

ENDOR.COIN 프로토콜

모두를 위해 접근 가능한 인공지능 예측

Endor Ltd. 제공

2018년 2월 18일

모두를 위해 접근 가능한 인공지능 예측

초록

Endor.coin 은 인공지능 데이터 분석으로의 접근을 민주화하고, 접근가능하게 하여, 신뢰가 필요없고 검열을 방지하는, 모두에게 유용한 예측 분석을 재발명하고 있습니다.

이는 비즈니스 사용자와 비전문 암호 토큰 보유자 모두 복잡한 예측 질문을 하고 몇 분 내에 높은 품질의 결과를 얻을 수 있는 세계 최초의 자동화된, 자주적인 예측 플랫폼을 제공하는 *Endor.coin* 프로토콜을 통해 달성됩니다. 이것은 오늘날 포춘- 500 대 기업을 위해 대부분 예약되어 있는 데이터 과학 분야를 민주화하는 것을 목표로 합니다. *Endor.coin* 은 MIT의 프로젝트 팀원 Prof. Alex Pentland와 Dr. Yaniv Altshuler가 개발한 사회 물리학의 새로운 과학에 기반을 둡니다.

- **끊임없이 확장되는 예측 카테고리:** *Endor.coin* 은 다양한 사전 정의 토큰 관련 예측(예: 수량 증가 및 변동성 감소 등을 위해 예측된 토큰)과 함께 출시됩니다. 이러한 예측은 플랫폼의 전용 *EDR* 토큰을 사용하여 구매할 수 있습니다. *Endor.coin* 은 그 다음, 사용자에게 *예측 요청* (RFP)을 전송하도록 허용하고, 구매용으로 실행되고 사용할 수 있게 될 새로운 유형의 예측을 제안함으로써 이것이 제공하는 예측의 선택 범위를 확대합니다.
- **고급 사용자를 위한 DIY API:** 최신 기술에 정통한 사용자와 전문가는 셀프 서비스 인터페이스를 제공받아 원하는 행동 패턴 정의를 쉽게 제공할 수 있게 되고 답례로 *그 같은 패턴의 "매우 흡사한" 예측을 자동으로 생성하는 Endor.coin 플랫폼을 갖게 됩니다.*
- **비공개 및 공개 데이터의 자동 융합:** *Endor.coin* 커머셜 고객(은행, 소매업체, 보험회사 등)은 그들의 독점 데이터 스트림을 플랫폼으로 쉽게 통합할 수 있게 되고 비공개 및 공개 데이터의 융합으로 수확된 높은 수준의 예측 통찰력을 제공할 수 있게 됩니다. *사회 물리학을 사용한 덕분에* 데이터는 자동적이고 사용자 친화적으로 통합되어 클리닝 또는 데이터 준비를 요구하지 않습니다.
- **데이터 프라이버시:** *사회 물리학의* 고객 측면에서 완전히 암호화된 데이터 사용 능력 덕분에 보장됩니다.
- **사람에 의한, 사람을 위한 예측:** *사회 물리학* 데이터를 사용하면 일단 사용자가 그들이 요구한 개인화 구성요소에 대한 비용만 지불하면, 테크놀로지 대기업들이 오늘날 보유한 예측 능력의 "99%"를 그들이 청구하는 비용의 1%로 제공합니다.

Endor.coin 은 Endor의 기술 플랫폼에 뿌리를 둡니다. 세계 경제 포럼이 최근 기술 개척자로 인정하고, 가트너가 선정한 주목할 만한 기업인 Endor는 MIT 분사 기업으로

Innovation Endeavors 와 같은 선두 투자자가 재정적으로 후원하고 있으며 코카콜라, 월마트, 마스터카드 등과 같은 포춘 500 대 기업과 함께 일하고 있습니다.

목차

1	가치 제안	4
2	소개	6
	2.1 비효율적이고 문제가 많은 시장.....	6
	2.2 민주화를 위해서는 분산화와 데이터 분리가 필요합니다.....	6
	2.3 기술 격차	
	2.4 Endor.coin 프로토콜 – 프로젝트 개요	
3	예측 과학 민주화	10
	3.1 사회 물리학 – MIT 개발 새로운 과학	10
	3.2 첫 번째 단계 : Endor.com – 기업용 자동 예측 엔진	13
	3.3 두 번째 단계 : Endor.coin 프로토콜 – 대중용 데이터 과학	17
	3.3.1 정의.....	18
	3.3.2 데이터 제공자	18
	3.3.3 예측 엔진	19
	3.3.4 사전 정의 예측 및 예측 요청 (RFP).....	19
	3.3.5 비공개-데이터 분석 및 셀프 서비스 API.....	19
	3.4 로드맵.....	19
4	신뢰가 필요없고, 검열을 방지하며 책임을 짐	22
	4.1 인공지능을 위한 책임성 및 진정성	22
	4.2 분산화된 프로토콜을 통한 검열 방지	23
	4.3 프라이버시 보호 데이터 분석: 암호화된 데이터 사용.....	24
	4.4 네트워크 효과.....	25
5	생태계 활성화	28
	5.1 공개 데이터 제공자	28
	5.2 학계 연구 그룹	29
	5.3 캐털리스트 – 애플리케이션 개발자	29
	5.4 데이터 주권	30
6	토큰 구현	31

6.1	블록체인 구조.....	31
6.2	스마트 계약	32
6.3	출시 시 Endor 의 역할.....	34
6.4	토큰 특권과 경제.....	35
6.5	수익금 사용	36
7	기술적 장점과 차별화	38
7.1	MIT 기원의 과학적 혁명 38	
7.2	실제 제품. 입증 기술	38
7.3	이용성, 사용자에게 주는 가치, 토큰 보유자를 위한 가치	39
8	팀	40
8.1	주요 팀원.....	40
8.2	어드바이저	43
	부록 A : 사회 물리학 설명	46
	부록 B : 부록 B – Endor 의 기업용 상용 사례	82
	부록 C : 사전 정의 예측의 Endor.coin 예	94
	부록 D : 지식 영역 클래스 API	95
	용어집	102

제 1 장

가치 제안

소비자 행동을 이해하고 예측하며 영향을 줄 수 있는 능력을 어느 기업에게 제공할 경우, 해당 기업은 다른 경쟁사에 비해 불공평하게 유리할 수 있습니다. 스마트 비즈니스 리더는 사업 실적을 개선하기 위해 고객 행동에 영향을 미칠 수 있는 아이디어가 많습니다. 이를 구현하기 위해서는 다음과 같은 질문에 답해야 합니다.

- 누가 당사의 최고 고객이며 어떻게 더 많은 최고 고객을 얻을 수 있는가?
- 누가 이 새로 출시한 제품을 사용해 볼 가능성이 있는가?
- 프로모션 의존성을 어떻게 줄일 수 있는가?
- 다음 매장은 어디에 열어야 하는가?
- 다음 달에 누가 제품 A에서 B로 전환할 것인가?

이 질문에 답하기 위해, 기업들은 데이터 과학 및 예측 분석과 같은 강력한 도구를 찾아야 합니다. 안타깝게도 이러한 도구의 실행을 위한 과정은 느리고, 문제가 많으며, 값비쌉니다.

- 즉, 많은 교육을 받고, 값비싸며, 구하기 힘든 데이터 과학자 및 박사들이 필요합니다.
- 4-6 번의 반복이 필요하다면, 각 반복마다 며칠에서 몇 주가 걸립니다.
- 각각의 새로운 비즈니스 질문에는 몇 주 동안에 걸친 새로운 모델의 구축이 필요합니다.
- 제품과 행동이 변경되면, 모델이 바뀝니다.

MIT의 새로운 *사회 물리학* 기술이 동력이 되는 *Endor.coin* 블록체인 기반 프로토콜은 최초의 분산화된, 신뢰가 필요 없고 검열이 방지되는 행동 예측 플랫폼으로 어떤 예측 질문에도 높은 수준의 답변을 몇 분 내에 제공합니다. 코딩이나 데이터 클리닝 또는 박사들로 이루어진 팀이 필요없습니다.

다음은 *Endor.coin* 프로젝트의 핵심 측면들입니다:

- **기술:** MIT *사회 물리학* 기술 [1]이 원동력이 되고, 트렌드 예측에서 최대 10 배 까지 더 높은 정확성은 물론 완전히 암호화된 데이터에 기초하여 예측 통찰력에서 생성하는 능력을 제공합니다 (금융 시장 [2-6] 및 관련 특허 [7, 8], 추가적인 리뷰 [9-11]에서 트렌드 예측에 관한 당사의 학문적 작업 참조).
- **혁신:** 짧은, 중간 범위 행동 패턴 (며칠에서 몇 주)의 자동 모델링에 중점을 두고 다른 가용 기술 (*FirstMark* 의 “데이터 주도 뉴욕” 이야기에서 Dr. Altshuler 의 이야기[12] 또는 금융 투자에 관한 대규모 실시간 분석 참조 [13])로 관찰되기 전에 그같은 신호를 검출.
- **업계 검증:** *Endor.coin* 은 *Endor.com* – MIT 분사기업 [14]이 개발한 기술에 기초하며 선두 투자자가 재정적으로 후원하며 [15], *코카콜라*[16], *마스터카드*[17], *월마트* 등과 같은 포춘 500 대 기업과 함께 일합니다. *Finnovate 2017* 에 출품된 *Endor* 의 제품 [18] 참조.
- **수상 경력과 인정:** *Endor* 는 *가트너 선정 주목할 만한 기업*[19]이며 *세계 경제 포럼* 이 “기술 파이오니어”로서 인정했습니다[20]. MIT 의 프로젝트 팀이 실시한 연구로 권위 있는 *DARPA Network Challenge* [21]와 *McKinsey 상*[22] 등 서너 개의 추가적인 상을 수상하였습니다.
- **팀:** 블록체인 의 세계 전문가들, MIT 컴퓨터 사이언스 및 인공지능 연구소, MIT 미디어 연구소 및 MIT 슬론 경영대학원 출신의 디지털 बैं킹 테크놀로지 및 예측 분석가들로 이루어진 팀이 주도합니다. 미국 공학학회 회원인 Prof. Pentland (공동 창립자)는 세계적으로 가장 많이 인용되는 과학자 [23] 중 한 명이며 최근에는 포비스에 의해 “세계 7 명의 가장 강력한 데이터 과학자” [24]로서 선정되었습니다. 이 팀 전체적으로 수 백 가지의 과학논문, 수 십가지의 상업용 특허, 8 권의 블록체인, 머신지능 및 데이터 프라이버시에 대한 책 등을 발간했습니다[25-32].
- **예측 분석 재창조:** 지난 몇 년 간 인공지능 및 머신 러닝 능력은 재력 있는 기업만 이용할 수 있었으나, *Endor.coin* 은 개인과 중소기업이 적은 비용으로도 우수한 첨단기술을 이용할 수 있게 했습니다. 예측 통찰력은 기여된 데이터의 집단적 분석에 기초하며 저비용으로 제공되는 한편 데이터 소유자가 그들의 데이터 프라이버시를 통제할 수 있게 허용합니다. *Endor.coin* 프로토콜은 새 데이터 소스는 물론 새 예측 엔진의 통합을 가능하게 하여 두 가지 네트워크 효과를 창출합니다. 즉, 참가자 증가 및 예측 당 비용을 줄이면서 예측 정확성을 높입니다.
- **신뢰가 필요 없고, 완전 분산화되어 믿을 수 있으며, 검열 방지:** *Endor.coin* 프로토콜은

완전히 분산화되어 예측 결과에 대해 완전한 책임을 집니다. 이를 통해 예측 중의 모든 조작이나 편견을 방지합니다. 또한 분산화된 개방적 성격의 프로토콜로 인해 모든 예측을 지원하며, 임의의 단일 권한에 의한 검열을 방지합니다.

제 2 장 도입

2.1 비효율적이고 문제가 많은 시장

우리의 세계에는 데이터를 수집하고, 보유하며, 주의 깊게 유지하는 기업들이 많이 있습니다. 또한 데이터를 처리하고 예측 모델을 구축하는 데이터 과학자와 머신 러닝 전문가들이 있습니다. 또한 많은 사람들이 미래를 예측하고자 합니다(여기에는 대기업의 고위급 임원에서부터 중간급 마케팅 또는 제품 관리자 그리고 비행기표의 최적 구입 시기를 알고자 하는 개인에 이르기까지 다양한 사람들이 포함됩니다). 오늘날, 후자가 질문에 대한 예측을 제공 받기 위해서는 세 가지 기능이 동일한 조직 내에 반드시 공존해야 합니다. 이는 곧, 예측의 99%가 대기업의 이해관계자에 의해 그리고 이해관계자를 위해 생성되고 있음을 의미합니다. 또한, 데이터 과학자의 수가 적고 고임금을 받기 때문에 그리고 예측을 생성하는 과정에서 각 프로젝트 당 몇 개월의 힘든 작업이 요구되기 때문에, 오늘날 이 프로세스를 수행하는 데 비용이 많이 듭니다. 이로 인해 예측에 관심이 있는 사람들이 진입하기 어려운 매우 높은 장벽(그리고 가격표)이 설정됩니다.

2.2 민주화를 위해서는 분산화와 데이터 분리가 필요합니다.

인공지능 및 예측 분석의 민주화는 '모든 이들'을 위한 접근 가능성을 이루고 다음과 같은 요건을 만족할 새로운 패러다임을 개발할 것을 요구합니다.

분리: 대기업과 연구소를 제외하고, *데이터 큐레이션과 데이터 과학, 의미 기원 질문 생성을 결합시키는* 높은 수준의 활동을 지속시킬 수 있는 비 상업적, 학구적 또는 비영리 조직은 거의 없습니다. 개인과 NGO, 중소기업은 일반적으로 그러한 차원 중 한 가지에만 집중하며, 그들은 조심스럽게 유지하는 데이터를 소유하거나 또는 제작하고, 강한 (및 고가의) 데이터 과학자를 고용하거나 또는 '바른 질문을 하는 데' 있어서 전문가입니다. 그러므로, 병목현상을 풀고 포춘 500 대 클럽 밖에서도 예측에 접근할 수 있기 위해, 이러한 세 가지 기본 요소를 근본적으로 분리해야 합니다.

진정으로 민주화된 예측 프로토콜에서는 데이터 제공자가 자유롭게 데이터(공개 또는 비공개)를 기여할 수 있는 한편, 기술 전문가들이 AI 및 예측 엔진(중단없이 접속이 가능하고, 프로토콜과 통합된)을 기여할 수 있어야 하고, 이를 통해 최종 사용자는 이러한 데이터 소스를 기반으로 하며 이러한 엔진으로 준비된 예측을 쉽게 소비할 수 있어야 합니다.

책임성: 첨단기술 자이언트의 데이터 과학 부서에서 책임성은 “모두가 동일한 상관 (예: 사업 정보 및 마케팅 예측 등을 담당하는 관련 C-레벨 책임자)을 위해 일하고 있다”는 사실에 의해 자동으로 제공되기 때문에 주요 특징으로서 요구되지 않습니다. 데이터 및 인텔리전트, 산술 등이 필요에 따라 계속해서 빌리는 민주화된 플랫폼에서는 책임성은 ‘공정한 경기’를 보장하고 단기간의 수익을 올리도록 장려하기 보다는 ‘장기간’ 모든 이해관계자의 장점 기능을 조화롭게 조율하기 위해 필수적이 되었습니다.

민주화: 분산화된 예측 프레임워크에는 두 가지 중요한 기여요소가 있습니다 – 한 가지는 엔지니어링과 관련이 있으며, 다른 한 가지는 검열 없음과 편향 방지와 관련이 있습니다. 많은 예에서 입증했듯이, 분산화된 솔루션은 더 쉽게 확장 및 확대될 수 있습니다. 더 많은 데이터 소스와 산술 자원, 여러 유형의 예측 엔진 추가 등은 모두 분산화된 솔루션에서 나온 큰 혜택입니다. 또한 분산화된 아키텍처는 이러한 예측이 명시적 검열 또는 암시적 검열이 대상이 되지 않음을 보증하는 유일한 아키텍처이며, 편향된 결과의 생성 또는 금전적 자원을 통해 실행됩니다. 또한, 효율적인 민주화는 새롭게 부상하는 네트워크 영향에서 열쇠이며, 사용자 당 비용을 줄이는 한편 참가자 수 증가와 함께 (데이터 제공자와 예측 엔진, 예측 소비자 모두) 끊임 없이 예측 정확성을 높입니다.

2.3 기술 격차

불행히도, 데이터 과학 민주화의 혜택은 꽤 일정 시간 동안 분명하게 나타날 지라도 그같은 프레임워크의 실제에서 구현은 조금도 과장하지 않고 큰 도전이 되고 있습니다. 주요 이유는 기술입니다 – 오늘날의 일반 과학으로는 “포괄적인 분산화된 행동 예측”을 지원할 수 없습니다 ... 오늘날 존재하는 기술인 뉴럴 네트워크 (또는 딥 러닝), 포괄적 프로그래밍, 결정 포레스트, SVM 등 모두는 ‘실제’ 머신 러닝 작업이 시작될 수 있기 전에 엄청난 양의 데이터의 위생과 프로세싱, 이해를 요구합니다. 이것이 산업이 오늘날 당면한 병목현상과 유능한 데이터 과학자의 급여 인상, 그들의 희소성의 이유입니다. 이 주제에 관한 세부적인 논의는 섹션 3.1 과 부록 A 에서 나옵니다.

어떤 데이터든 *자동으로* 소화할 수 있는 과학적으로 돌파구 *없이*, 비 전문가와 전문가는 똑같이 *예측 질문을 할 수 있으며* - 산업은 기존 패러다임에 갇혀 있었으며 회사가 새로운 데이터 과학자를 고용하는 속도, 또한 그 회사가 그들에게 기꺼이 지불할 6 자리수 급여로 명목현상에 단단히 묶여 있었습니다.

2.4 Endor.coin 프로토콜 — 프로젝트 개요

이러한 제약을 초월하기 위해, 새로운 과학을 발전시켜야 했습니다. *사회 물리학*은 Endor.coin 창립자 Dr. Yaniv Altshuler와 Prof. Alex “Sandy” Pentland가 개발했으며 인간 군중 행동 방식을 효율적으로 모델링하는 수학 이론입니다. 행동 데이터 소스에서 나오는 것으로 보여지는 한 세트의 수학 식을 통해, *사회 물리학* 이론은 행동 데이터 소스를 행동 클러스터의 한 세트로 자동 변환할 수 있게 합니다. 데이터 클리닝 또는 사전 프로세싱, 데이터 의미에 관한 이해 (또는 물을 질문) 를 요구하지 않습니다. 행동 클러스터의 이러한 수집은 “*지식 영역*”으로 알려져 있습니다.

Endor.coin 프로토콜은 행동 데이터가 행동 클러스터의 규범적 표상 안에서 취급될 때, 데이터 과학의 전통적 과정은 (마침내) 기본적 구성요소로 나뉘질 수 있고 각각을 분산화된 방식으로 여러 실행자에게 배정할 수 있게 된다는 사실에 기초합니다. 다음은 Endor.coin 프로토콜의 기본 윤곽과 주요 구성요소입니다.

규범적 데이터 표상: Endor.coin 네트워크에 기여하는 각 데이터는 “*지식 영역*” 규범적 표상으로 변환됩니다. 다양한 예측 엔진에 의해서 할 수 있으며 (아래 참조), 실행 비용은 엔진이 지불합니다. 일단 데이터가 이 변환을 겪을 면, 그것에서 추출된 다양한 행동 클러스터는 그 다음 다른 유형의 데이터로 생성한 클러스터와 한데 묶어 효율적인 자동 예측 과정이 이루어지게 합니다 (더 자세한 내용은 섹션 3.1 및 부록 A 참조).

데이터 제공자 분리: 데이터가 규범적 “*지식 영역*” 표상으로 변환됨에 따라, 데이터 제공자는 더 이상 이 분석의 이후 단계에 적극적으로 참여할 필요가 없습니다. 이 때문에 데이터 소유자는 그들의 데이터 (전체 또는 부분적으로)를 Endor.coin 네트워크에 통합하여, 생태계에서 자율적인 이해당사자로서 역할을 할 수 있습니다 – 데이터 품질 유지에 중점을 두고 누가 그것의 어느 부분에 접근할 것인지를 관리하며, 그것이 제공하는 미래 가치에서 재정적으로 혜택을 보는 데 중점을 둡니다.

예측 엔진 분리: 데이터 과학자들 사이에서 알려진 비밀은 데이터 과학 프로젝트에 소요된 시간 중 약 90%가 데이터 위생과 사전 프로세싱에 쓰인다는 것입니다. 이러한 단계를

최초로 자동화한 *사회 물리학*의 혁명적인 측면으로 인해 다양한 예측 엔진을 드디어 다양한 유형의 데이터 소스에 간극없이 연결할 수 있게 되었습니다. *Endor.coin* 네트워크를 연결하기 위해 예측 엔진 제공자에게 요구되는 유일한 한 가지는 *Endor.coin* 프로토콜 지원이며, 이는 데이터세트(선택적으로, 엔진에 의해 정의되는)를 소화하고 출력을 “*지식 영역*” 형태로 (전체 사양과 API 코드는 부록 D 참조) 제공하는 능력으로서 정의됩니다.

분산화된 실행: 블록체인 사용. 데이터 제공자는 데이터 (AWS 상에 저장)를 기부할 수 있고 *Endor.coin* 프로토콜을 사용하여 접근 가능합니다. 행동 클러스터는 다양한 예측 엔진에 의해 분산화된 방식으로 추출됩니다. 퀴리는 최종 사용자가 일정 예측 (*EDR* 수수료를 위해)을 발행하기 위한 *Endor.coin*. 스마트 계약을 요청함으로써 트리거됩니다. 각 예측을 위해, 최선의 행동 클러스터는 *Endor.coin* 예측 코드 (프로젝트의 GIT 계정 [33] 상에서 자유롭게 접근할 수 있는)에 의해 선택됩니다. 자금 중재는 데이터 기여자와 예측 엔진 사이에서 스마트 계약으로 예측을 위해 선택한 클러스터에 의거하여 관리됩니다. 이 중재는 예측 품질을 극대화하고 결과가 편향되지 않게 합니다.

