデータを独自に取得して分析する代償コストを大幅に削減して分析します。これはまた、**Endor.coin** ネットワークに参加するユーザーとデータ提供者

が増えればコストが下がるという、ポジティブなネットワーク効果ももたらしめます。
(詳細はセクション 4.4 を参照)。

第 3 章

予測科学の民主化

3.1 社会物理学 –MIT からの新しい科学

社会物理学 は、大規模なデータ分析と生物学の数学的法則を使用して人間の群衆の行動を理解する革命的な新しい科学であり、これにより **Endor** は従来の機械学習の限界を克服することができます。この新しい科学は、**Alex** 「サンディ」 **Pentland** 教授と **Yaniv Altshuler** 博士の研究によって、MIT で生まれました。独自の技術を使用し、**Endor** によって更に開発されました。その結果、行動が急速に変化し、進化しているときでも、あらゆる種類の人間の行動を説明し予測することができる強力なエンジンが得られました。

簡単に言えば、社会物理学は、人間の活動を表す全てのイベントデータ（通話記録、クレジットカードでの購入、タクシー乗車、ウェブ活動など）には、そのデータ中に特別な一連の人間活動パターンが確実に含まれているのです。これらの数学的不変性は、すべての人のデータタイプに共通であり、すべての人口統計において、出現する行動パターンを他の手法で観察する前に検出するフィルタとして機能します。

社会物理学の力の説明： 大手銀行のマーケティング部門が、近い将来、潜在的に融資を必要とする顧客に頻繁に電話しているとします。部門の収益は、オファーに実際に積極的に反応する顧客から直接得られます。この進行中のキャンペーンに関わるダイレクト・マーケティング費用は重要であり、適切な顧客に適切なタイミングで連絡することは肝心です。遅すぎると、すでに別の金融機関から融資を受けている可能性があります。早過ぎると、ニーズはまだ具体化していません。

このため、銀行は、これらの顧客が誰であるかを予測するための 2 つのツールを検討するでしょう。銀行のデータ科学チームによって社内で開発された機械学習モデルと **Endor** のエンジンです。ここでは、各ツールのオススメのポイントを簡単にご説明します。

機械学習モデルで検出された顧客グループは、銀行によるマーケティングオファーに積極的に反応する顧客（例：真陽性）と、そうでない顧客（例：偽陽性）から構成されています。たとえば、真陽性がモデルの結果の 10% であるとします。大規模な実験では、**Endor** 社会物理学エンジンによって 10% の大部分が検出されることが期待できることが



わかりました。(a) 機械学習モデルで検出された偽陽性の多くは **Endor** エンジンでは報告されませんでした。(b) **Endor** の結果には、従来のモデルでは検出されなかった多くの追加の真陽性が含まれます。その結果、**Endor** の精度/リコール・トレードオフが向上したため、販売力が大幅に改善されました。



どのように？一時的パターンの検出：人間の現実には、多くの小さな一時的な出来事と変化から成っています。社会物理学は、人間の行動の原動力を取り入れているため、社会環境の最近の変化のために特定の方法で行動する可能性のある少人数のグループを発見することができます。この為、社会物理学は人間の行動データの動的信号を識別することができます。従来の機械学習やディープラーニングではこの信号をノイズと区別することができない為、社会物理学の助けがなければ、統計的に意味のないものとなります。

機械学習とディープラーニング対社会物理学、どちらがどの目的に適していますか？データ科学とビッグデータ分析ツールを使用してビジネス上の質問に解答する場合、機械学習と社会物理学の両方とも有効なオプションです。以下の表は、その属性に基づいて適切なツールを特定するのに役立ちます。

なぜ社会物理学なのか？1970年代には、画像処理やテキスト認識などの「静的問題」のために、さまざまな数学的および統計的な機械学習技術が開発された歴史があります。このような問題は、相対的に少数の比較的安定した「信号」によって支配されています。熟練のテキスト認識モデルは、2016年のMITの学生の手書きのテキストを処理するときでも、アルバート・アインシュタインの個人的な手紙を分析するときでも、同様の成果を達成するでしょう。同様に、どちらも Siri でもグーグルの音声認識エンジンでも、JFKの有名な "ein bin Berliner" の高音質録音の翻訳は難しいでしょう。

	...には機械学習がより適している	...には社会物理学がより適している	なぜ？
データのタイプ	機械的/物理的駆動データ： 例： ・オイルドリルポンプの制御データを監視して誤動作を予測する ・顔認証 	人間の行動データ： 例： 誰がプレミアムサービスを購入するかを予測する金融取引の分析 	人間の行動は、不安定で、予測不能で、騒々しく複雑で動的です。数学的に言えば、人間の行動は、多数の「一時的」信号によって支配され、それぞれが個人の少数グループに影響を与えます。したがって、それを表す一貫して安定したモデルを生み出す人間のデータを「学習」することは非常に困難です。 Endor は、社会物理学を使用してそのような一時的信号を検出するため、特に人間ベースのデータに合わせて調整されます。

しかし、人間の行動は別の話です。多数の「動的信号」によって支配されているという事は、非常に動的で、非常に「断裂している」ということです。クレジットカードによる購入からミレニアル世代を検出するために訓練された従来の機械学習モデルは、時間の経過とともに精度が急速に悪化し、熟練の専門家による絶え間ないメンテナンスが必要になりました。ミレニアル世代の行動は、頻繁な（そして絶え間なく変化する）トレンドに従うので、データの中でこれを見つけることは、モデルの絶え間ない再トレーニングだけでなく、これらのトレンドを検出するための新しい機能の絶え間ない開発（すなわち、生データの一部ではない複雑な集合的行動特性）をも意味します。これは、セマンティクス・ドメインの専門家がデータ専門家と連携して作業することによってのみ実行できます。



機械学習アルゴリズムは **70** 年代以来
変わっていません。もともと静的問
題のために設計されました。

「人は死に、国家は繁栄と
凋落を繰り返すが、アイデ
ィアは生き続ける。」



「限界を受け入れれば、我々
はそれを越えていける」



以下は、人間データにおける行動予測のためのツールとして社会物理学を使用する主な利点です。

従来の機械学習		ディープラーニング (社会物理学を伴わない)	社会物理学	なぜ?
スモールデータセット	スモールデータセットを分析することは可能ですが、専門のデータ科学者が必要で時間がかかるプロセスです	すべての質問に解答するには大量のデータが必要	人間の行動に関連する質問に答えるためのデータはほとんど必要ありません。結果は自動的に生成されます (データ科学者が関与する必要はありません)	Endor は、社会物理学にすでに人間の行動データの原動力を組み込まれている為、結果を生成するために「ビッグデータ」を必要としません。したがって、非常に小さなデータセットであっても、正確な予測と実行可能な信号をすぐに生成することができます。

3.2 第¹段階 : Endor.com – 企業のための自動予測エンジン

これまでのセクションで説明したように、今日のデータ科学と利用可能な機械学習技術は、このような機能の使用が裕福な巨大テクノロジー企業のために維持されていることを示しています。成長の鈍い分野の専門家に数十万ドルの年俸を支払うことができる企業。しかし、そうした企業でさえ、データ科学と予測分析は、商品とは程遠いものでした。逆に、信頼性の高い予測を目的としたプロジェクトでは、通常 2~3 人の専門家チームが 4~6 回の反復をそれぞれ 1~2 週間で行うことを妨げており、最初のモデルが作成された後は通常

定期的なメンテナンスが必要になります。したがって、業界標準の平均的な予測プロジェクトの総コストを約**\$150**万と見込まれます。つまり、次のことを意味します。

- ほとんどの企業は、一握りのプロジェクト以上の資金を調達できません。
- 今日、ROI（投資収益率）をプラスにするためには、大きな利益を上げなければならず、ビジネス面のみに集中せざるを得ません。

	従来の機械学習	ディープラーニング (社会物理学を 伴わない)	社会物理学	なぜ?
機能対生 データ	生データの表現の正しい機能を定義して選択するためには、熟練したデータ科学者またはドメイン専門家が必要	機能無しでも生データを処理できますが、非常に狭い範囲の問題(*)に限定されています。	人間の行動に関わるあらゆる種類の予測問題に対して、機能無しで生データを処理できます。	機械学習では、生データを意味のある機能に変換するために、しばしば手作業による長いプロセスが必要です。これは通常、すべての問題のために、そして新しいタイプのデータに対して、ゼロから行われます。 ディープラーニングは機能作成を自動的に処理しますが、それでも大量のデータが必要であり、問題の複雑さに伴い必要なデータがさらに増加します。したがって、それは「単純な行動」に限られません。 また、ディープラーニングのダイナミクスは、一般的な企業では利用できない大量のデータを必要とするため、ディープラーニングは通常「静的問題」に限定されています。 社会物理学は、人間の行動の生データを自動的に人間の行動クラスターの正規形式に変換します。 Endor は、この正規の表現を使用して、データサイズに関係なく、すべてのデータタイプとすべての質問に対処し、統合された人間行動データセットを生成します。

	従来の機械学習	ディープラーニング (社会物理学を 伴わない)	社会物理学	なぜ?
ユーザーと必要な専門知識	機械学習の専門家は、通常はセマンティック機能作成をサポートするドメイン専門家の助けを借ります	ディープラーニング専門家	ビジネスユーザー 必要なのは、「あなたがもっと探したい人々」の例を提供することだけです。	機械学習では、大きなデータセットの基本的な動作を「学習」するか、以前のドメインの専門知識を活用する必要があります。 Endor はすでに人間の行動データの基本的なダイナミクスを組み込んでいます。
データの変化のペース	変化が遅いデータに限定されます。データを変更すると、機能を調整するためにドメイン	それはダイナミクスに対処することができますが、変化の遅いデータに限られます(人間	急速に変化するデータソースを簡単に分析することができます(ドメイン専門	Endor のエンジンは、人間の行動データに特化しているため、本質的に動的な性質のデータに作用します。 社会物理学は人間のデータセットに埋め込

	専門家の継続的な介入が必要になります。	の行動データに関しては厳しい制限です！)	家は不要)。	まれた一連の数学的不変性であるため、非常に短い時間セグメントを表す信号を検出することさえできます。言い換えれば、他の手法で観測可能になる前に新しく発生する変化を識別することができます。
分析の範囲	特定/限定		広域/人間の行動に関する質問	機械学習では、自動的に選択されたモデルの特徴を再学習する必要があるため、各データセットと質問について学習プロセスを繰り返す必要があります。社会物理学は、その質問固有のものではない基本的な人間行動原理に基づいています。

	ディープラーニング (社会物理学を従来の機械学習と伴わない)	社会物理学	なぜ?
データ・クリーニング	機械学習は、データのノイズや隙間に非常に敏感です。長く高額なデータクリーニング・プロセスを必要とするディープラーニングでは、データをディープラーニング・ツールで受け入れられる形式に変換するための注意深いプロセスが必要になることがよくあります。	データクリーニング不要	機械学習とディープラーニングの両方では、データ駆動の数学的パターンを使用して、ルールを推定し、信号を抽出し、予測を生成します。これには、データクリーニングの微妙なプロセスが必要です。 一方、社会物理学は、あらゆる人間の行動データセットに埋め込まれていることがわかっている数学的不変性である外部パターンを使用します。これにより、データノイズの影響が大幅に軽減されます。 また、社会物理学は生データを行動クラスタに変換し、データの隙間とノイズ（ほとんどが自動的に除外される）の影響をさらに低減します。
反復と微調整	データの各調整や問題の定義には、ドメイン専門家（ビジネスユーザーまたはアナリスト）と機械学習\ディープラーニングの専門家の共同作業が必要です。各反復には数週間かかることがあり、通常のプロジェクトは少なくとも4~6回の反復で構成されます。	インタフェースは、予測質問を修正できるビジネスユーザーまたはアナリストが使用するために設計されています。結果は新しい定義に自動的に適用されます。	Endor のエンジンは、人間の行動データに特化しているため、本質的に動的な性質のデータに作用します。 社会物理学は人間のデータセットに埋め込まれた一連の数学的不変性であるため、非常に短い時間セグメントを表す信号を検出することさえできます。言い換えれば、他の手法で観測可能になる前に新しく発生する変化を識別することができます。

- 企業は、(a) 技術的な成功プラス (b) ビジネスでのハイリターンになると考えるプ

プロジェクトを常に優先して、データ科学と言う超高価な武器を使用します。

今日のデータ科学が認識されている方法を中断することでこの問題を解決するために、新しい科学である社会物理学の専門企業として MIT のスピンオフ企業である Endor 社が設立されました。MIT での 3 年間の研究と、エリートチームの研究者とエンジニアによる更に 3 年間の開発に基づいて、Endor 社は世界で初めて完全自動の「データ科学サービス」エンジンを開発しました。これにより、企業は所有する行動データをオンボードに持ち込むことができ、迅速な統合（通常は数時間）の後に、「オブジェクト」（ユーザー、製品、クーポン、ロケーションなど）の将来の行動に関する無制限の予測質問を開始することができます。

年間コストが\$100 万未満で、企業は数十件の予測質問に対して数分で迅速な結果を得ることができるようになりました。これは単なる一次的な改善ではなく、パラダイムチェンジとなる可能性を秘めています。「予測プロジェクト」の実行が「グーグル」ほど簡単になり、潜在的プロジェクトへの優先順位付けを事実上不要となり、あらゆる決定が予測に基づいて成されるでしょう。

例 1 – コカコーラ社共同研究： 最近の Endor 社とコカコーラ社[16]のコラボレーションでは、さまざまな消費者行動に関する正確な予測を提供する社会物理学の能力が実証されました。これには、ブランド離れ、ブランドロイヤリティ、新製品の採用、マーケティング・キャンペーンへの対応などが含まれます。Endor の社会物理学エンジンは、数百万件の POS 取引を生入力（3 ヶ月間）として使用して、それぞれが実世界の行動グループを代表する 2,000 万件の「相関異常」を検出しました。各グループの正確な意味は不明ですが、グループは「行動推定」のために使用される：サンプルセット（例えば、新製品の早期採用者）を仮定すると、システムは行動グループのコレクションを使用して類似した人物（サンプルセットのメンバーに行動的に類似）を見つけ出します（新製品をすぐに体験したがるユーザーなど）。この方法論を使用して、15 の異なる予測質問がなされ、それぞれが要望に応じた予測レポートを生成しました。検証のために保存されたサンプル外のデータを使用して、高精度のレポートが実証されました。



「社会物理学はビッグデータの行動分析に関するものですが、まったく新しいレベルに引き上げます。我々は非常に幸運なことに **Endor** を見つけてそれに取り組んでいました」

Alan Boehme (アラン・ベーム) 博士 (コカコーラ社 CTO)

例 2 - ツイートの自動分析： 最近のテストでは、1,500 万件のツイートのメタデータが分析用の生データとして Endor エンジンに提供されました。さらに、顧客は、入力データに含まれていた ISIS の活動家として知られている 50 件のツイッター・アカウントの身元を明らかにし、Endor の能力をテストして、データ内に隠された更に 74 件のアカウントを検出しました。Endor のエンジンは、1 台のラップトップ上でタスクを完了しました（生データがシステムに導入されてから最終結果が得られるまでの時間を測定）、提供されている例から 80 件のツイッター・アカウントを「類似アカウント」として特定し、そのうち 45 件（56%）が 74 件の隠しアカウントのリストに含まれることが判明しました。重要なことに、これは誤報率が非常に低く（結果：偽陽性 35 件）、顧客は容易に特定された標的を人間専門家に調査させることができました。

「画期的な概念と真の技術革新。彼らが提示した結果は、他の競合するツールと比べても抜群です」

イスラエル諜報部隊 CIO

エコシステムのニーズが分散型プロトコルを可能にします： Endor は、社会物理学と強力な専門家チームを使用して、フォーチュン誌上位 500 社の大企業がどのようにして大幅にコストを削減し、大幅に向上するかを実証しました。ただし、この製品は主に大手の銀行や小売業者向けに提供されていますが、年間ライセンスとして \$25 万～\$120 万のコストがかかります。これはもちろん、個人向けではなく、ほとんどの「ロングテール」ビジネスに適用可能な取引ではありません。

したがって、別のソリューションの必要性が明らかになりました。誰もが社会物理学の新しい技術から恩恵を受け、それをリーズナブルな価格で行うことができるもの。したがって、そのソリューションは、

- 新規参加者が利用可能なリソースで構成され、自立しています。
- 強力なネットワーク効果を生み出し、参加者に継続的に参加するようインセンティブを与え、成長するにつれてより良く、より安くなります。
- 公平を保ち、本質的に偏っておらず、トラストレスである。この条件は他のサービスとは異なります。予測は、それに頼る人や組織の行動によって引き起こされる「外部の影響」と呼ばれるものを受けることが予想されるためです。

このコンセプトを念頭に置いて、Endor チームは革命の次のステップを提示することを誇りに思っています。Endorcoin プロトコル！

3.3 第²段階： *Endor.coin* プロトコル – 大衆のためのデータ科学

Endor の行動分析に影響を与えた社会物理学の業界での実施の成功に続き、*Endor.coin* プロトコルは、この能力をロングテール・ビジネスだけでなく、プロフェッショナルな個人にもたらすという目的で作成されました。予測分析を改革する *Endor.coin* プロトコルは、行動予測のために人工知能を民主化し、すべての人が利用できるエコシステムの作成を可能にします。さらに、プロトコルの完全に分散化された性質は、トラストレス、検閲耐性、説明責任を保証します。巨大テクノロジー企業の潜在的な操作によるデータと技術の制御から解放され、初めて、安全な枠組みで、手頃な料金で誰もが行動予測を利用できるようになりました。

3.3.1 定義

生データ：その *Endor.coin* プロトコルは、任意の時間関連データやトランザクションデータソース（例：コールデータレコード、**ERC20** ブロックチェーン、アプリ内購入など）をサポートします。データ提供者は、予測の基礎となる列を指定する必要があります。たとえば、**ERC20** を使用すると、将来の興味深い行動のアドレスやトークン自体を予測して発見することが可能です。ツイートデータを使用して、興味のあるツイッターID、ロケーション、ハッシュタグを見つけることができます。

処理済みデータ：行動クラスターの抽出を受けたデータを示し、「知識球体」として社会物理学の正規の表現に変換されました。

知識球体：1 つ以上の生データソースから抽出された行動クラスターの集合を示します。*Endor.coin* プロトコルは、任意の数の知識球体の分割または結合を可能にします。予測のタイプによって行動クラスターとその関連性が異なる為、*Endor.coin* プロトコルは特定の時点で最も関連性の高いクラスターを選択し、実際の予測の生成に使用される「知識球体」を生成します。

予測：*Endor.coin* プロトコルは、「グループ Y と行動上類似している可能性によってグループ X をランク付けする」と言う形で表現できるあらゆる質問に対応します。たとえば、グループ X にすべての **ERC20** トークンが含まれていて、グループ Y に発行量が劇的に増加したトークンが含まれている場合、予測結果には **ERC20** トークンのリストが含

まれます。最上位には、近い将来に発行量を増加させる可能性が最も高いトークンが含まれ、最下部にはこの行動を示す可能性が最も低いトークンが含まれます。

3.3.2 データ提供者

その *Endor.coin* プロトコルは、任意の行動および時間に関連する構造化データの統合をサポートします。データのオンボーディングは単純な API コールを使用して行われ、データ所有者がデータを提供しながら、どの列が非公開で、どの列が公開分析を利用できるようにするかを制御できます。*Endor.coin* ユーザーが非公開と定義したデータは、公開データストリームと共に自動的に統合され、非公開データと公開データが融合された高品質の予測分析を生成します。社会物理学の *知識球体の概念* (付録 D を参照) を使用することにより、データ統合は自動化され、簡単になります。クリーニングやデータ準備は必要ありません。

データ提供者はデータの分析の代金を予測エンジン・プロバイダに *EDR* トークンで支払うように請求されます。それらは、順番に、データから導き出された洞察が予測に使用されているときに報酬トークンになります。これは、高品質のデータストリームの寄稿とメンテナンスを促進します。

3.3.3 予測エンジン

その *Endor.coin* プロトコルは、重複する行動クラスタの特徴空間へのデータの投影に基づく「予測言語」を定義します。現在の *Endor* エンジンは、立ち上げ後すぐに使用できるようにする為、そのネットワークに接続された最初の予測エンジンになります。しかし、*Endor.coin* は、複数のタイプのエンジンで構成され、補完的な機能を提供し、パフォーマンスの精度を高め、信頼性を向上させるエコシステムの構築を目指し、新しい予測エンジンの開発を促進、支援、資金提供します。予測エンジンの数が増えるにつれて、様々な予測エンジンによって抽出されたクラスタの集合体から *Endor.coin* オープンソースプロトコルによって自動的に選択される最も関連性の高いクラスタに基づいて予測が作成される為、偏りのない予測結果が保証されます。

3.3.4 事前に定義された予測と予測リクエスト (RFP)

その *Endor.coin* プロトコルは、さまざまな事前に定義された予測を含む大きなカタログで開始されます。これらの予測は、プラットフォームの *EDR* トークンを使用して購入することができます。*Endor.coin* は、ユーザーが *予測リクエスト (RFP)* を送信すると、実行可能で購入に役立つ新しいタイプの予測を提案する事により、予測の選択肢を広げます。

プラットフォームによってサポートされている予測を徐々に拡張するこの方法は、新たにサポートされた予測の選択を最適化するために、(プロトコルユーザーのリクエスト

によって明らかにされる) 群衆の知恵を利用します。

3.3.5 非公開データ分析とセルフサービス API

後にリリースされる *Endor.coin* プロトコルは、高度なユーザー向けに完全な「DIY」API をサポートしています。技術に精通したユーザーやプロのデータ科学者は、独自のデータを簡単に搭載できるセルフサービス・インターフェイスを使用し、独自の新しいタイプの予測を作成できます。これらの予測は、非公開に定義することも、公開することもできます（*EDR* トークンが広く使用されるようになれば、予測に報酬が与えられます）。

3.4 ロードマップ

今日の一般的なソフトウェア市場では、当社のアプローチは、サービスとしてのプラットフォームや最近ではサービスとしてのブロックチェーンなどのサービスと比較することができます。その *Endor.coin* トークン（または *EDR*）は、プラットフォーム上の取引の運営に使用されます。*EDR* は、キーまたはソフトウェアのライセンスとして機能し、アプリケーションの領域を拡大する開発者のコミュニティを介して、パフォーマンスと規模を拡大するために、時間の経過と共により多くのトークンを使用できるようになります。さらに、管理者がトークンを分配し、使用状況を監視し、必要に応じてさらに多くのトークンを購入するためのダッシュボードが作成されます。

Endor.coin の進化の次のステップは、ますます成長するカタリストのコミュニティによってエコシステムを有機的に拡大することです。誰もが *EDR* で支払い、*RFP*（「予測リクエスト」）を送ることができ、それぞれの *RFP* に対処する課題は、（*EDR* を使用して誰もが利用できる）*Endor.coin* によって新たにサポートされた質問として埋め込む予測範囲の拡大に貢献するスマートコントラクトを通じてそれぞれの支払いから報酬を受け取るどのカタリストでも請け負うことができます。