데이터 주권: *Endor.coin* 네트워크에 기여하는 각 데이터 구성요소는 “공개” 또는 “비공개” 중 하나로 플래그될 수 있습니다 (동일 데이터 소스 내 일부는 공개로 다른 일부는 비공개로 표시됩니다). 공개 데이터 소스는 각 예측 엔진에서 접근 가능하며 미래 예측을 위한 행동 클러스터의 출처가 됩니다. 그 대신, 그러한 데이터 제공자는 프로토콜에 의해 선택되었을 때 *EDR* 토큰으로 보상을 받게 됩니다. 비공개 데이터 구성요소는 비록 암호화 방식일지라도 여전히 다양한 예측 엔진에 의해 접근 가능합니다. 하지만 그 같은 출처에서 추출한 클러스터는 그러한 예측을 요청한 사용자가 데이터를 위한 키를 제공할 때만 예측 출처로서 선택될 수 있습니다. 더 자세한 내용은 섹션 5.4. 참조

책임성 및 검열 방지: 블록체인을 사용하여, 예측은 무한정으로 저장되고 그러한 예측은 플랫폼의 명성과 그것을 위해 사용된 데이터, 그것을 분석하는 예측 엔진을 추론하는데 관심이 있는 자라면 누구나 접근할 수 있습니다. 또한 *Endor.coin* 프로토콜에는 각 예측의 생성 (과 이를 위해 지급할 자금의 중재) 을 위해 사용된 행동 클러스터 선택을 담당하는 오픈 출처 예측 코드 [33]가 들어 있기 때문에 편향되지 않고 정확성만 극대화되었음이 보증됩니다. 더 자세한 내용은 섹션 4.2와 4.1. 참조

모두를 위한 예측 효율성: 궁극적으로, *Endor.coin* 프로토콜은 최종 사용자에게 낮은 비용으로 우수한 예측을 얻게 해줍니다. 이는 과정의 자동화와 (높은 연봉의 정규직 데이터 과학자를 고용할 필요성 줄임), 간극 없이 다양한 유형의 행동 데이터 소스에

함께 '융합'되는 *사회 물리학*의 능력에 기반하고 있습니다. 이는 코카콜라와 같은 심지어 대기업 고객도 *Endor.coin* 네트워크로 이동함으로써 즉각적인 혜택을 얻을 수 있다는 의미입니다. 그들의 데이터를 “완전 비공개”로 플래깅함으로써, 독점 데이터의 결합에 기초한 예측 통찰력을 얻을 수 있으며, 시스템에 기여한 공개 데이터와 결합될 수 있고, 그러한 데이터를 자체적으로 획득하여 분석하는 비용에 비해 상당히 낮은 수수료를 지불합니다. 또한 많은 사용자와 데이터 제공자가 *Endor.coin* 네트워크에 참여함으로써 비용을 낮추는 등 (더 자세한 내용은 섹션 4.4 참조) 긍정적인 네트워크 효과를 제공합니다.

제 3 장

예측 과학 민주화

3.1 사회 물리학 - MIT 개발 새로운 과학

사회 물리학은 빅 데이터 분석과 생물학의 수학적 법칙을 사용하여 인간 군중의 행동을 이해하는 혁명적인 새로운 과학이며, Endor 에게 전통적 머신 러닝의 한계를 극복하게 해줍니다. 이 새 과학은 Prof. Alex “Sandy” Pentland 와 Dr. Yaniv Altshuler 의 연구를 통해 MIT 에서 나왔습니다. 이 새로운 과학은 Endor 가 독점 기술을 사용하여 더욱 발전시켰으며, 그 결과로 모든 종류의 인간 행동, 심지어 행동이 급속히 변화하고 진화할 때도, 이를 설명 및 예측할 수 있는 강력한 엔진이 개발되었습니다.

간단히, 사회 물리학은 인간 활동을 나타내는 모든 이벤트-데이터 (예: 전화 통화 기록과 신용카드 구매, 택시 탑승, 웹 활동)는 그 데이터 내 임베디드된 특별한 세트의 인간 활동 패턴이 들어있음을 보증한다는 전제에 기초합니다. 모든 인구통계 전체에서 모든 휴먼 데이터 유형에 공통된 이러한 수학적 불변성은 다른 기법으로 관찰할 수 있기 전에 새롭게 부상하는 행동 패턴을 검출하기 위한 필터로서의 역할을 할 수 있습니다.

사회 물리학의 힘 묘사: 대형 은행의 마케팅 부서가 잠재적으로 근시일 내 대출이 필요한 고객들에게 계속해서 전화를 건다고 상상해 봅시다. 부서의 수익은 제안에 실제로 긍정적으로 반응하는 고객 부분에서 직접 나옵니다. 이 계속 진행되는 캠페인에 들어간 직접 마케팅 비용이 상당히 크기 때문에, 적기에 딱 맞는 고객에게 연락하는 것이 매우 중요하며 너무 늦으면, 그들이 이미 다른 기관에서 대출을 받았을 수 있습니다. 너무 빠르면, 필요성이 미처 구체화되지 않습니다.

이를 위해, 은행은 그러한 고객이 누구인지 예측하기 위해 두 가지 도구를 고려할 수 있습니다. 은행의 데이터 과학팀이 자체 개발한 머신 러닝 모델과 Endor 의 엔진. 다음은 각 도구가 권고한 바를 간략하게 나타낸 내용입니다.

머신 러닝 모델로 검출된 고객 그룹은 은행의 마케팅 제안에 긍정적으로 반응할 고객 (예: 진정한 긍정적)은 물론 그렇지 않게 반응할 고객 (예: 가짜 긍정적)으로 구성되었습니다. 예를 들어, 참 긍정적 고객은 모델 결과의 10%라고 가정합니다. 광범위한 실험에서 그러한 10%의 대다수가 또한 Endor 사회 물리학 엔진에 감지될 것으로 나타나는데, 단 두 가지 주요한 차이점이 있습니다. (a) 머신 러닝 모델의 많은 가긍정적 고객은 Endor 엔진이 보고하지 않을 것이며 (b) Endor 의 결과에는 전통적 모델로는 검출되지 않는 많은 추가적인 참 긍정적 고객이 포함됩니다. 결과는 Endor 의 더 나은 정밀성 / 리콜 트레이드 오프 덕분에 영업 노력이 크게 개선되었습니다.





어떻게? 시간적 패턴 검출: 인간의 리얼리티는 많은 작은 일시적 이벤트와 변화로 구성되어 있습니다. 사회 물리학은 인간 행동의 기본적 역동성을 통합하고 그러므로 최근의 사회 환경 변화로 인해 일정 방식으로 행동할 가능성이 있는 인구 내 소 그룹을 벗겨내기 위해 더 잘 구축되어 있습니다. 사회 물리학은 그러므로 인간 행동 데이터 내 역동적 신호를 독특하게 식별할 수 있습니다. 사회 물리학의 도움 없이 그같은 신호에는 어떤 종류의 통계적 중요성도 부족하며, 그것들을 전통적 머신 러닝과 딥 러닝 방법의 노이즈와 구분되지 않게 만들기 때문 입니다.

머신 러닝 및 딥 러닝 대 사회 물리학 - 어느 것이 어느 목적에 더 나은가? 데이터 과학과 빅 데이터 분석 도구를 사용하여 비즈니스 쿼리를 풀 때, 머신 러닝과 사회 물리학 두 가지 모두 실행 가능한 옵션입니다. 아래 표는 그 속성에 기초하여 적절한 도구를 식별하는 데 유용합니다.

왜 사회 물리학인가? 70 년대로 거슬러, 다양한 수학적, 통계적 머신 러닝 기법은 역사적으로 이미지 프로세싱 및 문자 인식과 같은 ‘정적인 문제’를 위해 개발되었습니다. 그같은 문제는 상대적으로 적은 수의 상대적으로 안정적인 ‘신호’에 의해 차지되었습니다. 교육된 문자 인식 모델은 2016 년 MIT 학생의 손으로 쓴 텍스트를 프로세싱하고 Albert Einstein 의 사적인 편지를 분석할 때 유사한 성과를 달성합니다. 마찬가지로, Siri 또는

Google 의 음성 인식 엔진 어느 것도 J.F.K 의 유명한 “ich bin ein Berliner” 연설을 고음질 녹음으로 바꿔 기록하는 것이 어렵지 않을 것입니다.

머신 러닝이 다음에 대해 더 낫다:		사회 물리학이 다음에 대해 더 낫다:	왜?
데이터 유형	기계적 / 물리적 - 도출 데이터: 예: <ul style="list-style-type: none"> 오동작 예측을 위한 오일 드릴 펌프의 제어 데이터 모니터링 얼굴 인식 	인간 행동 데이터: 예: 프리미엄 서비스를 구매할 사람을 예측하기 위한 금융 거래 분석 	인간의 행동은 변덕스럽고 예측할 수 없으며 매끄럽지 않고 복잡하며 역동적입니다. 수학적으로 말해, 인간 행동은 많은 "일시적" 신호로 지배되고 각각은 소그룹의 개인에게 영향을 미칩니다. 그러므로, 이를 나타내는 일관되고 안정적인 모델을 생성하는 휴먼 데이터를 "학습하기"가 매우 어렵습니다. Endor 는 사회 물리학을 사용하여 그 같은 일시적 신호를 검출함으로써 특별히 인간 기반 데이터에 맞게 맞춤형으로 설계되었습니다.

하지만 인간 행동은 다른 이야기 입니다. 다수의 '역동적 신호'로 지배되어, 매우 역동적이고 '분열'되어 있습니다. 신용카드 구매에서 밀레니얼을 검출하도록 훈련된 전통적 머신 러닝 모델은 시간이 지나면서 정확성도 급속히 떨어져 유능한 전문가가 끊임 없이 유지보수하여 새로운 의미 지식을 그것에 통합시켜야 합니다. 밀레니얼의 행동은 빈번한 (그리고 끊임 없이 변하는) 트렌드의 영향을 받기 때문에 이를 데이터 내에 위치시키는 것은 끊임 없는 모델 재교육 뿐만 아니라 그러한 트렌드를 검출할 의도를 갖는 새로운 특징의 빈번한 개발을 (즉, 원데이터의 일부가 아닌 복잡한 종합적 행동 성질) 요구합니다. 이는 데이터 전문가와 함께 일하는 시멘틱 도메인 전문가의 결합된 작업을 통해서 이루어질 수 있습니다.



**머신 러닝 알고리즘은 70년대 이래
많이 변하지 않았습니다.
최초에는 정적인 문제에 대해
설계되었습니다.**

“사람은 죽을 수 있습니다.
국가는 흥하거나 망할 수
있습니다. 하지만 아이디어는



“일단 우리가 한계를
인정하면, 그것을 뛰어 넘을
수 있습니다.”

다음은 사회 물리학을 휴먼 데이터에서 행동 예측을 위한 도구로 사용할 때의 1 차적인 이점입니다.

	전통적 머신 러닝	딥 러닝 (w.o. 사회 물리학)	사회 물리학	왜?
적은 데이터 세트	적은 데이터 세트를 분석할 수는 있지만 전문가 데이터 과학자가 있어야 하며 많은 시간이 소모되는 과정입니다	모든 질문에 대량의 데이터 필요	인간 행동에 관련된 질문에 답하는 데 아주 적은 데이터 필요. 결과는 자동으로 생성됩니다(데이터 과학자가 개입할 필요 없음)	Endor 가 결과를 생성하는 데 "빅 데이터"는 필요하지 않습니다. 왜냐하면 사회 물리학이 이미 인간 행동 데이터의 근본적인 역동성을 도입하기 때문입니다. 따라서, 아주 적은 데이터 세트로도, 정확한 예측과 행동이 가능한 신호를 즉시 생성할 수 있습니다.

3.2 첫 번째 단계: Endor.com – 기업용 자동 예측 엔진

이전 섹션에서 논의한 바와 같이, 오늘날의 데이터 과학의 상태는 물론 사용 가능한 머신 러닝 테크놀로지는 재력있는 첨단과학 대기업을 위해 보존되었던 그 같은 능력의 사용을 요구합니다. 갈수록 희귀해지는 전문가에게 역대 연봉을 줄 수 있는 기업. 하지만, 이러한

기업들에게조차 기업 데이터 과학과 예측 분석이 상품화되기는 멀었습니다. 반대로 신뢰할 수 있는 예측 생성을 목표로 하는 프로젝트는 전형적으로 각각 서너 주가 걸리며 4~6 회 반복을 위한 2~3 명의 전문가 팀을 구성해야 하고, 첫 번째 모델 제작 후, 일반적으로 끊임 없는 유지관리를 요구합니다. 그러므로 산업 표준에서는 평균적 예측 프로젝트 한 건 당 전체 비용으로 약 \$1 백 5 십만 달러를 추정합니다. 이는 곧 다음을 의미합니다.

- 대부분의 기업들은 소량의 프로젝트 이상에 자금을 지원할 여유가 없습니다.
- 오늘날 ROI 를 플러스로 만들기 위해서는, 프로젝트들에 막대한 마진이 발생하고, 가장 중심적인 비즈니스 관점에만 집중해야 합니다.

	전통적 머신 러닝	딥 러닝 (W.O. 사회 물리학)	사회 물리학	왜?
특징 대 원 데이터	원데이터의 바른 특징 표상을 정의 및 선택하기 위해 숙련된 데이터 과학자 및/또는 도메인 전문가 필요	특징은 필요 없으며 원 데이터를 가공할 수 있지만 극히 좁은 유형의 문제로 제한(*)	인간 행동과 관련된 모든 유형의 예측 문제에서 특징은 필요 없으며 원 데이터 가공 가능	머신 러닝은 길고 종종 수작업의, 가공 전 데이터를 의미 있는 특징으로 변환하는 과정을 요구합니다. 모든 문제 및 모든 새로운 유형의 데이터에서, 이것은 전형적으로 아무 것도 없는 처음부터 시작됩니다. 비록 딥 러닝은 자동으로 만들어지는 특징을 다루지만, 여전히 많은 양의 데이터를 요구하고 데이터 요건은 문제의 복잡성과 함께 더 증가합니다. 그러므로, "간단한 행동"에 제한됩니다. 또한 딥 러닝의 역동성은 전형적인 회사에서는 불가능한 방대한 양의 데이터를 요구하기 때문에 딥 러닝 역시 일반적으로 "정적인 문제"으로 제한됩니다. 사회 물리학은 자동으로 원 인간 행동 데이터를 인간 행동 클러스터 기본형으로 변환합니다. 이 규범적 표상을 사용하여, Endor 는 데이터 크기와 상관 없이, 모든 데이터 유형과 모든 질문에 대응하여 통일된 인간-행동 데이터 세트를 생성하고 그 다음 예측 질문에 답하기 위해 딥 러닝의 힘을 사용합니다.

	전통적 머신 러닝	딥 러닝 (W.O. 사회 물리학)	사회 물리학	왜?
사용자 및 전문지식 필요	일반적으로 도메인 전문가의 도움을 받아, 의미적 특징을 만드는데 도움을 주는 머신 러닝 전문가	딥 러닝 전문가	비즈니스 사용자. 해야 할 일은 오직 한 가지 "더 자세히 알고자 하는 사람"의 예를 제공하는 것입니다.	머신 러닝은 근본적인 정상적 행동에 관한 큰 데이터 세트를 "학습"하거나 또는 이전 도메인 전문지식을 활용할 것을 요구합니다. Endor 는 이미 인간 행동 데이터의 근본적인 역동성을 담아 내고 있습니다.
데이터 변화 속도	느리게 변하는 데이터로 제한. 데이터 변화로 인해 특징의 향상 및 세부적 조정을 위해	역동성을 다룰 수 있는 반면, 느리게 변하는 데이터 (인간 행동 데이터를 다룰 때의 엄격한	빠르게 변하는 데이터 소스를 쉽게 그리고 자동으로 분석합니다(도메	Endor 의 엔진은 특별히 인간 행동 데이터에 맞게 맞춤형으로 구성되기 때문에 이에 본질적으로 역동적 성격의 데이터에서 작용합니다. 사회 물리학이 휴먼 데이터 세트 내 내재된 수학 불변성 세트이기 때문에, 극히 짧은 시간

	도메인 전문가의 향후 지속적인 개입이 필요합니다.	제약!)로 제한됩니다.	인 전문가 불필요).	세그먼트를 나타내는 신호 조차도 검출할 수 있습니다. 다시 말하면, 다른 기법으로 관찰할 수 있기 전에 새로 부상하는 변화를 식별할 수 있습니다.
분석 범위	특정 / 제한적		인간 행동에 관한 광범위한/모든 질문	머신 러닝을 위해, 학습 과정은 자동으로 선택된 모델 특징을 재학습해야 하기 때문에 각 데이터세트와 질문 별로 반복해야 합니다. 사회 물리학은 질문에 국한되지 않는 기본적인 인간 행동 원리에 기초합니다.

	전통적 머신 러닝	딥 러닝 (W.O. 사회 물리학)	사회 물리학	왜?
데이터 클리닝	머신 러닝은 데이터 내 노이즈와 겹에 특히 민감합니다. 길고 값 비싼 데이터클리닝 과정을 요구합니다. 딥 러닝은 종종 데이터를 딥 러닝 도구가 수용할 수 있는 포맷으로 전환하는 주의를 기울여야 하는 과정을 요구합니다.	데이터 클리닝을 요구하지 않음		머신 러닝과 딥 러닝 모두는 규칙을 추론하고 신호를 추출하며 예측을 하기 위해 데이터 기반 수학 패턴을 사용합니다. 이는 정교한 데이터 클리닝 과정을 요구합니다. 다른 한편, 사회 물리학은 외부 패턴 즉, 각 인간 행동 데이터 세트 내에 내재되어 있는 것으로 알려진 수학 불변성을 사용합니다. 이는 데이터 노이즈의 영향을 크게 줄입니다. 또한, 사회 물리학은 가공 전 데이터를 행동 클러스터로 변환하여, 데이터 격차와 노이즈 (대부분 자동으로 여과되어 제거됨)의 영향을 더욱 더 줄입니다.
반복 및 트위킹	데이터 내 각 트윗 또는 문제 정의는 도메인 전문가 (비즈니스 사용자 또는 분석가)와 머신 러닝 \ 딥 러닝 전문가의 공동 작업을 요구합니다. 각 반복은 몇 주가 걸릴 수 있으며 전형적 프로젝트는 최소 4 ~6 회의 반복으로 구성됩니다.	인터페이스는 예측 쿼리를 개정 및 수정할 수 있는 비즈니스 사용자 또는 분석가에 의해 사용되도록 설계되었습니다. 그 다음 결과는 자동으로 새 정의에 맞게 조정됩니다.	Endor의 엔진은 특별히 인간 행동 데이터에 맞게 맞춤형으로 구성되기 때문에 이에 본질적으로 역동적 성격의 데이터에서 작용합니다. 사회 물리학이 휴먼 데이터 세트 내 내재된 수학 불변성 세트이기 때문에, 극히 짧은 시간 세그먼트를 나타내는 신호조차도 검출할 수 있습니다. 다시 말하면, 다른 기법으로 관찰할 수 있기 전에 새로 부상하는 변화를 식별할 수 있습니다.	

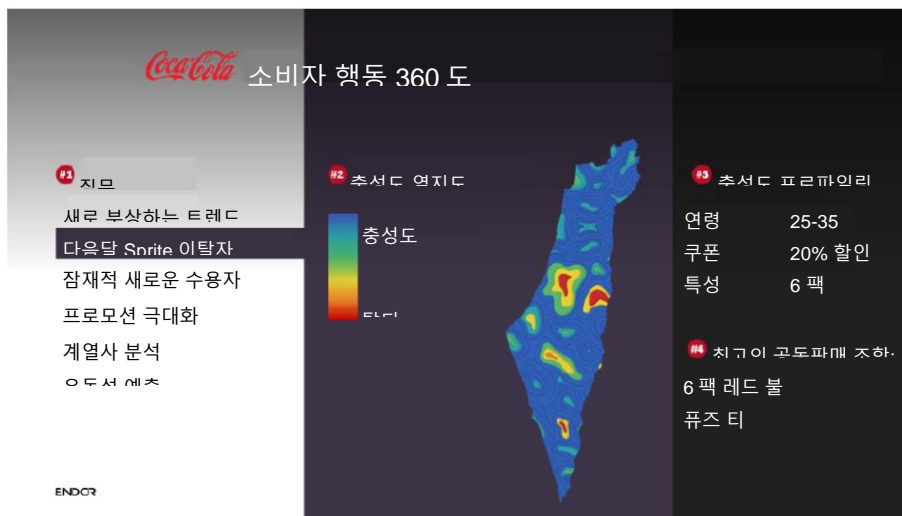
- 기업은 끊임 없이 우선순위를 매겨야 하고 매우 고가의 무기인 데이터 과학을 (a) 기술적 성공 + (b) 높은 사업 수익을 가져다 주리라 생각되는 프로젝트에만 조준해야 합니다.

MIT 분사기업인 Endor 는 새로운 과학인 사회 물리학 주위에서 오늘날 데이터 과학을 인식하는 방식을 교란시킴으로써 이 문제를 풀고자 수립되었습니다. MIT 에서 3 년 간의 연구와 또한 추가로 연구가와 엔지니어 엘리트 팀에 의한 3 년간의 개발에 기초하여, Endor 는 세계 최초의 완전 자동 “서비스로서의 데이터 과학” 엔진을 개발하여 기업들에게 그들이 소유한 행동 데이터에 탑승하게 허용하고, 빠른 통합 후 (전형적으로 단지 몇 시간 만 요구) 데이터에 담겨 있는 ‘객체’ (사용자, 제품, 쿠폰, 위치, 등)의 미래 행동에 관한 *무한한* 양의 예측 질문을 하도록 하게 합니다.

\$1 백만 달러 이하의 연간 비용으로, 기업은 수 십개의 예측 질문을 물어보고 몇 분 내 빠른 결과를 얻게 할 기회를 얻습니다. 이는 단지 선형 개선 만이 아니며 차라리 패러다임의

변화이며 '예측 프로젝트' 실행이 '구글링' 만큼 쉬워지게 되고, 모든 결정이 예측 지향적이 되기 때문에 사실 상, 잠재적 프로젝트의 우선순위 선정작업을 불필요하게 만들고 있습니다.