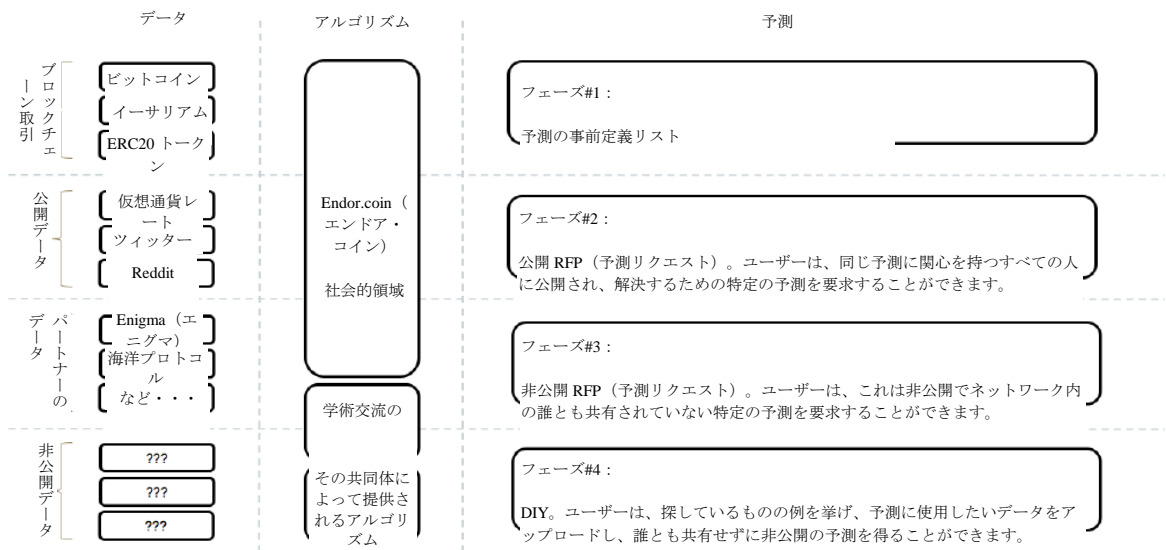
この第 2 段階では、*Endor.coin* がモジュール式アプリケーション・プラットフォームになり、そこではカタリストがブロックチェーンの力で予測ドメインを拡張することができます。この *Endor.coin* を有効にするには 再利用可能なソフトウェアコンポーネントの作成者にマイクロペイメントのオープンエンドストリームが与えられ、これを永続的に結合して再結合して、有効で高度なカスタマイズが可能な質問アプリケーションの拡張ライブラリを作成することができます。カタリストには、*Endor.coin* プラットフォームに追加する各コンポーネントのマイクロライセンスが付与されます。エンドユーザーは選択したアプリケーションを

インストールします。支払われるライセンス料は、そのアプリケーションで使用されているコンポーネントのマイクロライセンスの総額です。スマートコントラクトを通じて、**Endor.coin** にはエンドユーザーに課金し、関連する各カタリストに支払いを分配する責任があります。

時と共に **Endor.coin** は、技術力が高い人にしか利用できない機械学習を民主化し、企業家精神を持つコーダーやブロックチェーンでデータリッチウェブの構築やモバイル製品の開発をしようとしている企業のための最先端の開発プラットフォームになります。さらに、現在データ分析アルゴリズムを効率的に実行するために必要な取引量をサポートしていない既存のブロックチェーンにはこれらの機能が提供され、絶えず増大するカタリストのコミュニティの助けを借りて、質問のドメインを継続的に拡張するエンジンを使用してアプリケーションを構築します。

私たちのビジョンは、完全に自動化されトラストレスで分散型であり、エンドユーザーにとって完全な透明性を持った予測インフラストラクチャを作成することです。インセンティブは **EDR** トークンで支払われ、アプリケーションは基本的にプラグインにセットされています。肯定的なフィードバックループが予想されます。アプリケーションが多く構築されるほど、プラグインがシステムに追加され、よりユニークなコンポーネントが再利用できる状態になります。これは、**Endor.coin** エコシステム全体で相互に自己補強するネットワーク効果に貢献します。

究極的には、プロの専門家や企業がより複雑な細かい **RFP** を送ることができる、ハイエンド顧客向けのプラチナプラットフォームを提供します。このような複雑な質問を集めるために、提供者に **EDR** で報酬を支払うことによりを次のようにデータアクセスを拡張します：カタリストが企業からのデータ（時間制限付き）をリクエストする場合、各企業は希望価格を提案することができます。カタリストが支払いに同意した場合、企業はカタリストの公開鍵で暗号化されたデータへのリンクを送信します（その時点で、スマートコントラクトによって料金が支払われます）。レビューは、カタリストとデータ提供者の両者が公開できます（今後の評判のために）。**EDR** は、データ提供者への支払いと、プラチナアプリケーション開発者からの資金調達のための調停によって生成されます。



第4章

トラストレス、検閲耐性、説明責任

4.1 人工知能の説明責任と真正性

彼の記事「あなたがまだそれを知らないとしても、あなたがブロックチェーンベースの AI を望む理由」[34]で、執筆者の Jeremy Epstein (ジェレミー・エプスタイン) と読者との読者の娘がアマゾンの AI ガジェット Alexa (アレクサ) を持っているという会話を公開しています。この会話では、「ネットワーク中立性」の話題を中心に、Alexa はリスナーと豊富な関連情報を共有しました。ああ、エプスタインがその話題に精通していることから、彼は驚くべきことに、正確ではあるが潜在的に一方的な情報を発見しました。この経験は、ワープの速度で飛躍している AI パワーのいくつかの将来のリスクを強調しています。従来、ビッグデータや AI に関してはアマゾンのような大企業に大きなメリットがありました。

エプスタインの発言の引用：

レースがより多くのデータを集め、保存し、分析することだとすると、誰がポールポジションにいるのですか？つまり、米国の FANG (フェイスブック社、アップル社、ネットフリックス社、グーグル社)、中国の BAT (Baidu 社、アリババ社、Tencent 社)、裕福なフォーチュン誌上位 1000 社以内の企業、多国籍企業。

彼らは、より多くのデータを取得し、保存し、分析し、AI モデルを構築するための力と資本を持つ数少ない企業です。さらに、\$30 万~\$50 万の初任給を支払える数少ない企業であり、7~8 桁に及ぶトップレベルの給与は数百万~数千万ドルに登ります。あなたの息子や娘は NBA や NFL には向かないかもしれませんが、最高の AI 科学者になるかもしれません。

このすべての最終的な結果は、富裕層がより豊かでより強力になり、イノベーションへの障壁がさらに高まることです。

しかし、苦しんでいるのはイノベーションだけではありません。大企業の AI の閉鎖性は、社会が「ブラックボックス」を信頼しなければならないことを意味しています。

したがって、「予測の信頼性」を提供することは、次のすべてを提供するインフラストラクチャの有効性に依存します。

- **説明責任：** 予測の提供者は、振り返る事により、彼らの予測が正しかったことを証明することができます。予測の消費者は、あらゆるタイプの正確な予測を生成する予測提供者の効率（またはその欠如）を振り返って確実に推論することができます。言い換えれば、評判は操作が絶対に不可能です。
- **信頼性：** 予測の提供者に偽りはありません。予測消費者の誤解を招く偽装は絶対に不可能です。
- **偏りが無い：** 偏った予測によって達成された盲目的な利益ではなく、市場の力を利用して精度を上げる予測提供者間の競争は公正でなくてはなりません。
- **正確さ：** ネットワークへの金銭的な流入を提供するためには、概して言えば、予測は前述の 3 つの条件に関して肯定的フィードバックループを閉じるには十分正確でなければなりません。

その *Endor.coin* プロトコルは、上記のすべての条件を満たすインフラストラクチャを提供します。

- **Endor.coin プロトコルの説明責任と確実性：** ブロックチェーンを使用すると、予測は無期限に保存され、誰でも利用できます。時間依存予測は消費者にのみ送信され、一定期間後にリリースされるキーを使用して、暗号化済み未公開バージョンにも保存されます。
- **Endor.coin プロトコル・ゼロバイアス：** プロトコルは、各予測に対して、（分析エンジンやデータソースに関係なく）最も関連性の高い行動クラスタを選択します。したがって、プロトコルはデータ提供者と分析エンジン提供者によって制御されない独立体であるため、予測の偏りは本質的に防止されます。
- **正確性：** *Endor.coin* は MIT の社会物理学に基づいており、さまざまなユースケースに対して正確な予測を一貫して提供する事が業界で広く実証されています。

4.2 分散化プロトコルを通じた検閲耐性

検閲は厄介な仕事です。その明示的な表現が存在する場合、IP プロキシなどの「低レベル」の技術的ソリューションを使用して検出しやすくなり、予測の偏りが起こりやすくなります。一方で、暗示的な検閲はまったく別の問題です。グーグル社やフェイスブック社などの企業は、特定の種類の検索質問をブロックすることが知られています。これは何度も道徳的または法的理由のために行われていますが、どうしてこの事がいつも正しいと確信できるのでしょうか？グーグルが「簡単にはアクセスを許可できない」と判断した場合、グーグルに関連

ウェブページを提供するよう強制することはできますか？ **Alexa** が不適切だと思うリクエストに、**Alexa** が従うように強制することはできますか？ 答えは簡単です - いいえ。

この文脈における予測は、運営主体、またはそれらが対象とする規制当局による検閲の対象であるという意味で、検索と非常に似ています。AI と予測分析がどのように機能するかを見ると、潜在的な検閲に関して 3 つの重要なレイヤーがあります。

- **データリポジトリ**：データの整合性、完全性、および安全性を保証します（入力 は正確で信頼性が高く、操作や盗難の可能性はありますか？）
- **アルゴリズム/機械学習エンジン**：予測が中央の権限によって検査されていないこと、およびすべての予測リクエストが公平に実行されていることを確認しており、予測のトピックとは無関係なサービス品質の考慮事項があります。
- **質問インターフェース**：予測質問のアウトプットを確実に表現し、効果的に新しいデータをキャプチャし、サポートされている予測の制限を予測のトピックとは無関係にします。

集中的な予測ソースへの意思決定を信頼するならば、絶対的な確信をもって、上記のすべての条件が満たされていることを暗に仮定します。集中的な閉鎖モデルの予測スキームでは、シーンの裏側で何が起きているかを知らずに各レイヤーを信頼するように求められるとき、（プレーンな嫌悪性ではないにしても）そのような信頼性を正当化することは困難です。

Endor.coin プロトコルは、本質的に、すべての予測を完全に分散化されトラストレスな方法で実行できるようにすることによって、これらのアサーションを提供します。データが提供され、ネットワークにオンボードされると、関連する予測が実行され、*Endor.coin* プロトコルが自動的に最適化されます。

4.3 プライバシー保護データ分析：暗号化されたデータの使 用

新しい科学である社会物理学の使用は、データ解析と予測のための民主化されたエコシステムの形成に関して重要な利点を提供します。それは、社会物理学的計算が可能であるため、データのプライバシーや整合性を犠牲にすることなく、暗号化されたデータに対して行われます。具体的には、*Endor.coin* データを分析することで、データ寄稿者は、データそのものを公開することなく、暗号化された形式でデータ分析を収益化することができます。代わりに、このシステムは、「類似した予測」のための基礎としてその後で使用される「暗黙の行動クラスタ」（ハッシュまたは暗号化されたデータ空間にマッピングされる）の抽出を可能にし、ソースデータの性質と予測質問のセマンティクスを気にせずに正

確な行動予測をもたらします。

同様の技術は現在既に *Endor* で使われており、大手銀行や金融系の顧客と協力して、完全に暗号化されたバージョンでデータをオンボードで使用できるようにし、データのプライバシーと整合性を妥協することを不可能にしています。

4.4 ネットワーク効果

完全に分散化した *Endor.coin* プロトコルでは、結果的にネットワーク効果が他よりとりわけ大きくなり、その使用法が広がるにつれてそのメリットが増します。以下は、プロトコルの採用が増えると発生すると予想される主なスラストです。

ユーザーが増えれば、それぞれのコストが削減されます。大企業の顧客とは異なり、個人や抽象企業には主に公開データセットに基づく予測が必要です。*Endor.coin* プロトコルは、このようなデータセットを一度分析し、「知識球体」と呼ばれる一連の行動クラスターを生成します。このデータ構造は、このデータソースに基づく予測ユースケースのすべてをサポートすることができ、エンドユーザーが、行動クラスターの抽出と彼らが興味を持っている特定のユースケースのデルタである「個人化要因」をカプセル化するために支払う必要がある手数料はごく僅かです。

言い換えれば、与えられたデータセット（例：ビットコイン・ブロックチェーン）および特定のユースケースについてのコストは以下のように計算されます。

- **知識球体計算**： 総コストの約 **99%**で構成され、**1** 回計算されます。このタスクに必要なリソースは、このデータストリームのすべてのユーザー間で償却されます。
- **予測個人化要因**： 総コストの約 **1%**で構成され、ユーザー毎に計算（及び支払い）がされます。さらに、異なるユーザーが作成した同様の予測では、*Endor.coin* プロトコルは、最初に計算された予測から導かれた洞察を自動的に再利用し、同じコストの予測精度をさらに向上させます。

これは、 N 人のエンドユーザーのネットワークの運営コストの方程式が次のようになることを意味しています。

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

単一の商業的プレーヤーによる同じ予測のコストを示す C のためのものです。これは、最終的には、ユーザー数 N が増加すると予測の精度が向上する一方で、そのコストは大企業の顧客が使用する同じシステムのコストの約 **1%** になることを意味します。

最終的には、質問をする人が多いほど、コストの大部分はアクティブユーザーに按分されるため、受け取る結果がより良くなり、コストが低くなります。

データ提供者が増えると精度が向上します。データ提供者は、*Endor.coin* プロトコルの定義に従って、*EDR* トークンを使用してデータの実行に資金を提供し、データから抽出された行動クラスタに基づいた「知識球体」を作成する必要があります。この初期実行コストは、データの品質とさまざまな予測への貢献度に応じて、部分的に、均等に、あるいは重要な関心を持ってエンドユーザーによってデータ提供者に返済されます。この資金調停は、さまざまな分析エンジンによって検出された行動クラスタの集合体全体にアクセスしたり、さまざまなデータセットにアクセスしたり、関連性の高い上位クラスタを選択したりする事ができる *Endor.coin* プロトコルによって処理されます。実行トークンは、最終的な予測への貢献に応じて、データセット提供者に比例して配当されます。

このメカニズムの結果、高品質データの所有者はデータソースのサポートを継続するようインセンティブを与えられます。（また、品質と有効性をさらに向上させます）。低品質データの提供者は、コストが返済されないため、淘汰されます。この経済モデルは、市場原理を利用して、予測プロトコルで利用可能なデータソースが可能な限り最高の状態に保たれることを自動的に保証されます。したがって、新しいデータ提供者は、総コストを増加させることなく、予測精度全体を向上させることができます。

予測エンジンを追加すると予測効率が向上します。*Endor.coin* プロジェクトの第 1 段階では、行動クラスタ抽出の最初のプロバイダーとして機能する *Endor* 予測エンジンが使用されます。ただし、追加の予測エンジンが *Endor.coin* プロトコルをサポートする予定です。このような新しい予測エンジンの導入は、データ提供者とエンドユーザーの両方に大きな効果をもたらすことが期待されます。異なるテクノロジーを利用するエンジンは、同じデータソースから異なるタイプのクラスタを生成することが期待されます。これは、*Endor.coin* プロトコルをサポートする予測エンジンに使用されるさまざまなテクノロジーが増加すると、新しい予測が要求されたときにプロトコルが選択できるようになるクラスタの種類も増えることを意味します。

これには 3 つの主な効果が期待されます：

- **精度の向上、新しい予測のサポートの強化**：新しいタイプのクラスタが利用可能になると、*Endor.coin* プロトコルは、リクエストされた各予測の生成に使用するクラスタをよりよく選択できるようになります。既存の予測では、（クラスタリング技術の使用に由来する「直観的洞察」の利用可能性のために）精度が向上します。しかしながら、これはまた、隣接する時間に実行された予測間での洞察を共有するため、システムによってサポートされる様々な予測を増加させ、その後全体的な精度をさらに向上させ、その時点まで、精度が最適でないためにシステムによってサポートされていなかった新しい予測が、実行するのに費用対効果に優れていることを示唆しています。

- **予測 1 件あたりのコスト削減：** システムによってサポートされる予測量が増加するにつれて、予測を購入するエンドユーザーの数が増加します。つまり、初期データ分析の購入に使われる EDR トークンのプール全体が急速に増加し、続いてデータ提供者が受け取る資金量が増加し、エンドユーザーは予測 1 件あたりのコストを減らすことができます。
- **新しいデータソースの経済的持続可能性：** 最終的に、新しいクラスタリング技術の有効性は、これらの新技術がそのデータソースの分析と利用可能なデータソースの分析との互換性が良好な場合に、特定のタイプのデータソースの貢献を経済的に実現可能にする可能性があります。そのような場合、そのデータソースは突然 *Endor.coin* プロトコルの実行可能な情報源になり、プラットフォームへの最初の統合コストを返済し、その後も利用可能なデータソースの量を増やすことによって、そこから利益が派生します。

第 5 章

エコシステムの有効化

その *Endor.coin* プロジェクトは、データ所有者、開発者、データ科学の専門家、中小企業、個人ユーザーの間で相乗的な協力関係を構築し、多面的なエコシステムの確立を目指しています。各プレーヤーが資産（データ、資金、追加すべき新しい予測に関するアイデアなど）で貢献できるようにすることで、*Endor.coin* コミュニティは、自己永続的な肯定的フィードバックのビジネスループを可能にし、促進していきます。以下は、このエコシステムの主な側面に関する簡単な説明です。

5.1 公開データ提供者

Endor.coin プロトコルは、公開データストリームの寄稿を当初から奨励し、第 1 段階で「登録データ提供者」によって提供され（簡易的なコンプライアンスと品質保証のプロセスを経て、エコシステムが最高品質のデータでブートストラップされるようになる予定です）、そして後に、分析のために利用可能なデータを充実させたいと思う個人または企業によって、こうする事によって *EDR* で報酬を受け取ります。

潜在的なデータ寄稿者の例は次のとおりです。

- **データパートナー**： *Enigma*、*Twine*、*Thasos*、*Ocean Protocol* などのプロジェクトでは、データストリームを（部分的または完全に）*Endor.coin* プラットフォームに分析用に接続します。
- **ソーシャル・チャンネル**： ツイッター、*Reddit* など。
- **データスクレーパー**： 公開が可能な様々なソースからのデータの取得、クリーニング、構造化、およびセマンティックのリッチ化。
- **ブロックチェーン・プロトコル**： *ORBS* または同様のもの。データをダウンロードして解析し、分析のために使用できるようにするノードの確立によって達成されます。

Endor.coin は、そのような潜在的貢献者である数社とすでにパートナーシップ交渉をしています。

5.2 学術研究グループ

Endor.coin プロトコルは、生データのクラスタリングのための複数の分析エンジンの統合をサポートしているため、当社はこの分野における有力な学術研究グループとの協力を促進するために、収益の一部を活用する予定です。このアクティビティ中に開発されるテクノロジーは、拡大する *Endor.coin* エコシステムに容易に統合でき、予測の精度を高めるために *EDR* トークンを使用している場合は、常にその開発者に報酬が与えられます。

こうした努力は、業界および学界のデータ科学、機械学習およびブロックチェーン技術の世界的リーダーで構成されている *Endor.coin* 科学諮問委員会（セクション 8.2 の詳細リストを参照）の指導のもとに先導されます。

プロジェクトの創設者によって *MIT* で開発された *社会物理学* の科学革命を担う *MIT* のスピノフ企業として、当社は *Endor.coin* を用いて、データ分析やデータ市場の分野でのブロックチェーン・プロジェクトとデータの洞察と予測の普及と利用のためのインフラストラクチャを研究する学会の主要研究グループを網羅した強力で持続可能なアライアンスの創造に努めます。

5.3 カタリスト – アプリケーション開発者

Endor.coin が扱う質問の範囲を広げるために、ドメインは無限であり、「カタリスト」コミュニティである社会物理学エンジンを使用してアプリケーションを構築したい人に限定されます。カタリストの力で、*Endor.coin* はモジュール化されたアプリケーション・プラットフォームになり、カタリストはブロックチェーンの力で予測ドメインを拡張することができます。*Endor.coin* が扱う質問の範囲を広げるために、ドメインは無限であり、エンジンを使用してアプリケーションを構築したい人に限定されます。「カタリスト」*Endor.coin* コミュニティは、カタリストがブロックチェーンの力で予測領域を拡張するモジュラー式アプリケーション・プラットフォームになります。*Endor.coin* は、再利用可能なソフトウェアコンポーネントの作成者にマイクロペイメントのオープンエンド・ストリームを可能にします。このソフトウェア・コンポーネントは、永久に結合され、再結合され、有効な高度なカスタマイズ可能なアプリケーションの拡張ライブラリを作成します。カタリストには、フレームワークに追加する各コンポーネントのマイクロライセンスが付与されます。エンドユーザーは選択したアプリケーションをインストールします。支払われるライセンス料は、そのアプリケーションで使用されているコンポーネントのマイクロライセンスの総額です。スマートコントラクトを通じて、*Endor.coin* にはエンドユーザーに課金し、関連する各カタリストに支払いを分配する責任があります。すべてが自動化され、エンドユーザーには透明性を確保します。インセンティブはトークンで支払われます。肯定的なフィードバックループが予想されます。アプリケーションが多く構築されるほど、プラグインがシステムに追加され、よりユニークなコンポーネントが再利用で

きる状態になります。これは、*Endor.coin* エコシステム全体で相互に自己補強するネットワーク効果に貢献します。

5.4 データ主権：

Endor.coin は、貪欲な企業の利益のためにプライバシーを奪う世界において、ゲームを変え、富を利用可能なデータの提供者に返還することに全力を尽くしています。*Endor.coin* の簡単なサインアップは、*EDR* トークンによって有効にされるさまざまなシナリオに従って展開できるプロセスを開始します。メンバーは、自分のデータに基づいて個別の洞察を得ることができます。グループとしてチームを組むことで、中小企業と連携して大きなデータプールを集約し、より個人化されたサービスを得ることができます。例えば、最適な価格で適切な時期に適切な製品にズームインすることができます。

人々は、自分が作成したデータを所有、管理、利益を得るという基本的な考え方に基づいて、*Endor.coin* は現在、裕福な企業だけが利用できる洞察力をオープンにし、これまで不利益を被ってきた小規模のロングテール・ビジネスに *Endor.coin* メンバーのデジタルデータを集約して自分たちで利用できるようにし、データ所有者の許可を得て、高い価値のあるデータ分析を提供し、その分析によってもたらされた非常に効率的なサービスだけでなく、データが他の方法で利用される場合の金銭的利益をそのメンバーに還元します。この個人データの独立性に基づくアプローチは、*Endor.coin* の価値と信頼の両方を損なうことなく、企業と人々の両方に大きな価値をもたらします。*Endor.coin* は、企業や他の組織がデータに対する責任を排除し、優れた顧客データを利用し、顧客関係を構築するためにサインアップすることで、メンバーシップとブランド参加を利用して負の買収費用に充てることでビジネスをサポートします。

会員が増加し続けるにつれて、*Endor.coin* は会員が自らのデータから生活のさまざまな側面についての洞察と理解を得るのに役立つ一連のツールとサービスを提供します。「人々の、人々によるデータ」という使命をもって、*Endor.coin* は、フェイスブック社、グーグル社、**Uber** 社などの多くのデータベース大企業などによって今までに侵害されてきた「プラットフォーム経済」の繁栄についての個人の生活を資本化する固有の権利を個人に還元します。実績のある企業の専門知識、法的思考、技術ノウハウ、価値創造力を消費者に統合した多面的なアプローチにより、*Endor.coin* は、企業がデータの洞察を利用するために会員に直接報酬を支払うことで、顧客との良好な意思を醸成できる合意的なデータ関係を作成することにより、「データの違法コピー」問題を解決するユニークなビジネスモデルを提供します。

プロセスの整合性を保証するために、*Endor.coin* トラストを作成します。これは独立して管理された存在であり、*Endor.coin* との相乗的な関係で動作し、会員の利益、価値創造、データセキュリティを最大限に引き出します。*Endor.coin* にサインアップすると、

データはトラストによって監督され、自動的にトラスト会員になります。**Endor.coin** トラストは、その会員に安全性、セキュリティ、プライバシー、およびデータの価値を確実に最大限に高めることを多くの方法を提供します。トラストは、個人データの管理、価値の実現およびプライバシーの基準を設定および実施することにより、会員のデータを保護します。