예 I - 코카콜라 공동 연구: Endor 와 코카콜라 간 최근의 협업에서[16] 사회 물리학으로 광범위한 다양한 소비자 행동을 정확하게 예측할 수 있다는 사실이 입증되었습니다. 여기에는 브랜드 결합 및 충성도, 신 제품 채택, 마케팅 캠페인에 대한 반응, 기타 여러가지가 포함되었습니다. 수 백만의 POS 거래를 원 입력으로서 사용한(3 개월 기간을 나타냄), Endor 의 사회 물리학 엔진은 거의 2 천만 개의 '상관관계가 있는 변칙'을 검출했으며, 각각은 한 개의 단일 실제 세계 행동 그룹을 나타냅니다. 각 그룹의 정확한 의미는 모르지만, 그룹은 그 다음 '행동 외삽법'에 사용됩니다. 샘플 세트(예: 신제품 얼리어답터)가 주어지면, 시스템은 님은 꿀 즉, 샘플 세트 구성원과 유사하게 행동하는 사용자(예: 곧 신제품을 경험할 가능성이 있는 사용자)를 찾는 데 행동 그룹 컬렉션을 사용합니다. 이 방법을 사용하여 15 개의 다른 예측 질문에서 답을 찾았으며 각각은 요구 시 예측 보고서를 생성합니다. 검증을 위해 보관된 샘플 밖 데이터를 사용하여 보고서의 높은 정확성을 입증하였습니다.



“사회 물리학은 빅 데이터 내 행동 분석에 관한 것이지만 완전히 새로운 레벨로 올려줍니다. Endor 를 알게 되고 함께 일하게 된 것은 우리에게 매우 행운입니다.”

Dr. Alan Boehme, CTO, 코카콜라 컴퍼니

예 II - 트윗 자동 분석: 최근의 시험에서, Endor 엔진에 1 천 5 백만 개의 트윗 메타 데이터를 분석을 위한 원데이터로 제공했습니다. 또한 고객은 입력 데이터 안에 들어 있는 /S/S 활동가로 알려진 50 개의 Twitter 계정 신원을 드러냈으며, Endor 가 그 데이터 안에

숨겨져 있는 추가적인 74 개 계정을 검출할 수 있는지 테스트했습니다. Endor 의 엔진은 이 작업을 한 대의 랩톱으로 단지 24 분 만에 완료했으며 (원데이터를 시스템에 입력한 시간부터 최종결과가 나올 때까지 측정), 80 개의 Twitter 계정을 제공한 예와 ‘닭은 꿀’로 식별하였고, 그 중 45 개 (56%)가 74 개의 숨겨진 계정 목록 중 일부임이 밝혀졌습니다. 중요한 점으로, 이는 극히 낮은 허위 경보율(35 개의 허위 긍정 결과)을 제공했기 때문에 고객은 쉽게 식별된 타겟 조사를 인간 전문가에게 의뢰할 수 있었습니다.

“혁명적인 컨셉과 진정한 기술적 돌파구. 그들이 제시한 결과는 다른 경쟁 도구와 일치하지 않습니다.”

이스라엘 정보국의 CIO

생태계 활성화 분산화된 프로토콜을 위한 필요: Endor의 엔진은 사회 물리학과 강력한 전문가 팀을 이용하여, 포춘 500대 기업이 덜 지불하고 더 얻는 방법을 입증하고 있습니다. 하지만, 이 상품은 대형 은행과 소매업체를 위해서만 가능했으며 여전히 연간 라이선스 비용이 \$25만 달러 ~ \$120만 달러에 달합니다. 물론 이것은 대부분의 ‘롱 테일’ 비즈니스는 물론 언급할 필요도 없이 개인에게 적용되는 거래는 아닙니다.

그러므로, 분명 또 다른 솔루션이 필요합니다. *누구나* 사회 물리학이라는 새로운 기술로 혜택을 받고 또한 저렴한 비용으로 이를 가능하게 할 솔루션 그러므로 이러한 솔루션은 다음과 같아야 합니다.

- 특별 참가자에 의해 사용 가능하게 될 자원을 구성하며, 자립적이어야 합니다.
- 참가자에게 계속해서 참여하게 하는 인센티브를 주며, 성장함에 따라 더 낮고 더 저렴해지는 강한 네트워크 영향 생성.
- 공정성을 보존하고 내재적으로 편향되지 않으며 신뢰가 필요없게 합니다. 예측은 그것에 의존하는 개인 또는 조직의 활동에 의해 야기되는 '외부 영향'이라 부를 수 있는 것을 갖고 있다고 예상되기 때문에 이 요건은 다른 서비스와 같지 않습니다.

이 컨셉을 염두에 두고 Endor 팀은 대변혁 즉, *Endor.coin* 프로토콜의 다음 단계를 제시하게 된 것에 자부심을 갖습니다!

3.3 두 번째 단계 : *Endor.coin* 프로토콜 – 일반대중용 데이터 과학

Endor의 행동 분석에 고무되어 산업에서 사회 물리학이 성공적으로 구현된 후 이 능력을 롱 테일 비즈니스는 물론 전문적인 개인에게 제공하는 것을 목표로 *Endor.coin* 프로토콜이 생성되었습니다. 예측 분석 재발명, *Endor.coin* 프로토콜은 행동 예측을 위한 인공지능을 민주화하여, 모두가 접근가능한 생태계를 생성할 수 있습니다. 또한, 프로토콜의 완전히 분산화된 성격으로 신뢰가 필요없고 검열을 방지하며, 책임성을 보증합니다. 최초로, 행동

예측은 모두가 저렴한 수수료로 안전한 프레임워크 내에서 오늘날 데이터와 기술을
컨트롤하는 기술 대기업의 잠재적 조작 없이 사용할 수 있게 됩니다.

3.3.1 정의

원 데이터: *Endor.coin* 프로토콜은 시간 관련 또는 거래 데이터 소스 (예: 통화 데이터 기록, ERC20 블록체인, 인 앱 구매, 등)를 지원합니다. 데이터 제공자는 예측에 기초가 되는 부분을 나타내야 합니다. 예를 들어, ERC20 를 취하면, 흥미 있는 행동 또는 토큰 그 자체의 주소를 예측 및 발견할 수 있습니다. 트윗 데이터는 흥미로운 Twitter ID 또는 대신 흥미로운 위치나 해시태그를 찾는데 사용할 수 있습니다.

처리된 데이터: 행동 클러스터 추출을 겪고 사회 물리학 규범적 표상을 ‘지식 영역’으로서 변환된 데이터를 나타냅니다.

지식 영역: 한 개 이상의 가공 전 데이터 소스에서 추출한 행동 클러스터 컬렉션을 나타냅니다. *Endor.coin* 프로토콜은 지식 영역을 여러 개수로 분리 또는 합칠 수 있습니다. 행동 클러스터는 예측 유형에 따라 서로 다른 관련성을 갖고 있기 때문에, 주어진 예측에 대해 *Endor.coin* 프로토콜은 주어진 시간에 더 큰 관련성이 있는 클러스터를 선택하여 그 다음 실제 예측을 생성하는 데 사용할 ‘지식 영역’을 생성합니다.

예측: *Endor.coin* 프로토콜은 “랭크 그룹 X 대 그룹 Y와 유사하게 행동할 가능성”의 형식을 띄는 질문을 지원합니다. 예를 들어, 그룹 X에 모든 ERC20 토큰이 들어 있고 그룹 Y에 최근 양이 크게 증가한 토큰이 들어 있으면, 예측 결과에 ERC20 토큰 목록이 포함될 것이며, 상부에는 근시일 내 가장 통계적으로 양이 증가할 가능성이 있는 토큰이 있고 하부에는 이 행동을 나타낼 가능성이 가장 적은 토큰이 나올 것입니다..

3.3.2 데이터 제공자

Endor.coin 프로토콜은 행동적, 시간 관련 구조화된 데이터의 통합을 지원합니다. 데이터 온보딩은 간단한 API 콜을 사용하여 이루어지고 데이터 소유자에게 데이터에 기여할 수 있게 하는 한편 어느 행을 비공개로 남기고 어느 것은 공개 분석으로 접근가능하게 할지 관리할 수 있게 합니다. *Endor.coin* 사용자가 비공개로 남겨두도록 정의한 데이터는 여전히 자동으로 공개 데이터-스트림에 통합되어 프라이버시와 공개 데이터의 융합으로부터

수확될 높은 수준의 예측 통찰력을 생성합니다. 사회 물리학의 *지식 영역* (부록 D 참조) 데이터 통합에 대한 개념은 자동으로 친화적으로 사용되며 클리닝 또는 데이터 준비를 요구하지 않습니다.

데이터 제공자는 데이터 분석을 위해 *EDR* 토큰을 예측 엔진 제공자에게 지불해야 합니다. 그들은 다시 자체 데이터에서 나온 통찰력이 예측에 사용된 경우 *EDR* 토큰으로 보상받습니다. 이는 높은 수준의 데이터 스트림에 기여 및 유지관리 하도록 인센티브를 제공하게 됩니다.

3.3.3 예측 엔진

Endor.coin 프로토콜은 데이터를 부분적으로 중복된 행동 클러스터의 형상 공간에 투영한 것에 기초한 ‘예측 언어’를 정의합니다. 현재 Endor 엔진은 출시 후 즉시 사용 가능하게 만들기 위해 그 네트워크에 플러그인할 첫 번째 예측 엔진이 됩니다. 하지만, *Endor.coin* 은 새로운 예측 엔진을 촉진하고 도우며 자금을 지원할 것이며, 복수 유형의 엔진으로 구성된 생태계 형성을 목표로 하며, 보안 능력을 제공하고, 성능 정확성을 강화하며 신뢰성을 증진시킬 것입니다. 예측은 가장 관련성 있는 클러스터에 기초하여 생성되고 *Endor.coin* 오픈 출처 프로토콜에 의해 클러스터 콜렉션에서 자동으로 선정되고, 다양한 예측 엔진에 의해 추출될 것이기 때문에 예측 엔진 수의 증가는 또한 결과의 비편향된 성격을 보장합니다.

3.3.4 사전 정의 예측 및 예측 요청 (RFPs)

Endor.coin 프로토콜은 다양한 사전정의 예측을 담고 있는 큰 카탈로그와 함께 출시됩니다. 이러한 예측은 플랫폼의 *EDR* 토큰을 사용하여 구매를 위해 접근할 수 있을 것입니다. *Endor.coin* 은 그 다음, 사용자에게 *예측 요청 (RFP)*을 전송하도록 허용하고, 구매용으로 실행되고 사용할 수 있게 될 새로운 유형의 예측을 제안함으로써 이것이 제공하는 예측의 선택 범위를 확대합니다.

플랫폼에 의해 지원되는 예측을 점차적으로 확대해 나가는 이 방법은 집단지성 (프로토콜 사용자의 요청에 의해 나타나듯)을 사용하여 새로 지원되는 예측 선택을 최적화할것입니다.

3.3.5 비공개-데이터 분석 및 셀프 서비스 API

Endor.coin 프로토콜의 이후 출시품에는 상급 사용자를 위한 완전한 ‘DIY’ API 지원이 포함됩니다. 최신 기술에 정통한 사용자와 전문 데이터 과학자는 셀프 서비스 인터페이스를 사용하여 쉽게 독점 데이터에 탑승하여 자신들만의 고유한 새로운 유형의 예측을 생성할 수 있게 됩니다. 이러한 예측은 개인정보로서 정의되거나 또는 공개적으로 공유할 수 있습니다 (널리 사용되게 되면, 예측 개발을 *EDR* 토큰으로 보상).

3.4 로드맵

오늘날 일반 소프트웨어 시장에서, 당사의 접근방식은 서비스로서 플랫폼 또는 더 최근에는 서비스로서의 블록체인과 같은 제안과 비교될 수 있습니다. *Endor.coin* 토큰 (또는 *EDR*) 은 플랫폼 상에서 거래를 촉진하기 위해 사용됩니다. *EDR* 은 키 또는 소프트웨어 라이선스 역할을 하며, 시간이 흐르면서 더 많은 토큰이 애플리케이션 영역 확대를 꾀하는 개발자 커뮤니티를 통해 성능과 규모를 증가시키기 위해 사용될 수 있습니다. 또한 당사 관리자가 토큰을 배포하고 사용 용도를 모니터링하며 필요 시 더 많은 토큰을 구매할 수 있도록 대시보드가 만들어 질 것입니다.

Endor.coin 의 진화에서 다음 단계는 계속 늘어나는 캐탈리스트 커뮤니티에 의한 생태계의 유기적인 확대입니다 . 모두가 EDR 로 지불된 RFP(“예측 요청”)를 제공할 수 있고, 각 RFP 를 처리하는 문제는 캐탈리스트가 수행할 수 있으며 이 캐탈리스트는 예측범위를 확장하는 데 기여하기 위한 스마트 계약을 통한 각 결제로부터 보상을 받습니다. 이 때 이 예측범위는 Endor.coin (모두가 EDR 을 사용하여 액세스할 수 있는)에 의해 새롭게 지원되는 쿼리로서 내장됩니다.

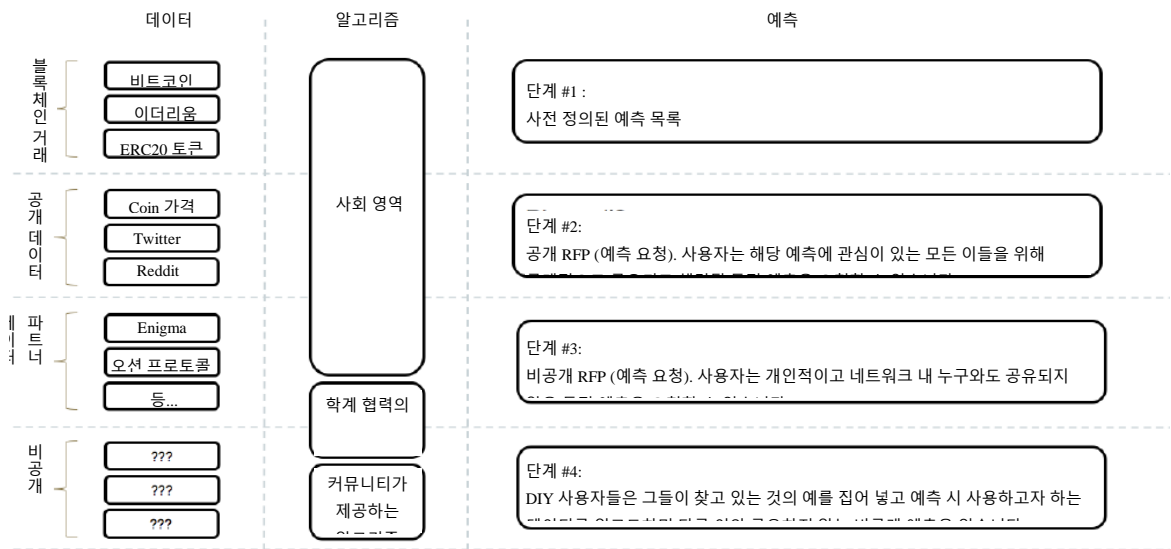
이 두 번째 단계에서 *Endor.coin* 은 모듈식 앱 플랫폼이 되고 여기에서 캐탈리스트는 블록체인의 힘을 통해 예측 도메인을 확장시킬 수 있습니다. 이를 가능하게 하기 위해 *Endor.coin* 은 유용하고 고도로 맞춤화된 쿼리 앱의 계속해서 커지는 라이브러리를 생성하기 위해 영속적으로 결합 및 재결합할 수 있는 재사용 가능 소프트웨어 구성요소의 저자들에게 소액 결제 오픈 엔드 스트림에 힘을 실어줍니다. 캐탈리스트에게는 *Endor.coin* 플랫폼에 추가한 각 구성요소에 대해 Micro Use 라이선스가 부여됩니다. 최종 사용자는 그들이 선택한 앱을 설치합니다. 지급될 라이선스는 그 앱에 의해 사용되는 구성품의 모든 Micro Use 라이선스의 합입니다. 스마트 계약을 통해 *Endor.coin* 은 최종 사용자에게 청구하고 결제액을 개입한 캐탈리스트 각자에게 배분할 책임이 있습니다.

시간이 흐르면서 *Endor.coin* 은 블록체인에 기초한 데이터가 풍부한 웹 및 모바일 제품을 구축하고자 하고 그래서 지금까지는 고도로 숙련된 이들만이 접근할 수 있는 머신 러닝을 민주화 하고자 하는 기업가 컴퓨터 프로그래머와 기업을 위한 최고의 개발 플랫폼이 됩니다. 또한, 오늘날 데이터 분석 알고리즘을 효율적으로 실행하기 위해 필요한 거래 량을 지원하지 않는 기존 블록체인은 쿼리 도메인을 계속해서 확장하는 엔진을 사용하는 애플리케이션을 구축할 끊임 없이 성장하는 캐탈리스트 커뮤니티의 도움을 통해 그러한 능력을 제공할 수 있게 됩니다.

우리의 비전은 최종 사용자에게 완전히 투명하고, 완전 자동화되어 신뢰가 필요 없는 분산화된 예측 기반 인프라를 만드는 것입니다. EDR 토큰과 애플리케이션에 지불되는 인센티브는 필수적으로 일종의 플러그인입니다. 포지티브 피드백 루프가 예상됩니다: 애플리케이션을 더 많이 구축할 수록, 더 많은 플러그인이 시스템에 추가되고 더 많은 고유 구성요소가 재사용할 준비가 됩니다. 이는 *Endor.coin*

생태계 전체에서 상호간에 스스로 강화된 네트워크 효과에 기여합니다.

궁극적으로 우리는 하이 엔드 고객을 위해 플래티늄 플랫폼을 전달하며 여기서 전문가와 비즈니스는 더 복잡한 정교한 RFP 를 제출할 수 있습니다. 그 같은 복잡한 쿼리를 전달하기 위해, 제공자에게 다음과 같이 *EDR* 로 보상함으로써 당사의 데이터 액세스를 확장합니다. 캐탈리스트가 객체로부터 데이터를 요청할 때(시간 제한) 각 객체는 가격을 제안합니다(시간 제한). 캐탈리스트가 지급액에 동의하면, 객체는 캐탈리스트 공개 키로 암호화한 데이터로 연결되는 링크를 전송합니다(가격이 스마트 계약을 통해 지출되는 시점). 캐탈리스트와 데이터 제공자 모두는 리뷰를 발표할 수 있습니다 (미래 명성을 위해) . *EDR* 은 데이터 제공자의 결제액과 플래티늄 애플리케이션 개발자에게서 받은 기금 간 차익거래로 생성됩니다.



제 4 장

신뢰가 필요없고 검열을 방지하며 책임을 짐

4.1 인공지능을 위한 책임성 및 진정성

그의 논문 “*아직 잘 알고 있지 않지만 우리가 블록체인 기반 AI 을 원하는 이유*” [34]에서, 저자 Jeremy Epstein 은 그와 그의 어린 딸이 Amazon 의 AI gadget Alexa 와 나눈 대화를 그의 독자들과 공유합니다. 이 대화는 Alexa 가 그의 청취자와 엄청난 양의 관련 정보를 공유하는 동안 ‘네트워크 중립성’이란 주제 중심으로 천천히 흘러 갑니다. Epstein 의 이 주제에 대한 정통함에 경험한 Alas 는 놀랍게도 정보가 비록 정확하긴 했지만 잠재적으로 한 쪽에 치우쳐 있음을 발견했습니다. 경험은 우리가 워프정도 속도로 돌진해 가고 있는 AI 으로 움직이는 미래의 리스크 일부를 강조합니다. 또한 Amazon 과 같은 큰 기업이 전통적으로 빅 데이터와 AI 에 있어서 큰 이점을 갖고 있음을 상기시켜 줍니다.

Epstein 인용:

가능한 한 많은 데이터를 수집 및 저장, 분석하는 경기가 있다고 하면, 누가 이길 가장 유리한 입장에 있을까요? 맞습니다, 미국의 FANGs (Facebook, Apple, Netflix, Google)과 중국의 BATs (Baidu, Alibaba, Tencent), 부유한 포춘 1000 대 기업, 기타 다국적 기업입니다.

더 많은 데이터를 얻고, 그것을 저장 및 분석하며 그것들 상부에 AI 모델을 구축할 권한 범위와 자본을 가진 유일한 기업들입니다. 그 외에, \$300,000 ~ \$500,000 범위의 초봉과 7 자리, 8 자리 수까지로 늘어나는 최고 연봉을 제시할 수 있는 유일한 기업들입니다. 자녀가 NBA 또는 NFL 에서는 그렇게 할 수 없겠지만 최고 AI 과학자가 될 수도 있으며 그렇다면 정말 잘 한 것입니다.

이것 모두의 순 효과로 부자는 심지어 더 부자가 되고 더 막강해지고 혁신 장벽이 심지어 더 높아집니다.

하지만, 시달림을 받는 것은 혁신만이 아닙니다. 큰 회사 AI 의 패쇄적 성격은 사회가 신뢰를 “검은 상자” 안에 넣어야 한다는 의미입니다.

그러므로 “*예측 진정성*” 제공은 다음 모두를 제공할 기반시설의 가용성에 의존합니다.

- **책임성:** 예측 제공자는 돌이켜 생각해 보며 그들의 예측이 맞았음을 입증할 수 있습니다. 예측 소비자는 모든 유형의 정확한 예측 생산하는 데 있어서의 예측 제공자의 효율성 (또는 부족)을 돌이켜 보며 확실하게 추론할 수 있습니다. 다시 말해, 명성은 조작할 수 없어야 합니다.
- **진정성:** 예측 제공자는 자신들이 그렇다고 말하는 사람들입니다. 소비자를 호도하는 예측의 원인을 의인화할 수 없어야 합니다.
- **비편향:** 예측 제공자 간 공정하게 경쟁해야 하며, 편향된 예측을 통해 달성된 암시적 이득보다는 정확성에 인센티브를 주기 위해 시장의 힘을 활용할 수 있습니다.
- **정확성:** 예측은 일반적으로 말해, 앞의 세 가지 요건과 관련하여 긍정적인 피드백 루프를 닫을 금전적 유입을 네트워크에 제공하기에 충분할 정도로 정확해야 합니다.

Endor.coin 프로토콜은 위의 요건 모두를 만족할 기반시설을 제공합니다.

- ***Endor.coin* 프로토콜 책임성 및 진정성:** 블록체인 사용. 예측은 영구적으로 저장되고 누구나 이용할 수 있습니다. 시간에 민감한 예측만 소비자에게 전송할 때, 암호화되었지만 공개 버전으로, 일정 시간 후 풀리는 키와 함께 저장됩니다.
- ***Endor.coin* 프로토콜 제로 편향:** 프로토콜은 각 예측마다 가장 관련성 있는 행동 클러스터를 선택합니다(분석 엔진 또는 데이터 소스와 상관 없이). 그러므로, 프로토콜은 데이터 제공자는 물론 분석 엔진 제공자가 제어하지 않는 별개의 개체이기 때문에 편향은 기본적으로 방지됩니다.
- **정확성:** *Endor.coin* 은 많은 다양한 사용 사례에서 정확한 예측을 일관되게 제공하는 것으로 산업계에서 널리 입증된 MIT 의 *사회 물리학* 기술에 기초합니다.