第 6 章

トークンの実装

6.1 ブロックチェーン構造

データストレージ： ブロックチェーンは汎用データベースではありません。 *Endor.coin* には、分散型オフチェーン分配ハッシュテーブルストレージがあり、ブロックチェーンを介して利用できます。ブロックチェーンにはデータのリフェレンスは保存されますが、データ自体は保存されません。非公開データは、転送とストレージに先立って、(AES-256 - Amazon S3 サーバーサイド暗号変換または AWS IAM メカニズムを使用して) 暗号化され、アクセス制御プロトコルがブロックチェーンにプログラムされます。 *Endor.coin* は、オフチェーン・ネットワークの非公開/公開データセットと同様に、既存のブロックチェーンにも接続できるように設計されています (最終的に *Endor.coin* AWS インフラストラクチャにアップロードされる構造化形式にエクスポートできるデータベース、中央ストレージなどの任意の形式で保存されます)。将来のリリースでは、EDR トークンのデータを販売できる外部プレーヤーにデータレイヤーが開かれます。データは認証され、*Endor.coin* オフチェーン・インフラストラクチャにアップロードされます。ここで、データの所有者が利用できるようになります。また、「公開」とマークされている場合は、予測アルゴリズムが関連する顧客とみなされます。価格はデータ提供者によって決定され、この予測リクエストに従って *Endor.coin* プロトコルによって自動的に調整され、同じデータソースのコストをユーザーが分担します。

消費予測： ブロックチェーンは、重く複雑な取引を処理できません。重い計算を実行するためには、同じオフチェーンの計算ネットワークが使用されます (さまざまな予測エンジンに必要です)。結果が利用可能になると、エンドユーザー用公開ブロックチェーンを通じてブロードキャストされます (予測をリクエストしたユーザーのキーを使用して認証されます)。並行して、同じ結果が別の臨時キーを使用して暗号化され、公開ブロードキャストされます。このキーは、時間事前定義後にリリースされ、ユーザーが事前に要求していなくても、結果を後の段階で認証することができます。このメカニズムは、予測の「鮮度」に関連する様々なコストを複数の臨時キーを作成する事で調整することができます。

データの処理：コード実行は、AWS または類似の環境（例：GOLEM プロジェクト）で実行される独自の超弾性計算レイヤーの使用に基づいて、ブロックチェーン（公開予測、公開 RFP オーダー、非公開 RFP オーダー）と *Endor.coin* インフラストラクチャーで実行するために分割されます。データの処理と動作クラスタの抽出には複雑で高価な実行環境が必要なため、*Endor.coin* プロトコルの今後のリリースでもプロトコルの実行レイヤーが開かれます。これにより、新しいタイプのステークホルダーが、集中的な実行に特化した *Endor.coin* エコシステムに参加することが可能になります。分析されるデータがより大きく、クラスタ抽出アルゴリズムがより重い（計算上）と、実行に必要な *EDR* トークンで支払う価格は高くなります。

Payment: a user can pay for every prediction request an amount of *EDR* depending on the prediction complexity, dynamically defined by the *Endor.coin* Protocol, per the available resources and the demand at the time of request, and the number of users asking similar questions. 事前定義された質問は比較的安定した価格で維持されると予想されますが、RFP（予測リクエスト）は比較的高いコストで開始され、（定義上、少数のユーザーから始まるため）その後、関連性のあるユーザーのコミュニティとして徐々に減少していきます。このコミュニティの規模が大きくなればなるほど、最終的に誰でも支払いが少なくなり、新たな予測質問の作成者にネットワーク内の人々に「言葉を広げる」インセンティブが与えられます。

6.2 スマートコントラクト

Endor.coin プロトコルは、エンドユーザーにいくつかの基本的なプリミティブを提供します。

- `GetPrediction (prediction_def)` - フェーズ 1 で実行
- `PutPredictionReq (prediction_def)` - フェーズ 2 で実行
- `PutData (data_def)` - フェーズ 2 で実行
- `RunCustomPrediction(data_alg_prediction_def,price)` - フェーズ 3 で実行

これらのプリミティブは、顧客が *Endor.coin* プラットフォームで利用可能な予測を動的なコストで取得し、後でデータをアップロードして販売し、予測目的で使用できるようにします。プリミティブは、*Endor.coin* プロトコルのデフォルトのユースケースをカバーしていますが、今後のリリースでは、スマートコントラクトの展開をサポートすることで、より複雑な操作を `Get` と `Put` の上に設計できます。プロトコルが非公開データとカスタム非公開予測へ向けて進行するにつれて、最初の公開予測バージョンとして導入される基本的スマートコントラクト上にセキュリティと複雑性の追加レイヤーが使用可能になります。スマートコ

ントラクトにより、*Endor.coin* ユーザーは、トークンの使用、予測リクエスト、市場でのデータ検索ができるステータフルなプログラムを作成し、データ品質の妥当性を検証することができます。ユーザーは、スマートコントラクトの機能呼び出しをトリガーする取引を元帳に送信することにより、スマートコントラクトとインタラクティブになります。スマートコントラクトシステムは、*Endor.coin* 特有の操作（証明検証）、データアップロード（後の段階で公開または非公開で使用される）と、より一般的なスマートコントラクト特有の契約をサポートするように拡張されます。

getPrediction(prediction_def) : ユーザーが *EDR* トークンを支払って *Endor.coin* ネットワークに保存されている予測を取得できるようにします。顧客は、（ネットワークにオーダーを伝播することによって）検索予測市場オーダーブックに売り注文をすることによって、**Get Protocol** を開始する事ができます。予測提供者からの買い注文と一致すると、クライアントは予測への非公開臨時リンクを受け取ります。受け取った際に、両当事者は取引注文書に署名し、取引が成功したことを確認するためにブロックチェーンに送信します。予測消費者は、リクエストした予測の対価を支払います。予測は *Endor.coin* ウェブポータルからダウンロードすることができます。

putPredictionReq (prediction_def) : *Endor.coin* プラットフォームに送信することで、追加の予測タイプを提案することができます。このインターフェイスは、*Endor.coin* ウェブポータルで利用できます。提案された予測は *Endor.coin* のすべてのユーザーに表示され、ユーザーはそれらのランク付けを行うことができます。最も高く評価された予測は、継続的に成長している予測カタログに追加されます。予測リクエストは予測発行者の公開アドレスで識別されます。予測がカタログに追加されると、予測発行者は予測が他のユーザーによって消費されるたびに報酬を受けます。

putData (data_def) : データ提供者とパートナーのために使用されます。データ提供者は、データが予測に使用されている時は常に *EDR* トークンによって報酬を受け、*Endor.coin* プロトコル予測アルゴリズムによって自動的に価格が計算されます。この機能の変種は、*Endor.coin* プロジェクトの第 2 段階で利用できます。プロトコルの第 3 回リリースでは、ユーザーが独自のデータストリームをアップロードでき、「非公開」にマークし、それらをプラットフォーム上で使用可能な公開データストリームと融合させる「do-it-yourself」モードの API があります。クライアントは、（ブロックチェーンにオーダーを送信することによって）ストレージ市場のオーダーブックに売り注文を送信し、データ検証機関（例：予測エンジンプロバイダ）からの買い注文を待つことによって、**Put** プロトコルを開始します。クライアントは、予測エンジンによる行動クラスターの抽出に資金を提供する必要がありますが、エンドユーザーが使用するたびに販売価格を決定することができます。

runCustomPrediction (data_alg_prediction_def、price) : この方法は、プロトコルの第 3 回リリースで実行され、*Endor.coin* カタログには含まれていない予測の作成をサポートします。これが「DIY」モードです。入力には、例として、目的の動作の記述、または公開または独占データを参照したロジックの説明が含まれます。利用可能などのエンジンでリクエストを処理できるため、最高価格が定義されます。手数料は、予測に貢献した異なるプレイヤー（例：予測エンジンおよびデータ提供者）の間で調停されます。

6.3 開始時の *Endor* の役割

前のセクションでの詳細説明の通り、*Endor.coin* プロトコルは、多面的なプレイヤー（データ提供者、分析エンジン開発者など）間の相互通信を可能にします。この意味でプロトコルは、当事者間の予測ソースではなく、むしろイネイブラーと言語であり、「エッジ」で価値を創造することを可能にします。*Endor.coin* プロトコルは、さまざまな価値提供者間の容易なコミュニケーションを提供する TCP / IP プロトコルと同様に、データ提供者が商品を提供し、予測エンジンを使用して分析サービスを提供し、相乗効果の提供のために使用することができます。

このため、*Endor.coin* は、プロトコルの実行に焦点を当て、周辺のエコシステムの最適な構成を確保し、成長の加速を最大限に引き出すために、独立した法人の下で活動します。

しかし、*Endor.coin* プロジェクトは、プロトコルの生産グレード・ワークフローの迅速な確立を保証するために、顧客が可能な限り早い時期にさまざまな正確な予測を消費できるようにし（1 日目でも）、*Endor* をいくつかの重要な役割を果たす主役にします。その後、新しいプレイヤーが *Endor.coin* プロトコルを採用すると、*Endor* の役割は大きく減少し、*Endor.coin* エコシステムの新しい参加者によって導入される革新と価値につながります。

Endor.coin の立ち上げ時に *Endor* が提供する主な役割は次のとおりです。これらの各活動は、*EDR* トークンの取引で *ICO* 中に得た収益から配分される専用資金を必要となります。両社は現在、このコラボレーションの詳細について交渉中です。

公開データ提供者 : *Endor* は、さまざまなデータストリームを収集してクリーンアップし、分析の準備ができたデータとして *Endor.coin* インフラストラクチャにアップロードし、プロダクションレベルのインフラストラクチャを実行するために、その **Dev** チームを使用します。これらには、ビットコイン・ブロックチェーン、**ERC20** ブロックチェーンとその他の独自のブロックチェーンが含まれます。さらに、これは今後、ツイッター、**Reddit** 等のソーシャルネットワーク・フィードをサポートします。*Endor* は、このサービスを **HR** 経費をカバーする「コストプラス」ベースで提供します。

処理済みデータ提供者：*Endor* は、*Endor.coin* プロトコルをサポートする最初の分析エンジンとして機能し、取引データストリームから行動クラスターを抽出する機能を提供します。*Endor* は、AWS 経費をカバーする「コストプラス」ベースでこのサービスを提供します。

需要提供者：*Endor* は既存の（そして新しい）法人顧客からの切望する予測をチャネリングする需要提供者として行動します。*Endor* のカスタマーサクセス・クルーは、法人顧客が支払ったドルで *EDR* トークンを購入して、欲しい予測を生成します。

アプリケーションとパートナーシップ開発：Endor.coinのエコシステムをさらに拡大し、Endor.coin ベースの予測の普及を促進するため、Endorは、パートナーシップを通じて、または新しいタイプのビジネスを創り出すことによって、新しい予測ベースのビジネスの創造に積極的に取り組んでいきます。予測を行い、それらを公開する - カタリストとしての役割を果たします。これにより、Endorの事業開発チームだけでなく、その大規模で高品質の諮問委員会を活用することになります。

6.4 トークンの特権とトークン経済

前の章で説明したように、Endor.coinによって発行されたEDRトークンは、Endor.coinエコシステムによって提供されるさまざまなサービスの支払いメカニズムとして利用されます。EDRトークンの主な用途は次のとおりです。これらのサービスの一部は、その成長を促進するためにプロジェクトの初期段階で無償で提供されることにご注意ください。

事前定義またはDIY予測の購入：期待されるEDRトークンの主な用途は、予測の消費のための支払いメカニズムになる事です。これには、専用のセルフサービスAPIを使用してユーザーがオーダーする、今後増加するカタログで利用可能な事前定義予測や（後の段階での）「DIY」予測が含まれます。支払いは、主にクラウドコンピューティングリソース（例：Amazon AWS、ゴースト等）をカバーし、プロレートベースで予測作成に使用されるデータの提供者に10%から25%の支払いが割り当てられます。予測コストは、それを生成するために必要なクラウドリソースのコストに固定されます。

予測生成プロセスの主要な要素が異なる予測間で共有されるため、需要の急増に伴う急な変動に伴い、ユーザー数の増加に伴って漸的に減少するコスト関数が作成されます。

予測コストの10%から25%がデータ提供者に支払われます。

データ分析のためのクラウドリソース（AWSなど）は、クラウドコストの約80%～90%をアクティブユーザーの数で割ったものです（つまり、予測の全体コストから55%～80%をアクティブユーザー数である n で割った値）。

個人化された予測（AWSなど）のためのクラウドリソースのクラウドコストの約10%～20%（つまり、予測コスト全体の10%～15%）。

したがって、予測コストの高品質支払いモデルは次のようになります。

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

CPREDICTIONは、Endor.coinプラットフォームの起動時の単一予測のコストを表し、

n は、アクティブなプラットフォームユーザーの数を表します。

RFP（予測リクエスト）の送信：カタログでまだサポートされていない特定の予測に関心のあるユーザーは、**RFP** メカニズムを使用して新しい予測を作成することができます。そのような場合、ユーザーは、熟練した **Endor.coin** チームメンバーによる予測の作成および最適化に必要な時間と、必要なクラウドリソースをカバーする **EDR** トークンを支払います。この意味で、**RFP** は熟練したデータエンジニアにとってはクラウドコストと市場の平均給与の組み合わせに固定されます。

予測用 QOS プリファレンス・リクエスト：予測を購入するとき、または **RFP** メカニズムを介して新しい予測をリクエストするとき、ユーザーは予測のプレミアムアクセス権を購入するか、プレミアム以外の購入者にリリースされる数時間前に受け取るか、**RFP** メカニズムの場合は、事前定義された期間、独占的アクセス権を保持します。このようなサービス品質要素では、予測のコストに予め定義された係数を掛けて購入することが可能です。この意味では、このプレミアムサービスは、予測の基本コストの倍数であるため、クラウドリソースと市場給与のコストにもつながることになります。

プラットフォームの新しいデータストリームの送信：**Endor.coin** エコシステムの新しいデータストリームを提供したいデータ提供者は、データの初期分析に必要なクラウドリソースと **Endor.coin** プロトコルへの適応のコストを支払う必要があります。これは、予測の全体的な質を高め、高品質のデータソースの統合に積極的なインセンティブを提供するために必要です。この初期手数料は、データ寄稿者に、予測のための行動クラスターのソースとして **Endor.coin** プロトコルによってサポートされているデータを持つ権利を付与し、その後、予測を作成するために **EDR** トークンが使用されるときに、**EDR** トークンで補償される権利も含まれます。

6.5 収益の使途

EDR トークンの購入から得られた収益は、**Endor.coin** プロトコルソフトウェアインフラストラクチャを実装し、マーケティングとの戦略的コラボレーションによる初期成長と採用を支援し、必要なクラウドリソースを資金提供し、必要な法的管理費用をカバーします。

収益の主要部分は、最先端のクラウドとブロックチェーンのエンジニアだけでなく、データ科学、機械学習、社会物理学の世界の専門家で構成される **Endor.coin** R&D チームに割り当てられます。

Endor.coin プロジェクトとチームの重要な役割は、できるだけ多くの主要プレイヤーによる *Endor.coin* プロトコルの迅速な採用を保証することです。これが達成する方法の1つは、学会と共同研究活動を行い、*Endor.coin* プロトコルとインフラストラクチャへの主要な研究を初日から可能にする事です。収益の最大 10%は、そのような活動を支援するために、世界有数の研究機関に割り当てられます。

収益の最大 30%は、*Endor.coin* プロトコルと周辺のエコシステムの適切なブートストラップを促進し、加速するために必要な予測エンジンや ETL コネクタなどの独自の技術の購入に割り当てられます。

第7章

技術的優位性と差別化

7.1 MIT ファーナスからの科学革命

マサチューセッツ工科大学（MIT）の *社会物理学* の最新技術[1]の提供により、*Endor.coin* は世界で最も先進的な行動予測技術を活用しています。10年前に MIT メディア研究所で始まったこの科学の画期的な進歩は、行動データ分析の分野に革命をもたらし、有名な *DARPA ネットワークチャレンジ*[21]の受賞を含め、テクノロジーの分野で多くの業績を積み上げ、個人投資家のコミュニティのリターンを高め[9、13]、効率的に未知のサイバー攻撃の存在を予測することに成功しました[35]。この技術は、何百もの科学論文、数十の特許、これらのテーマに特化した数冊の書籍を出版している学会と業界の専門家チームによって開発されました。

そして今、この革命は、大手小売業者、銀行、大手テクノロジー企業に限られた特権だった機能を専門家でもそうでない人でも利用できるように、ブロックチェーン技術を用いて一般公開されています。社会物理学の詳細は、セクション 3.1 と付録 A をご参照ください。

7.2 現実の製品。実証済みの技術

セクション 3.2 で説明した通り、Endor は大手銀行や大手小売業者に予測のボトルネックをオープンにする SaaS ソリューションを提供するように設計された製品として社会物理学の実装を実行する MIT のスピンオフ企業[14]です。当社は大手投資家の支援を受けており、フォーチュン誌企業上位 500 社の *コココーラ社*[16]、*マスターカード社*[17]、*ウォルマート社*などとのプロジェクトに取り組み、さまざまなユースケースに対して正確な予測を自動的に生成する能力を実証しました。

ブロックチェーンと仮想通貨分析技術の価値は、[27]などで実証されました。Endor の製品は、*Money-2020* や *Finnovate 2017* [18]などの主要なイベントで紹介されています。Endor は *ガートナー社認定クールベンダー* [19]であり、*世界経済フォーラム*によって「技術のパイオニア」と認定されました[20]。

7.3 ユーザーに対するユーザビリティと価値、トークン所有者のための価値

A key uniqueness of *Endor.coin* is that the *EDR* Token will be usable on day 1 of the token launch – offering token holders complete access to the pre-defined predictions. さらに、ベータユーザーのグループがトークンの開始直後に選ばれ、事前定義予測に予測リクエストを追加する事ができます。

さらに、将来的に予測が手頃な価格で使用できると信じるトークン所有者は、*EDR* トークンの購入と保有のインセンティブを与えられ、期待通りに使用量が増えると、予測のコストが増加し、トークンの価格が上昇し、*EDR* トークンを早期に購入した予測ユーザーは報酬を受ける事ができます。

第 8 章 チーム

8.1 主要チームメンバー

Yaniv Altshuler 博士

Altshuler 博士は *Endor* の CEO であり、MIT メディア研究所の研究員であり、Pentland 教授と共に、群衆の行動をモデル化する新しい科学である「社会物理学」を開発しました。At MIT, the technology was used to win the prestigious *DARPA Challenge* [21], help the Government of Singapore improve its ability to predict traffic jams, and assist a community of thousands of financial investors to improve their financial returns [13]. *Endor* では、このテクノロジーを用いて、大手銀行や大手小売業者[14]に、効果的な様々なユースケースでの群衆行動の正確な予測を提供しました。

Altshuler 氏は、MIT での研究と *Endor* の設立前は、*IBM* の研究員であり、スーパーコンピュータのパフォーマンスを向上させるための新しい最適化技術を開発しました。Altshuler 博士は 2011 年以来ブロックチェーン研究を活発に行っており、60 以上の科学論文を執筆し、15 件の特許を出願しています。Altshuler 氏の業績は、*ファイナンシャルタイムズ* [10]、*ハーバードビジネスレビュー* [11]などの著名メディアで紹介されています。彼の最近の著書には、「ソーシャルネットワークのセキュリティとプライバシー」[26]と「スウォームとネットワークインテリジェンスの検索」[25]があります。

Alex 「Sandy」 Pentland 教授 (科学諮問委員会委員長)

MIT メディア研究所起業プログラムおよび MIT コネクション科学・人間工学ラボ・ディレクター。Pentland 教授は、世界で最も賞賛されている科学者の 1 人であり [23]、最近、*フォーブス*によって、*Google* 創設者や米国政府 CTO とともに、「世界で最も強力な 7 人のデータ科学者」に挙げられました[24]。彼は、個人の業績により、*ハーバードビジネスレビュー* [22]の *マッキンゼイ賞*、*DARPA* の *インターネット 40 周年記念賞* (21)、*ブランディーズ賞*など数多くの賞を受賞しています。

同氏は、*グーグル*、*AT&T*、*日産*、*国連事務総長*による諮問委員会の創設メンバーであり、数多くの企業や *Data Transparency Lab*、*Harvard-ODI-MIT DataPop Alliance*、*Institute for Data Driven Design* などの社会的企業を共同設立した起業家です。彼は *米国国立科学アカデミー*のメンバーであり、*世界経済フォーラム*のリーダーでもあります。彼の最近の著書に「社会物理学」と「正直なシグナル」があります。

Stav Grinshpon (スタヴ・グリーンシュポン)

Grinshpon 氏は、ベテランの IT 業界専門家であり、*SAP* や *AT&T* などの企業で 18 年間のプロダクトとマネージメントの経験を持っています。Grinshpon 氏はサイバー防衛の世界的な専門家であり、イスラエル 8200 技術部門で技術リーダーとして 8 年間働いていました。Grinshpon 氏は、データ分析に焦点を当てた 3 つの特許の権利者であり、*Endor* の R&D 活動を率いていました。

David Shrier (デビッド・シュリエ)

David Shrier は、金融革新に関する世界的に認められた権威であり、オックスフォード大学のオンラインプログラムであるオックスフォード・フィンテックとオックスフォード・ブロックチェーン戦略、MIT の未来コマースを主導しています。彼は、「サイバーセキュリティの新しいソリューション」、「金融テクノロジーの最前線」、「Trust :: Data 」[27、29、32]など、フィンテック、ブロックチェーン、サイバーセキュリティに関する多くの書籍を出版しています。

Shrier 氏は、MIT の研究員による機械学習企業 *Distilled Analytics* の CEO であり、金融サービス業界を行動分析で変革しています。また、AI 主導のコラボレーション・テクノロジー・プラットフォーム・プロバイダである *Riff Learning* の会長も兼任しています。David 氏は、サイド・ビジネス・スクール、オックスフォード大学の准フェローであり、MIT メディア研究所の講師を務めています。彼はまた、ブロック・チェーンとデジタルアイデンティティ (ミレニアム・アドバイザーズ) についてのドバイ政府のコンサルタントであり、テクノロジー・トレンド、ブロックチェーンベースのデジタル商品取引所である *Cleer.digital*、戦略における中間市場の信用流動性プロバイダーでもあります。

David 氏は、*WorldQuant University* の諮問委員会のメンバーです。*WebQuant University* は、完全に無料で、金融工学のオンライン修士号を提供する公認プログラムです。彼はこれまで、欧州委員会に対し、デジタル技術を中心とした技術革新の商業化を勧告してきました。彼は現在、米国証券業界の自己規制機関である *FINRA* のフィンテック産業委員会のメンバーであり、金融サービスに影響を及ぼす新規開発に

関するカウンセリングを行っています。

Mihaela Ulieru (ミハエラ・ウリエル) 教授

この人工知能の専門家は、*世界経済フォーラム*[36]のブロックチェーン・チャンピオンです。彼女の分散型インテリジェントシステムの研究は、製造、物流、国土安全保障に革命を起こし、組織的技術としてのブロックチェーンのガバナンスのための強力な基盤を作り出しました。カリフォルニア大学バークレー校の **Ulieru** 教授は、世界経済フォーラムのグローバルアジェンダ協議会、シンガポールの科学工学研究評議会、カナダ科学技術革新評議会のメンバーです。

Goren Gordon (ゴレン・ゴードン) 博士

Gordon 博士はテルアビブ大学のキュリオシティ研究所の代表であり、コンピュータ・キュリオシティの最先端モデルを開発しています。ディープラーニングとニューラルネットワーク最適化の専門家である **Gordon** 氏は、量子物理学の博士号と神経生物学の博士号を取得しています。ゴードン博士は、MIT メディア研究所で **Cynthia Breazeal** (シンシア・ブレアジール) 教授と共同研究を行い、好奇心を抱いたロボットが好奇心を抱いた子どもたちとどのように交流するかを研究しました。ゴードン氏は科学教育にも興味があります。この為、量子物理学を遊びを通して子供に教える「量子コンピュータゲーム」を開発しました。ゴードン氏はまた、量子物理学、脳、学際的思考題について人気講演者です。