4.2 분산화된 프로토콜을 통한 검열 저항

검열은 다루기 곤란한 일입니다. 명시적 발현이 존재할 때 검출하기 쉽기 때문에 IP-프록시 등과 같은 '낮은 레벨'의 기술 솔루션을 사용하여 우회하기 쉽습니다. 다른 한편 암시적 검열은 다른 이슈입니다. Google 및 Facebook 과 같은 기업은 일정 유형의 검색 쿼리를 차단하는 것으로 알려져 있습니다. 많은 경우에 이는 올바른 윤리 또는 법적 이유로 그렇게 되었지만, 그것이 항상 바른지 어떻게 확신할 수 있습니까? 만약 "쉽게 액세스하지 말아야 한다고" 결정되었을 때 우리가 Google 에게 관련 웹페이지를 제공하도록 "압박"할 수 있습니까? 우리가 Alexa 에게 부적절하게 여겨질 수 있는 우리의 요청 중 일부에 따르면 "압박"할 수

있습니까”? 답은 간단합니다 – 아니요.

이 맥락에서 예측은 검색과 매우 유사하며, 이런 면에서, 그것은 운영 객체 또는 규제기관에 의한 검열 대상이 됩니다. AI 와 예측 분석의 작동 방식을 살펴보면, 잠재적 검열과 관련하여 세 가지 필수적 층이 있습니다.

- **데이터 저장소:** 데이터의 무결성과 완전성, 보안 보장 (입력이 정확하고 신뢰할 수 있으며, 조작되거나 또는 도난당할 수 있는가?)
- **알고리즘/머신 러닝 엔진:** 예측은 중앙 당국의 검사를 받지 않으며, 모든 예측 요청은 공평하게 실행되며, 예측 주제와 상관 없이 서비스 품질이 고려됩니다.
- **쿼리 인터페이스:** 예측 쿼리의 출력을 신뢰할 수 있게 나타내며, 효과적으로 새로운 데이터를 잡아 내며, 예측 주제와 상관 없이 증거로 지지되는 예측으로 제한합니다.

만약 어떤 이가 그의 의사결정을 신뢰하기 위해 중앙 집중식 예측 출처로 간다면, 그는 절대적인 확신으로 위의 모든 요건이 만족되었음을 암시적으로 가정합니다. 이 중앙 집중식, 폐쇄 모델 예측 계획에서, 그 장면 뒤에 무엇이 올지 모르는 채 각 층을 신뢰하는지 라는 질문을 받았을 때 그 같은 확신은 정당화 하기가 힘이 듭니다(그대로 속은 것이 아닌 한).

Endor.coin 프로토콜은 예측을 완전 분산화되어 신뢰가 필요 없는 방식으로 실행되게 함으로써 그러한 주장을 본질적으로 제공합니다. 일단 데이터가 네트워크에 기여하고 거기에 탑승하면, 어떤 관련 예측도 그것을 자동으로 극대화하는 *Endor.coin* 프로토콜로 실행할 수 있습니다.

4.3 프라이버시 보호 데이터 분석: 암호화한 데이터 사용

사회 물리학이란 완전 새로운 과학의 사용은 데이터 분석 및 예측을 위한 민주화된 생태계 형성에서 주요 이점을 제공합니다. 사회 물리학 산술이 암호화한 데이터에서 이루어지기 때문에 데이터 프라이버시 또는 무결성을 훼손하지 않고 데이터 사일로 전체에서 산술을 가능하게 합니다. 특히, *Endor.coin* 의 고유한 데이터 분석 방식으로 데이터 기여자는 데이터를 노출하지 않고도 데이터 분석을 암호화한 형태로 화폐화할 수 있습니다. 대신, 시스템은 이후 '땀은 꿀 예측'을 위한 기초로서 사용할 '암시적 행동 클러스터'(해시 또는 암호화한 데이터 공간으로 매핑된)를 추출할 수 있어, 그 결과, 명백하게 소스 데이터의

성격이고 예측 질문의 의미인 정확한 행동 예측이 생성됩니다.

오늘 날 *Endor* 는 이미 유사한 기술을 사용하여 대형 은행 및 금융 고객과 함께 일하며 그들에게 그들 데이터를 완전 암호화한 버전으로 온보드할 수 있게 하고, 데이터 프라이버시와 무결성 훼손을 불가능하게 만듭니다.

4.4 네트워크 효과

Endor.coin 프로토콜의 완전 분산화는 서너 네트워크 영향 내 다른 것 중에서도 용도가 넓혀짐에 따라 장점이 증가하는 결과를 낳았습니다. 다음은 증가된 프로토콜 채택에 의해 촉발될 수 있는 주요 요지입니다.

사용자 추가는 각자 지불해야 할 비용을 줄여줍니다. 대형 커머셜 고객과 달리, 개인과 중소기업은 주로 공개 데이터 세트에 기초한 예측을 요구합니다. *Endor.coin* 프로토콜은 한 번 분석할 그 같은 데이터 세트를 사용 가능하게 하여 “지식 영역”으로 알려진 행동 클러스터의 콜렉션을 생성합니다. 이 데이터 구조는 이 데이터 소스에 기반한 예측 이용 사례 모두를 지원하며, 최종 사용자에게 “개인화 인자” 즉, 행동 클러스터 추출과 그들이 관심을 갖는 특정 이용 사례 간 델타를 압축하는 대가로 적은 금액의 수수료를 지불할 것을 요구합니다.

다시 말해, 주어진 데이터 세트 (즉, 비트코인 블록체인)와 특정 이용 사례에서, 비용은 다음과 같이 계산합니다.

- **지식 영역 계산:** 전체 비용의 약 99%로 구성되며 한 차례 계산합니다. 이 일에 필요한 자원은 이 데이터 스트림의 모든 사용자 간에 분할 상환합니다.
- **예측 개인화 인자:** 전체 비용의 약 1%로 구성되며 각 사용자 별로 계산 (지급)합니다. 또한 다른 사용자가 한 유사 예측의 경우, *Endor.coin* 프로토콜은 자동으로 먼저 계산한 예측에서 도출한 통찰력을 다음 예측 계산 중에 사용하여 동일 비용에 대한 예측 정확성을 더 개선합니다.

이는 N 명의 최종 사용자 네트워크를 지배하는 비용 식은 다음과 같을 것임을 의미합니다.

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

C 는 단일 커머셜 기업에 의한 동일 예측에 대한 비용을 나타냅니다. 이 궁극적으로 이는 사용자 수 N 이 증가함에 따라 예측의 정확성이 증가하는 한편 비용은 대형 커머셜 고객이 사용한 동일 시스템 비용의 약 1%를 갈망한다는 의미입니다.

궁극적으로, 질문을 묻는 사람이 많을수록 그들이 받는 결과가 더 나아지고 대부분의 비용이 전체 적극적 사용자 수로 나뉘지기 때문에 비용이 낮아집니다.

데이터 제공자 추가로 정확성 증가: *Endor.coin* 프로토콜의 정의에 의거하여, 데이터 제공자는 데이터에서 추출한 행동 클러스터에 기초한 “지식 영역” 생성을 위해, 데이터 실행 자금을 *EDR* 토큰을 사용하여 제공해야 합니다. 이 초기 실행 비용은 최종 사용자가 데이터 제공자에게 데이터 품질에 따라 그리고 다양한 예측 쿼리에서의 기여도에 따라 부분적으로 또는 동일하게, 큰 이자와 함께 상환합니다. 이 기금 중재는 다양한 데이터 세트 상에서 다양한 분석 엔진에 의해 검출된 행동 클러스터의 전체 콜렉션에 접근할 수 있고 가장 높은 관련성이 있는 상위 클러스터를 선택할 수 있는 *Endor.coin* 프로토콜이 관리합니다. 실행 토큰은 그 다음 그러한 데이터 세트 제공자에게 최종 예측에서 그들의 기여도에 비례하여 전달됩니다.

이 메커니즘의 결과는 높은 수준의 데이터 소유자에게 계속해서 데이터 소스를 지원(그리고 심지어 품질과 가용성을 더욱 증진)하도록 인센티브를 주는 한편, 뒤 떨어진 품질의 데이터 제공자는 사용한 비용을 복구할 수 없기 때문에 도태됩니다. 이 경제는 시장의 힘을 활용하여 예측 프로토콜이 사용할 수 있는 데이터 소스에 가능한 한 최고의 적합성을 유지하도록 자동으로 보장합니다. 그러므로 신규 데이터 제공자 만이 전반적 비용을 증가시키지 않고 전반적 예측 정확성을 높입니다.

예측 엔진 추가는 예측 효율성을 높입니다: *Endor.coin* 프로젝트의 첫 번째 단계는 행동 클러스터 추출의 첫 번째 제공자 역할을 하기 위해 *Endor* 예측 엔진에 의존합니다. 하지만, 지금으로서는 추가적인 예측 엔진이 *Endor.coin* 프로토콜을 지원할 것으로 기대됩니다. 그같은 새로운 예측 엔진 추가는 데이터 제공자는 물론 최종 사용자에게 크게 긍정적인 영향을 미칠 것으로 기대됩니다. 다른 기술을 사용하는 엔진은 동일한 데이터 소스에서 다른 유형의 클러스터를 생성할 것을 예상됩니다. 이는 *Endor.coin* 프로토콜을 지원하는 예측 엔진에 사용된 다양한 기술이 증가함에 따라 새 예측이 요청될 때 그것에서 선택하기 위해 프로토콜에서 사용하게 될 다양한 클러스터 역시 증가할 것이라는 의미입니다.

이는 세 가지 주요 효과를 얻기 위해 기대됩니다.

- **개선된 정확성과 새로운 예측을 위한 증가된 지원:** 새로운 유형의 클러스터가 사용 가능해짐에 따라 *Endor.coin* 프로토콜은 요청된 각 예측을 생성하는 데 사용할 클러스터를 더 잘 선택할 수 있게 됩니다. 기존 예측에서는 이로 인해 정확성이 증가합니다 (클러스터링 기술 사용으로 나온 “직교 통찰력”의 가용성으로 인해).

하지만, 이는 또한 차선의 정확성으로 인해 시스템에 의해 지지되지 않았던 시점까지는 그랬으며 이제는 인접 시간에 실행된 예측 간 통찰력 공유덕분에 새로운 예측은 비용 효율적으로 실행할 수 있게 되었으며, 시스템으로 지지되는 예측이 다양하게 증가되고, 이후 전체적인 정확성이 강화되었음을 암시합니다.

- **예측 당 감소된 비용:** 시스템으로 지원되는 예측량 증가로 또한 유료 예측 최종 사용자의 수가 증가하며 또한 바로 이어, 초기 데이터 분석에 지급되었던 가용 EDR 토큰의 전체적인 풀 역시 증가하고 이후, 데이터 제공자가 수령하는 자금액 역시 증가하는 한편, 최종 사용자를 위한 예측 당 비용이 낮아집니다.
- **새 데이터 소스의 경제적 지속가능성:** 마지막으로, 새로운 클러스팅 기술의 가용성은 또한 새로운 기술이 그러한 데이터 소스의 분석에서 가용 기술 보다 호환성이 더 나은 경우, 일정 유형의 데이터 소스의 기여를 경제적으로 실행 가능하게 만들 수 있습니다. 이러한 경우에서, 그러한 데이터 소스는 갑자기 *Endor.coin* 프로토콜을 위한 중요한 정보 출처가 될 수 있으며, 초기 플랫폼 통합 비용을 상환하고 또한 이후, 가용 데이터 소스의 양이 그것에서 나오는 혜택과 함께 증가합니다.

제 5 장

생태계 활성화

Endor.coin 프로젝트는 다각적인 생태계 수립을 목표로 하며, 데이터 소유자와 개발자, 데이터 과학 전문가, 중소기업, 개인사용자 간 시너지 협업을 생성하고자 합니다. 이들 각 기업에게 그들의 자산 (데이터, 기금, 추가해야 할 새로운 예측 관련 아이디어 등)을 기여하게 함으로써, *Endor.coin* 커뮤니티는 자기 영속적 포지티브 피드백 비즈니스 루프를 가능하게 하고 촉진합니다. 다음은 이 생태계의 주요 측면에 관한 간략한 설명입니다.

5.1 공개 데이터 제공자

Endor.coin 프로토콜은 1 일차 부터 공개 데이터 스트림의 기여에 인센티브를 줄 것입니다. 데이터는 첫 번째 단계에서는 ‘등록 데이터 제공자’ 가 제공하고, (간단한 법 준수 및 품질보증 과정 진행, 생태계가 최고 품질 데이터로 자구력을 가질 수 있도록 보장하기 위해), 이후에는 분석을 위해 데이터를 풍요롭게 하길 원하고 그렇게 함으로써 *EDR* 로 보상을 받는 개인 또는 회사 누구나가 제공합니다.

잠재적 데이터 기여자의 예 :

- **데이터 파트너:** 분석을 위해 그들의 데이터 스트림 (부분적으로 또는 전체적으로)을 *Endor.coin* 플랫폼에 연결하는 *Enigma, Twine, Thasos, Ocean 프로토콜* 등과 같은 프로젝트.
- **소셜 채널:** *Twitter, Reddit*, 기타.
- **데이터 스크레이퍼:** 다양한 가용 공개 출처에서 데이터를 수확 및 클리닝, 구조화, 의미적으로 풍부하게 만들.
- **블록체인 프로토콜:** *ORBS* 또는 유사 시스템 . 데이터를 다운로드 및 분석할 수 있고, 분석을 위해 액세스할 수 있게 만드는 노드 수립.

Endor.coin 은 이미 서 너 군데의 그 같은 잠재적 기여자와 파트너십을 협의하고 있습니다.

5.2 학계 연구 그룹

Endor.coin 프로토콜이 원 데이터의 클러스화를 위해 복수 분석 엔진의 통합을 지원하고 있기 때문에, 회사는 이 분야를 선도하는 학계 그룹의 협업을 장려하기 위해 일부 수익금을 활용할 의도를 갖습니다. 이 활동 동안 개발될 기술은 쉽게 증가하는 *Endor.coin* 생태계에 통합시키고 이후 그것이 예측 정확성을 향상시키는 데 사용될 때 마다 개발자에게 *EDR* 토큰으로 보상할 수 있습니다.

이러한 노력은 산업 및 학계 데이터 과학과 머신 러닝, 블록체인 기술의 세계적인 리더들로 구성된 *Endor.coin* 과학 자문위원회 (섹션 8.2 내 세부목록 참조)의 지도하에서 진두지휘 될 수 있습니다.

MIT 분사기업으로서 *MIT* 에서 프로젝트의 창립자에 의해 개발된, 사회 물리학의 과학적 혁명에 의존하는 이 회사는 *Endor.coin* 이 선도하고, 학계와 데이터 분석 분야와 데이터 시장에서의 블록체인 프로젝트, 데이터 통찰력과 예측의 확산 및 이용을 위한 기반시설에서 선도적인 연구 그룹을 포용하는 강하고 지속가능한 동맹의 생성을 위해 노력을 다하고자 합니다.

5.3 캐털리스트 – 애플리케이션 개발자

Endor.coin 이 다룰 수 있는 질문 범위를 늘리기 위해, 도메인은 무한하며 사회 물리학 엔진을 사용하여 애플리케이션을 구축하고자 하는 이들 즉, '캐털리스트' 커뮤니티에 의해서만 제한됩니다. 이에 힘 입어 *Endor.coin* 은 모듈식 앱 플랫폼이 되고 그 위에서 캐털리스트는 블록체인의 힘을 통해 예측 도메인을 확장시킬 수 있습니다. *Endor.coin* 이 다룰 수 있는 질문 범위를 늘리기 위해, 도메인은 무한하며 사회 물리학 엔진을 사용하여 애플리케이션을 구축하고자 하는 이들 즉, '캐털리스트' 커뮤니티에 의해서만 제한됩니다. *Endor.coin* 은 모듈식 앱 플랫폼이 되고 그 위에서 캐털리스트는 블록체인의 힘을 통해 예측 도메인을 확장시킬 수 있습니다. *Endor.coin* 은 유용하고 고도로 맞춤형화된 앱의 계속해서 커지는 라이브러리를 생성하기 위해 영속적으로 결합 및 재결합할 수 있는 재사용 가능 소프트웨어 구성요소의 저자들에게 소액 결제 오픈 엔드 스트림을 활성화합니다. 캐털리스트에게는 그들이 프레임워크에 추가한 각 구성요소에 대해 Micro Use 라이선스가 부여됩니다. 최종 사용자는 그들이 선택한 앱을 설치합니다. 지급될

라이선스는 그 앱에 의해 사용되는 구성품의 모든 Micro Use 라이선스의 합입니다. 스마트 계약을 통해 *Endor.coin* 은 최종 사용자에게 청구하고 결제액을 개입한 캐탈리스트 각자에게 배분할 책임이 있습니다. 모든 것이 자동이고 최종 사용자에게 투명합니다. 인센티브는 토큰으로 지급됩니다. 애플리케이션은 플러그인 세트에서 필수적이며, 포지티브 피드백 루프가 예상됩니다: 애플리케이션을 더 많이 구축할 수록, 더 많은 플러그인이 시스템에 추가되고 더 많은 고유 구성요소를 재사용할 준비가 됩니다. 이는 *Endor.coin* 생태계 전체에서 상호간에 스스로 강화된 네트워크 효과에 기여합니다.

5.4 데이터 주권

탐욕스러운 기업의 이익을 위해 우리의 프라이버시를 소비하는 세상에서, *Endor.coin* 은 게임을 바꾸고 부를 데이터 접근권을 제공하는 이들에게 돌려주기 위해 최선을 다합니다. 간단히 *Endor.coin* 에 가입함으로써 EDR 토큰으로 가능한 다양한 시나리오에 의거하여 펼쳐질 수 있는 과정을 시작할 수 있습니다. 구성원은 그들 자신의 데이터에 기초한 개인적 통찰력을 얻을 수 있거나 또는 하나의 그룹으로써 팀을 구성하여 더 큰 데이터 풀을 형성하고, 중소기업과 함께 일하며 예를 들어 최적의 시기에 최선의 가격으로 알맞은 제품에 주목하는 등 더 개인화된 서비스를 얻을 수 있습니다.

사람들은 그들이 생성한 데이터를 소유 및 관리하고 그것으로부터 혜택을 얻을 권리를 갖고 있다는 근본적 신념에 기초하여 *Endor.coin* 은 현재 재력있는 투자자에게만 가능한 통찰력의 힘을 아직까지는 사회적 혜택을 받지 않은 중소기업에 열어주어 이제 그들은 *Endor.coin* 구성원의 디지털 데이터를 데이터 소유자의 허가로 자신들의 용도에 맞게 종합하여 높은 가치의 데이터 분석을 제공하고 그것은 다시 그들의 구성원에게 다시 그 분석의 결과로서 높은 효율적인 서비스은 물론 데이터를 다른 방식으로 사용했을 때의 금전적 계인을 가져다 줍니다. *Endor.coin* 으로 지원되는 이 개인 데이터 독립성 기반 접근방식은 기업과 사람 모두에게 서로에 대한 가치 손실이나 신뢰 손실 없이 막대한 가치를 사용 가능하게 합니다. *Endor.coin* 은 기업과 기타 조직이 데이터 관련 법적 책임을 줄이고 그들 고객에 관한 우수한 데이터에 접근하고, 고객관계를 발전시키기 위해 가입함에 따라 네거티브 취득 비용으로 멤버십과 브랜드 참여를 유도함으로써 사업을 지원합니다.

멤버십이 계속해서 증가하기 때문에, *Endor.coin* 역시 구성원이 그들 자신의 데이터에서 삶의 여러 측면에 관한 통찰력과 이해를 얻을 수 있도록 도울 도구와 서비스를 제공합니다. “사람의, 사람에 의한, 사람을 위한 데이터”라는 미션으로 *Endor.coin* 은 지금까지는 Facebook, Google, Uber 등과 같은 많은 다른 데이터 기반 대기업에 의해 위반되고 있는, 개인에게 그들의 삶을 활용할 고유 권리를 다시 돌려 주고, “플랫폼 경제”를 번창시킵니다. 입증된 기업 전문지식과 미래 지향적 법적 사고와 기술적 노하우, 소비자를 위한 가치 창조를 통합하는 다각적인 접근방식을 통해, *Endor.coin* 은 데이터 통찰력으로 액세스를 허용하는 구성원에게 직접 보상함으로써 기업에게 그들 고객과 선의의 관계를

형성하는 합의적 데이터 관계를 생성함으로써 “데이터 불법복제” 문제를 해결하는 고유한 비즈니스 모델을 제공합니다.

이 과정의 무결성을 보장하기 위해, 우리는 *Endor.coin* 과 시너지 관계로 일하며 구성원을 위한 혜택과 가치 창조, 데이터 보안을 최대화하는 독립적으로 관리되는 객체인 *Endor.coin Trust* 를 창설하고 있습니다. 사람들이 *Endor.coin* 에 가입하면, Trust 는 그들의 데이터를 감독하고 그들은 자동으로 Trust 구성원이 됩니다. *Endor.coin Trust* 는 구성원 데이터의 안전과 보안, 프라이버시, 가치를 보장 및 최대화하는 많은 방식으로 구성원을 대변합니다. Trust 는 개인 데이터 관리 및 가치 구현, 프라이버시 기준을 설정 및 강화함으로써 구성원의 데이터를 보호합니다.