Arie Matsliah (アリー・マトスリア) 博士

Matsliah 博士は、グラフ分析理論の世界的な専門家であり、*グーグル社*、*IBM 社*、*インテル社*、*Lyft 社*などの巨大企業で16年間働き、成功した **Menlo-Park** ベースのスタートアップ企業である **TripActions** のチーフアーキテクトを務めました。**Matsliah** 博士はまた、アルゴリズム、複雑さ、量子コンピューティングの基礎研究に焦点を当てた40以上の科学論文を発表しました。

Shahar Somin-Gavrielov (シャハール・ソミン・ガブリエロフ)

Somin-Gavrielov 氏は、統計学学習理論の専門家であり、ヘブライ大学の修士号（優等学位）を取得しています。**Somin-Gavrielov** 氏は、データ科学の深層理論実現と実践的な業界経験を持ち合わせた熟練した研究者です。彼女は過去に、イスラエルの8200情報部門の優秀なアナリストでした。

Edo Eisenberg (エド・エイゼンバーグ)

金融リスク・マネージメントの専門家。以前は *モルガンスタンレー* と *バークレイズ* で個人向けクレジット・ポートフォリオをマネージメントしていました。**NICE** で9年間の経験があり、米国の銀行トップ10のうち7行に導入された不正防止ソリューション

ションの設計と実装を行いました。デューク MBA プログラム修了。

Lior Regev (リオール・レゲフ)

Regev 氏は、テクノロジーに対する大きな情熱を持った経験豊富なソフトウェアエンジニアです。Regev 氏は、分散型システム、クラウドアーキテクチャ、SaaS 製品で幅広い経験を持つ元インテリジェンス技術リーダーです。

Liat Yitshaki

Yitshaki 氏は、公法訴訟の専門家です。彼女は法と倫理の修士号を取得しており、ロンドン大学で優秀な成績を修め LLB を取得しています。これまで Yitshaki 氏は、McKinsey&Co.で勤務し、英国政府の上級法律顧問としても働いていました。

8.2 顧問

Alexander Lipton (アレクサンダー・リプトン) 教授

リプトン教授は、バンク・オブ・アメリカのグローバル定量ソリューションの最高業務責任者、メリルリンチのグローバル定量および与信分析部門の最高業務責任者を務めました。それ以前は、リプトン教授は、シタデルの定量的研究の責任者であり、クレディ・スイスのエクイティ・デリバティブの責任者でした。リプトン氏は現在 EPFL で金融工学の教授を務めており、MIT コネクション科学センターのフェローでもあり、近著に Scientific American 誌の記事「テクノロジーが崩壊した金融システムの修復にどのように役立つか」[37]があります。

Ron Gross (ロン・グロス)

Ron Gross 氏は、イスラエル・ビットコイン協会の創設者であり、役員を務めています。彼は 2011 年以来ビットコインとブロックチェーンの世界で活躍しており、世界初の ICO である Mastercoin (現 Omni) の常任理事を務めました。この元グーグル社員の成功者は、Commerce Science 社の主任アーキテクトを務めていました。

Nuria Oliver (ヌリア・オリバー) 博士

MIT 卒のオリバー博士は、ボーダフォン社のデータ科学研究ディレクター、Data-Pop Alliance の主任データ科学者を務めています。以前は、Telefonica R&D 社の科学ディレクター、マイクロソフト社の研究員でした。オリバー氏は、ビッグデータ分析を民間セクターの企業が公共の利益のために共有するコラボレーションの一形態として使用されるデータ・フィランソロピーの世界的な専門家です。オリバー氏は、90 以上の科学論文と書籍を執筆しており、40 件以上の特許出願をしています。

Daniel Tunkelang (ダニエル・タン克蘭) 博士

世界的データ科学専門家である Tunkelang 氏は元グーグル社員、IBM 社員であり、アップル社、Salesforce Etsy 社、Yelp 社、Pinterest 社などの大手テクノロジー企業の顧問を務めています。これまで Tunkelang 氏は、LinkedIn の検索のエンジニアリング・ディレクター、後には常任データ科学者として、また Endeca 社（オラクル社が買収）の主任科学者として働いていました。Tunkelang 氏は MIT で修士号、カーネギー・メロン大学で博士号を取得しています。

Guy Zyskind (ガイ・ジスキンド)

シークレット・スマート・コントラクトのためのエニグマ・プロトコルを作成した企業であるエニグマ社、そして洗練されたツールとデータを使用して誰もが仮想通貨ヘッジファンドを開始できるプラットフォームであるカタリスト社の設立者兼 CEO。ガイ氏は、エニグマ社設立前は MIT の大学院生であり、ブロックチェーン技術を研究し教えていました。ガイ氏は数点の学術論文を執筆し、最近では、エニグマのホワイトペーパーや「個人情報分散化：個人データを保護するためのブロックチェーンの使用」など、プライバシーとブロックチェーンに関する著作があります。ガイ氏は、電気工学とコンピュータ科学において、MIT で修士号、テル・アビブ大学で学士号を取得しました。

Michael Bronstein (マイケル・ブロンシュタイン) 教授

起業家にして主要な研究者であるブロンシュタイン教授は、ハーバード大学の研究フェロー、ルガノ大学の情報学教授を務めています。ブロンシュタイン氏は、主要なジャーナルや会議で 100 冊以上の出版物、20 件以上の特許、研究論文「非剛体の数値幾何学」の著者であり、4 冊の書籍を編集しています。ブロンシュタイン氏は、欧州研究評議会 (ERC) 助成金を 3 回授与された数少ない研究者の 1 人であり、グーグル優秀研究賞、ハーバード大学の Radcliffe 奨学金、Rudolf Diesel 産業奨学金も授与されました。世界経済フォーラムで、彼は 40 歳未満の世界の 40 人の主要研究者の 1 人に選ばれました。

ブロンシュタイン教授は産業アプリケーションに積極的に関わっています。同氏は、Novafora 社 (Turner Broadcast 社にライセンス供与) の共同設立者であり、技術担当副社長を務めています。また、3D センシング技術会社の Invision 社の共同設立者であり、主要開発者の一人でした (同社はインテル社に買収され、ブロンシュタイン氏は現在、知覚コンピューティングの研究科学者を務めています)。ブロンシュタイン氏は Videocites 社の共同設立者で技術顧問でもあります。

Wei Pan (ウェイ・パン) 博士

Pan 博士は MIT の卒業生であり、元 Google 社員であり、ビッグデータ分析、機械学習、複合システムに関する世界的専門家です。彼は、社会システムの理解とモデルと、群衆のビッグデータの測定を通じて金融市場のダイナミクスとマクロ経済を理解しよ

うとする「現実ヘッジ」と呼ばれる分析アプローチの発明者です。現在は、ニューヨークを拠点とするスタートアップ企業である *Thasos* グループの共同創設者兼主任科学者を務めています。Pan 氏は以前、*Fidelity Investments* 社で主任エコノミストのもとで働いていました。そこでは、システム研究とフラッシュクラッシュに焦点を当てていました。

Igor Gonta (アイガー・ゴンタ)

Gonta 氏は、エンジニアリングおよびコンピュータ科学を専攻した MIT 卒業生であり、*BlackRock* 社で商品およびリスクマネージメントを担当しています。Gonta 氏は過去に、ソーシャルメディアの会話分析に基づいたリアルタイムの株式センチメント生成エンジンである *Market Prophit* の CEO を務めました。Gonta 氏は同社の創設者であり、その後、大手ヘッジファンドに売却されました。彼の初期のキャリアにおいて、Gonta 氏はバークレイズの商品販売担当最高業務責任者、ゴールドマンサックスの商品セールス担当副社長を務めました。

Thomas Hardjono (トーマス・ハードジョーノ)

Hardjono 氏は分散型アイデンティティ、ブロックチェーン、スマートコントラクトを専門とするセキュリティ専門家です。Hardjono 氏は、*VeriSign* 社と MIT Kerberos Consortium の執行役員を務め、コンピュータのセキュリティと暗号化に特化した 5 冊の書籍を出版しています[38-42]。

付録 A-社会物理学

説明済み

1. 社会物理学の働きは？

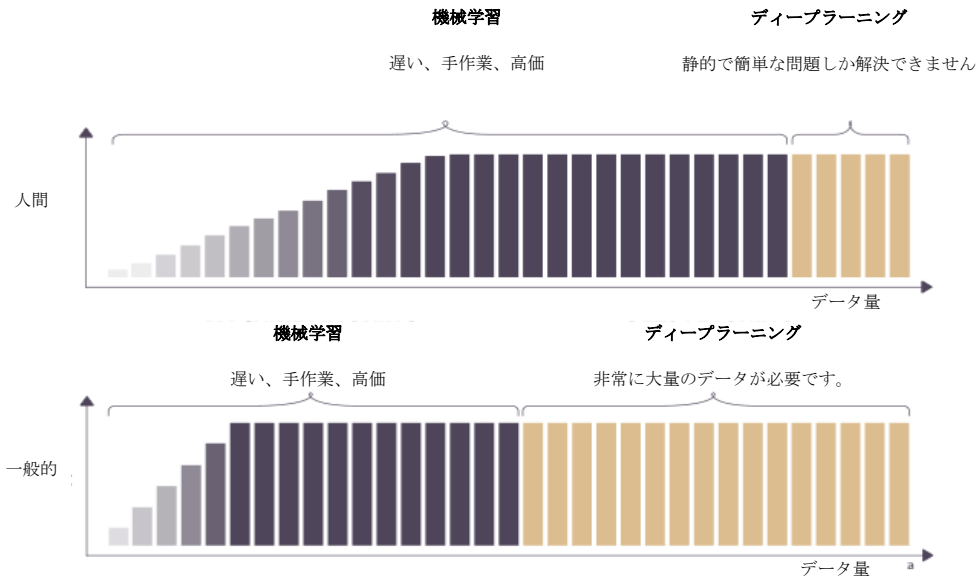
明確化：この文書は、社会物理学、その背後にある高水準の原則、および Endor の技術的な実装についての包括的な見解を提供することを意図しています。しかし、Endor によって使用される社会物理学の法則の特定の数学的定式化に関する特定の技術的詳細、およびこれらの法律違反を検出するために使用されるメカニズムの具体的な実施は、IP の考慮により省略します。

情報化時代には、企業はビジネスオペレーションに関する多くの情報源からのあらゆるタイプのデータを収集します。データには、画像や動画、テキスト、ツイート、取引、利用状況ログが含まれます。しかし、大部分のデータは、単一の基礎をなすソース（人々）から発生します。

したがって、例えば、ツイートやブログの投稿は人間によって人間のために書かれています。購入取引情報や通話情報には、物や他人への人間の欲求が表れます。使用状況やアプリログからは、人々がコンピュータやモバイルデバイスとどのようにやり取りするかが判ります。

人間の行動から派生したデータは「乱雑」であり、動的で複雑で多方面に渡っています。このようなデジタルデータチャンネルに記録されている人間の行動は、時間の経過とともに大幅に変化し、複雑なソーシャルネットワークの影響を受け、高度にマルチモーダルなデータストリームで伝えられます。これらの特性は、事業活動を改善するために顧客の行動を分析し、理解し、予測したい企業にとって大きな課題となっています。

近年、データ科学者は、この複雑性に対処するために、「重量級の」統計手法と機械学習アルゴリズムを採用し始めています。新しい「ディープラーニング」技術などの強力なツールは、行動パターンを分類し、異常を検出し、将来のトレンドを予測できるように、データを収集し、その属性を分析します。しかし、歴史的に画像処理や文字認識などの「静的な問題」のために開発されたツールは、人間の行動データに簡単に対処できません。動的で複雑で多方面に渡るデータストリームを学習することは非常に困難であり、ほとんど不可能です。

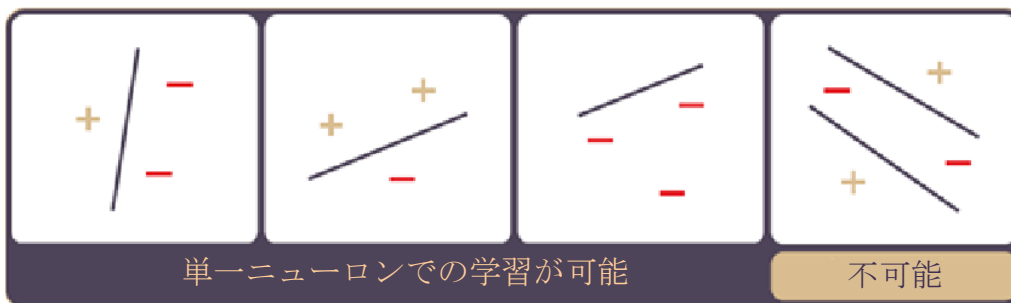


Endor の社会物理学エンジンは、まるで違う方法で動作します。入力データからパターンを引き出す代わりに、人間の行動データには、同じスペース十分な人数の人が行動するたばに出現する数学的な関係という共通の「社会行動法」のセットが含まれています。これらの法則は、データの種類、それを作成したユーザーの人口統計、またはデータサイズに関係なく、群衆の行動のさまざまな統計的特性が時間とともに進化する方法を規定しています。Endor は、これらの法則をデータ分析エンジンに統合しました。データ分析エンジンは、入力情報（通話、タクシー乗車、財政投資など）として提供されている生データに含まれるすべての人の社会的属性を効率的に抽出します。

2.1 → なぜ社会物理学が必要なのですか？

抽象的に、学習の問題やオブジェクトを分類したり、将来のイベント予測を生成する能力には、分析するデータが必要があり、更にそれを分析するアルゴリズムが必要です。必要なデータの量は、いくつかの要因によります。

1. **問題の内部の複雑性** - 問題はさまざまな形や大きさで現れ、解決の難しさも異なります。問題の「難しさ」または「複雑性」は、学習アルゴリズムが問題をうまく学ぶために最低限必要とする「強度」を指します。学習アルゴリズムが十分に強くなければ、問題のインスタンスを正しく学習することはできません。例えば、「パーセプトロン」（つまり、単一のニューロンからなる最も単純なニューラルネットワーク）は、「XOR」機能（つまり、ブールの「排他的 OR」機能）を決して学ぶことができないことを容易に示すことができます。なぜなら、パーセプトロンの働き（これは入力空間を線形に分割すると考えられる）に埋め込まれている XOR 関数が、このような表現には単純すぎるからです。



XOR 関数の例が右のグラフに表示されています (出典: ウィキペディア)

問題の基礎となるモデルの学習複雑性は、問題の Vapnik-Chervonenkis 次元 (VC 次元) と呼ばれます。問題の学習複雑度が高ければ高いほど、それを正しく学習するためには予測アルゴリズムが強くなければなりません。適切にモデル化するためにはそのようなアルゴリズムが必要とするデータが多くなります。例えば、線形の複雑性を有する関数 $y = f(x)$ (ランク 1 の多項式を使用してよくモデル化することができる) は、定義により、正確にモデル化するためには、より高い次数の多項式を必要とする関数よりもサンプリングポイントが少なくなります。

アルゴリズムの学習効率

適切なモデルパラメータを確認するためには、それぞれ異なる量のトレーニングデータとエキスパートドメイン知識が必要な、多くの学習アルゴリズムがあります。たとえば、単純回帰では大量のデータと多くの問題特有の機能がうまくいくのに対して、ディープラーニングは膨大な量のデータを必要としますが、ドメイン機能を自動的に学ぶことができます。したがって、学習プロセスの全体的な効率、つまり、基礎となるモデルを学習するために必要なデータの量、およびアルゴリズムがどれだけ適切な表現を学習するかを定量化することができます。

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

M=モデルのビット数は、データを公式化またはモデル化するために必要なビット数であり、つまりパラメータの数は、モデルの完全な説明に必要となります。この意味で、**M**は、問題のコルモゴロフ複雑性に類似しています。これは、データの複雑性の既知の理論的尺度であり、与えられたデータを生成できる最短のチューリング機械を指しています。モデルは問題によって決定され、変更することはできません。つまり、単純な問題の場合、モデルのビット数は殆どありません。

R_A(M) =モデル特有の表現のビット数は、アルゴリズム **A** によって学習されたビット数であり、基礎となるモデル **M** を表します。この数は、アルゴリズムの自動特徴検出を表し、手動でアルゴリズムにプログラムする必要があるドメイン固有の知識のエキスパートの数に反比例します。

D_A(M) =データのビット数は、アルゴリズム **A** がモデル **M** を学習するために必要とするビット数です。

η_A=アルゴリズム **A** の学習効率、すなわち、アルゴリズムの自動表現学習とその学習品質を保持するために必要なデータ量との比。高効率アルゴリズムは、少量のデータで適切な表現を学ぶことができますが、効率の低いアルゴリズムでは、手作業で特徴を作り、膨大な量のデータを調整する必要があります。

例えば、問題クラス **M** 一部のアルゴリズムでは、学習の質を維持するために、他のアルゴリズムよりも多くのデータが必要になります。

- ロジスティック回帰アルゴリズムでは、通常、エキスパートによる機能の手動コーディングと、大量のデータを手元の特定の問題に合わせて微調整する必要があります。

$$R_M(A) \ll 1$$

$$D_M(A) \gg 1$$

$$\eta_A \ll 1$$

- ワンショット学習アルゴリズムでは、エキスパートドメインの機能も必要ですが、いくつかの例を使用して、基礎となるモデルを微調整し、予測することができます。

$$R_M(A) \ll 1$$

$$D_M(A) \ll 1$$

$$\eta_A \sim 1$$

- ディープラーニング アルゴリズムは最も有益な特徴を自動的に学習することができますが、膨大な量のデータを必要とします。

$$R_M(A) \gg 1$$

$$D_M(A) \gg 1$$

$$\eta_A \sim 1$$

- Endor のアルゴリズムは、社会物理学を使用して、データのわずかなサンプルから関連する動作特徴を自動的に抽出します。

$$R_M(A) \gg 1$$

$$D_M(A) \ll 1$$

$$\eta_A \gg 1$$

2. 問題の変化率 - 正確な予測を生成するために必要なデータの量に影響するもう 1 つの要素は、問題の基礎となるモデルの変化率です。いくつかの問題は静的なもので、基本的なパラメータが変更されたり、まれにしか変更されません。たとえば、画像内の顔は数か月にわたって変化しません。顔は顔です。一方で、有償サービスからの濫造につながる基本的な行動パターンは、競合他社によるマーケティングキャンペーンの成功への対応として、数か月にわたる社会的変化を経て徐々に、または数日で急速に変化する可能性があります。このダイナミクスを以下のように定量化します。

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

τ は問題の持続性または強さであり、基礎となるモデルが変化する速度を表し、 ∂T はモデルが ∂M ビットだけ変化する持続時間を表します。

例えば、 $\tau = 1 \text{ 日} / 10 \text{ ビット}$ は、 $\tau = 1 \text{ ヶ月} / 10 \text{ ビット}$ と比較して 1 日の期間（すなわち、変化率の速さ）に劇的に変化することを意味し、これはモデルの変化がはるかに遅いことを意味します。動的モデルは、効果的に、全ての期間で異なるモデルを提示し、モデルを再トレーニングまたは再学習する必要があります。

予測分析の運用上の実装：変化するモデルの追跡

したがって、与えられた問題に対する予測分析の実行可能性の主な課題は、モデル化されたオブジェクトの行動変化に対処するのに十分な情報を得ることです。1回の単位あたりに蓄積できる情報ビットの量を It と表します。

非常に大きなソーシャルネットワークや検索エンジン（グーグルやフェイスブックなど）のオペレーターは、短期間、膨大な情報を蓄積することがあります。しかし、これは顧客の行動予測に関心を持つ大多数の企業にとっては不可能です。

さらに、大量の情報を毎日得る大規模プレーヤーであっても、(a) 複雑すぎる、または (b) 変化が速すぎる、またはもちろん (c) (a) と (b) の両方の問題を正確にモデル化するのは難しいでしょう。

しかし、我々には企業が運用予測分析ソリューションを実行する能力を決定する簡単な方程式があります。予測したい問題のために、採用する学習アルゴリズムと継続的に取得する新しいデータを使用してこの原則を満たすことができる企業には、この目標を達成するための運用プロセスの構築に成功する企業もあれば、失敗する企業もあります。

学習原理の基本操作：

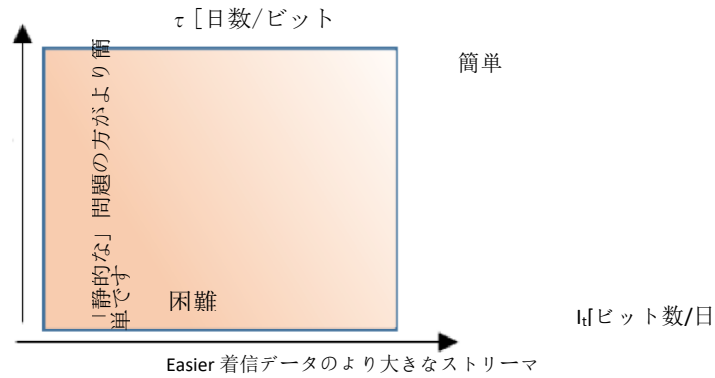
$$It \times \tau \times \eta A >$$

この原則の実際の意味は、機能的な予測システムの構築に失敗した企業にどちらを選択すべきかを指示する事です。

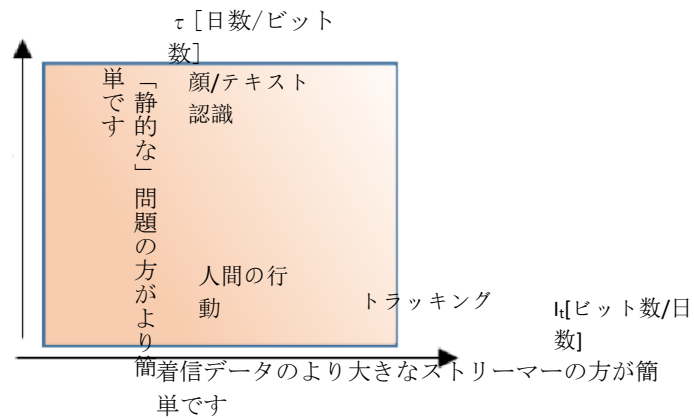
1. 関連データの取得量をより多くする事で、データ収集帯域幅を改善します。
または

2. より静的な問題に焦点を当てる。または
3. 学習アルゴリズムをより効率的にリソートします。

この単純な関係は、次の図に示すように、問題の複雑さが増し、その持続性が低下するにつれて、1日あたりの必要な情報量がより多くなることを意味します。



ここでは、これを使って分析すると、いくつかの共通予測問題がどのように見えるかを示します

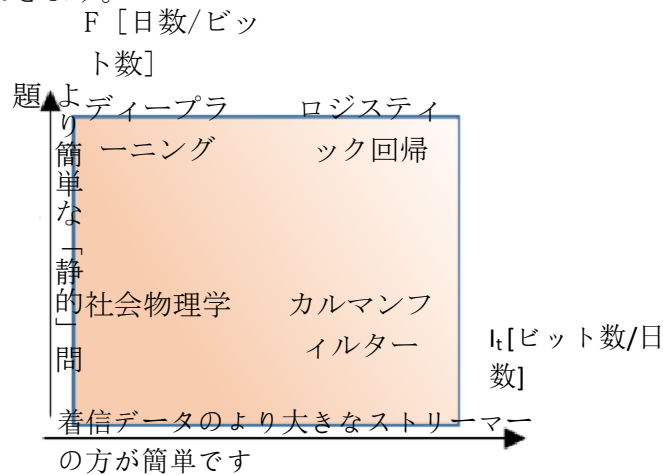


- **顔認識**：この問題は、一般的に言えば、顔は変化しないため、非常に低い変化率、すなわち $\tau \gg 1$ と特徴付けられます。したがって、 ηA が比較的小さく非効率な学習アルゴリズムであっても、それはまだ非常に小さく、問題を最初に学ぶためには多くの情報を必要としますが、再トレーニングには多くの情報を必要せず、それほど大きな変化はありません。
- **モバイルターゲットの操縦のトラッキング**：モバイルターゲットのトラッキングは、非直近の過去のデータは役に立たなくなるターゲットの高速移動（または、ナビゲーション・アルゴリズムによる検出を避けるために導入された軌道または動的で局所的な急速な変化）のために、 $\tau \ll 1$ によって特徴付けられます。しかし、このよう

な問題は、 $lt - 1/\tau \gg 1$ を与える着信データの大きなストリームによって特徴付けられます。これは、ターゲットの位置の動的性質が、着信データストリームによって学習アルゴリズムに完全に伝達されることを意味します。これにより、 $\eta_A \sim 1$ の比較的簡単なアルゴリズムで問題を効率的に解決できます。

- **人間の行動**：人間の行動は非常に動的であり、 $\tau \ll 1$ であり、比較的速く変化する要素を含むことを意味します。さらに、人間の行動を予測したい企業は、顧客に関する追加情報を得る（そして行う）ことができますが、それは常に実際の行動変化の部分的な誘導です。比喩的に言えば、これはプラトンの洞窟のたとえ話に似ています。我々の知覚情報は、非常に粗いレンズを通過するため、いくつかの側面しか捕らえることができません。我々の形式論では、これは $l_t \ll 1/\tau$ に変換され、その後、 $l_t \geq 1$ の意味になります。これは、人間の行動を効率的に予測するためには、非常に効率的な $\eta_A \gg 1$ アルゴリズムを採用しなければならないことを意味します。

同様に、この問題をモデル化して、それぞれの問題に最適なソリューションのテクニックの違いを見ることができます。



この図は、前述の基本的操作学習を示しています。

原理 $l_t \times \tau \times \eta_A > 1$. より静的な問題と、それに関するデータが多くなればなるほど、グラフの右上隅に近きます（そして予測が

より正確になる可能性が高くなります）。また、右上に最も近いのは正確な予測を生成するために使用する低効率のアルゴリズムです。

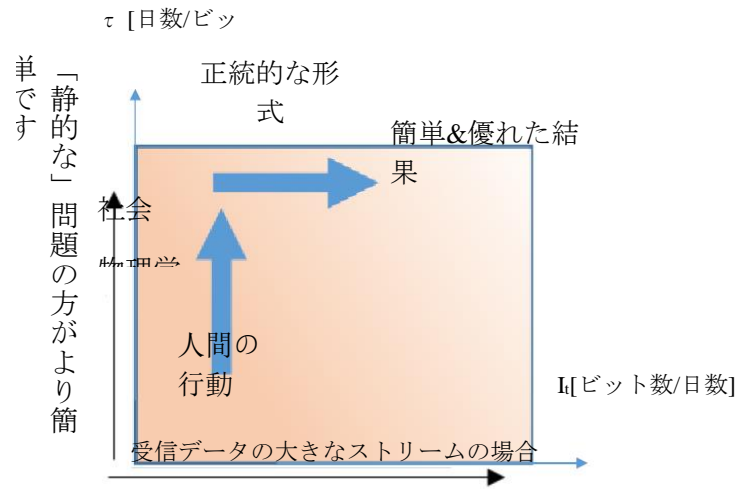
膨大な量の情報を必要とする技術は動的な問題ではうまく機能しないため、ディープラーニングは画像処理やジェスチャー認識などの基本構造が急速に変化しない問題に最適です。対照的に、情報を迅速に処理できる単純なアルゴリズム（例：カルマンフィルター）は、動的な問題に対処できますが、予測に成功するには高いスループットの情報が必要です。ロジスティック回帰などの従来の機械学習へのアプローチは、問題が非常に静的で、着信トレーニングデータの量が比較的多いシナリオでは効果的です。

終わりのない追跡のソリューションとしての社会物理学

Endor は、2つの観点からモデル変更の為の継続的な大量のデータの取得の必要性の問題に取り組んでいます。

- **動的問題の静的モデルへの変換**：社会物理学の法則では、データのタイプと起源は不変性があり、不可知的です。したがって、それらを使って未加工データを社会物理学の正統的表現空間に投影すると、（社会物理学の法則固有の静的性質の為に）元の問題は元のモデルが静的であった新しいクラスの問題のインスタンスに変換されます。これにより、元の問題の実際の $T \ll 1$ は非常に大きな $T_{SP} \gg 1$ に変換されます。
- **スモールデータから「ビッグデータ」を作成する**：前述のように、顧客の行動を予測するという課題に直面したとき、ほとんどの企業は大きなハードルに直面します。所有する関連データの量が不十分なのです。これは、動的問題（頻繁なマーケティング・ユースケース）や新しい要素の導入（新しい製品への反応の予測や新しい種類の入力データの使用など）に関連する問題の場合に特に当てはまります。Endor は、タイプ、サイズ、ソースに関係なく、顧客から受け取ったすべてのデータストリームを社会物理学の形式に変換します。この標準形式を使用することにより、Endor は、すべてのクライアントおよびすべての質問からのすべてのデータを、一貫して高いペースで成長する単一の非常に大きなデータベースに統合することができます（つまり、 $k \gg 1$ で）。これは、社会物理学の正統な形式（特定の質問や顧客ではない）のインスタンスを分析することで、単一で不変のディープラーニング・ネットワークをトレーニングするために使用されます。したがって、たとえ各顧客が提供できるデータが非常に限られていても、エンド・ユーザーは膨大な（正統な形式の）データを蓄積することができます。

Endor のエンジンは、人間の行動予測の最も困難な問題を、社会物理学を介してゆっくりと変化するものに変換し、ビッグデータを（正統な形式への変換を介して）「簡単に効率的な」問題にし、ディープラーニングツールを使用して解決します。これは次の図に示されています。



2.2 社会物理学：概要

前のセクションでは、セマンティクスに依存しない静的数学的不変量を使用して、人間の行動を予測するという固有の課題を克服するために社会物理学をどのように使用できるかを簡単に説明しました。これがどのように実現可能かをよりよく理解するためには、ニュートンの第 2 法則や運動量保存法などの物理学の法則を考慮しなければなりません。オブジェクトは、外的な力が作用しない限り、初期パスを維持します。このような隠された力の存在を推測するために、「データを学ぶ」、その統計的な属性を理解する、または同様の方法で行動する多くのシステムをテストする必要はないことにご注意ください。物理学の法則を用いれば、法則に違反していれば異常だとわかる為、すぐに検出して、「見えない手」の結果と解釈することができます。突然方向を変える物体が検出された場合、力が作用したとすぐに推測することができます。変化が似ていると、同じ力がすべてに作用した可能性が最も高くなります。この単純な実現は、物理学の法則を理解していなければ不可能です。

社会物理学は物理学の法則よりも絶対的で厳格ではありませんが、コンセプトは似ています。何か社会物理学の法則に違反した場合、それは現実世界の貴重な財産または帰属のデータ表現である「興味がある」として直ちに認識することができます。これには、学習、ベンチマーク、ベースライン、その他のデータ科学や機械学習ツールは必要ありません。社会物理学の法則違反は、それを生成したデータタイプに関係なく、非常に高速かつ非常に堅牢な方法で検出できます。

2.3 ENDOR：社会物理学による

2.1.1. 正統的表現へのデータ変換

先に論じた「行動クラスター」（すなわち、社会物理学の不変性に反する検出物）を抽出し、それらを「知識球体」に集約することにより、生の入力データ（人間に由来する限り、どのような形や形をとっていても）正統的な形式に変換されます。この形式は、社会物理学の法則に反している人々、すなわち、一定の不変性に対して、行動に「高すぎる同期変化」を示す人々のクラスタリングを表します。これは、同様の方法で特定の時間に方向を変える物理的なオブジェクトに似ています。これらの変更を引き起こす力は見えませんが、

オブジェクトがすべて単一のソースの影響を受けている可能性が高いと推定できます。同様に、行動クラスタの形でのデータの標準的な表現には、同じく隠れた「社会的な力」によって影響を受けた可能性が最も高い人々のグループが含まれており、共通の「現実世界の特性」を共有しています。新しい生データが利用可能になると、それは顧客（通常は毎日または毎週）によって **Endor** に送信され、追加の行動クラスタが自動的に抽出されます。

この表現による利益は3倍になります：

- **自動**：社会物理学の数学的不変量を用いて、人間の行動に由来するデータは、入力データの種類にかかわらず、申告や分析の必要がない行動クラスタの集合に自動的に変換されます（例：クレジットカードによる購入、タクシー乗車などの顧客が所有する固有のデータ）。これは、入力データのノイズやギャップに対する極めて高い弾力性と相まって、未知のタイプの汚れた生データを一様な動作クラスタに変換するプロセスが初めて完全に自動化されることを意味します（詳細は「ノイズに対するロバスト性のセクションをご参照ください」）。
- **統一**：任意のドメイン、人口統計またはセマンティクスのデータを取り除くことによって、行動クラスタを含む残りの情報は、以下の「質問」段階のために理想的に形成されます。実は、**Endor**はこの統一表現を使用して、それが存在しなければ「ビッグデータを作成」することができ、データの元の所有者にとっては不可能だったディープラーニングの手法をプロセスの照会段階で活用できます。これは、**Endor** ディープラーニング・エンジンが、多くの顧客からの多種類のデータに由来する行動クラスタの全てを単一フォームに変換することを可能にします。
- **新たなトレンド**：単純に言えば、社会物理学の不変性は、時間の経過とともに群衆の特定の統計的性質が進化する方法を記述します。この時間的側面により、**Endor**は、最近発生したばかりのダイナミクスを容易に検出することができます。これまでは、従来手法を使用して統計的有意性が高いほどに十分な観測をできるデータを生成するのに十分な時間がありませんでした。社会物理学の使用により、これらの追加シグナルが検出されるという事実に加えて、これらは通常、最近のトレンドに関する情報を含んでいるため、さまざまなビジネス上の質問にとって非常に重要なシグナルになります。

「古いタイプ」の機械学習では、事前定義された特徴を使用し、比較的少量の一般データから関連情報を抽出していました。しかし、その結果の多くは選択した特徴に依存していました。ディープラーニングは、最も関連性の高い特徴を単独で識別しますが、膨大な量のデータを必要とします。それぞれのデータタイプと質問では、関連する特徴を再度検索する必要があるため、より多くのデータが必要です。

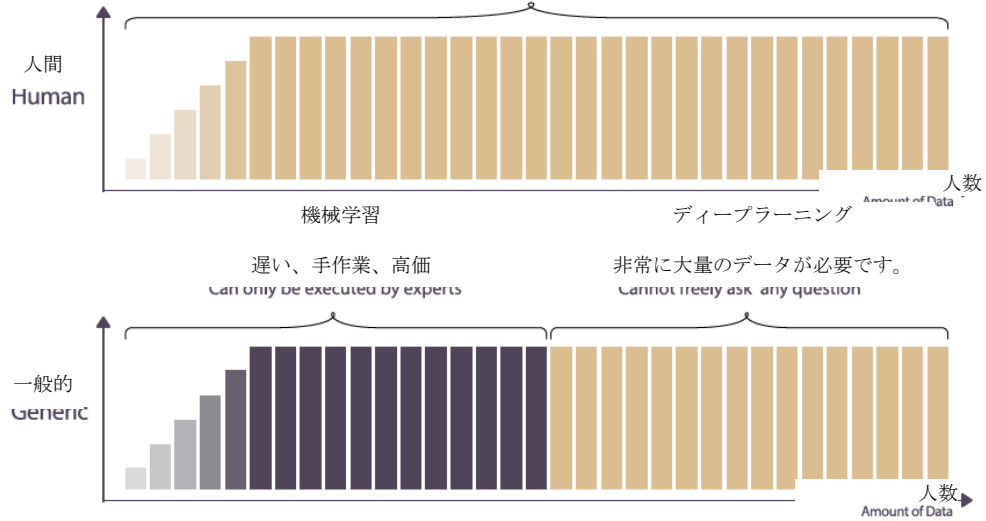
社会物理学は、あらゆるタイプの人間の行動データを、行動に基づいて人間のクラスタの正統な形式に変換します。これは、データ量の大小に関わらず機能します。さらに、社会物理学の正統な形式のおかげで、**Endor**はデータサイズに関係なく、すべてのデータタイプとすべての質問を取り込み、質問への解答にディープラーニングの力を使用した1つ

の巨大な人間の行動データセットを生成できます。

社会物理学→ディープラーニング

完全自動で、あらゆる質問に答えることができます。専門家でもなくても使用できます

社会物理学は、あらゆるタイプの人間行動データを、人間の行動クラスタから正統的に変換します。Endotorは社会物理学の正統的な表現を使用して、データサイズに関係なく、すべてのデータから生成された大きな人間行動データセットを使用し、ディープラーニングの力を活用してあらゆる質問に答えます。



現在の予測分析パラダイム（トップチャート）が社会物理学を通じて可能になる新しいパラダイム（下の図）。

2.3.2 正統的表現の質問（「知識球体」）

知識球体には、すべてのユーザーについて検出された全体的な情報が含まれているため、シンプルなデータ検索の容易さとスピードで、あらゆる質問に対する予測を提供できます。どんな質問 単純なデータ検索の容易さとスピード：知識球体の最初の作成には、通常、10 億件の記録からなる典型的なデータの場合、1~4 時間かかります。このプロセスが完了すると、同じ知識球体を使用して、数分間で数十件の質問に答えることができます。事前の知識のドメインや関連する特徴の抽出は必要ありません。

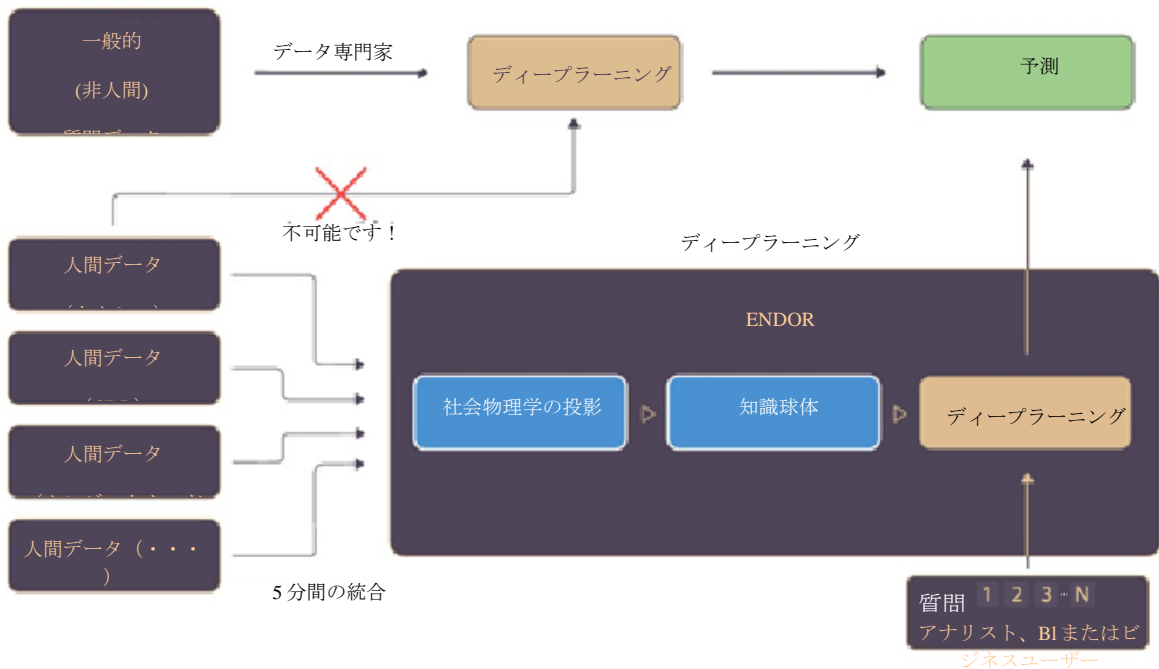
質問は、任意のサイズの「例」（積極的にラベル付けされた ID）を提供することによって送信されます。Endor のエンジンは知識球体を使用して回答を生成します。ユーザーのランク付けされたリストです。関連する質問と「行動的に類似している」可能性が最も高いユーザーです。「トレーニング」/「学習」段階ではありません。「結果を解釈する」必要はありません。データ科学者が特定の質問ごとに、この自動プロセスと同等のを行うには、従来の方法では何週間から何ヶ月もの努力が必要でした。

例えば、近い将来チャーンする可能性が最も高いユーザーに関する質問は、以前チャーンした人の ID を含むリストに記述されます。あるいは、どの新規顧客がプレミアムアカウントに変更される可能性が高いかを特定しようとする質問は、最近このプレミアムサービスに変更した顧客を含むリストに記述されます。ただし、両方の質問で同じ知識球体を使用するため、ソートの再トレーニングの必要はありません。

このような行動予測の質問のほとんどは、データが不足していることや基礎モデルが頻繁に変更されている為、ディープラーニングを使用して解決することが非常に困難ですが、Endor はすべての知識球体と質問のコレクションを使用してこの問題を回避します。したがって、Endor は、実際に予測を提供する内部ディープラーニング・コンポーネントに必要なビッグデータを「作成」します。これらの予測は、たとえそれが非常に少量のデータに基づいていても、特定の知識球体に基づいて作成されます。

次の図は、従来のディープラーニング（上部フロー）のフローを Endor のフロー（下部フロー）と比較して示しています。人間以外の行動データ（例えば、何百万もの画像）について、ディープラーニングは、データ専門家による適切なトレーニングを受けて、高品質の予測を生成する事ができます。現在のディープラーニング・ツールは、経験豊富なエンジニアの使用を想定して設計されています。他の人間のデータ（例：タクシー乗車データ）は、基本モデルが頻繁に変化する為、ディープラーニング・モデリングでは容易には解決できません（この問題についてはセクション 2.1 を参照）。

しかし、行動クラスタのセットの形式に変換されると、タクシー乗車データはセマンティクスを取り除かれ、通話記録、クレジットカード購入記録などの人間データから得られた行動クラスタと実質的に同一になります。さらに、それぞれのタイプのデータを提供する多くの異なる顧客（つまり、多数の E コマース・プラットフォームが各自で独自の購入・ウェブ活動データをアップロードする）が存在するため、ディープラーニング・モデルのトレーニングが可能になります。これは、モデルが生データや問題についてトレーニングされていない場合に有効ですが、多数の行動クラスタ（一貫した表現のデータの大きな集合体）は、静的な基礎モデルによって特徴付けられています（すなわち、社会物理学の法則は数学的不変であり、生データの行動力学と比べると静的ですが、本来は非常に動的です）。



2.4 → データセキュリティと匿名性

2.4.1 → Endor によって保存されたデータ

Endor のソリューションは、さまざまな行動予測質問に自動的に解答する機能に加えて、高度なデータセキュリティと必要なデータフィールドの匿名化機能を提供します。これまでのセクションで詳述したように、予測は、（各ユーザーが共通の社会的または行動的特性を持つことが保証された）多数のユーザーグループを含む、知識球体（セマンティクスのない行動クラスタの集合体）を使用して行われます。そのような情報が元の生データに存在する場合、この段階でシステムによって使用されるいかなるデータにも、機密または個人情報が含まれていないことが容易に分かります。次の例に示すように、システムで使用できる唯一の情報はユーザーID です。

行動_クラスタ₁ = (ID₁, ID₁₇, ID₂₃, ...)

行動_クラスタ₂ = (ID₁₄₂, ID₄₂₈₇, ID₉₇₁₁, ...)

...

行動_クラスタ_{748,329} = (ID₅, ID₃₇, ID₂₁₈, ...)

この情報でさえ、オンボーディングデータ（下記参照）に基づいて、顧客が生データレベルで ID をハッシングすることによって容易に隠すことができることにご注意ください。

2.4.2 → Endor へのデータ・オンボーディング

前のセクションで示したように、生の入力データから行動クラスタを抽出する方法は、社会物理学の不変性に反するデータパターンを示すユーザーグループの検出に依存します。これは、ユーザーの活動の同期的な性質を表す特定の統計的性質の動態をトラッキングすることによって行われます。前述のように、この実行ではデータに含まれる実際の値は必

要ありませんが、代わりに完全なハッシュ置換によって行うこともできます。これにより、顧客は優れた予測能力の恩恵を受けながら、十分にハッシュされたデータセットを **Endor** に提供することができます。さらに、**Endor** はセマンティクスに対して 100%不可知である為、データフィールドの名前もハッシュできます。

このような財務記録のハッシュの例を以下に示します。

ハッシュ前のヘッダ:				
口座番号	支店	性別	取引の種類	送信先アカウント
ハッシュ前のデータ記録:				
183972	291/30	男性	転送	382732
183972	291/30	男性	残高照会	未発表
382732	291/30	女性	転送	439001
...				
ハッシュ後のヘッダ:				
フィールド 1	フィールド 2	フィールド 3	フィールド 4	フィールド 5
ハッシュ後のデータ記録:				
AjF32sdx	Q2KPbv3A	Wsqp289X	q8Vb3MAs	Je2qx92n
AjF32sdx	Q2KPbv3A	Wsqp289X	q8Vb3MAs	x3PNm78A
Je2qx92n	Q2KPbv3A	m28SbA12	q8Vb3MAs	yL19B4GQ
...				