제 6 장

토큰 구현

6.1 블록체인 구조

데이터 저장: 블록체인은 일반용 데이터베이스가 아닙니다. *Endor.coin*에는 블록체인을 통해 접근할 수 있으며 데이터 자체가 아닌 데이터 레퍼런스가 저장된 분산화된 오프 체인 배포 해시 테이블 스토리지가 있습니다. 비공개 데이터는 전송 및 보관에 앞서 암호화 (AES-256 – Amazon S3 서버측 암호화 또는 AWS IAM 메커니즘을 사용하여) 되며, 액세스 관리 프로토콜이 블록체인 내에서 프로그램 됩니다. *Endor.coin*은 오프 체인 네트워크(데이터베이스, 중앙 스토리지 등의 형식으로 보관, 결과적으로 *Endor.coin* AWS 기반시설에 업로드되는 구조화된 포맷으로 내보낼 수 있음)에서 기존 블록체인은 물론 비공개 및 공개 데이터 세트에 연결하도록 설계되었습니다. 미래 출시품에서, Data Layer는 EDR 토큰을 받고 데이터를 판매할 수 있는 외부 기업에게 개방될 것입니다. 데이터는 인증되고 데이터 소유자가 사용할 수 있으며, ‘공개’라는 마크가 있는 경우, 예측 알고리즘이 관련이 있다고 간주하는 고객 역시도 사용할 수 있는 *Endor.coin* 오프 체인 기반시설에 업로드됩니다. 가격은 데이터 제공자가 결정하고 *Endor.coin* 프로토콜에 의해 역동적으로 이 예측에 대한 수요에 의거하여 자동으로 조정되어 사용자에게 동일 데이터 소스를 위한 비용을 공유하게 할 인센티브를 줍니다.

예측 소비: 블록체인은 복잡하고 활발한 거래는 취급할 수 없습니다. 동일 오프 체인 산출 네트워크는 힘든 산출 (다양한 예측 엔진에 의해 요구됨)을 가동하기 위해 사용합니다. 일단 결과가 나오면, 최종 사용자가 사용하도록 공개 블록체인 전체에서 방송됩니다 (예측을 요청했던 사용자의 키를 사용하여 인증됨). 동시에 동일한 결과는 다른, 임시 키를 사용하여 암호화하고 일반에게 방송합니다. 이 키는 사전정의 시간 다음에 배포되며 결과를 이후 단계에서 심지어 원래 그것을 요구하지 않았던 사용자에게 의해 인증할 수 있게 합니다. 이

메커니즘은 그 같은 복수의 임시 키를 생성하는 방식으로 예측의 '신선도'에 따라 비용을 달리하여 조정할 수 있습니다.

프로세싱 데이터: 코드 실행은 블록체인 상의 실행(공개 예측, 공개 RFP 주문, 비공개 RFP 주문)과 AWS 상에서 또는 유사한 환경(예를 들어 GOLEM 프로젝트)에서 가동하는 독점 초탄성 산출 레이어 사용에 기초한 *Endor.coin* 기반시설 상의 실행으로 나뉘집니다. 프로세싱 데이터와 추출 행동 클러스터가 복잡하고 비싼 실행 환경을 요구하기 때문에 *Endor.coin* 프로토콜의 이후 출시품은 또한 프로토콜의 실행 층을 열 것입니다. 이로써, 새로운 유형의 이해관계자 즉, 집중적 실행 전문 이해관계자가 *Endor.coin* 생태계에 합류할 수 있게 됩니다. 분석할 데이터가 클수록 또한 클러스터 추출 알고리즘이 활발할수록(산출 면에서) 그것을 가동하는 데 필요한 *EDR* 토큰 가격이 높아질 것입니다.

결제: 사용자는 매 예측 요청 별로 예측 복잡성에 따라 *Endor.coin* 프로토콜에 의해 역동적으로 정의되는 가용 자원과 요청 시간 기준 수요, 유사한 질문을 묻는 사용자 수에 의거한 *EDR* 금액을 지불할 수 있습니다. 사전정의 쿼리는 상대적으로 변함 없는 가격으로 유지될 것이라 기대되는 한편 RFP (예측 요청)는 상대적으로 높은 비용 (정의 상 적은 수의 사용자로 시작)에서 시작하여 관련 사용자 커뮤니티가 형성됨에 따라 점차적으로 인하됩니다. 이 커뮤니티가 클 수록 각자가 결국 지불하는 금액은 적어져, 새 예측 쿼리 생성자에게 그들 네트워크 내 사람들 사이에 '말을 확산'시키는 데 대한 인센티브를 줍니다.

6.2 스마트 계약

Endor.coin 프로토콜은 최종 사용자에게 몇 가지 기본적인 기초 요소를 제공합니다.

- `GetPrediction(prediction_def)` – 단계 1 에서 실행
- `PutPredictionReq(prediction_def)` – 단계 2 에서 실행
- `PutData(data_def)` – 단계 2 에서 실행
- `RunCustomPrediction(data_alg_prediction_def,price)` – 단계 3 에서 실행

이러한 기초 요소로 고객은 *Endor.coin* 플랫폼에서 역동적 비용으로 사용할 수 있는 예측을 검색하고 나중에 데이터를 예측 목적에 사용되도록 업로드하고 판매할 수 있습니다. 기초 요소가 *Endor.coin* 프로토콜을 위한 기본설정 이용 사례를 커버하는 한편, 미래 출시품은 스마트 계약의 전개를 지원함으로써 `Get` and `Put` 위에 설계된 훨씬 더 복잡한 연산을 가능하게 합니다. 프로토콜이 비공개 데이터와 맞춤형 비공개 예측으로 진행되면서,

기본 스마트 계약 위 추가적인 보안 및 복잡성 레이어가 덮여 최초의 공개 예측 버전으로서 소개될 수 있을 것입니다. 스마트 계약을 통해 *Endor.coin* 사용자는 토큰을 소비하고, 예측을 요청하며 시장에서 데이터 검색에 사용됨은 물론 데이터 품질 증거를 입증하는 프로그램을 작성할 수 있습니다. 사용자는 거래를 장부에 전송함으로써 스마트 계약과 상호작용할 수 있을 것이며 이는 계약 내 기능수 통화를 트리거합니다. 스마트 계약 시스템은 확장되어 *Endor.coin* 의 특정 작업 (증거 검증)을 지원하며, 특히, 데이터 업로드 계약(이후 단계에서 공개 또는 비공개용으로 사용할)은 물론 더 포괄적인 스마트 계약을 지원합니다.

getPrediction(prediction_def): 사용자는 *Endor.coin* 네트워크에 저장된 예측을 *EDR* 토큰을 지불하고 검색할 수 있습니다. 고객은 입찰 주문을 Retrieval Predictions Market 주문서 (주문서를 네트워크에 전함으로써)를 제출함으로써 Get Protocol 를 시작합니다. 예측 제공자의 요청주문이 일치하면, 고객은 예측에 연결되는 비공개 임시 링크를 받습니다. 받을 때, 양 당사자는 거래 주문서에 서명하여 블록체인에 제출하여 교환이 성공적으로 이루어졌음을 확정합니다. 예측 소비자는 요청했던 예측에 첨부된 것과 동일한 가격을 지불합니다. 예측은 *Endor.coin* 웹포털에서 접근할 수 있게 됩니다.

putPredictionReq(prediction_def): 사용자에게 추가적인 예측 유형을 *Endor.coin* 플랫폼에 제출함으로써 제안할 수 있게 합니다. 인터페이스 역시 *Endor.coin* 웹포털에서 접근할 수 있을 것입니다. 제안된 예측은 *Endor.coin* 사용자 모두가 볼 수 있으며 그것에 관한 평점을 매길 수 있습니다. 대부분 높은 평점을 받은 예측은 계속 성장하는 예측 카탈로그에 추가됩니다. 예측 요청은 예측 발행인의 공개 주소로 식별될 것이며 일단 예측이 카탈로그에 추가되면, 예측 발행인은 다른 사용자가 그 예측을 사용할 때 마다 보상 받습니다.

putData(data_def): 데이터 제공자와 파트너를 위해 사용. 데이터 제공자는 그들의 데이터가 예측에 사용될 때 마다 *EDR* 토큰 으로 보상을 받고 가격은 *Endor.coin* 프로토콜 예측 알고리즘에 의해 자동으로 계산됩니다 . 함수의 이 변종은 *Endor.coin* 프로젝트의 두번째 단계에서 접근 가능합니다. ‘DIY’ 모드를 위한 API 인 프로토콜을 세 번째 출시하는 동안, 고객은 자신의 독점 데이터 스트림을 업로드할 수 있고 그것들에 ‘비공개’란 마크를 달 수 있으며, 플랫폼 상에서 액세스 할 수 있는 공개 데이터 스트림과 융합시킬 수 있습니다. 고객은 스토리지 마켓 주문 장부에 입찰주문을 제출함으로써 (주문을 블록체인에 제출함으로써) Put Protocol 을 게시하고, 요청주문이 데이터 검증자(예: 예측 엔진 제공자)에게 배정되길 기다립니다. 고객은 예측 엔진에 의한 행동 클러스터 추출에 자금을 제공해야 하지만, 최종 사용자가 사용할 때 마다 얻고 싶은 가격을 결정할 수 있어야 합니다.

runCustomPrediction(data_alg_prediction_def, price): 이 방법은 프로토콜의 세 번째 출시 시 실행될 것이며 *Endor.coin* 카탈로그에 들어 있지 않은 예측을 ‘DIY’ 모드로 생성하도록 지원합니다. 입력에는 원하는 행동에 대한 설명이 예를 제시하는 방식으로 들어

있거나 또는 공개 또는 독점 데이터를 지칭하는 로직에 대한 설명이 들어 있습니다. 요청은 어느 엔진이건 가능한 엔진이 취급할 수 있으므로 최고가는 정의되었습니다. 수수료는 예측에 기여하는 여러 당사자(예: 예측 엔진과 데이터 제공자)간에 조정됩니다.

6.3 출시 시 Endor 의 역할

이전 섹션에서 자세히 설명했듯이, *Endor.coin* 프로토콜은 다양한 양상을 지니는 당사자(예: 데이터 제공자, 분석 엔진 개발자, 등) 간 상호소통을 가능하게 합니다. 이런 의미에서, 프로토콜은 그 자체로는 예측의 출처가 아니고 차라리 '가장 자리'에서 가치가 창출되도록 허용하는 *이너블러* 또는 *언어*입니다. 다양한 가치 제공자 간 용이한 소통을 제공하는 TCP/IP 프로토콜과 유사하게, *Endor.coin* 프로토콜은 데이터 제공자에게는 그들의 상품을, 예측 엔진에게는 분석 서비스, 캐털리스트에게는 시너지 가치 제시품의 창출에 사용할 수 있는 방식을 제시할 수 있게 합니다.

이를 위해, *Endor.coin* 은 별도의 객체 하에서 프로토콜 구현에 중점을 두고 운영될 것이며, 주위 생태계의 최적 구성을 확보하고 성장 가속화를 최대화합니다.

하지만, 프로토콜의 생산 등급 워크 플로우가 신속하게 수립되도록 보장하고, 고객에게 가능한 한 일찍 (심지어 1 일차에) 다양하고 정확한 예측을 소비하도록 허용하기 위해, *Endor.coin* 프로젝트는 *Endor* 를 몇 개의 주요 역할을 하는 역할자로서 선택적으로 구성할 것입니다. 나중에, 새로운 기업이 *Endor.coin* 프로토콜을 채택하면서, *Endor* 의 역할은 감소할 것이라 기대되며, *Endor.coin* 생태계의 새로운 참가자가 도입할 혁신과 가치에 자리를 내줄 것입니다.

다음은 *Endor* 가 *Endor.coin* 출시 시 제공했던 주요 역할입니다. 이러한 활동 각각은 전용 자금지원을 요구할 것이며, 그것들은 *EDR* 토큰 교환으로 ICO 동안 받은 수익금에서 배정됩니다. 기업들은 현재 이 협력의 정확한 세부사항을 협의하고 있습니다.

공개 데이터 제공자: *Endor* 는 자체 Dev 팀으로 생산 등급 기반시설을 실행하여 다양한 데이터 스트림을 수확 및 클린하고 분석 준비가 데이터로서 *Endor.coin* 기반시설에 업로딩합니다. 여기에는 비트코인 블록체인과 ERC20 블록체인, 몇 가지 다른 독점 블록체인이 포함됩니다. 또한 이후 *Twitter*, *Reddit* 등과 같은 소셜 네트워크 피드로 지원될 것이며, *Endor* 는 이러한 서비스를 '비용 플러스' 기초로 제공하여 HR 경비를 충당합니다.

가공된 데이터 제공자: *Endor* 는 *Endor.coin* 프로토콜을 지원하고 거래 데이터 스트림에서 행동 클러스터를 추출하는 첫 번째 분석 엔진으로서의 역할을 합니다. *Endor* 는 이러한 서비스를 '비용 플러스' 기초로 제공하여 AWS 이러한 서비스를 '비용 플러스' 기초로 제공하여

HR 경비를 충당합니다.

수요 제공자: *Endor*는 기존 (및 새) 기업 고객들의 예측에 대한 갈망을 채널링하는 수요 제공자로서의 역할을 합니다. *Endor*의 고객 성공 직원은 기업 고객이 지불하는 미국 달러로 *EDR* 토큰을 구매하여 원하는 예측을 생성합니다.

애플리케이션 및 파트너십 개발 : *Endor.coin* 생태계를 더 확장하고 예측에 기초한 *Endor.coin* 을 더 빠르게 침투시키기 위해 , *Endor* 는 새로운 예측 기반 사업을 파트너십을 통해 또는 새로운 유형의 예측을 형성하여 캐털리스트로서의 역할을 하여 대중에게 배포함으로써 생성하기 위해 적극적으로 노력합니다. 이는 *Endor*의 사업 개발팀은 물론 많은, 높은 수준의 자문단에 의해 활용됩니다.

6.4 토큰 특권과 경제

이전 장에서 논의했듯, *Endor.coin* 에서 발행한 *EDR* 토큰은 *Endor.coin* 생태계가 제시하는 다양한 서비스를 위한 결제 메커니즘으로서 활용됩니다. 그러한 서비스 중 일부는 성장 촉진을 위해 프로젝트 초기 단계에는 무료로 제공될 수 있음에 유의하며, 다음은 *EDR* 토큰의 주요 용도입니다.

사전 정의 또는 DIY 예측 구매: *EDR* 토큰의 주요 용도는 예측 소비를 위한 결제 메커니즘이 될 것으로 기대됩니다. 끊임 없이 성장하는 카탈로그에서 사용 가능한 사전 정의 예측은 물론 (더 이후의 단계에서) 사용자가 전용 셀프 서비스 API 를 사용하여 주문하는 “DIY” 예측이 포함됩니다. 결제는 주로 클라우드 산출 자원 (예: *Amazon AWS*, *Golem* 등)의 10% ~ 25%를 감당할 것이며 예측 생성에 사용된 데이터의 제공자에게 안분 비례로 결제액으로서 배정됩니다. 예측 비용은 이를 생성하는 데 필요한 클라우드 자원의 비용에 연동됩니다.

예측 생성 과정의 주요 구성요소는 다른 예측들 사이에서 공유되기 때문에, 사용자 수 증가로 또한 갑작스런 수요 증가에 수반된 임시 스파이크로 점근적으로 감소하는 비용 함수가 만들어 질 수 있습니다.

예측 비용의 10% ~ 25% 사이에서 책정되는 데이터 제공자를 위한 결제액.

데이터 분석을 위한 클라우드 자원 (AWS 또는 유사) 클라우드 비용의 약 80% ~ 90%,

적극적인 사용자 수로 나눔 (즉, 전체 예측 비용에서 55% ~ 80%, 적극적인 플랫폼 사용자 수인 n 으로 나눔).

개인화한 예측을 위한 클라우드 자원 (AWS 또는 유사한) 클라우드 비용의 약 10% ~ 20% (즉, 전체 예측 비용에서 10% ~ 15% 사이).

그러므로 예측 비용을 위한 수량적 결제 모델은 다음과 같습니다.

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

$C_{PREDICTION}$ 은 Endor.coin 플랫폼 출시 시 단일 예측 비용을 나타내고 n 은 적극적인 플랫폼 사용자 수를 나타냅니다.

RFP (예측 요청) 제출: 카탈로그에 의해 아직 지원되지 않은 특정 예측에 관심이 있는 사용자는 RFP 메커니즘으로 새로운 예측 생성을 신청할 수 있습니다. 이 경우에, 사용자는 EDR 토큰을 지불하여 숙련된 *Endor.coin* 팀원에 의한 예측 생성 및 최적화에 필요한 시간 비용은 물론 그것에 요구되는 클라우드 자원을 보상합니다. 이런 의미에서, RFP 는 클라우드 비용과 숙련된 데이터 엔지니어의 평균 시장 급여의 결합에 연동됩니다.

예측을 위해 QOS 선호도 요청 : RFP 메커니즘을 통한 예측 구입 또는 새 예측 요청 시, 각 사용자는 예측으로의 프리미엄 액세스를 구매하여 비 프리미엄 구매자에게 발표되기 서너 시간 전에 예측을 받아보거나 또는 RFP 메커니즘의 경우, 사전 정의 기간 동안 독점적 액세스를 유지할 수 있습니다. 그 같은 서비스 품질 구성요소는 예측 비용에 사전 정의 인자를 곱한 비용으로 구매할 수 있습니다. 이런 의미에서, 이 프리미엄 서비스 역시 기본 예측비용의 배수가 되기 때문에 클라우드 자원 비용과 시장 급여에 연동됩니다.

플랫폼을 위한 새 데이터 스트림 제출 : *Endor.coin* 생태계를 위한 새로운 데이터 스트림에 기여하고자 하는 데이터 제공자는 데이터의 초기 분석과 *Endor.coin* 프로토콜에 맞는 적응에 요구되는 클라우드 자원 비용을 지불해야 합니다. 이는 전반적인 예측 품질을 향상시키기 위해 높은 수준의 데이터 소스 통합 시 긍정적인 인센티브를 제공하기 위해 요구됩니다. 이 초기 과정 수수료로서, 데이터 기여자에게 예측을 위한 행동 클러스터 출처로서 *Endor.coin* 프로토콜에 의해 지원되는 그들의 데이터를 소유할 권리를 주고 이후, EDR 토큰이 예측 생성에 사용될 때 그것에 대해 보상 받을 권리를 부여합니다.

6.5 수익금 사용

EDR 토큰 구매에서 나온 수익금은 *Endor.coin* 프로토콜 소프트웨어 기반시설 구현과 마케팅 및 전략적 협력을 통한 초기 성장과 채택을 지원하고 요구되는 클라우드 자원을 위한 자금 지원, 요구되는 법적, 행정적 경비를 지출하는데 사용됩니다.

수익금의 주요 부분은 첨단 클라우드 및 블록체인 엔지니어는 물론 데이터 과학과 머신러닝, 사회 물리학의 세계적 전문가로 구성된 *Endor.coin* R&D 팀에게 배정됩니다.

Endor.coin 프로젝트와 팀의 중요한 역할은 가능한 한 많은 선두 기업이 *Endor.coin* 프로토콜을 빠르게 채택하도록 보장하는 것입니다. 이를 달성하는 방법 중 하나는

학계와의 공동 연구 활동을 구성하여, 선도적 연구기관에 *Endor.coin* 프로토콜과 기반시설에 액세스할 수 있게 하는 것입니다. 수익금의 최대 10%까지는 세계 선두 연구소와 함께 그 같은 활동을 지원하는 데 배정됩니다.

수익금의 최대 30%까지는 예측 엔진 또는 ETL 커넥터 등과 같은 관련 독점 기술의 구매 가능성에 배당되는데, 그러한 기술이 *Endor.coin* 프로토콜과 그 주변 생태계의 적절한 독자적인 힘을 촉진 및 가속화하는 데 필요하기 때문입니다.

제 7 장

기술적 장점 및 차별화

7.1 MIT 기원 과학적 혁명

MIT의 novel *사회 물리학* 기술 [1]에 의한, *Endor.coin*은 세계에서 가장 앞선 행동 예측 기술을 사용합니다. 2010년대 초 MIT 미디어 랩에서 시작한 이 과학적 돌파구는 행동 데이터 분석 분야에서 대변혁을 일으켰으며 권위 있는 *DARPA Network Challenge* [21] 수상을 포함하여 기술적 업적의 최전선에 있으며, 소비 투자자 커뮤니티의 투자 수익을 높이고 [9, 13]과 효율적인 미상의 사이버 공격의 존재를 성공적으로 예측하고 있습니다 [35]. 이 기술은 이러한 주제에 관한 수 백개의 과학 논문과 수 십개의 특허, 서너 편의 책을 출간한 산학 전문가팀에 의해 개발되었습니다.

그리고 지금, 이 혁명은 일반 대중에게 다가가고 있으며 전문가와 일반인 모두 똑같이 지금까지는 대형 소매 기업과 은행, 첨단기술 대기업의 단독 특권이었던 능력에 접근할 수 있도록 블록체인 기술을 사용하고 있습니다. 사회 물리학에 관한 세부적인 논의는 섹션 3.1 과 부록 A 에서 나옵니다.

7.2 실제 제품. 입증된 기술

앞서 섹션 3.2 에서 개략적으로 소개했듯이, *Endor*는 MIT 스피인 오프 [14]로 대형 은행 및 소매업체에 그들의 예측 병목현상을 해소해 줄 SaaS 솔루션을 제공하도록 설계된 하나의 상품으로써 사회 물리학 구현을 실시할 책임을 지고 있습니다. 회사는 주요 투자자의 후원을 받았으며 [15], 또한 *코카콜라* [16]와 *마스터카드* [17], *월마트*등과 같은 포춘 500대 기업과 함께 일하며 다양한 이용 사례에서 정확한 예측을 자동으로 생성하는 능력을 성공적으로 입증하고 있습니다.

블록체인과 암호화 화폐 분석을 위한 기술의 가치는 [27]와 다른 곳에서 입증했습니다. *Endor*의 제품은 *Money-2020*와 *Finnovate 2017* 등에서 같은 선도적 장소에서 선보였습니다 [18]. *Endor*는

가트너 선정 주목할 만한 기업 [19]이며 세계 경제 포럼 이 “기술 파이오니어”로서 인정했습니다 [20].

7.3 이용성, 사용자에게 주는 가치, 토큰 보유자를 위한 가치

Endor.coin 의 주요 독창성은 *EDR* 토큰은 토큰 출시 1 일에 사용할 수 있고 토큰 보유자에게 사전 정의 예측으로 완전한 접근을 제공합니다. 또한 베타 사용자 그룹은 토큰 출시 후 바로 선정될 것이며 사전 정의 예측 외에 여러 예측을 요청할 기회를 갖게 됩니다.

또한, 미래 예측에 저렴한 비용으로 액세스해야 한다고 믿는 토큰 소유자는 *EDR* 토큰을 일찍 구매한 예측 사용자는 금전적 보상을 받게 될 것이라 기대됨에 따라 *EDR* 토큰을 구매하고 보유할 인센티브를 갖게 됩니다.

제 8 장 팀

8.1 주요 팀원

Dr. Yaniv Altshuler

Dr. Altshuler 은 군중 행동을 모델링하는 새로운 과학인 “사회 물리학”을 발전시킨 Prof. Pentland 와 함께 하는 MIT Media Lab 의 연구 제휴사인 *Endor* 의 CEO 입니다. MIT 는 이 기술로 권위 있는 *DARPA Challenge* [21]에서 우승을 하고 싱가포르 정부의 교통 체증 예측 능력 개선에 도움을 주었으며, 수 천명의 금융 투자자로 구성된 커뮤니티에게 그들의 재정적 이익을 개선할 수 있도록 도왔습니다 [13]. *Endor* 에서, 이 기술은 많은 다양한 이용 사례에서 군중의 행동을 정확하게 예측하는 데 사용했으며, 대형 은행과 소매업체에서 효율적으로 활용되었습니다 [14].

MIT 에서의 복무 및 *Endor* 설립 전, Altshuler 는 *IBM* 의 연구원으로 슈퍼컴퓨터 성능 강화에 사용된 새로운 최적화 기법을 개발했습니다. 2011 년부터 블록체인 연구에 적극적이었던, Dr. Altshuler 는 60 편 이상의 과학 논문을 발표했고 15 개의 특허를 신청했습니다. Altshuler 의 논문은 *파이낸셜 타임즈* [10]와 *하버드 비즈니스 리뷰* [11] 등과 같은 인기 있는 학술서에 실렸습니다. 최근 출판된 저서로는 ‘*소셜 네트워크에서 보안과 프라이버시*’ [26]과 ‘*검색에서 스웱과 정보망*’ [25]이 있습니다.

Prof. Alex “Sandy” Pentland (과학자문 위원회 의장)

MIT Media Lab 기업가 프로그램은 물론 MIT 컨넥션 과학 및 인간 역동성 실험실 책임자. Prof. Pentland 는 세계에서 가장 많이 인용되는 과학자 중 한 명이며[23] 최근에는 Google 창립자이자 미국 최고기술 책임자와 함께 포비스에 의해 “세계 7 명의 가장 강력한 데이터 과학자” [24]로서 선정되었습니다. *하버드 비즈니스 리뷰*로부터 *McKinsey 상* [22]과 *DARPA 로부터 인터넷 제 40 주년* [21], 그의 프라이버시 관련 업적으로 인한 *Brandeis 상* 등 수 많은 상을 수상했습니다.

그는 *Google* 과 *AT&T*, *Nissan*, 유엔 사무총장, 수 십개 기업을 공동 창립한 일련의 기업가는 물론 데이터 투명성 실험실과 Harvard-ODI-MIT *데이터팝 연맹*, 데이터 주도 설계 연구소 등과 같은 사회적 기업을 위한 고문단의 창립 위원입니다. 그는 *미국*

공학학회 회원이며 세계 경제 포럼 내 리더입니다. 그의 최근의 저서는 ‘사회 물리학과
‘정직한 신호입니다.

Stav Grinshpon

Mr. Grinshpon 은 SAP 와 AT&T 등과 같은 기업에서 18 년의 제품 및 관리 경력을 갖고 있는 베테랑 기술 산업 전문가입니다. Grinshpon 은 세계적인 사이버 방어 전문가로 이스라엘 8200 기술 유닛에서 기술 리더로서 8 년간 역임했습니다. Grinshpon 은

David Shrier

David Shrier 는 세계적으로 인정 받는 금융혁신 권위자이며 옥스포드 대학의 온라인 프로그램인 Oxford 핀테크와 Oxford 블록체인 전략은 물론 MIT 의 미래 커머스를 이끌고 있습니다. 그는 핀테크와 블록체인, 사이버 보안에 관한 사이버 보안을 위한 새로운 솔루션과 금융 기술 프론티어, *트러스트::데이터* [27, 29, 32]를 포함한 수 많은 책을 출판했습니다.

Shrier 는 금융 서비스 산업을 행동 분석으로 변모시키고 있는 MIT 연구에 뿌리를 둔 머신 러닝 회사인 Distilled Analytics 의 CEO 이며 AI 주도 협업 기술 플랫폼 제공자인 Riff Learning 의 회장입니다. David 는 옥스포드 대학 전술한 경영 대학원의 Associate Fellow 이고 MIT Media Lab 의 강사입니다. 그는 또한 두바이 정부에게는 블록체인과 디지털 신원, 중동 신용 유동성 제공자인 Millennium Advisors 에게는 기술 트렌드, 블록체인 기반 디지털 상품 거래소인 Cleer.digital 에게는 전략에 관해 자문하고 있습니다.

David 는 완전 무료의 인가 받은 온라인 금융 엔지니어링 석사 학위를 제공하는 프로그램인 WorldQuant 대학의 자문단에 있습니다. 그는 이전에 유럽 위원회에 디지털 기술에 초점을 둔 혁신 상용화에 관한 자문을 제공했습니다. 그는 현재 미국 보안 산업의 자기 규제 조직인 FINRA 를 위한 FinTech 산업위원회의 위원이며, 금융

서비스에 영향을 주는 새로운 개발에 관한 컨설팅을 하고 있습니다.

Prof. Mihaela Ulieru

컴퓨터 조작 지능 전문가이며 *세계 경제 포럼의 블록체인 챔피언입니다*[36]. 분산지능시스템에서 그녀의 연구는 제조 및 물류, 국토안보를 대변혁한 후 기관 기술로서 블록체인에서의 거버넌스를 위한 강한 토대를 생성했습니다. 캘리포니아 버클리 대학 졸업생인 Prof. Ulieru 는 세계 경제 포럼의 글로벌 아젠다 위원회와 싱가포르 과학공학 연구위원회, 캐나다 과학 기술 혁신 위원회의 위원입니다.

Dr. Goren Gordon

Dr. Gordon 는 그가 산출적 호기심의 첨단 모델을 개발했던 텔 아비브 대학의 Curiosity Lab 의 소장입니다. 딥 러닝과 뉴럴네트워크 최적화를 선도하는 전문가인 Gordon 는 양자물리학 박사학위와 신경 생물학 박사학위 소지자입니다. Dr. Gordon 은 호기심 있는 로봇이 호기심 있는 아동과 상호작용하는 방식을 연구했던 MIT Media Lab 에서 Prof. Cynthia Breazeal 와 함께 일하며 그의 또 다른 의학 학위를 이 경험을 활용하고 있습니다. Gordon 은 또한 과학 교육에 관심이 있어 이 때문에 양자 물리학을 놀이를 통해 아동들에게 가르치는 “양자 컴퓨터 게임”을 개발하였습니다. Gordon 은 또한 양자물리학과 두뇌, 학제간 사고 주제에서 인기 있는 강연자입니다.

Dr. Arie Matsliah

Dr. Matsliah 는 그래프 분석 이론의 세계적 전문가이고 16 년 간 *Google, IBM, Intel, Lyft* 등과 같은 대기업에서 근무했으며 성공적인 Menlo-Park 기반 스타트업인 *TripActions* 에서 수석 건축가입니다. Dr. Matsliah 는 또한 알고리즘과 복잡성, 양자 계산에서 근본적인 연구를 중심으로 40 편 이상의 과학 논문을 발표했습니다.

Shahar Somin-Gavrielov

Ms. Somin-Gavrielov 은 통계적 학습 이론 전문가이며 Hebrew 대학에서 석사학위 (유등으로)를 받았습니다. Somin-Gavrielov 은 노련한 연구가이며 데이터 과학의 심도 있는 이론적 구현을 핸즈 온 산업 경험에 결합시킵니다. 과거, 이스라엘 8200 인텔리전스 유닛에서 훈장을 받은 분석가였습니다.

Edo Eisenberg

금융 리스크 관리 전문가이며 이전에는 *Morgan Stanley* 및 *Barclays* 에서 소매 신용 포트폴리오를 관리했습니다. 추가로 9 년은 *NICE* 에서 경력을 쌓았으며 미국 상위 10 대 은행 중 7 개 은행에서 전개된 사기 방지 솔루션을 설계 및 구현했습니다. Duke MBA 프로그램 졸업생.

Lior Regev

Mr. Regev 는 기술에 대한 열정이 많은 노련한 소프트웨어 엔지니어입니다. Regev 는 전 지능기술 리더였으며 분산 시스템과 클라우드 아키텍처, SaaS 제품에서 방대한 경험을 갖고 있습니다.

Liat Yitshaki

Ms. Yitshaki 는 공법 소송 전문가입니다. 그녀는 런던 대학에서 법률 및 윤리학 MA 와 우등 졸업으로 LLB 를 받았습니다. 과거에 Yitshaki 는 *McKinsey & Co.*에서 근무했으며 또한 선임 법률고문으로서 영국정부에게 자문을 제공했습니다.

8.2 어드바이저

Prof. Alexander Lipton

Professor Lipton 은 *Bank of America* 의 글로벌 정량 솔루션의 관리 이사와 *Merrill Lynch* 에서 글로벌 정량 및 신용 분석그룹의 관리 이사를 역임했습니다. 이 전에 Prof. Lipton 은 *Citadel* 에서 정량연구 책임자였으며 *Credit Suisse* 에서 주식연계 파생상품 책임자 였습니다. Lipton 은 현재 *EPFL* 에서 금융 엔지니어링 교수이며 또한 *MIT* 커넥션 과학 센터의 동료이며, 최근 출간된 “*기술이 망가진 금융 시스템을 어떻게 고칠 것인가*” 라는 제목의 *Scientific American* 논문의 저자입니다[37].

Ron Gross

Ron Gross 는 이스라엘 비트코인 협회의 창립자이자 이사입니다. 그는 2011 년 이래로 비트코인과 블록체인 세계에서 적극적이었으며 *Mastercoin* (현재 *Omni*) 의 전무 – 세계 최초의 ICO 였습니다. Gross 는 Google 의 전 직원이며 또한 *Commerce Science* 에서 최고 건축가로서 복무하고 있습니다.

Dr. Nuria Oliver

Dr. Oliver 는 MIT 졸업생이고 *Vodafone* 에서 데이터 과학 연구 이사이며, *Data-Pop Alliance* 에서 선임 데이터 과학자입니다. Dr. Oliver 는 *Telefonica R&D* 의 과학 이사였으며 *Microsoft* 의 연구원이었습니다. Oliver 는 빅 데이터 분석을 민간 분야 기업이 공개의 혜택을 위해 공유하는 협업의 형식으로 사용하는 데이터 박애에 관한 세계적 전문가입니다. Oliver 는 90 편 이상의 과학 논문과 저서 챕터의 저자이며 40 개 이상의 특허를 출원했습니다.

Dr. Daniel Tunkelang

데이터 과학의 세계적 전문가, Tunkelang 은 과거 Google 은 물론 IBM 에서 일했으며 현재 *Apple*, *Salesforce*, *Etsy*, *Yelp*, *Pinterest* 과 같은 *테크 자이언트에게 자문을 제공하고 있습니다.* 이전에, Tunkelang 은 *LinkedIn* 에서 검색 엔지니어링 이사는 물론 이곳의 *거주 데이터 과학자였고 Endeca* (Oracle 에 의해 인수)에서 최고

과학자였습니다. Tunkelang 은 MIT 에서 석사학위 카네기 멜론 대학에서 박사학위를 받았습니다.

Guy Zyskind

비밀 스마트 계약을 위한 *Enigma* 프로토콜과 누구에게나 정교한 톨과 데이터를 사용하여 암호 해지 기금을 시작하도록 허용하는 플랫폼인 캐탈리스트 뒤에 있는 회사인 *Enigma* 의 창립자이자 CEO. *Enigma* 전에 Guy 는 MIT 졸업생으로 블록체인 기술을 연구하고 가르쳤습니다. Guy 는 서너 편의 학위 논문을 집필했으며 최신의 대부분은 *Enigma* 백서와 “분산 프라이버시: 개인 데이터를 보호하기 위한 블록체인 사용”를 포함하여, 프라이버시와 블록체인에 관한 논문입니다. Guy 는 MIT 에서 M.S., 텔 아비브 대학, 전자공학 및 컴퓨터학과에서 B.S.를 취득했습니다.

Prof. Michael Bronstein

연쇄 창업가이자 선두 연구원인 Prof. Bronstein 은 하버드 대학 연구원이고 루가노 대학의 정보학 교수입니다. Bronstein 은 선두 저널과 컨퍼런스에 기고된 100 편 이상의 간행물과 20 개 이상의 특허, 연구논문 “*Numerical Geometry of Non-Rigid Shapes*”을 집필했고 4 권의 책을 편집했습니다. Bronstein 은 유럽연구회(ERC) 보조금을 3 회 받은 극히 적은 연구원 중 한 명이며 또한 *Google Faculty 연구* 상과 하버드 대학에서 *Radcliffe fellowship* , *Rudolf Diesel 산업 펠로우십* 받았습니다. 그는 *세계 경제 포럼*에 의해 40 세 이하의 세계 40 명의 선두 연구원 중 한 명으로 선정되었습니다.

Prof. Bronstein 은 산업 애플리케이션에 적극 개입하고 있습니다. 그는 *Novafora* (Turner Broadcast 에게 라이선스 제공)을 공동 창업했으며 또한 그곳에서 기술 부사장을 역임했고, 3D 센싱 기술 회사인 *Invision* (*Intel* 이 인수, 이곳에서 Bronstein 은 현재 지각 컴퓨팅을 위한 연구 과학자로서 근무하고 있음)의 공동 창업자이고 주요 개발자 중 한 명. Bronstein 은 *Videocites* 의 공동 창업자이자 기술 자문입니다.

Dr. Wei Pan

Dr. Pan 은 MIT 졸업생이며 Google 에서 근무했으며 빅 데이터 분석과 머신 러닝, 복합 시스템에 관한 세계적 전문가입니다. 금융 시장의 역동성과 거시 경제학을 사회 시스템에 대한 이해와 모델링, 사람과 군중에 대한 빅 데이터 측정치를 통해 이해하려는 "리얼리티 헤징"으로 알려진 분석 접근방식의 발명자입니다. 현재, Dr. Pan 은 뉴욕 기반 스타트업인 *Thasos 그룹*의 공동 창업자이자 최고 과학자입니다. Pan 은 이전에 *Fidelity Investments* 에서 최고 경제학자 아래에서 일했으며, 이곳에서 그는 시스템 연구와 플래시 크래시에 중점을 두었습니다.

Igor Gonta

MIT 의 엔지니어링 및 컴퓨터학 졸업생인 Mr. Gonta 는 *BlackRock* 에서 상품 및 리스크 관리를 맡습니다. 과거 Gonta 는 소셜 미디어 대화 분석에 기초한 실시간 주식 세그먼트 생성 엔진인 *Market Prophit* 의 CEO 를 역임했습니다. Gonta 는 이 회사의 창립자이고 이후 그는 이 회사를 더 큰 헤지 펀드에 매각했습니다. 초기에 Gonta 는 *Barclays* 에서 상품 판매 관리 이사는 물론 *Goldman Sachs* 에서 상품 판매 부사장을 역임했습니다.

Thomas Hardjono

Hardjono 는 분산화된 아이덴티티와 블록체인, 스마트 계약 전문 보안 전문가입니다. Hardjono 는 *VeriSign* 과 MIT Kerberos Consortium 의 전무였으며 컴퓨터 보안과 암호화에 관한 5 편의 저서를 발간했습니다 [38–42].

부록 A – 사회 물리학 설명됨

1. 사회 물리학의 작동방식은?

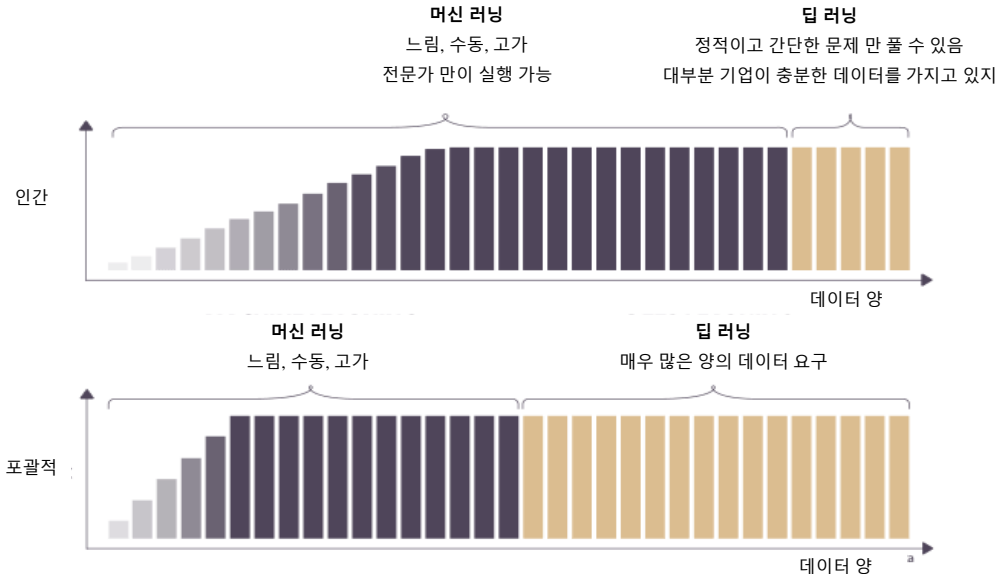
설명: 이 문서는 사회 물리학과 그 뒤에 있는 높은 수준의 원리는 물론 Endor 에 의한 기술적 구현에 관한 종합적인 관점을 제공하려 합니다. 하지만, Endor 에 의해 사용된 사회 물리학 법칙의 특정 수학 공식은 물론 그러한 법칙 위반을 검출하기 위해 사용된 메커니즘의 특정 구현에 관한 일부 기술적 세부사항은 IP 고려사항으로 인해 의도적으로 생략했습니다.

정보 시대에서, 기업은 그들의 사업 운영에 관한 모든 유형의 데이터를 수 많은 출처에서 수집합니다. 데이터는 이미지와 동영상, 문자, 트위터, 거래, 이용 로그를 포괄합니다. 하지만, 대부분의 데이터는 단일 기본 출처에서 기원합니다: 사람.

그래서, 예를 들어, 트위터와 블로그 게시물은 인간이 인간을 위해 작성하고 구매 거래와 전화 통화 정보는 인간의 물건과 다른 사람에 대한 욕망을 담고 있으며, 사용 및 앱 로그는 사람이 컴퓨터 및 모바일 기기와의 상호작용 방식을 말해 줍니다.

인간 행동에서 나온 데이터는 “영망”이며, 역동적이고 복잡하며 매우 다양합니다. 그 같은 디지털 데이터 채널에 기록된 대로 인간의 행동은 시간이 지나면서 극적으로 변하고 근본적으로 복잡한 소셜 네트워크에 영향을 받고 다중모달 데이터 스트림 내에서 전달됩니다. 이러한 특징은 사업 운영을 개선하기 위해 고객 행동을 분석 및 이해, 예측하고자 하는 기업에게는 상당히 큰 도전을 부여합니다.

최근에는, 데이터 과학자는 “영향력 있는” 통계적 방법과 머신 러닝 알고리즘을 사용하여 이 복잡성을 극복하려 하기 시작했습니다. 새로운 “딥 러닝” 기법을 포함한 이러한 강력한 툴은 행동 패턴을 분류하고 특이성을 검출하며 미래 트렌드를 예측하기 위해 데이터를 수집하고 그것들의 속성을 분석합니다. 하지만, 역사적으로 이미지 프로세싱과 문자 인식 등과 같은 “정적인 문제”를 위해 개발된 그 같은 툴은 인간 행동 데이터를 쉽게 극복할 수 없으며, 역동적이고 복잡하며 다양한 데이터 스트림은 배우기가 극히 어렵고 종종 거의 불가능합니다.

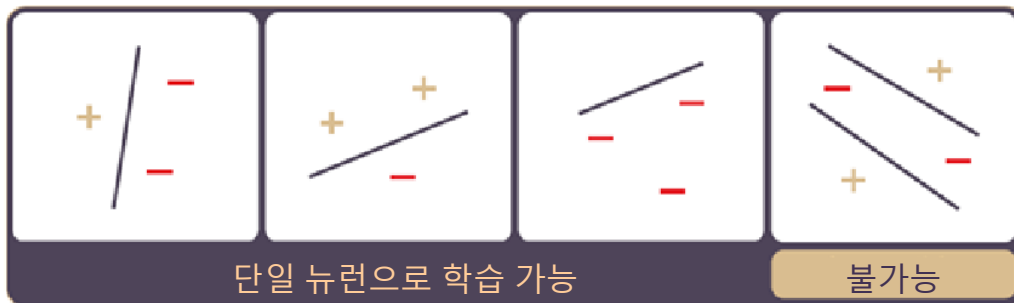


Endor 의 사회 물리학 엔진은 완전히 다른 방식으로 작용합니다. 입력 데이터 자체로부터 패턴을 도출하는 대신에, 모든 인간 행동의 데이터가 공통적인 "사회 행동 법칙" 내에 포함된다는 보장 발견을 토대로 합니다. 이 법칙은 대규모 사람이 동일한 장소에서 운영될 때마다 나타나는 수학적 관계입니다. 이러한 법칙은 데이터 유형 또는 그것을 생성한 사용자의 인구통계, 데이터 크기와 상관 없이 시간에 따른 균중 행동의 다양한 통계적 성질의 진화 방식을 지배합니다. Endor 는 이러한 법칙을 데이터 분석 엔진에 통합하고 있으며, 이 엔진은 입력으로서 제공될 원 데이터(예: 전화 통화, 택시 탑승, 금융 투자, 등)에 포함된 모든 이들의 근본적인 사회적 속성을 효율적으로 추출합니다.

2.1 왜 사회 물리학이 필요한가?

추상적으로, 학습 문제 또는 객체 분류 또는 미래 이벤트를 위한 예측 생성 능력은 분석할 데이터와 그것을 분석하는 알고리즘을 요구합니다. 요구되는 데이터 수량은 몇 가지 요소에 의해 좌우됩니다.