2.5 社会物理学：数学的説明

2.5.1 → フレームワーク

最初に、Endor エンジンの基本原理を一般的な数学的形式論で紹介します。これには、(1) コンピュータビジョンの指向と (2) ソーシャルグラフベースの 2 つの可能な実装例を説明します。最初の例は、(人間に関係ない) センサーベースのデータで使用される Endor の欠点を示していますが、後者では社会物理学の概念と人間の行動を予測するメトリックを示しています。

注：この説明では、社会物理学の原則とその Endor エンジンでの使用方法についての数多くの数学的実例を紹介します。しかし、IP の考慮により、社会物理学の法則の数学的詳細はこの説明から除外しました。

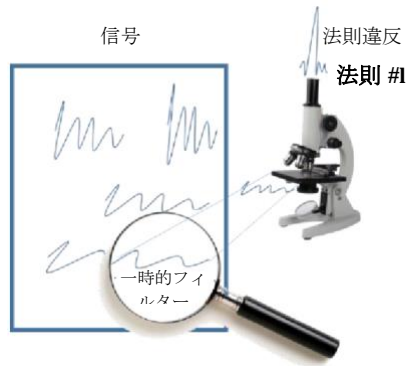
$d(x, t)$ を一時的データストリームとすると、 x は単一のデータ点を表します。

$L(\cdot)$ を生データ $d(x, t)$ を法則表現に変換する法則演算子とします。

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

法則自体は、法則演算子をアприオリ定数 \mathbf{C} (数、配電クラス、例えば電力法など) と等価とする方程式として定式化されています。この \mathbf{C} は、法則によって表される不変量を表します。

実例のために、時間の経過とともにホワイトノイズ信号の出力の変化が非常に小さなしきい値に限定されるという事への理解を表す仮定的な法則を想像できます。この法則に反すると、例えば、突然強い出力スパイクを示す信号の形を取ります。この法則の違反を見つける為に、信号をフィルタリングし、異常な出力を検出するメカニズムを構築することができます。当然、このメカニズムを使用することは、問題の法則 (ホワイトノイズの起源) に従う事が判っている信号の場合にのみ意味を持ち、そうしたスパイクは「違反」として分類できません。



また、放射線吸収材料の異常な「塊」を検出するために構築された X 線装置にも例えられます。この例では、X 線装置（およびそれを操作して結果画像を解読する技術者）は、2次元の情報ストリームを分析し、フィルムをヒットした一貫した X 線ビームをアサートする不変性の違反を突き止めるための「違反検出器」として機能し、等吸収の画像が生成されます（多かれ少なかれ、フィルムの品質およびビームの一貫性に応じて）。

この不変性、あるいは法則は、既知の物理的な事実であり、この物質に X 線吸収元素が存在することを示す「違反」を見つけようとする意思で、第 3 の物質を通過している間に、一貫した X 線ビームを人工的に高品質のフィルムに誘導し、第 3 の物質を通過させながら医療アプリケーションに使用されます。例えば、骨は、この場合、「データ」（すなわち、露光されたフィルム）で検出される「実世界現象」を表し、その不均一な放射線吸収によって示され、それ自体が上記の法則になります。



上記の原則を実施するための具体的かつ形式的かつ数学的な例は、セクション 2.5.2 と 2.5.3 に記載されています。

社会物理学の場合、そのような違反は、入力データが人間の活動から生じたものであるという前提の下、社会物理学の法則に従い、分析的かつ統計的にはほとんど不可能であるとされている入力データ内に特定のダイナミクスを生成した人々のグループを構成します。

2.5.2 → データ・サブセットが法則に反するかの検証

Endor のエンジンは、法則のオペレーターといくつかの社会物理の法則の定数を実装しています（具体的な詳細は IP の考慮のために記載されていません）。このセクションでは、法則違反の検証について説明します。他の多くの計算上の問題と同様に、特定の信号が法則違反であることの検証は、このような干渉の検出とは根本的に異なることに注意することが重要です。このセクションでは前者のみが議論されます。すなわち、"潜在的な干渉"が本当に法則違反であるかどうかの検証です。大規模な高次元データソースを効率的にスキャンし、そのような臨時検証を実行できる Endor の検索アルゴリズムの詳細については、付録 A: 社会物理学の説明 ²¹ は、ここでは説明しません。

検証する法則の数学的公式は、常に以下の形式で与えられます。

$$L_{x,\tau}(d) = C$$

法則を明示的に公式化すると、それからの偏差を ξ で示す偏差を以下のように測定することによって、局所偏差を検証することができます。

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

ここで、 Δx は X の部分空間を表し、 Δt は一時的ウィンドーを表します。この偏差は、 X のすべての部分空間および任意の期間について計算することができ、所定の期間に部分空間が関連法則にどれだけ違反しているかの尺度を生成します。

この測定値を予め定義されたしきい値 ξ と比較することにより、<法則に違反する部分空間を検出することができます。

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

違反しきい値 ξ は、しきい値を超えて法則から留保する信号の自然発生が非常に起こりにくいように選択されます。これにより、あるデータサブセットが、データそのもののセマンティクスに関する事前知識がなくても、統計的有意性が十分に高い為、法則違反を自動検証することができます。

時間と空間の両方で信号が変化すると、異なる一時的ウィンドーが法則違反として検出され、異なる部分空間を作成する可能性があることにご注意ください。Endor は、 $\Delta t = 1$ 日、7 日、30 日、90 日の（データからではなく、法則から導出される）一時的ウィンドーの事前定義固定セットを使用します。

データが非常に動的である場合、より長い一時的ウィンドーが偏差グループを生成する可能性は低くなります。データが静的である場合、より短い一時的ウィンドーが偏差グループを生成する可能性は低くなります。それにもかかわらず、ノイズは一貫した法則違反を生み出すことはできません（より正式な用語では、ノイズが法則の十分大きな違反を生み出す確率はゼロに近く、これがしきい値 ξ が選択される方法であるという事実のため）。

知識球体は、すべての関連する一時的ウィンドーのすべての法則からのすべてのグループ偏差の集計です。

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

この知識球体は、データセットごとに 1 回計算されます。このプロセスは質問に影響されるのではなく、生データから生成される内部動作構造です。抽象的な観点から、Endor の社会物理学エンジンは、匿名化された生データを行動に関連する標準的な表現に「圧縮」します。

2.5.3→例

以下は、データの違反パターン検出のための法則の使用を示す 2 つの例です。説明のために、このメカニズムを実証するための既知の数学的現象を使用します。

例 1：ビジョン

$d(x, t)$ を特定のフレーム t における特定の画素 x の色とします。ここで、同じ色の法則を定義して、任意の部分領域のすべての色があらかじめ定義された色 C でなければならないことを示します。

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

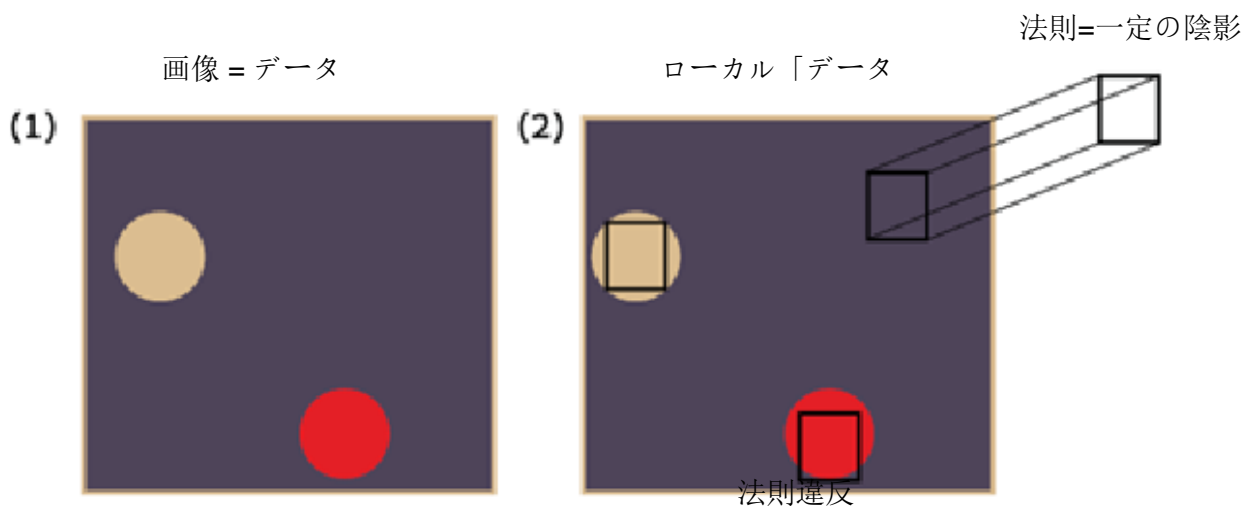
この演算子は、領域 x と時間窓 t をとり、この入力データの平均色を計算します。

この例は、否定による法則の概念を示しています。ビデオ内のピクセルに関する先験的な既知の法則はありません（確かに、すべてのサブリージョンが同じ平均色であると仮定しているわけではありません...）。この **Low** 演算子と定数 $C = \text{ライトブルー}$ を考慮して、 Δx を $N \times N$ ピクセルサイズの正方形のウィンドーとして定義し、単一のフレーム $\Delta t = 1$ を使用することができます。

これは、ライトブルーからの $N \times N$ 個のウィンドーのローカル偏差として以下をもたらします。

2次元画像では、この偏差をあらかじめ定義されたしきい値と比較して、ライトブルーからのパッチの偏差を測定するために簡単に適用できます。偏差がこのしきい値を超えるパッチは、「ライトブルーの法則」に反した四角形内のすべてのピクセルを表す「類似の特性を有するピクセルのクラスター」として分類されます。

画像は常に「滑らか」である（すなわち、小さな尺度では単調であり、「ピーク」がない、または極大値\極小値である）と仮定します。この前提の下では、次の「データ」 - 2つのパッチ（青色と赤色）を持つ滑らかな色の「正当な」背景を含む画像を観察することができます。多数の小さな四角形のランダム領域をサンプリングすると、これらの2つの「違反」が簡単に見つかります。



次に、「質問」を赤色ピクセルの形式で求めることができます。そのようなピクセルは、赤色違反に含まれていると特定され、結果として赤色パッチが返されます。

これらの例は、従来のコンピューティング・ビジョンやディープラーニングの方法で適切に対処することができるため、シンプルなビジョンベースのアプリケーションにとって有利であることを示唆するのではなく、社会物理学の数学とメカニズムを説明することです。

例 2：スケールフリーネットワーク

この例では、 $\mathbf{d}(\mathbf{x}, \mathbf{t})$ は、グラフのノードを表す \mathbf{x} を抽象的に表します。法則演算子は次のように公式化された次数分布演算子です。

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

このベクトル演算子は、各ノードの次数に対して 1 を生成します。この演算子の結果を

すべてのグラフのノードにわたって合計すると、グラフの度数（ノード数で割った度数分布に相当）の正統性ベクトルが得られます。

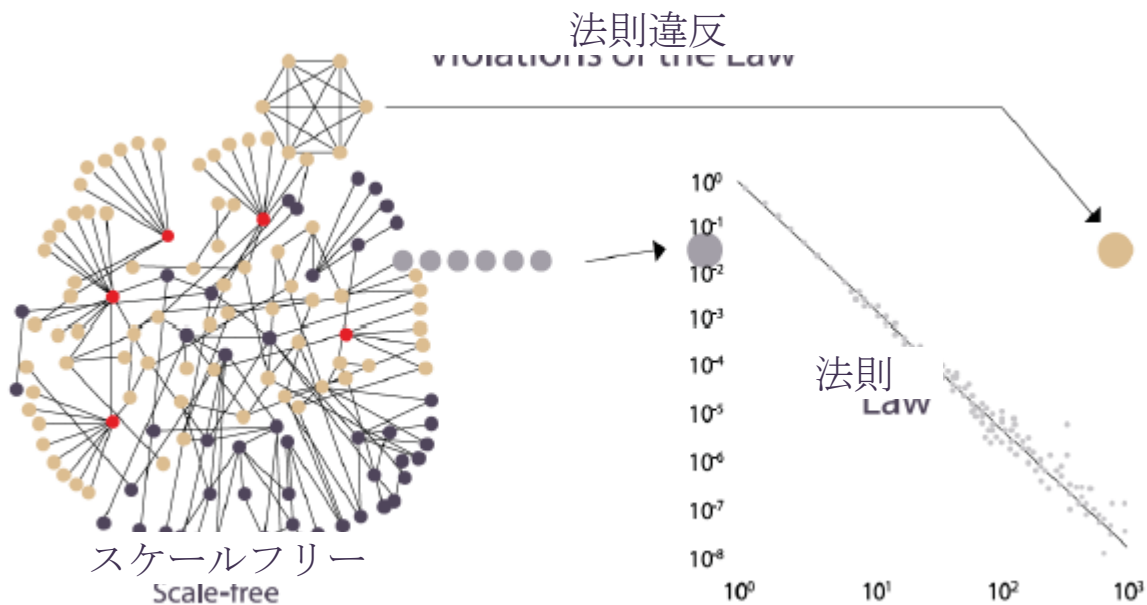
この例では、グラフがスケールフリーのネットワークであると仮定します。したがって、べき乗則の次数分布を法則定数を適用することができます（いくつかの正規化定数 α に対して）

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

この法則定数は、次のように公式化することができます。

$$L_{X,T}(d) = \frac{1}{|X| |T|} \int_X \int_T L_n(d(x,t)) dx dt = C_n$$

この法則は、全体のグラフが、そのノードのすべての次数のべき乗則の分布に従うべきであることを意味する。しかし、実際のスケールフリーの多くのグラフでは、そのような分布からの局所的な偏差が大きくなる可能性があります。これは、例えばクリーク（すなわち、完全に接続されたサブグラフ）またはチェーン（すなわち、2つを超える隣接ノードを有するノードのない接続されたツリーを形成するノードのサブセット）の周りに生じます。このような法則の逸脱または違反は、次の図に示されており、構造表現（左）と隣接行列（右）の両方で表現されています。



法則が与えられた場合、そのような違反は、以下のようなさまざまな措置によって容易に検証できます。

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

□

この偏差は、差の累積二乗を測定します（一方、そのような尺度の別の例は、両方の確率分布の *KL-発散* であり得ます）。ここで、 Δx はグラフのすべての可能な部分グラフを表します。明らかに、入力グラフのすべての可能な部分グラフをスキャンすることは実現不可能です。これは、「*非多項式ハード問題*」として知られる「*難しい問題*」クラスのメンバーです。この意味で、法則違反の検証（法則の詳細を知る必要があります）と法則違反（各法則のために特別に開発された一連の独自の技術が必要です）を検出することを区別することが重要です。このような検出技術は、社会物理学の数学的法則に合わせて作られた **Endor** の独自技術の一部です。

スケールフリーの例に戻って、このようなローカル干渉をグラフで見つけるための効率的な手法を持っていると仮定すると、次のように公式化できるサブグラフの集合が得られました。

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

そして、この法則が示す知識球体は、以下のように定義されます。

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

社会物理学の法則に反する集合体を含む知識球体は、同じドメインのオブジェクトのリストに定義された、任意のラベル付きサンプルの「類似したもの」を検出するために使用できます。この例では、グラフノードのリストが与えられると、他の全てのグラフノードには、ラベル付きサンプルと共有するクラスタの数に応じて採点することができます。あるいは、集団採点の形で出力を生成するために、検出されたクラスタおよびラベル付けされたサンプルのリストを入力としてのみ使用する限り、異なる採点メトリクスを使用することができます。このような指標や採点メカニズムを「採点者」と呼びます。これらについては、次のセクションで詳しく説明します。

~~異なる一時的ウィンドーは、グラフノード間の大きく異なる関連性を表す異なる知識球体を生成する。~~ 社会物理学の説明 異なる一時的ウィンドーは、グラフノード間の大きく異なる関連性を表す異なる知識球体を生成することにご注意ください。さらに、これらのクラスタは、~~単一~~ シャ

ルメディアやソーシャルネットワークなどの外部データソースから生成されたものではなく、顧客自身の内部取引データソースから生成される動作接続を表します。これにより、**Endor** は、利用可能なデータに明示されていない暗黙的な動作クラスタを検出できます。

2.5.4 → 質問への解答：シンプルな「採点者」

知識球体が利用可能になると、ユーザーはセマンティックな意味を持つ（おそらく非常に小さい）質問リストのラベル付きデータを提供することによって、ビジネス関連の質問を開始することができます。

$$y. \in X.$$

このセクションでは、「採点者」（行動クラスタ+質問を使用してランク付けされた母集団リストを出力として生成する関数）の2つの例を挙げます。

最初にご説明する採点者は、次のスコアを $\tilde{y} \in X$ 計算する各候補の単純な共同クラスタ集計採点者です。

~

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

可能性のある各「候補」は、ラベル付きデータと共存する知識球体内のクラスタ数によって採点されます。この単純な採点者は、 X のメンバーと入力ラベル付きリストの動作類似点の数をカウントすることを目的としています。ラベル付き入力リスト内のオブジェクトとクラスタを共有するオブジェクトのリストは、次のように表されます。

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

たとえば、色付きの点と例1の「同じ色の法則」において、点の周りの色付き領域は入力として与えられます。ノードが与えられた「スケールフリーグラフの法則」では、出力は、局所的に次数分布に反するノードを含むすべてのサブグラフ内のノードになります。

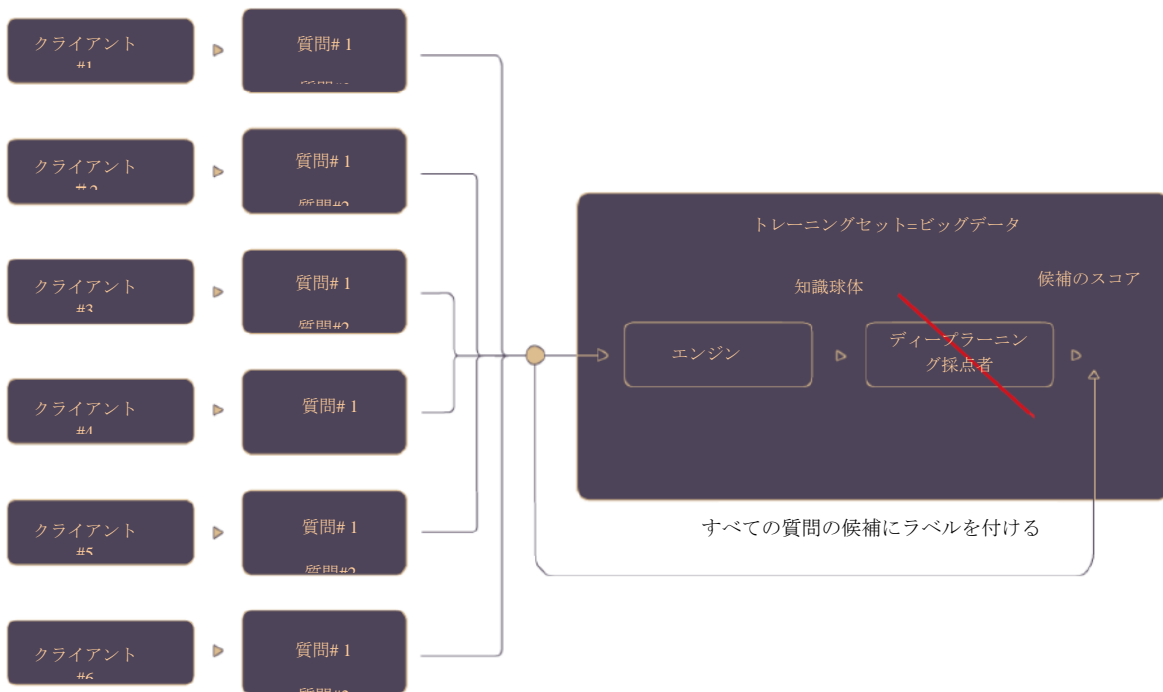
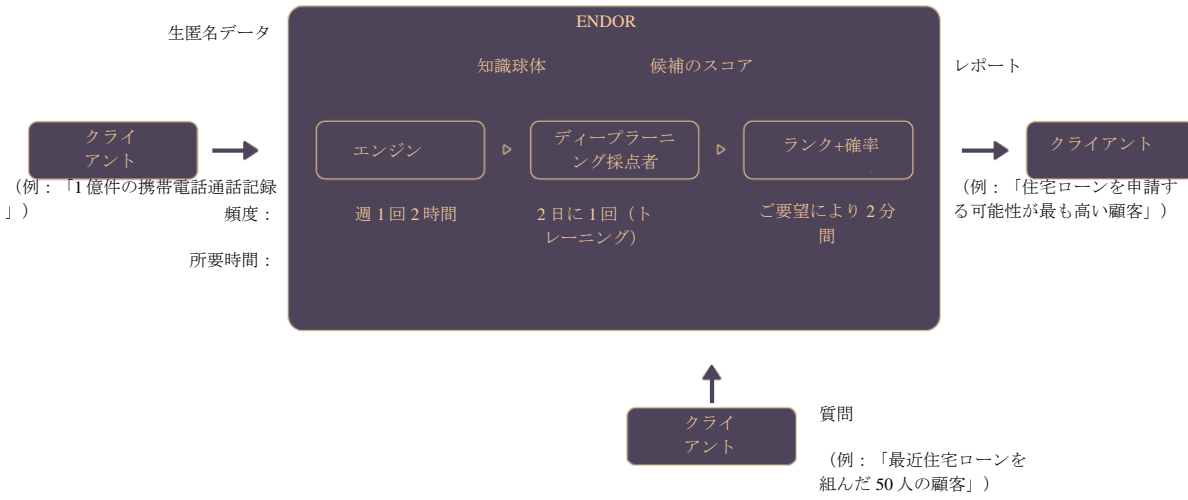
2.5.5 → 質問への解答：ディープラーニング採点者

前セクションで説明した採点者は、社会物理学知識球体を構成するクラスタを使用して、この質問の性質を事前に知らなくても「入力例」（質問）の高品質な「類似予測」を作成する事ができます。このシンプルな例は、社会物理学の正統なデータ表現の利点を示しています。しかし、正確な予測を生成するために、**Endor** は最先端のディープラーニング・アルゴリズムに基づいて、より堅牢な採点者を開発しました。

上記のように、人間の行動が関係している場合、ディープラーニングを効率的に使用するには、そのトレーニング段階で膨大な量のデータが必要です。これを生データを使用して達成することは困難です。ラベル付き質問には通常、含まれるラベルが少ないため、一般的に1つの質問あたりのインスタンス数は少なく、1クライアントあたりの質問数が比較的少ないためです。このタイプの複雑でダイナミックなラベルデータは、ディープラーニングにはあまり適していません。**Endor** の社会物理学エンジンは、すべてのクライアントからのすべてのデータセットとすべての質問のすべてのインスタンスを単一の正統な形式の大規模な集合体（人々のクラスタで構成される知識球体、およびこれらのクラスタの参照例）に変換することで、この限界を克服します。これらのソースからのすべてのデータを組み合わせることで、**Endor** は、ラベルと知識球体があれば、各個人を採点するディープラーニング・ネットワークをトレーニングするためのラベル付きトレーニングセットを作成することができます。このプロセスを実行するとすぐに、高度にトレーニングされたディープラーニング・モデルが生成され、新しいデータソースと新しい質問から生成された新しいクラスタセットを効率的に処理することができます。

このトレーニングされたモデルは、クラスタの集合体として表される（多くの）インスタンスのデータに対してトレーニングされているため、新しい質問や新しいタイプのデータセットに対して定期的に再トレーニングする必要はありません（社会物理学正統表現形式）。この方法を使用することにより、**Endor** は社会物理学とディープラーニングの両方のメリットを兼ね備えています。社会物理学は、社会物理学の法則に反する匿名化された生データを正統な形式に変換しますが、ディープラーニング・アルゴリズムは作成された知識球体と（恐らく少数の）ラベル付きデータに基づいて候補を採点します。

このフローを次の図に示します。



2.6 → ノイズへの頑強性

Endor 予測プラットフォームの主な強みの1つは、データギャップやノイズの多いデータ

に対する高い復元力です。従来のデータ分析プロジェクトは、以下のように、望ましくないデータセグメントを検出し、望ましい状態まで軽減する「データサニテーション」フェーズから始まります。

- データギャップ（データが完全または部分的に欠落している期間）
- 生データ内の文字化けの混入
- 複数の形式でデータに表示されるカテゴリ名などの意味的なあいまい性
- 数値の正規化問題
- 数値の必要なビンニング
- ...

機械学習アルゴリズムは通常、ノイズの影響を非常に受けやすく、データには通常ノイズが多いため、データの問題を緩和するフェーズは一般的にコストが高く、時間がかかります。なぜなら、現実の世界では従来の機械学習技術によって解析可能にするために生の形式（金融取引、通話など）でデータが生成されるのに対し、それは「特徴」または「プロパティ」と呼ばれる集約形式に変換しなければならないからです。この集約は、少数の間違ったデータ値、または異なるユーザーの異なる量の値によって大きく影響を受ける可能性があります。

Endor は、前のセクションで詳細を説明したように、生データ自体を分析することによってこの問題を克服します。さらに、Endor のエンジンは、予測に使用できるパターンの検索でデータの統計分析を実行するのではなく、社会物理学の法則（データの外部の数学的な不変性）を使用する為、その影響を受けません。このアプローチには、以下の基本的な考え方から生じるいくつかの重要な利点があります。

ノイズは、「任意に出現できない」データパターンを作成することはできません

（後者は無視できる程度の確率で存在すると分析されたデータパターンとして定義されます）

この観察が、事前のクリーニングなしで人間のデータからのデータの洞察を自動抽出する方法を理解するために、抽出プロセスは社会物理学の法則に反するデータオブジェクトのグループを検索してください。つまり、我々が証明できるデータパターンを表示するグループは、（社会物理学によって可能になった数学的分析を用いて）データ内に自然に出現する。これは、社会物理学の説明

つまり、ノイズがエンジンからの洞察を隠すことがあります。定義上、法則に反するとして検出されるデータパターンを作成することは（ほとんど）ありません。ノイズの多いデータが社会物理学の法則に反することはありません。

3. 結果

このセクションでは、Endor 予測システムの利用方法の様々なユースケースを紹介します。4 種類の予測質問に正確に答えるために、大規模な財政投資プラットフォームの 7 日間分の活動を含むデータを使用した予測プロセス全体の詳細な説明（グーグル・テンソルフロー・ディープラーニング・プラットフォームとの比較分析を含む）。

1. Kaggle Challenge をクラックするために Endor の完全自動での使用例
2. Endor が 24 時間以内の POS データを使用して多数のビジネス上の質問の正確な予測を提供したコカコーラ社の事例研究。

3.1.7 日間にわたる財政活動を使用して、4 つの予測質問に正確かつ自動的に解答します。

このセクションでは、Endor システムを使用した全体的な予測プロセスを説明します。

- 使用された生データの説明
- 社会物理学表現へのデータ変換（すなわち、一連の行動クラスタを構成する知識球体）
- 4 つの予測質問の定義
- 社会物理学形式の質問の兆候
- 総合的な予測精度
- グーグル・テンソルフロー・ディープラーニング・プラットフォームとの比較

3.1.1. データ :

この例で使用されたデータは、一般財政投資プラットフォームに由来し、投資コミュニティのメンバーのすべての投資取引を含みます。データは匿名化され、MIT で研究目的で公開されました（データはリクエストに応じて共有可能）。

次に、データセットの概要をご説明します。

-
- 7 日間のデータ
付録 A: 社会物理学の説明

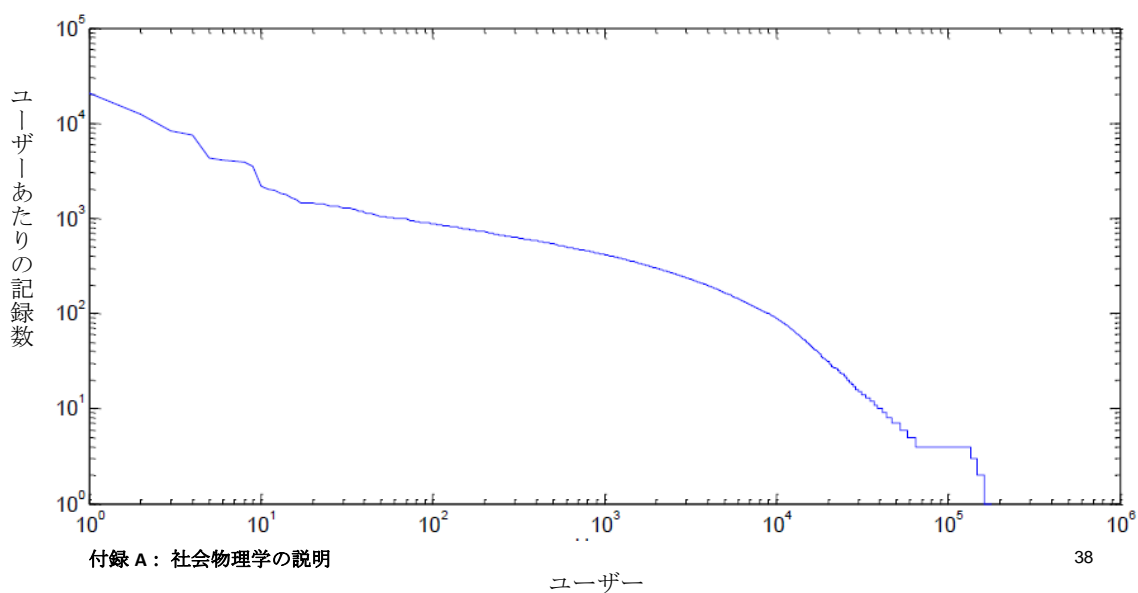
- 3,719,023 行
- ユニークユーザー178,266 人
- データの文脈的または意味論的解釈は与えられませんでした
- 12 のデータフィールド：

フィールド名	タイプ	#固有値
時間 1	時間	501573 固有
時間 2	時間	4 固有
ユーザーID	INT32 → 分類別	178266 固有
記録 ID	INT32 → 分類別	1053574 固有
プロパティ 1	INT32 → 分類別	24 固有
プロパティ 2	INT32 → 分類別	7 固有
プロパティ 3	INT32 → 分類別	134 固有
プロパティ 4	INT32 → 分類別	77527 固有
プロパティ 5	INT32 → 分類別	10 固有
プロパティ 6	INT32 → 分類別	27 固有
プロパティ 7	INT16 → 分類別	9 固有
プロパティ 8	二重数値	2772 固有

データセットの以下の重要な点に注意する必要があります

- データは、生の非集約形式で与えられ、ユーザーレベルのイベントが含まれていました。
- データの文脈的または意味論的解釈は与えられませんでした
- データサニテーションは行われませんでした。データにはノイズ、ギャップ、重複記録などが含まれていました。
- このデータには、著しく不均一なユーザーごとの記録の分配が含まれていました（ユーザーの 78% の取引回数は 10 回以下の為、記録は 200 回以下のはずが 4,000 回以上あります。中央値はユーザーあたり記録 4 回です）：

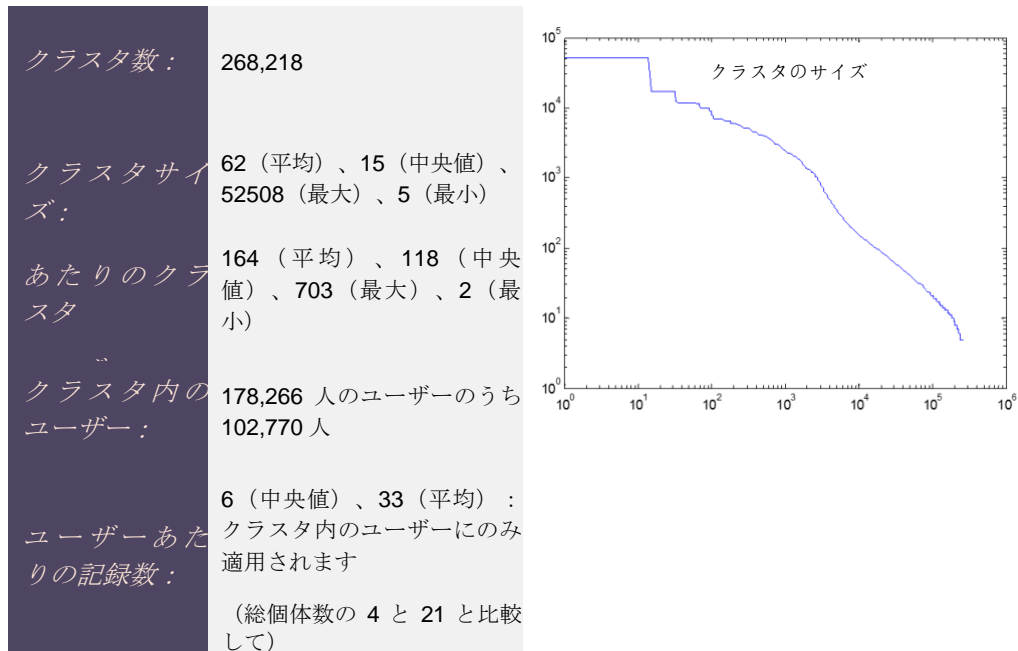
ユーザーあたりの記録数のログ・ログ表現 10



3.1.2. 自動クラスタ抽出

データの最初の分析時に、Endor システムは、データ動力学が社会物理学の数学的不変性に反するユーザーのグループである「行動クラスタ」を検出し抽出します。これらのクラスタは、データのすべての列に基づいていますが、入力としてシステムに提供されたデータであるため、最後の 7 日間に限ります。

以下は、システムによって検出されたデータセットの動作クラスタの概要です。



3.1.3. 予測質問

次の予測質問は定義されました：

- **新しいユーザーが「クジラ」になる**：過去 2 週間に参加したユーザーは、今後 90 日間で少なくとも \$500 の手数料を生み出します
- **還元活性**：先週中アクティブだったユーザーは、今後 30 日間で活動が 50% 減少するでしょう (ただし、チャーンではなく、取引は継続します)
- **「クジラ」のチャーン**：現在アクティブな「クジラ」(過去 90 日間の活動によって定義)で先週アクティブだった人は、今後 30 日間アクティブではなくなります

- **アップルの株式を初めて買うことになるでしょう**：アップルの株式を買うことは一度もありませんが、今後 30 日間で初めて買うでしょう

見てわかるように、上記の質問のすべては、（「検索数」または「例」の生成に使用された）過去のデータと（予測の検証に使用される）将来のデータの両方の 7 日間のデータセットを超えるデータを参照しています。7 日間の期間外の情報をシステムに提供することを避けるため、質問は `users_IDs` 値を含むリスト形式で作成されました。例：

- **質問名**：新しいユーザーが「クジラ」になる
- **検索数**：過去 2 週間以内に参加したユーザー (*End_of_data* より前) を含む *user_IDs* のリスト。
- 「例」：「クジラ」(つまり、*End_of_data* 以前の 90 日前に \$500 以上の手数料を生み出したユーザー) として知られているユーザーを含む *user_ID* のリスト。
- **見つける必要があるターゲットのリスト**：*End_of_data* 以降の 90 日以内に 500 ドル以上の手数料を生み出したユーザーを含む、「検索数」のサブセットである *user_ID* のリスト。このリストは検証目的でのみ使用され、システムには提供されませんでした。

3.1.4. 知識球体の質問の明示

検索質問の定義は、行動クラスタの抽出と、質問の定義とは独立して行われる知識球体の生成とは完全に直交していることに注意が必要です。したがって、システムによって検出されたクラスタ内の質問の表現を分析することは興味深い事です。クラスタには、次のような事実にもかかわらず、質問の定義に関連する情報が含まれていますか。

- クラスタは完全に自動で抽出され、データについてのセマンティックではない情報を用いて、
- 質問はクラスタが抽出された後に定義され、このプロセスに影響しませんでした。

この分析は、非常に高濃度の「サンプル」を含むクラスタの数を測定することによって、言い換えれば、「統計的に期待されるより多くの例」を含むクラスタを探すことによって行われます。そのようなクラスタの数が多い場合(同じ母集団を無作為にサンプリングするときに受け取る量よりも大幅に多い場合)、このプロセスの貴重な意味論的洞察を完全に自動的に抽出する能力が証明されます。

次の表は、同じ量のサンプルを含む「ランダムクラスタ」の数と比較して、特定の量の「サンプル」を含む動作クラスタの数を比較した、この観察結果を示しています。ランダムクラスタは、顧客の集団から無作為にサンプリングされた N 個のグループのユーザーの集合体を指し、 N がシステムによって検出された行動クラスタの数に等しく、これらのランダムにサンプリングされたグループのサイズがシステムによって検出されたクラスタのサイズに等しくなるようにします。これは、ランダムサンプルと比較して「ターゲットユーザー」の一貫性の高いクラスタを大幅に含むため、行動クラスタによってカプセル化された情報を実証するために使用されます(それらは「質問」の定義に先立ってシステムによって検出されます)。

必要な「サンプル」数は、「ベースライン」の単位で示されます。つまり、「活動量が減って

る」質問の「ベースラインの 5 倍」（ベースラインが約 11%）は、「サンプル」と 55%重複するクラスタを意味します（つまり、メンバーの 55%が「活動が減った以前のユーザー」のリストにも表示されます）。

多くのターゲット顧客を含むクラス		#ランダムクラス	#行動クラス
活動量が減っている	ベースラインの 2 倍	0	98
	ベースラインの 5 倍	0	11
「クジラ」のチャーン	ベースラインの 2 倍	212	1678
	ベースラインの 5 倍	67	525
	ベースラインの 10 倍	21	114
決して買わなかった	ベースラインの 2 倍	3962	60044
	ベースラインの 10 倍	1332	25542
	ベースラインの 20 倍	415	9090
新しいクジラ	ベースラインの 2 倍	38	1898
	ベースラインの 5 倍	0	65

3.1.5.→ 予測結果

次の表は、4つの質問の予測精度を示しています

- ベースライン：母集団のランダムサンプルに必要なターゲット顧客の平均であり、無作為な推測の正確性を表します。
- 候補：検索数のサイズ。たとえば、「クジラになる新しいユーザー」では、候補の数は新規ユーザーの数を指します。
- トップ 100：予測レポートのトップ 100 メンバーのリクエストされたターゲット顧客の一部（同様に、トップ 250 とトップ 500）。

	ベースライン	トップ 100	トップ 250	トップ 500
「クジラ」になる新しいユーザー <small>過去 2 週間に参加したユーザーは、今後 90 日間に少なくとも \$500 の手数料を生み出します</small>	7.5% <small>(2270 人の候補者のうち 170 人)</small>	37%	28.8%	21%
活動量が減っている <small>現在のアクティブユーザーのうち、今後 30 日間で活動量を 50% 減らす人 (ただし、チャーンではあ</small>	11.4% <small>(2233 人の候補者のうち 255 人)</small>	21%	23%	20.2%

「クジラ」のチャーン 先週アクティブだった現在アクティブな「クジラ」は、今後 30 日間にアクティブでなくなるで	1.66% (4141 人の候補者のうち 69 人)	10%	10.8%	9.2%
アップルで初めて取引する アップル株を購入したことのないユーザーは、今後 30 日間で初めて購入します	0.5% (161382 人の候補者のうち 839 人)	14%	12%	10%

見てわかるように、予想通り我々が予測リストにより深く関わるようになると、予測精度が低下します。

3.1.6. テンソルフローとの比較

このセクションでは、Endor システムとグーグルのテンソルフローで得られた予測結果の比較を示します。テンソルフローは、他のディープラーニング・ライブラリと同様、議論中のものに似たデータを扱う際にいくつかの困難に直面していることに注意が重要です。

- ユーザーあたりの記録数の分布が極端に不均一な場合、データの一部の正統化が必要になります。
 - いくつかの手作業は、データ科学を多少は理解している個人によって行われます。
 - 時間の投資とデータの所有者または提供者へのアクセスを必要とするデータのセマンティクスへの多少の理解
- 陽性サンプルと陰性サンプルの分布が極端に不均一な単一クラスの分類は、結果のオーバーフィットを招きやすく、軽微な調整が必要となる傾向があります。これは、ディープラーニングの専門家の関与を再度必要とします（機械学習やデータ科学の役得を持たないビジネス、プロダクト、またはマーケティングの専門家が使用できる Endor システムとは異なります）。

テンソルフローに基づいたソリューションを作成するために 2 週間かかるディープラーニングの専門家に、データを処理するのに十分な専門知識を持っているかを尋ねました。作成されたソリューションには、次の補助的なテクニックを使用します。

- データシーケンスを顧客あたりの記録 200 回に調整し、中立記録が 200 回未満のユーザーのストリームをパディングします。
- 1,000 人の顧客（50%が陽性ラベルと判明、50%が不明）を有する 200 のトレーニングセットを作成し、これらのトレーニングセットを使用してモデルをトレーニングする。
- 2つの出力ニューロン（陽性、陰性）を有する配列分類（128 の LSTM を有する RNN）を使用し、全体の結果は 2つのスコアの差があります。

以下の表は、これらの手法を使用して得られた結果（赤色）と Endor の予測（青色）を比較したものです。

	ベースライン	トップ 100	トップ 250	トップ 500
「クジラ」になる新しいユーザー 過去 2 週間に参加したユーザーは、今後 90 日間に少なくとも \$500 の手数料を生み出します	7.5% (2270 中 170 候補)	37% 21%	28.8% 27.2%	21% 19.6%
活動量が減っている 現在のアクティブユーザーのうち、今後 30 日間で活動量を 50% 減らす人（ただし、チャーンではありません）	11.4% (2233 中 255 候補)	21% 8%	23% 18.8%	20.2% 19.4%
「クジラ」のチャーン 先週アクティブだった現在アクティブな「クジラ」は、今後 30 日間にアクティブでなくなるでしょうか？	1.66% (4141 人の候補者のうち 69 人) (21156 例)	10% 11%	10.8% 12.4%	9.2% 8.4%
アップルで初めて取引する アップル株を購入したことのないユーザーは、今後 30 日間で初めて購入します。	0.5% (161382 人の候補者のうち 839)	14% 1%	12% 0.8%	10% 1%

観測:

- Endor は 4 つの質問のうち 3 つでテンソルフローを上回り、4 番目の質問で同じ精度になります。
- Endor の優位性は、タスクが「より困難」になるにつれてますます明白になります。つまり、トップ 500 ではなく、トップ 100 に焦点を当てています。
- 「動的ではない質問」（静的な信号が検出されやすくなる「クジラになる」、「チャーン」、「活動量が減る」）と「初めてアップルで取引する人」の質問は明確に区別されます。(a) はより動的で、(b) はベースラインが非常に低いため、後者の場合、Endor は 10 倍正確です！
- 前述のように、ここで示すテンソルフローの結果は、ディープラーニングの専門家による 2 週間の手動での改善が必要ですが、Endor の結果は 100% 自動です。

3.2. 3時間でKAGGLE CHALLENGEを自動的にクラッキング

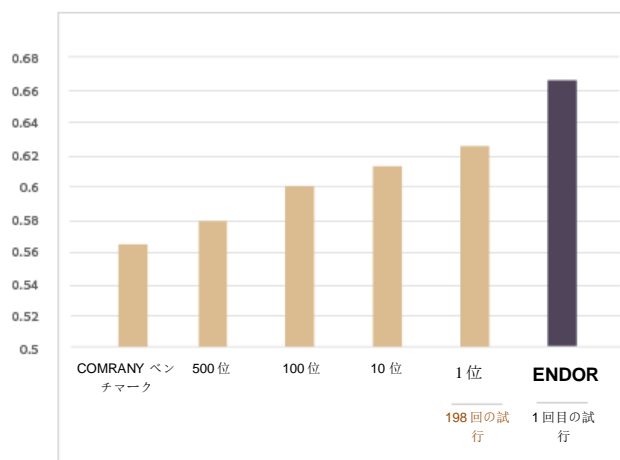
別の例では、Endor のシステムを「価値ある買い物客の獲得」で知られる Kaggle 社のコンペで公開されているデータでテストしました。このデータには、数十万の顧客を指し示す約 3 億件の POS 記録がありましたが、特定のプロモーションクーポンを受け取ったユーザーが定期的な顧客になると予測されました。

最初のチャレンジは 952 チームを網羅し、3 ヶ月間続きました。

チャレンジの参加者に与えられた生データに完全に自動的に実行される Endor マシンは、最初に優勝したチームを上回る予測を出すことができました。



- 952 の競合するデータ科学チーム
- 3 ヶ月のチャレンジ



ENDOR :

少ないクリック数、1 位

付録 B：企業のための Endor の一般的なユースケース

あなた自身の個人的なオラクルを持つのはどんな感じでしょうか？グーグルやフェイスブックのような強力なデータエンジンにアクセスするのはどうでしょう？あなたが最も強力なエンジンを使用して Endor の日付を入れて、今ある技術が提供できる最も信頼性の高い回答を得ることができたらどうですか？高度な機械学習を民主化することによって、ブロックチェーンの予測をすべての人が利用できるようにするというビジョンを掲げて、*Endor.coin* は、これまで専門のデータ専門家からなる内部チームを備えた巨大テクノロジー企業だけが使えたこの能力をあなたに提供します。

以下は、Endor の企業顧客から要求された一般的な予測ユースケースの概要を示しています。Endor が現在提供している主なセグメント別にユースケースを分類します。

- 個人向け銀行
- 保険会社
- 小売および E コマース

しかし、*社会物理学*を使用することで、新しいユースケースが容易にサポートできます。Endor のセールスエンジニアの所要時間は平均数時間です。



ENDOR



ユースケース：個人向け銀行

ENDOR

一般的なユースケース



もっと売る

- 購入する傾向
- クロスセル
- 売り上げアップ
- 大量消費者
- 新しいサービス - アーリーアダプター

質問例

- 購買傾向：過去 3 ヶ月間の資格のあるすべてのアクティブな顧客のうち、来週の電話をして \$ X 以上のローンを組む人は？誰がクレジットカードを持たずにローンが組めますか？短期借入？長期借入？
- クロス・セール：過去 6 ヶ月間に製品 X を使用した顧客のうち、誰が次の月にプロモーション Y を受けた製品 Z を使用開始するか
- 大口消費者：過去 30 日間の新規顧客のうち、今後 3 ヶ月間に月額 \$ X を超える可能性が高いのは誰ですか？
- 新サービス：過去 3 か月間の全てのアクティブな顧客のうち、2 週間前に開始された新サービスに登録するのは誰ですか？

一般的なユースケース

もっと売る



その他の例

- 過去 90 日間のアクティブなクレジットカード利用者のうち、2 枚目のカードを所有している場合、\$ / % X 以上支払いが増えるのは誰ですか？
- 過去 90 日間アクティブではなかったクレジットカード利用者のうち、商品 A から B に変更した場合、誰がカードを使用し始めますか
- 過去 6 ヶ月間のアクティブな顧客のうち、今後 3 ヶ月以内に残高を繰越すのは誰ですか？
- 過去 6 ヶ月間のすべてのアクティブな顧客のうち、来月に Tasa 0 を受け取るのは誰ですか？
- 全てのアクティブなゴールドまたプラチナアカウントのうち、誰が来月ファイルチャージにカードを含めるのでしょうか？
- 過去 6 ヶ月間に利息が増えている顧客のうち、今後 3 ヶ月間に残高を清算するのは誰ですか？
- 当社の旅行ウェブサイトでは飛行機のチケットを購入した人の中で、誰がホテル予約に興味がありますか？

- プラチナまたはゴールド・ポートフォリオの顧客のうち、今後 2 ヶ月以内に高額商品をポイント交換するのは誰ですか？
- プラチナカードの高価値顧客のうち、来月カードをキャンセルするのは誰ですか？
- 過去 12 ヶ月間のアクティブな顧客のうち、今後 2 ヶ月以内に \$ Y 以上の商品を購入する可能性のある顧客
- 過去 6 ヶ月間に製品 Y を購入した顧客のうち、今後 2 ヶ月以内に製品 Y に \$ Z 以上を費やす可能性が高いのは誰ですか
- 昨年クレジット払いをしていない顧客のうち、来週に連絡すれば来月に新しいクレジットを始める顧客

一般的なユースケース



マーケティング効率

- プロモーションの効率
- チャンネル間の粘着性
- カスタマー・ロイヤルティ

質問例

- プロモーション効率：アクティブな顧客のうち、テキストメッセージを受け取り、プレミアムサービス X に登録するのは誰ですか？呼びますか？プロモーション Y？
- デジタルカード：リストに掲載されたウェブサイトですくなくとも Y 回購入した顧客のうち、来月回収サービスを利用して購入する顧客
- 過去 1 年間に少なくとも 1 つの免税品を購入し、過去 X ヶ月間に海外で何も購入していない顧客のうち、3 ヶ月以内に誰が海外でクレジットカードを使いますか？
- すべてのアクティブな顧客のうち、今後 30 日以内にニューヨークまたは英国に飛行機で渡航する人
- 過去 90 日間に加入した顧客のうち、誰がカードを有効化していませんか？また、\$10 のボーナスを提供すれば有効化するの誰ですか？\$20？
- 過去 6 ヶ月間の有効なカードの所有者のうち、来月カード使用を止める可能性が高いのは誰ですか？

ENDOR

一般的なユースケース

マーケティング効率



その他の例

- 毎月の消費者クレジットの広告メールを開かない顧客のうち、(1) 電話をする (2) 来週 SMS を送信する場合、新規クレジットを始める人
- 3,000~24,000 ポイントを所有している既婚者の顧客のうち、明日メールを送った場合、今後 4 日間でフレグランスとポイントを交換するのは誰ですか？
- 飲食店でカードを使用する顧客で、与信限度額が\$100 万を超えていて、地域 X に住んでいる人のうち、来週 Y%以上を商品 X へ支出する人
- 来週ガソリンを特別割引にしたら、翌月に当社のカードを使ってガソリンスタンドで支払う顧客は誰ですか？
- 過去 6 ヶ月間に未払いの債務を再交渉した顧客のうち、明日 SMS で残額を支払うようにリマインダーを送っても、来月延滞する人誰ですか？
- 過去 2 ヶ月間カードを使用していない顧客のうち、来週ガソリンを特別割引にすれば、もう一度カードを使用し始める人は誰ですか？

ENDOR

一般的なユースケース

デジタル換算



- オンライン・アクティビティの増加
- 支店からオンラインへの移行
- ウェブサイト訪問者のターゲット変更
- 初回購入者、リピーター、リテンション
- 大量消費者