1. **문제의 내부 복잡성** – 문제는 여러 모양과 크기로 되어 있으며 어떤 것은 의심할 여지 없이 다른 것에 비해 더 어렵습니다. 문제의 “난이도” 또는 “복잡성”은 문제를 성공적으로 학습하기 위해 학습 알고리즘에 반드시 있어야 하는 최소한의 “강도”를 지칭하며, 학습 알고리즘이 충분히 강하지 않으면, 문제의 사례를 바르게 학습할 수 없게 됩니다. 예를 들어, “퍼셉트론” (즉, 단일 뉴런으로 구성된 가장 간단한 신경망)은 결코 “XOR” 함수 (즉, 불 방식의 “배타적 논리합” 함수)를 배울 수 없음을 쉽게 알 수 있습니다. 이유는 퍼셉트론이 작용하는 방식 안에 있으며 (입력 공간을 선형으로 나누는 것으로 상상할 수 있음) XOR 함수는 이 방식으로 나타내기에는 “너무 복잡합니다.”



XOR 함수의 예는 우측 차트에 있습니다(출처: 위키피디아)

근본적인 모델의 문제에서 학습 복잡성은 종종 문제의 Vapnik–Chervonenkis 차원 (또는 VC 차원)으로서 지칭됩니다. 문제의 학습 복잡성이 높을 수록, 성공적인 학습을 가능하게 하기 위한 예측 알고리즘 필요성이 더 강해지고 적절히 모델링을 위해 그 같은 알고리즘이 요구하는 데이터가 더 많아 집니다. 예를 들어, 선형 복잡성을 갖는 함수 $y = f(x)$ (랭크 1의 다항식을 사용하여 더 잘 모델링할 수 있다는 의미)는 정확한 모델링을 위해 더 높은 정도의 다항식을 요구하는 함수 보다 더 적은 샘플링 포인트를 요구합니다.

알고리즘의 학습 효율성

많은 학습 알고리즘이 있으며, 각각은 모델 매개변수를 적절하게 파악하기 위해 다른 양의 교육 데이터와 전문가 도메인 지식을 요구합니다. 예를 들어, 단순 회귀분석은 많은 양의 데이터와 많은 문제와 관련된 특징을 요구하고, 딥 러닝은 방대한 양의 데이터를 요구하는 한편 자동으로 도메인 특징을 학습할 수 있습니다. 그러므로 우리의 논의를 위해, 학습 과정의 전반적 효율성을 수량화할 수 있습니다. 즉, 얼마나 많은 데이터가 근본적인 모델 학습을 위해 필요하며, 알고리즘이 얼마나 많이 스스로 적절한 표상을 학습해야 하는지 파악할 수 있습니다.

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

M = 모델의 비트는 데이터를 형성하거나 또는 모델로 만드는 데 요구되는 비트 수로 예를 들어

모델의 완벽한 설명에 필요한 매개변수 수입니다. 이러한 면에서, **M** 은 콜모고로프의 문제의 복잡성 - 데이터 복잡성의 알려진 이론적 측정치로 주어진 데이터를 생성할 수 있는 가장 짧은 튜링 기계를 지칭-과 유사합니다. 모델은 문제로 결정하며 변경할 수 없습니다. 즉, 간단한 문제이면 모델에는 몇 개의 비트만 있습니다.

R_A(M) = 모델-특정 표상의 비트는 알고리즘 **A** 가 근본적인 모델 **M** 을 나타내기 위해 학습한 비트 수입니다. 이 수는 알고리즘의 자동 특징 검출을 나타내고 전문가가 수작업으로 알고리즘에 프로그래밍해 놓아야 하는 도메인 특정 지식의 수와 반비례합니다.

D_A(M) = 데이터 비트, 알고리즘 **A** 이 모델 **M** 를 학습하기 위해 필요한 비트 수.

η_A = 알고리즘 **A** 의 학습 효율성, 즉, 알고리즘의 자동 대표 학습과 학습의 질을 유지하기 위해 요구되는 데이터 양 간의 비율입니다. 고 효율성 알고리즘은 적절한 표상을 적은 양의 데이터로 학습할 수 있고 저 효율성 알고리즘은 수동으로 공들여 만든 특징과 그것들을 조율하기 위한 방대한 양의 데이터를 요구합니다.

예를 들어, 문제 등급 **M** 에 주어졌을 때 일부 알고리즘은 학습의 질을 보존하기 위해 다른 것에 비해 더 많은 데이터를 요구합니다.

- 로지스틱 회귀 알고리즘은 일반적으로 전문가에 의해 특징을 수동으로 코딩하고 그것을 특정 현안 문제에 정교하게 조율하기 위해 많은 양의 데이터를 필요로 합니다.

$$R_M(A) \ll 1$$

$$D_M(A) \gg 1$$

$$\eta_A \ll 1$$

- **원 샷 학습 알고리즘**은 또한 전문가 도메인 특징을 요구하지만 근본적인 모델을 정교하게 조율하고 예측을 생성하기 위해 단지 적은 수의 예만 사용합니다.

$$R_M(A) \ll 1$$

$$D_M(A) \ll 1$$

$$\eta_A \sim 1$$

- **딥 러닝 알고리즘**은 가장 정보적인 특징을 자동으로 학습할 수 있지만 learn 그렇게 하기 위해 방대한 양의 데이터를 요구합니다.

$$R_M(A) \gg 1$$

$$D_M(A) \gg 1$$

$$\eta_A \sim 1$$

- **Endor의 알고리즘**은 **사회 물리학**을 사용하여 자동으로 관련 행동 특징을 단지 적은 데이터 샘플에서 추출합니다.

$$R_M(A) \gg 1$$

$$D_M(A) \ll 1$$

$$\eta_A \gg 1$$

2. **문제 변화율** – 정확한 예측을 생성하기 위해 요구되는 데이터 양에 영향을 주는 또 다른 인자는 문제의 근본적인 모델의 변화율입니다. 일부 문제는 정적이어서 그것의 근본적인 매개변수는 변하지 않거나 또는 거의 드물게 변합니다. 예를 들어, 이미지에서 얼굴은 몇 개월 동안 변하지 않습니다. 얼굴은 얼굴입니다. 다른 한편, 유급 서비스 밖의 과당매매로 이끄는 근본적인 행동 패턴은 시간에 따라 사회적 변화를 통해 수 개월에 걸쳐 점진적으로 변하거나 또는 경쟁자의 성공적인 마케팅 캠페인에 대한 대응으로 며칠 사이에 급격하게 변할 수 있습니다. 우리는 이 역동성을 다음과 같이 정량화하였습니다.

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

τ 는 문제의 지속성 또는 강도이며 근본적인 모델의 변화율을 나타내며 여기서, ∂T 는 모델이 ∂M 비트에 의해 변화하는 지속기간을 나타냅니다.

예를 들어, $\tau = 1 \text{ 일} / 10 \text{ 비트}$ 는 훨씬 더 느린 모델 변화를 나타내는 $\tau = 1 \text{ 개월} / 10 \text{ 비트}$ 와 비교하여 모델이 하루 동안 극적으로 변화한다 (즉, 빠른 변화율)는 의미입니다. 효과적으로, 역동적 모델은 매 기간마다 다른 모델을 제시하고 모델의 재교육 또는 재학습을 요구합니다.

예측 분석의 운영상의 구현 : 변경 모델 추적

그러므로 주어진 문제에 예측 분석의 실행 가능한 구현에서 주요 도전은 모델화된 객체의 행동 변화를 극복하기에 충분한 정보를 얻는 것입니다. 1 회 누적할 수 있는 정보 비트의 양은 It 로 나타냅니다.

매우 큰 소셜 네트워크 또는 검색 엔진 (예: Google 또는 Facebook)의 운영자는 종종 방대한 양의 정보를 상대적으로 짧은 시간 내에 모을 수 있습니다. 하지만 고객의 행동 예측에 관심이 있는 대부분의 기업은 그렇게 할 수 없습니다.

또한 심지어 매일 방대한 양의 정보를 획득하는 대형 기업도 (a) 너무 복잡하거나 또는 (b) 너무 빠르게 변하거나 또는 물론 (c) 앞의 두 가지가 결합된 문제를 정확히 모델링하는 게 어렵다는 점을 발견합니다.

그러므로 우리는 운영상의 예측 분석 솔루션을 구현할 수 있는 회사의 능력을 결정하는 하나의 단순식을 가리킬 수 있습니다. 기업이 사용하는 학습 알고리즘과 그들이 계속해서 획득하는 새로운 데이터를 사용하여 이 원리를 만족시킬 수 있는 기업은 성공적으로 이 목표를 달성하는 운영 과정을 구성할 수 있지만 그렇게 하지 못하는 기업은 성공에 도달하지 못할 것입니다.

기본 연산 학습 원리:

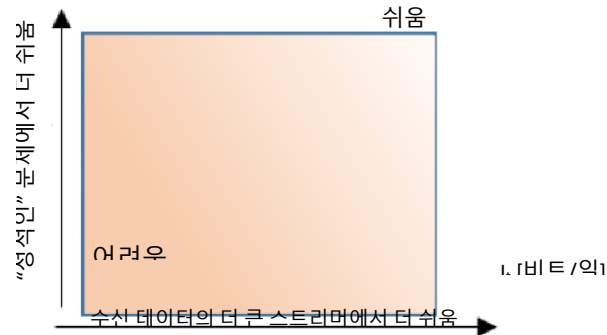
$$It \times \tau \times \eta A > 1$$

이 원리가 실질적으로 함축하는 바는 운영상 기능하는 예측 시스템을 구성하려는 시도에서 실패한 기업은 다음 중 하나를 실행해야 한다는 의미입니다.

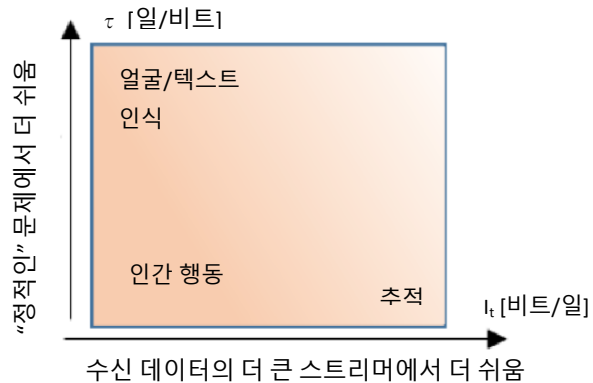
1. 데이터 수집 대역폭을 개선하여 하루 당 더 많은 양을 관련 데이터를 획득하거나 또는
2. 더 정적인 문제에 중점을 두거나 또는

3. 더 효율적인 학습 알고리즘에 의지해야 합니다.

이 단순 관계식은 다음과 같은 차트에서 묘사되듯, 문제의 복잡성이 증가하고 그것의 지속성이 감소함에 따라 하루에 필요한 정보가 더 많아짐을 의미합니다.



여기서는 몇 가지 공통된 예측 문제가 다음으로 분석할 때 어떻게 보이는나 입니다.

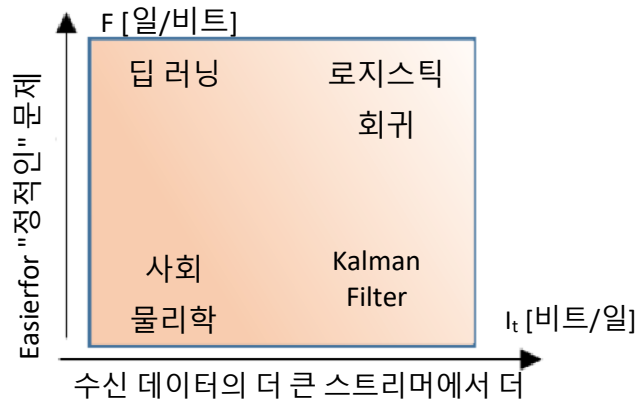


- **얼굴 인식:** 일반적으로 말해 얼굴은 변하지 않으므로 이 문제는 매우 낮은 변화율, 즉 $\tau \gg 1$ 로 특징지어 집니다. 그러므로, 심지어 ηA 가 상대적으로 적은 비효율적인 학습 알고리즘에서도, 이 값은 여전히 매우 적어 그래서 처음 문제를 배우기 위해서는 많은 정보가 필요하지만 너무 많이 변하지 않으므로 재교육 시에는 그렇게 많은 정보가 필요하지 않습니다.
- **동작 모바일 타겟 추적:** 빠른 타겟 이동 (또는 궤도 내 빠른 변화 또는 검출을 회피하기 위해 네비게이션 알고리즘에 의해 채택된 역동적 국부 변화) 으로 즉각적이지 않은 과거 데이터를 쓸모 없게 만들기 때문에 이동 타겟 추적은 $\tau \ll 1$ 으로 특징지어 집니다. 하지만, 그같은 문제 또한 일반적으로 큰 흐름의 수신 데이터로 특징지어지고, 타겟 위치의 역동적 성격은 입력 데이터 스트림에 의해

학습 알고리즘으로 완전히 전달된다는 의미를 갖는 $k \sim 1/\tau \gg 1$ 을 제공합니다. 이는 문제를 효율적으로 풀기 위해 상대적으로 더 간단한 알고리즘을 $\eta_A \sim 1$ 으로 가능하게 합니다.

- 인간 행동:** 인간 행동은 극도로 역동적으로 상대적으로 빠르게 변하는 구성요소를 담고 있다는 의미인 $\tau \ll 1$ 을 갖습니다. 더구나, 인간 행동을 예측하길 원하는 기업이 고객에 관한 추가적인 정보를 얻을 수 (또한 얻고) 있더라도, 그것은 항상 실제 행동 변화의 부분적 편류입니다. 은유적으로, 감지기에서 인지된 정보가 매우 조악한 렌즈를 통과하여 단지 몇 가지 측면만 잡아 낼 수 있는 플라톤의 동굴의 비유와 유사합니다. 형식주의에서, 이는 $I_t \ll 1/\tau$ 으로 해석되고 이것은 이후 $I_t \geq 1$ 을 의미합니다. 이것은 인간 행동을 효율적으로 예측하기 위해서는 $\eta_A \gg 1$ 의 매우 효율적인 알고리즘을 반드시 활용해야 한다는 의미입니다.

마찬가지로, 우리는 다른 솔루션 기법이 어떻게 각 문제에서 가장 잘 이용되고 다음 관계식으로 모델링 되는지 볼 수 있습니다.



이 차트는 이전에 언급한 기본 연산 학습 원리를 설명합니다. $I_t \times \tau \times \eta_A > 1$. 문제가 더욱 정적이 될수록, 그리고 우리가 가진 데이터가 더욱 많아질수록 차트의 우측 상단 구석으로 더 가깝게 됩니다(그리고 우리의 예측이 더 정확해질 가능성이 더욱 높아집니다). 또한 우리가 우측 상단에 가까이 있을수록 정확한 예측을 생성하기 위해 우리는 덜 효율적인 알고리즘을 활용해야 합니다.

교육을 위해 방대한 정보량을 요구하는 기법은 역동적 문제에서는 잘 작용하지 않지만 딥 러닝은 근본적인 구조가 이미지 프로세싱과 몸짓 인식과 같은 빠르게 변하지 않는 문제에서 가장 잘 작용합니다. 반대로, 예를 들어 Kalman Filter 와 같이 정보를 빠르게 처리할 수 있는 간단한 알고리즘 은 역동적 문제를 다룰 수 있지만 성공적으로 예측하기 위해서는 높은 정보 처리율을 요구합니다. 로지스틱 회귀와 같은 전통적 머신

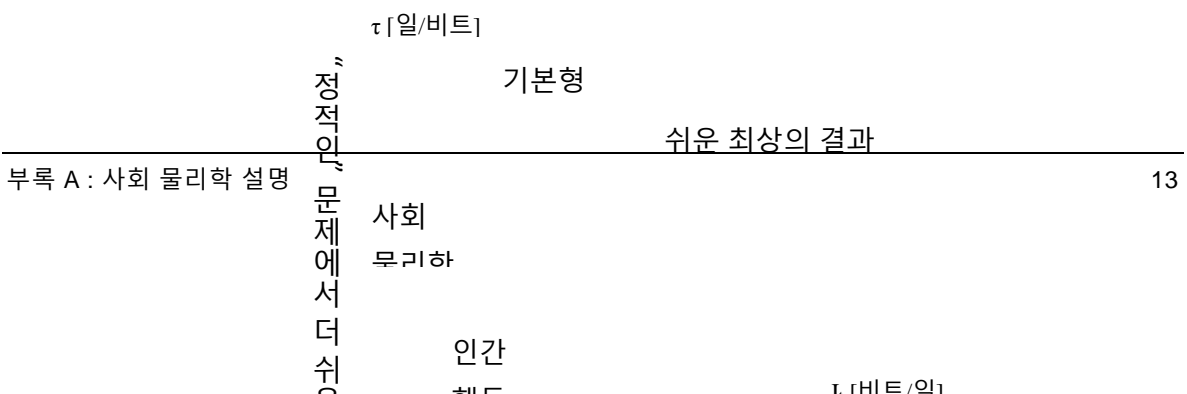
러닝 접근방식은 문제가 또한 상당히 정적인 경우 상대적으로 많은 양으로 들어오는 교육 데이터가 있는 시나리오에서 효율적일 수 있습니다.

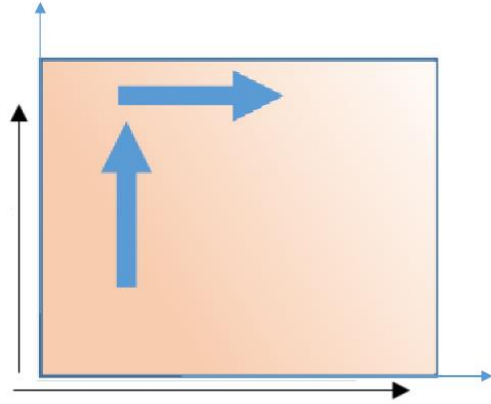
결코 끝나지 않을 추적에 대한 솔루션으로서 사회 물리학

Endor 는 두 가지 직교 관점에서 모델을 변하시키기 위해 r 끊임 없이 많은 양의 데이터를 획득해야 하는 필요성에 반기를 듭니다.

- **역동적 문제를 정적인 모델로 전환:** 사회 물리학의 법칙은 불변하고 데이터 유형 및 기원에 구속 받지 않습니다. 그러므로, 가공 전 데이터를 사회 물리학 규범적 표상 공간에 투영하기 위해 그것들을 사용하는 것은 원래의 문제를 새로운 등급의 문제의 사례로 변환하며, 그 문제의 근본적인 모델은 정적입니다 (사회 물리학 법칙의 고유 정적인 성격 덕분에). 이는 원래 문제의 실제 $r \ll 1$ 를 극히 큰 $r_{sp} \gg 1$ 가 되도록 변환합니다.
- **작은 데이터에서 “빅 데이터” 생성 :** 위에서 언급했듯이, 고객의 행동 예측이라는 도전에 당면할 때, 대부분의 기업은 주요 장애물을 맞이 합니다. 즉, 그들이 소유한 관련 데이터의 양이 종종 불충분합니다. 이는 특별히 역동적 문제 (마케팅 이용 사례에서 종종) 또는 새로운 구성요소의 도입과 관련된 문제 (신제품에 대한 반응 예측 또는 새로운 종류의 입력 데이터 활용 등)에서 사실입니다. Endor 은 고객에게서 받은 모든 데이터 스트림을 유형 및 크기, 출처와 상관없이 사회 물리학 기본형으로 전환합니다. 이 기본형을 사용함으로써, Endor 는 모든 고객에게서 온 데이터와 그들의 모든 쿼리를 일관되게 높은 속도로 성장하는 (즉, $k \gg 1$) 하나의 매우 큰 데이터베이스로 통합할 수 있습니다. 그 다음 이 데이터베이스는 사회 물리학 기본형 (특정 쿼리 또는 고객이 아닌)의 예 분석에 관해 훈련된 단일 불편의 딥 러닝 네트워크를 훈련하는 데 사용합니다. 이에, 비록 각 고객은 매우 제한적인 데이터를 제공하더라도, Endor 는 방대한 양의 (기본형) 데이터를 모을 수 있습니다.

Endor 의 엔진은 인간 행동 예측에서 가장 어려운 문제를 느리게 변하는 문제 (사회 물리학을 통해)와 빅 데이터 (기본형으로의 변환을 통해)로 변환하여 그 결과 “쉽고 효율적인” 문제가 되고, 그 다음 딥 러닝 도구를 사용하여 풀 수 있습니다. 이는 다음의 차트에서 나타납니다:





2.2 사회 물리학: 개요

이전 섹션에서는 정적인 수학 불변성에 구속 받지 않는 의미를 사용하여, 인간 행동을 예측하는 데 있어서의 내재적 도전을 극복하는 데 사회 물리학을 어떻게 사용하는지를 대략적으로 설명했습니다. 이것이 어떻게 실행 가능한 지 더 잘 이해하기 위해서 물리 법칙 – 예를 들어 뉴턴의 제 2 법칙 또는 운동량 보존 법칙을 생각해 보십시오. 객체는 외부 힘이 그것에 작용하지 않은 한 초기 경로를 유지합니다. 그것은 숨겨진 힘의 존재를 추론하기 위해, 데이터를 학습하거나, 통계적 속성을 이해하거나, 유사한 방식으로 작동하는 많은 시스템을 시험할 필요가 없음을 명심하십시오. 물리적 법칙이 주어진 이래로, 그 법칙의 위반은 비정상적이고 즉시 검출 및 일부 "보이지 않는 손"의 결과로서 및 해석될 수 있습니다. 갑작스럽게 방향을 변경하는 객체를 발견하면, 어느 한 힘이 그것에 작용했다고 즉시 추론할 수 있습니다. 유사한 변경이 나타내면, 동일한 힘이 그들 모두에게 가해졌을 가능성이 높습니다. 이 간단한 구현은 물리 법칙에 대한 이해를 통해서만 가능합니다.

사회 물리학이 물리 법칙에 비해 훨씬 덜 절대적이고 엄격성이 떨어지지만 개념은 유사합니다. 무언가가 사회 물리학 법칙을 위반하면, 그것은 즉각적으로 “흥미있는” 것으로서 자격을 갖게 되고, 실제 세상에서 일부 가치 있는 자산 또는 속성의 데이터 발현이 됩니다. 이것은 학습 또는 벤치마킹, 기준선, 다른 데이터 과학, 머신 러닝 도구를 요구하지 않습니다. 사회 물리학 법칙 위반은 그것이 생성한 데이터의 유형과 상관 없이 극히 빠르고 매우 탄탄한 방식으로 검출할 수 있습니다.