質問例

- 2ヶ月間アクティビティがなかったオンライン顧客のうち、今後7日間で電話を受ければアクティブになるのは誰ですか？インセンティブ X?
- 従来の取引（支店、コールセンター）を行うすべての顧客のうち、来週プロモーション Xを受ければ誰がデジタル取引に移行しますか？
- 過去7日間の新規オンライン・ユーザーのうち、最初の3か月間に\$ X以上の取引を行う可能性が高いのは誰ですか？
- 過去3ヶ月間にデジタルチャンネルで新製品の契約をした顧客のうち、プロモーション Xを受けて、来月中に次の製品の契約をするのは誰ですか？
- 過去3ヶ月のウェブサイト訪問者のうち、その人にターゲットを変更した場合、オンラインバンキングに登録する可能性が高いのは誰ですか？

ENDOR

一般的なユースケース

デジタル換算



その他の例

- 先月登録した新規オンラインユーザーのうち、今後3ヶ月以内にオンラインを継続的に使用するユーザーは？誰がしませんか？
- オンラインバンキング利用客のうち、来月オンラインバンキングの利用を中止する人は誰ですか？
- 先月登録した新規オンラインユーザーのうち、来週インセンティブ Xを貰えば継続的なオンラインユーザーになる人
- すべてのオンライン顧客のうち、機能 Xを使用しているが Yを使用していない人で、連絡すれば翌月 Yを使用する人
- 過去3か月間に新しいデジタル製品（X、Y、Z）の契約をした顧客のうち、プロモーション Xを受ければ、翌月に2つ目の製品の契約をする人
- 過去3か月間にデジタルチャンネルで製品を購入した顧客のうち、1ヶ月以内に退会する顧客
- 過去3ヶ月間のすべてのウェブサイト訪問者（クッキーID）のうち、その人にタ

ターゲットを変更すればオンラインバンキングに登録する可能性が高い人

- 銀行商品を1つだけ所有している顧客のうち、来月中に2つ目の銀行商品を購入する可能性のある顧客
- 当社のウェブサイトを訪れたことがない顧客のうち、明日メールを送信すれば、来週訪問するのは誰ですか？



ENDOR



ユースケース：保険

一般的なユースケース

毎日、どのように予測しますか？決めましたか？性能を向上させる？

あなたは次のような質問に答える必要があります



販売

- 誰が補助的補償 X を追加しますか
- 来週アプローチしたら、誰がアップセル/クロスセルする可能性が高いか
- 発効日前にキャンセルした人のうち、再購入する人
- 1週間前にリリースされた新製品/サービスを誰が購入しますか
- 来週誰が新しいプランを申し込みますか
- 新規顧客のうち、誰が大口消費者/プレミアム顧客になりますか



マーケティング

- 誰がプロモーション X に反応しますか
- 誰が最初の 90 日以内に契約を終了しますか
- 誰がモバイルサービス (アプリ) を利用しますか
- 誰が 2 ヶ月以内にニューヨーク & メキシコに旅行しますか？
- 誰が友人を紹介しようですか？



デジタル

- 誰が \$ / % X 以上アクティビティを増やしますか
- 誰が新しいデジタルサービスを利用しますか
- アクティブではない人のうち、誰がアクティブになりますか
- 誰が支店からオンラインに切り替えますか
- 誰がプロモーション X に反応しますか
- その人にターゲットを変更したら切り替えるウェブサイト・ユーザーは誰ですか

一般的なユースケース



もっと売る

- 購入する傾向
- クロスセル
- アップセル
- 大口消費者
- 新サービス - アーリーアダプター

質問例

- 購買傾向：すべての潜在的顧客のうち、ポリシー X を追加する可能性が高いのは誰ですか？既存顧客のうち、誰が補助的補償を追加する可能性が高いですか？
- クロスセル：プラン X の顧客のうち、プロモーション Z を受けて製品/サービス Y を購入する可能性があるのは誰ですか？
- 大口消費者：過去 30 日間の新規顧客のうち、今後 3 ヶ月間に月額 \$ X を超える可能性が高いのは誰ですか？
- 新サービス：過去 3 か月間の全てのアクティブな顧客のうち、2 週間前に開始された新サービスに登録するのは誰ですか？

一般的なユースケース



データ化

- オンライン・アクティビティの増加
- 支店からオンラインへの移行
- ウェブサイト訪問者のターゲット変更
- 初回購入者、リピーター、リテンション
- 大口消費者

質問例

- 2ヶ月間アクティビティがなかったオンライン顧客のうち、今後7日間で電話を受ければアクティブになるのは誰ですか？インセンティブ X?
- 従来の取引（支店、コールセンター）を行うすべての顧客のうち、来週プロモーション Xを受ければ誰がデジタル取引に移行しますか？
- 過去7日間の新規オンライン・ユーザーのうち、最初の3か月間に\$ X以上の取引を行う可能性が高いのは誰ですか？
- 過去3ヶ月間にデジタルチャンネルで新製品の契約をした顧客のうち、プロモーション Xを受けて、来月中に次の製品の契約をするのは誰ですか？
- 過去3か月のウェブサイト訪問者のうち、その人にターゲットを変更すると、オンラインで補償を追加する可能性があるのは誰ですか？

ENDOR

一般的なユースケース



マーケティング効率

- プロモーション効率
- 新製品
- カスタマーロイヤルティと顧客紹介
- ターゲットとするプロモーション

質問例

- プロモーション効率：アクティブな顧客のうち、テキストメッセージを受け取れば、プレミアムサービス Xに登録するのは誰ですか？呼びますか？プロモーション Y?
- すべてのアクティブな顧客のうち、今後30日以内に目的地 Xに旅行するのは誰ですか？
- 紹介：すべてのアクティブな顧客のうち、インセンティブ Xをもらえれば友人を紹介する可能性のある人は誰ですか？
- 新製品：全てのアクティブな顧客のうち、来週アプローチしたら、新しいデジタルサービス Xを利用する可能性がある誰ですか？

一般的なユースケース



リテンション

- ポリシーのリニューアル
- 終了
- 購入歴のある顧客

質問例

- 終了：ポリシーXの新規顧客のうち、最初の90日間に契約を終了する可能性があるのは誰ですか？
- 購入歴のある顧客：発効日前にキャンセルしたすべての顧客のうち、誰が再契約しますか？
- リニューアル：過去60日間に新しいポリシーを適用したすべての顧客のうち、発効日前にキャンセルするのは誰ですか？

ENDOR



ユースケース：小売および電子商取引

ENDOR

一般的なユースケース

毎日、どのように予測していますか？決めましたか？性能を向上させる？

あなたは次のような質問に答える必要があります



販売

- 誰が来月 X を購入しますか
- 1 週間前に発売された製品を購入する可能性が高いのは誰ですか
- 誰がプレミアムに切り替えますか
- 誰が大口消費者になりますか
- どの X 顧客が Y を購入しますか



マーケティング

- 誰がプロモーション X に反応しますか
- どのキャンペーン継続しますか
- どの顧客がオンラインで購入しますか
- 誰が当社のモバイルアプリを使用しますか
- 誰が \$X 以上アクティビティを増やしますか



e-コマース

- どの Cookie ID を切り替えますか
- 誰がプロモーション X に反応しますか
- 誰が初回購入者になりますか
- 誰がアクティビティを増やしますか
- どのアクティブでない顧客が、プロモーション X を受けてアクティブになりますか



店舗/製品

- どの店舗が、製品 Y の売上が \$X に達しますか
- どの店舗が下回りますか
- どの製品が来月 20% 以内で増加しますか



戦略

- 早期警告システム
- どこで脱会されますか
- 製品 X はどこで追い風になりますか

一般的なユースケース



もっと売る

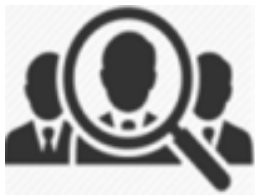
- 購入する傾向
- クロスセル
- アップセル
- 将来の大口消費者
- 新製品 - アーリーアダプター

質問例

- 過去 3 か月間のアクティブな顧客のうち、30 日以内に 352553 を購入する可能性が高いのは誰ですか？
- 過去 30 日間に製品 219 を購入したすべての顧客のうち、今後 90 日間に製品 890 を購入する可能性が高いのは誰ですか？
- 過去 3 ヶ月の製品 32236 のアクティブな顧客のうち、今後 7 日間でプレミアムにアップグレードする可能性があるのは誰ですか？
- 過去 6 ヶ月のアクティブな顧客のうち、3 週間前に発売された新製品を購入する可能性が高いのは誰ですか？
- 過去 30 日の新規顧客のうち、最初の 3 ヶ月間に合計\$1 万以上を購入するのは誰ですか？

ENDOR

一般的なユースケース



マーケティング効率

- より良いセグメンテーション
- プロモーション効率
- チャネル間の粘着性
- カスタマーロイヤルティ

質問例

- 過去 6 ヶ月間の消費金額上位 10%の顧客のうち、消費金額が 2 ヶ月で少なくとも 20%減少するのは誰ですか？
- 過去 6 ヶ月間の消費金額が下位 20%の顧客のうち、消費金額が 3 ヶ月で少なくとも 30%増加するのは誰ですか？
- 全ての顧客のうち、来週 20%割引クーポンを貰えば、製品 6347 を購入するのは誰ですか？
- すべての顧客のうち、来週テキストメッセージを送信すれば製品 6347 を購入するのは誰ですか？
- 過去 90 日間にウェブサイトで購入した全ての顧客のうち、来月モバイルアプリを

使用する可能性が高いのは誰ですか？

ENDOR

一般的なユースケース



アクティビティの増加

- 効率的なユーザー獲得（ターゲット変更）
- 初回購入者
- リピーター
- アクティビティの増加

質問例

- すべてのウェブサイト訪問者（クッキー）のうち、来週にグーグルを介してターゲットを変更すれば、顧客になる可能性が高いのは誰ですか
- すべてのウェブサイト訪問者（クッキー）のうち、最初の3か月間に合計\$1万以上を購入するのは誰ですか？
- 過去6ヶ月間のアクティブな顧客のうち、来月中にFTBになるのは誰ですか？
- 過去7日間に初めて食料品を注文した顧客のうち、60日以内に2回の新規注文を行うのは誰ですか？

他の多くの...



製品/店舗

- どのサービスの売上が、来月20%以内で増加しますか
- どのサービスの売上が今後3ヶ月以内に50%以上減少する可能性が高いですか
- どの店舗が今後3ヶ月以内にカテゴリXの売上の50%以上に達しますか



オペレーション

- 誰が明日、Xのコールセンター/テクニカルサポートに連絡しますか
- 明日、誰が支店Xで取引しますか



戦略

- 来月、どこからブランドX離れが起きますか（逆風）
- どこからブランドYへの追い風が吹きますか

「事前定義メニュー」はありません。

どのビジネスユーザーも、ビジネスを成長させるために必要なことを質問します

付録 C : Endor.coin の事前定義済み予測の例

Endor.coin の壮大なビジョンは、保険業界、銀行業、電子商取引、健康など、ブロックチェーンによるセクターの変革を目指しています。当社は、取引をリードすることを求める仮想通貨所有者をサポートするため、仮想通貨洞察の前例のない予測プラットフォームを提供します。*Endor.coin* は、ブロックチェーンの信頼性が証明された分散型インフラストラクチャであり、他の既存のテクノロジーまだ成し遂げていない未来予測を提供します。

以下は、起動時に *Endor.coin* プラットフォームでサポートされる最初の事前定義されたブロックチェーン予測の例です。*RFP* (予測リクエスト) メカニズムを使用して、コミュニティがリクエストする新しいユースケースが徐々に追加されます。

仮想通貨アドレス予測：この予測は、事前定義されたアドレスリスト（例：「過去 1 ヶ月以内に少なくとも 1 回の取引履歴があるビットコインアドレス」）を受け取り、直近の事前定義された行動（例：「過去 1 週間に少なくとも 0.1 ビットコインを受け取ったアドレス」）との行動上の類似性に従ってランク付けします。結果リストには、事前定義された動作に最も類似した（そして、将来は同じ動作を統計的に表示する可能性が高い）アドレスが最上位に含まれ、結果リストの最下位には事前定義された動作に類似しないアドレスが含まれます。

- **アクティブ・アドレス**：過去 1 か月に少なくとも 1 回はアクティブだったすべてのアドレスのうち、最近、取引数が大幅に増加したアドレスに最も類似したアドレスはどれですか？
- **重取引アドレス**：過去 1 か月間に少なくとも 1 回はアクティブだったすべてのアドレスのうち、過去 1 ヶ月間の取引総額が 10 BTC 以上のアドレスに最も類似しているのはどれですか？
- **非アクティブになるアドレス**：過去 1 か月に少なくとも 1 回はアクティブだったすべてのアドレスのうち、最近、合計取引量が 50%減少したアドレスに最も類似するのはどれですか？

トークン予測： この予測は、トークンの事前定義リスト（例：「過去 1 ヶ月に少なくとも \$10,000 万ドルの取引量があるトークン」）を受け取り、直近の事前定義された行動（例：「月間平均取引がを 2 倍に増加したトークン」）との行動上の類似性に基づいてランク付けします。結果リストは、事前定義された行動に最も類似する（したがって、将来は行動を示す可能性が統計的に高い）トークンが上位に表示され、事前定義された行動に最も類似しないトークンは、リストの最下位に表示されます。

- **収益性の高いトークン：** 先月の取引金額が少なくとも \$ 100 万のすべてのトークンのうち、先月、価格が BTC 標準で 50%以上上昇したトークンに最も類似するのはどれですか？
- **収益性の低いトークン：** 先月の取引金額が少なくとも \$ 100 万のすべてのトークンのうち、先月、価格が BTC 標準で 50%以上下落したトークンに最も類似するのはどれですか？
- **不安定なトークン：** 先月少なくとも \$500 万の取引量があったすべてのトークンのうち、先月最も不安定だった 10 のトークンに最も類似しているのはどれですか？
- **安定しているトークン：** 先月少なくとも \$500 万の取引量があったすべてのトークンのうち、先月最も安定していた 10 のトークンに最も類似しているのはどれですか？
- **取引量の増加しているトークン：** 先月の取引金額が \$2 千万以上のすべてのトークンのうち、先月の取引量が 2 倍になったトークンに最も類似しているのはどれですか？
- **取引量の減少しているトークン：** 先月の取引金額が \$2 千万以上のすべてのトークンのうち、先月の取引量が 50%に減少したトークンに最も類似しているのはどれですか？

付録 D - 知識球体クラス : API アクセス

この付録には、*Endor.coin* プロトコルの基本ブロックである「知識球体」のデータ構造の仕様と、*Endor.coin* ネットワークに接続するために、将来の分析エンジンの実装のためのアクセスを提供するクラスの完全な説明が含まれます。

使用法の説明： クラスタ・オブジェクトは **3** つのエンティティで構成されています：

- スパース・マトリックス M の寸法（「検索可能なオブジェクト」 \times 「行動クラスタ」）は、検索可能なオブジェクトと行動クラスタ間の接続性を表し、 M_i があり、 $j=1$ になります（インデックス $i \in$ クラスタ j の検索可能なオブジェクトの場合）。検索可能なオブジェクトは、トークン、ウォレットアドレス、ロケーション、電話番号、またはその他のタイプのオブジェクトを指します。

データに含まれ、予測の基礎となります。

- 配列 A_M は、各検索可能オブジェクト SO をスパース・マトリックス M のインデックスにマッピングします。
- データフレーム D_M は、各予測エンジンによって定義され、計算された、その他のクラスタプロパティを含んでいます。このような特性は、例えば、クラスタのサイズ、内部接続対外部接続の比率、このクラスタを生成するために予測エンジンによって使用される内部モジュールなどであり得ます。

使用のためにクラスタオブジェクトを抽出して構築するには、以下のファイルが必要です。

- スパースマトリックス M を構築するために：
- マッピング配列 AM を構築するには：
- クラスタプロパティを構築するには データフレーム DM ：

上記のファイルは、以下に示すクラスタ抽出者クラスによって参照される特定のパスに配置されます。

<クラスタ_ファイル_パス>

使用例： 関連する 5 つのクラスタファイルをクラスタファイルパスに配置した後、次のコードを使用してクラスタを抽出できます。

この例では、「ポップ・トゥ・クラスタマップ」はスパースマトリックス M を指しています。「クラスタポップ」は、クラスタプロパティを含むデータフレーム D_M を指します。「翻訳ポップ」は、各検索可能なオブジェクトを、スパースマトリックス A_M のインデックスにマッピングする配列を指します。

クラスタ抽出者クラス： 以下に、知識球体 API の完全な説明を示します。

i

参考文献

[1] ウィキペディア - 社会物理学 (2017)。

URL https://en.wikipedia.org/wiki/Social_physics

[2] W. Pan、Y. Altshuler、A. Pentland 「金融取引ネットワークにおける社会的影響と群衆の知恵を解読する：プライバシー、セキュリティ、リスクと信頼」 (PASSAT)、2012年国際会議と2012年国際会議ソーシャルコンピューティング (SocialCom)、IEEE、2012、pp. 203-209

[3] Y.-Y. Liu, J. C. Nacher、T. Ochiai、M. Martino、Y. Altshuler 「オンライン金融取引のプロスペクト理論」 PloS one 9 (10) (2014) e109458

[4] Y. Altshuler、W. Pan、A. Pentland 「社会拡散モデルを用いたトレンド予測」 arXiv.org、2011

[5] P. M. Krafft, J. Zheng, W. Pan, N. Della Penna, Y. Altshuler, E. Shmueli, J. B. Tenenbaum, A. Pentland, 「分散型ベイジアン推論としての人間の集団知能」 arXiv preprint arXiv : 1608.01987

[6] Y. Altshuler、A. Pentland、G. Gordon、「社会行動バイアスと知識マネジメントの最適化、ソーシャルコンピューティング、行動・文化モデリング、予測」 Springer、2015、pp258-263

[7] Y. Altshuler、A. Pentland 「最適性能のためにネットワークを調整する方法および装置」 米国特許第 8,914,505号 (12月16日2014)

URL <https://www.google.com/patents/US8914505>

[8] W. Pan、Y. Altshuler、A. Pentland、N. Aharony 「ネットワークにおける行動予測と修正の方法と装置」 米国特許第 9,098,798号 (8月4日2015)

URL <https://www.google.com/patents/US9098798>

[9] 「群衆の知恵を得るためのソーシャルネットワークのチューニング」 (MIT メディアラボ・ウェブサイト (2017))

URL <https://www.media.mit.edu/research/highlights/tuning-social-networks-gain-wisdom-crowd>

[10] 「市場の洞察：市場でのツイッター効果に目覚める」 (ファイナンシャルタイムズ) (2017)

URL http://web.media.mit.edu/~yanival/Markets_Insight.htm

「エコーチャンバーを超えて」 (ハーバード・ビジネス・レビュー) (2017)

URL <https://hbr.org/2013/11/beyond-the-echo-chamber>

[12] 予測分析の再考 (ファーストマークのデータ駆動) (2017)

URL <http://firstmarkcap.com/insights/rethinking-predictive-analytics/>

[13] MITによるソーシャルメディアにおける投資家からの\$100万出資テスト (2013)

URL <https://tinyurl.com/MIT-1M-USD>

[14]Endor - 「予測分析のグーグル」の発明（2017）

URL <http://news.mit.edu/2017/endor-inventing-google-predictive-analytics-1220>

[15]Endor の主要投資家 - イノベーションの取り組み（2014）

URL <http://www.innovationendeavors.com>

- [16] A. Boehme、Y. Altshuler 「消費者行動を予測するための社会物理学の使用：ネットワーク科学」
(NetSci) 2017
- [17] 「マスターカードはスタートパスアクセラレータプログラムに 5 つの新しいスタートアップをもたらす」 (2016)
- URL <https://tinyurl.com/MasterCard-Endor>
- [18] Endor – *Finnovate Fall 2017* (2017).
- URL <https://www.youtube.com/watch?v=69rUQloq-qA>
- [19] Endor - ガートナー社公認クールベンダー (2017)
- URL <https://www.gartner.com/doc/3727117>
- [20] Endor - 世界経済フォーラムにおいてテクノロジーのパイオニアと認定 (2017)
- URL <http://widgets.weforum.org/techpioneers-2017/>
- [21] DARPA ネットワーク・チャレンジ (2011)
- URL <http://archive.darpa.mil/networkchallenge/>
- [22] 2012 年マッキンゼー賞受賞 (2012)
- URL <https://hbr.org/2013/04/the-2012-mckinsey-award-winners>
- [23] グーグル社奨学金 - Alex "Sandy" Pentland 教授 (2017)
- URL <https://scholar.google.com/citations?user=P4nfoKYAAAAJ&hl=ja>
- [24] Tim O'Reilly : 7 人の世界で最も強力なデータ科学者 (2017)
- URL <http://www.forbes.com/pictures/lmm45emkh/6-alex-sandy-pentland-professor-mit/>
- [25] Y. Altshuler、A. Pentland、AM Bruckstein 「スウォームとネットワークインテリジェンスの検索」
Springer (2017)
- [26] Y. Altshuler、Y. Elovici、AB Cremers、N. Aharony、A. Pentland 「ソーシャルネットワークにおけるセキュリティとプライバシー」 Springer Science&Business Media (2012)
- [27] H. Shrobe、DL Shrier、A. Pentland 「サイバーセキュリティのための新しいソリューション」 MIT Press (2018)
- [28] J. Clippinger、D. Bollier 「ビットコインからバーニングまで、そしてそれを超えて：デジタル社会におけるアイデンティティと自律の探求」 ID3 および Off The Common Books (2014)
- [29] T. Hardjono、D. Shrier、A. Pentland 「トラスト :: データ：アイデンティティとデータ共有のための新しい枠組み (2016 年)
- [30] A. Pentland、T. Heibeck 「正直なシグナル：それはどのように我々の世界を形作るか」 MIT press (2010)
- A. Pentland 「社会物理学：どのようにこの良いアイデアが普及したか - 新しい科学の教え」 Penguin (2014)

- D. Shrier、A. Pentland 「金融技術のフロンティア：未来の商業の冒険 ブロックチェーンとデジタルバンキングから予測市場に至るまで、そしてそれを超えて」 出版社：Visionary Future (2016)
- [33] *Endor.coin* プロトコル *GIT* (2017)
URL <https://github.com/orgs/EndorCoin>
- [34] たとえまだそれを知らなくても、何故あなたはブロックチェーンベースの AI を望むのか (2017)
URL <https://tinyurl.com/blockchain-based-AI>
- [35] Y. Altshuler, N. Aharony, A. Pentland, Y. Elovici, M. Cebrian 「現実を盗む：犯罪者がデータ科学者になるとき（またはその逆）」 *Intelligent Systems*、*IEEE* 26 (6) (2011) 22-30. doi : 10.1109 / MIS2011 年 7 月 8 日
- [36] M. Ulieru 「ブロックチェーン：それは本当に世界を変えることができるか」 世界経済フォーラム
- [37] 「テクノロジーはどのように崩壊した金融システムの修復に役立つのか」 (2017)
URL <https://tinyurl.com/technology-fixing-our-financia>
- [38] J. Pieprzyk、T. Hardjono、J. Seberry 「コンピューターセキュリティの基礎」 Springer Science&Business Media (2013)
- [39] T. Hardjono、LR Dondeti 「マルチキャストおよびグループセキュリティ」 Artech House (2003)
- [40] T. Hardjono、LR Dondeti 「無線 LAN と MANS のセキュリティ」 (*Artech House Computer Security*) Artech House, Inc. (2005)
- [41] J. Seberry、T. Hardjono 「インドネシアとマレーシアの Bahasa の解読に向けて」 (1989)
- [42] SG Ong、J. Seberry、T. Hardjono、ADF Academy 「マンダリン (Pinyin) の解読に向けて」 (1991)

免責事項: このホワイトペーパーは、議論のみを目的としています。

Endor.coin は、このホワイトペーパーで到達した結論の正確さを保証しません。

著作権 ©2018 Endor.coin