2.3 ENDOR: 사회 물리학으로 추진되는

2.1.1. 규범적 표상으로서의 데이터 변환

이전에 논의한 “행동 클러스터” (즉, 검출된 사회 물리학 불변성 위반)를 추출하고 그것들을 하나의 “지식 영역”으로 종합함으로써, 원 입력 데이터(어떤 모양 또는 형태이든, 인간에게서 기원을 둔)는 기본형으로 변환됩니다. 이 형태는 사회 물리학 법칙을 “함께” 위반한 사람들 다시 말해, “너무 높은 동시적 행동 변화”를 나타내는 사람들, 일정한 불변성에 반하는 사람들의 클러스팅을 나타냅니다. 특정 시간에 유사한 방식으로 방향을

바꿀 수 있는 한편 그러한 변화를 일으키는 힘은 보이지 않는 물리적 객체와 유사하며, 우리는 높은 가능성으로 그러한 객체 모두가 단일 출처에 의해 영향을 받았을 것이라 추론할 수 있습니다. 마찬가지로, 행동 클러스터 형태로 된 Endor 의 규범적 데이터 표상에는 동일한 숨겨진 “사회적 힘”에 의해 영향을 받고 따라서 공통된 “실제 세상 특성”을 공유하고 있을 가능성이 높은 사람들의 여러 그룹이 담겨져 있습니다. 새로운 가공 전 데이터가 나올 때 마다, 고객은 그러한 데이터를 Endor 에 보내 (일반적으로 매일 또는 주 단위로), 자동으로 추출된 행동 클러스터가 추가됩니다.

이러한 표상이 가져다 주는 혜택은 3 가지로 나뉩니다.

- **자동:** 인간 행동에서 기원한 데이터가 신고 또는 분석할 필요 없는(예: 전화 통화 기록, 신용카드 구매, 택시 탑승 또는 고객이 소유하는 다른 유형의 독점 데이터) 입력 데이터의 유형과 상관 없이 자동으로 행동 클러스터 콜렉션으로 변환되기 때문에, 사회 물리학의 수학 불변성 사용. 입력 데이터 내 노이즈와 격차로 극히 높은 탄성력과 결합하여, 이는 알려 지지 않은 유형의 지저분한 가공 전 데이터를 균일한 행동 클러스터로 변환하는 과정이 최초로 완전 자동화 되었음을 의미합니다 (추가적 정보는 “노이즈에 강함” 섹션 참조).
- **균일:** 도메인 또는 인구통계, 의미 측면의 데이터를 벗겨냄으로써, 행동 클러스터를 담고 있는 나머지 정보는 아래 “문의” 단계를 위한 이상적인 형태를 띵니다. 사실, 이 균일한 대표를 사용하여, Endor 는 아무 것도 없을 때 “빅 데이터”를 생성하여 이 과정의 문의 단계에서 데이터의 원래 소유자에게는 불가능했던 딥 러닝 기법을 사용할 수 있게 합니다! 이는 Endor 딥 러닝 엔진이 많은 유형의 데이터와 많은 고객에게서 유래하고, 모두 단일 형태로 전환된 행동 클러스터로 접근할 수 있기 때문에 가능합니다.
- **새로 부상하는 트렌드:** 간단히 말해, 사회 물리학 불변성은 인간 군중의 특정 통계적 성질이 시간이 따라 진화되는 방식을 설명합니다. 이 시간중심 측면으로 Endor 는 새로 부상하는 행동 변화, 최근에 되어서야 발생한 역동성을 쉽게 검출할 수 있으며, 대부분의 경우, 전통적 방법을 사용해서는 충분한 통계적 중요성으로 검출할 수 있는 충분히 관찰 가능한 데이터를 생성하기에는 충분한 시간을 갖지 못했습니다. 사회 물리학 사용을 통한 이러한 추가적인 신호를 검출할 수 있다는 사실 외, 그것들은 최근의 트렌드에 관한 정보를 담고 있으므로 그것들은 일반적으로 다양한 비즈니스 질문에서 높은 중요성을 갖는 바로 그 신호입니다.

“옛날 학교” 머신 러닝은 사전 정의 특징으로 작용했고 상대적으로 적은 양의 포괄적 데이터에서 관련 정보를 추출할 수 있었습니다. 하지만, 많은 결과는 선택된 특징에 의해 좌우됐습니다. 딥 러닝은 가장 관련성이 큰 특징을 스스로 식별하지만 엄청 난 양의 데이터를 요구합니다. 각 데이터 유형과 질문은 관련 특징을 다시 찾기를 요구하고 따라서

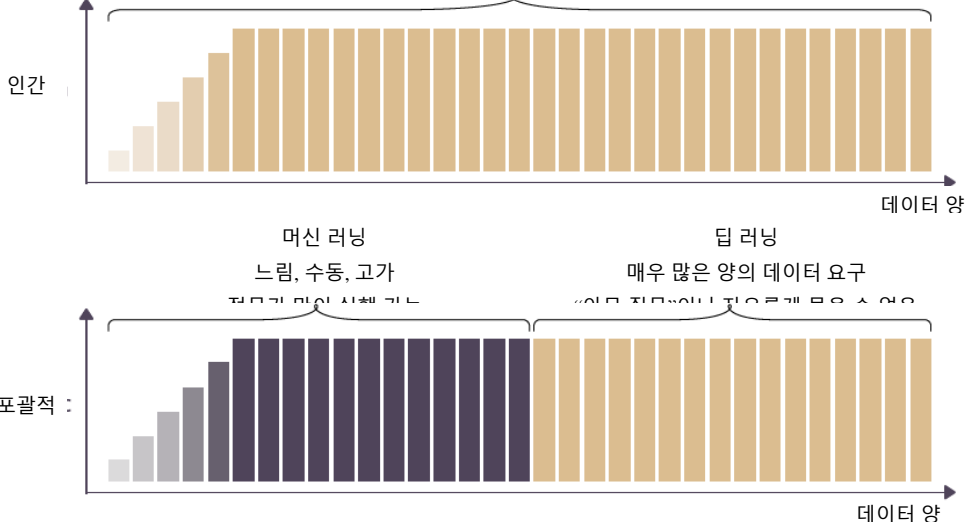
더 많은 데이터를 요구합니다.

사회 물리학은 모든 유형의 인간 행동 데이터를 행동에 기초한 인간 클러스터의 기본형으로 변환합니다. 이는 적은 양과 많은 양의 데이터 모두에서 작용합니다. 또한, 사회 물리학의 기본형 덕분에, Endor 는 데이터 크기와 상관 없이, 모든 데이터 유형과 모든 질문를 섭취하고 하나의 거대한 인간-행동 데이터 세트를 생성하고 그 다음 예측 질문에 답하기 위해 딥러닝의 힘을 사용합니다.

ENDOR 사회 물리학 → 딥 러닝

완전 자동으로 이루어지며 어떤 질문에도 답할 수 있습니다. 비전문가 역시 사용할 수 있습니다

사회 물리학은 모든 유형의 인간 행동 데이터를 인간 행동 클러스터 기본형으로 변환합니다. 사회 물리학의 기본형 표상을 사용하여, Endor는 데이터 크기와 상관 없이, 모든 데이터 유형과 모든 질문을 처리하여 데이터 크기가 커도 데이터 배치를 생성하고 그 다음 예측 질문에 대한 기계 학습 모델을 생성합니다.



사회 물리학 (하부 차트)를 통해 가능한 현재 예측 분석 패러다임 (상부 차트) 대 새 패러다임.

2.3.2 규범적 표상 문의 (“지식 영역”)

지식 영역에는 모든 사용자에게 관련 검출된 전반적 정보가 들어 있기 때문에, 어떤 질문에 대한 예측도 용이하고 빠른 간단한 데이터 검색으로 제공할 수 있습니다. 지식 영역의 초기 생성은 일반적으로 10 억개의 기록으로 구성된 전형적인 데이터의 경우 1~4 시간 정도 걸립니다. 이 과정이 완료된 후, 동일 지식 영역을 사용하여 수 분 내 수십개의 질문에 답합니다. 사전 도메인 지식 또는 관련 특징 추출이 필요하지 않습니다.

쿼리는 크기와 상관 없이 “예” (긍정적으로 라벨 표시된 ID)를 제공함으로써 제출합니다. Endor의 엔진은 그 다음, 지식 영역을 사용하여 답을 생성합니다: 관련 쿼리에 “행동면에서 유사한 가능성이 가장 큰 사용자들로 구성된 순위 표시 목록. “교육” / “학습” 단계가 없으며 “결과를 해석”할 필요가 없습니다. 각 특정 질문에 대해서 이 자동 과정과 같은 과정은 종래의 방법을 사용하여 데이터 과학자의 수 주에서 수 개월이 걸리는 노력을 요구할 수 있습니다.

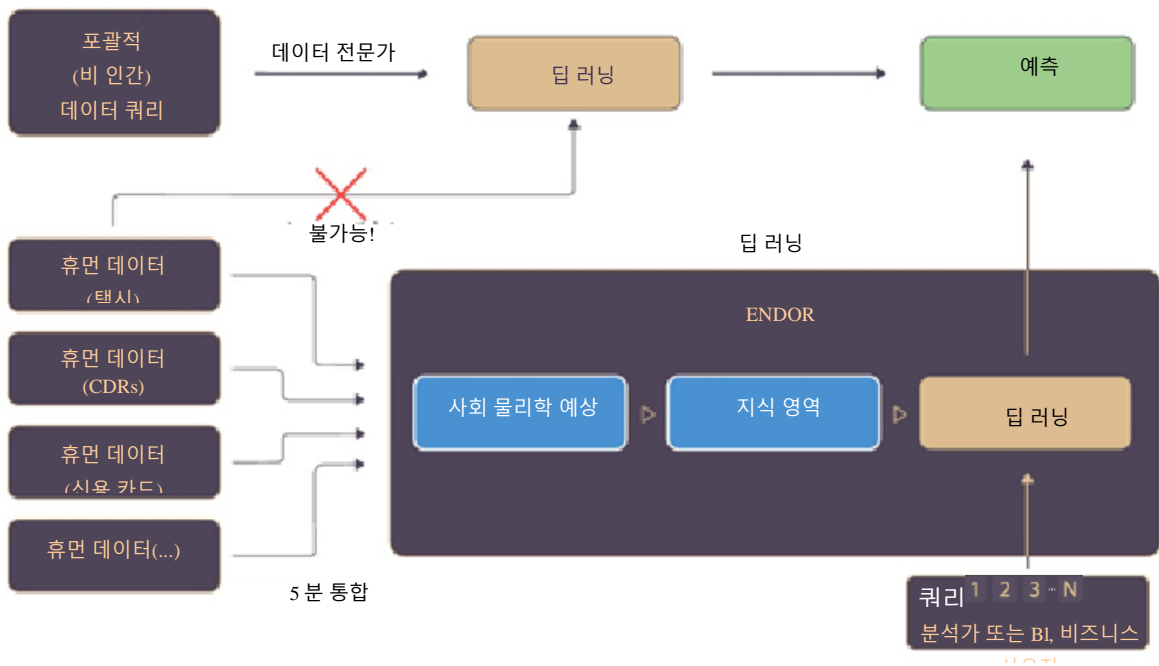
예를 들어, 가까운 미래에 이탈될 가능성이 가장 큰 사용자에게 관한 쿼리는 이전 이탈되었던 사용자의 신원이 포함된 목록으로 설명됩니다. 대신, 어느 고객이 프리미엄 계정으로 전환할 가능성이 있는지 식별하려고 했던 쿼리는 최근 이 프리미엄 서비스로 전환한 고객들 목록으로 설명될 수 있을 것입니다. 하지만 두 쿼리 모두 동일한 지식 영역을 사용할 것이며 어떤 종류의 재교육도 요구하지 않습니다.

대부분의 그같은 행동 예측 질문은 딥 러닝 (충분한 데이터의 부족과 근본적인 모델의 빈번한 변경)을 이용하여 풀기가 매우 어려울 수 있는 반면, Endor는 모든 지식 영역과 쿼리 수집을 사용함으로써 이 이슈를 회피한다는 점을 유의하십시오. 따라서, Endor는 실제 예측 제공을 담당하는 내부 딥 러닝 구성요소를 위해 필요한 빅 데이터를 “생성”합니다. 이러한 예측은 비록 매우 적은 양의 데이터에 기초하더라도 특정 지식 영역에 기초합니다.

다음 차트는 전통적 딥 러닝 (위 차트)의 플로우를 나타내며 Endor의 플로우와 비교됩니다 (아래 차트). 현재 딥 러닝 도구는 그 같은 도구를 사용한 경험이 있는 엔지니어가 사용하도록 설계되었기 때문에 데이터 전문가가 적절한 교육을 실시한 경우, 비 인간 행동 데이터 (예: 수 백만 장의 이미지)의 경우, 딥 러닝은 고 품질 예측을 생성할 수 있을 것입니다. 다른 것에 관한 휴먼 데이터(예: 택시 데이터)에는 딥 러닝 모델링으로 쉽게 해결할 수 없을 정도로 자주 변하는 근본적인 모델이 있습니다 (이 이슈에 관한 심도 있는

논의는 섹션 2.1 참조).

하지만, 행동 클러스터 한 세트의 형식으로 변환되었을 때, 택시 데이터는 그 의미를 빼앗기며, 전화 통화 기록 또는 신용 카드 구매, 다른 유형의 휴먼 데이터에서 얻은 행동 클러스터와 표상 면에서 사실상 동일해집니다. 또한 이러한 각 유형의 데이터에 기여하는 많은 여러 고객이 있기 때문에 (즉, 많은 e-커머스 플랫폼, 각각은 자신의 구매 및 웹 활동 데이터를 업로딩함), 딥 러닝 모델 교육이 이제 가능해졌습니다. 모델에게 원데이터 또는 문제에 관해서가 아니라 다수의 행동 클러스터 즉 균일한 표상을 띄는 대형의 데이터 컬렉션에 관해 교육하기 때문에 이것이 가능하며, 또한 정적인 근본적인 모델로서 특징지어집니다 (즉, 매우 역동적인 원 데이터의 행동 역동성과 비교되는, 수학 불변량이며 그러므로 정적인 사회 물리학 법칙).



2.4 데이터 보안 및 익명성

2.4.1 Endor 저장 데이터

끝없는 다양한 행동 예측 질문을 자동으로 푸는 능력 외는, Endor의 솔루션은 높은 레벨의 데이터 보안과 요구되는 데이터 필드를 익명화하는 능력을 제공합니다. 이전 섹션에서 자세히 살펴 보았듯이, 예측은 지식 영역 - 의미 없는 행동 클러스터 콜렉션, 많은 수의 사용자 그룹 (각각은 하나의 공통된 사회적 또는 행동 특징을 가짐)을 사용하여 이루어집니다. 민감한 또는 개인 정보가 만약 그 같은 정보가 원래의 원 데이터에 존재하면, 이 단계에서 시스템이 사용하는 데이터의 일부가 더 이상 아님을 쉽게 알 수 있습니다. 시스템에서 사용 가능한 유일한 정보는 다음의 예에서 보듯 사용자의 ID입니다.

Behavioral_cluster₁ = (ID₁, ID₁₇, ID₂₃, ...)
Behavioral_cluster₂ = (ID₁₄₂, ID₄₂₈₇, ID₉₇₁₁, ...)
...
Behavioral_cluster_{748,329} = (ID₅, ID₃₇, ID₂₁₈, ...)

심지어 이 정보는 원 데이터 차원에서 고객에 의해, 데이터 탑재 시 (아래 참조) ID를 해싱함으로써 쉽게 숨길 수 있음에 유의해야 합니다.

2.4.2 Endor에 데이터 온보딩

이전 섹션에서 제시했듯이, 행동 클러스터를 원 입력 데이터에서 추출하는 데 사용한 방법은 사회 물리학 불변성을 위반하는 데이터 패턴을 나타내는 사용자 그룹 검출에 의존합니다. 사용자 활동의 동기적 성격을 나타내는 일부 통계적 성질의 역동성을 추적함으로써 할 수 있습니다. 이러한 구현을 위해, 전술한 바는 데이터 내 들어 있는 실제 값을 요구하지 않고, 대신 완전히 해시된 교체에 의해 이루어 질 수 있습니다. 이를 통해 고객은 Endor에게 완전히 해시된 데이터 세트를 제공하는 한편 여전히 그것의 우수한 예측 능력으로 혜택을 받을 수 있습니다. 또한 Endor는 구속 받지 않는 100% 의미이기 때문에 데이터-필드의 이름 또한 해시할 수 있습니다.

금융 기록을 위한 그 같은 해싱의 예는 아래와 같습니다.

해싱 전 헤더:				
계정 번호	지점	성별	거래 유형	목적지 계정
해싱 전 데이터 기록:				
183972	291/30	남성	이체	382732
183972	291/30	남성	잔고 조회	해당사항 없음
382732	291/30	여성	이체	439001
...				
해싱 후 헤더:				
필드 1	필드 2	필드 3	필드 4	필드 5
해싱 후 데이터 기록:				
AjF32sdx	Q2KPbv3A	Wsqp289X	q8Vb3MAs	Je2qx92n
AjF32sdx	Q2KPbv3A	Wsqp289X	q8Vb3MAs	x3PNm78A
Je2qx92n	Q2KPbv3A	m28SbA12	q8Vb3MAs	yL19B4GQ
...				

2.2 사회 물리학: 수학적 설명

2.5.1 프레임워크

먼저 Endor 의 엔진의 기본 원리를 그것의 포괄적 수학적 형식주의로 소개합니다. 그 다음에는 가능한 구연의 두 가지 예가 나옵니다. (1) 컴퓨터 비전 중심과 (2) 소셜 그래프 기반. 첫 번째 예는 센서 기반 데이터 (즉, 인간과 관련 없는)으로 사용했을 때의 Endor 의 문제점을 보여주고 두 번째는 사회 물리학의 개념과 인간 행동 예측에서의 이점을 보여줍니다.

비고: 이 논의 전체에서, 우리는 사회 물리학의 원리와 Endor 엔진에서 사용되는 방식을 위해 많은 수학적 설명을 제공합니다. 하지만, 일부 사회 물리학 법칙에 관한 수학적 세부사항은 IP 고려사항으로 인해 이 논의에서 생략합니다.

$d(x, t)$ 를 일시적 데이터 스트림이라 하고 여기서 x 는 단일 데이터 포인트를 나타냅니다.

$L(\cdot)$ 은 원 데이터 $d(x, t)$ 를 법칙 대표성으로 전환하는 법칙 연산자입니다.

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

법칙 그 자체는 법칙 연산자를 연역적 상수 \mathbf{C} (역법칙 등과 같은 숫자 또는 분포 등급이 될 수 있습니다)와 동일시하는 식으로서 생성됩니다. 이 \mathbf{C} 는 법칙으로 나타낸 불변량을 나타냅니다. .

설명 목적 상, 시간에 지남에 따른 흰색 노이즈 신호 출력의 변화가 매우 적은 임계치로 제한된다는 이해를 나타내는 가설적 법칙을 상상할 수 있습니다. 이 법칙 위반은 예를 들어, 갑작스러운 강한 출력 스파이크 신호 형태를 띌 수 있습니다. 신호를 여과하고 그 같은 변칙 출력을 검출할 수 있는 메커니즘을 이 법칙 위반을 찾아내기 위해 수립할 수 있습니다. 스파이크와 같이 다른 신호에 대해서는 “위반”으로 분류할 수 없기 때문에, 자연스럽게 그 같은 메커니즘의 사용은 문제가 되는 법칙 (즉, 흰색 노이즈의 기원)에 의해 지배되는 것으로 알려진 신호의 경우에만 수공이 갈 것입니다.



또 다른 예로 방사 및 흡수 재료의 변칙적인 "결집"을 검출하기 위해 제작된 "X 선 기계"가 있습니다. 이 예에서 X 선 기기 (와 그것을 작동하고 결과 이미지를 복호화하는 기술자)는 2 차원 정보 스트림을 분석하고, 필름에 투사되는 일관성 있는 X 선 빔은 동일하게 흡수되는 이미지를 생성해야 한다 (다소, 필름의 품질과 빔의 일관성에 좌우됨)는 사실을 주장하는 불변성 위반 위치를 파악하기 위한 "위반 검출자" 역할을 합니다.

이 불변성, 또는 법칙은 알려진 물리적 사실이며, 일관성 있는 X 선 빔을 고품질 필름으로 인공적으로 유도하는 한편, 이 물질 내 X 선 흡수 구성요소의 존재를 나타낼 "위반"을 찾고자 하는 열망으로 제 3 의 물질을 거쳐 지나감으로써 의료 분야에서 사용됩니다. 예를 들어 뼈는 비균일한 방사선 흡수로 보여지는 "데이터" (즉, 노출된 필름)에서 검출되고 그 자체로 전술한 법칙의 위반인 이 경우에서 "실제 세상의 현상"을 나타내는 물질입니다.



위의 원리를 구현하기 위한 구체적인 공식 수학적 예는 섹션 2.5.2 와 2.5.3.에서 나옵니다.

사회 물리학의 경우, 그 같은 위반은 입력 데이터가 인간 활동에서 나왔으므로 사회 물리학 법칙을 충실히 지킨다는 가정 하에서 분석 상 통계적으로 개연성이 낮은 것으로 알려진 입력 데이터에서 일부 역동성을 생성하는 사람들의 그룹으로 구성될 수 있습니다.

2.5.2 데이터-서브셋은 법칙 위반임을 검증

Endor 의 엔진은 서너 개의 사회 물리학 법칙 (이의 구체적 세부사항은 IP 고려사항으로 인해 설명하지 않음)의 법칙 연산자와 상수를 구현합니다. 이 섹션에서는 법칙 위반 검증에 관해 설명합니다. 많은 다른 산출 문제에서와 같이, 특정 신호가 법칙 위반을 구성함을 검증하는 것은 그 같은 간섭 검출과 근본적으로 다르다는 것을 염두해 두는 것이 중요합니다. 이 섹션에서는 전자 만이 논의되었으며 즉, "잠재적 간섭"이 정말로 법칙 위반인지 아니지 여부의 검증이 주어졌습니다. 효율적으로 대규모의 높은 차원의 ²¹ 데이터 소스를 스캔하고 그 같은 즉각적인 검증을 수행할 수 있는 Endor 의 검색 알고리즘의

부록 A: 사회 물리학 설명

세부사항은 여기서는 논의하지 않을 것입니다.

우리가 검증하는 법칙의 수학적 공식화는 언제나 다음과 같은 형식으로 주어집니다.

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

명시적 법칙 공식화가 주어질 때, 그것과의 국부적 편차는 다음과 같이 ξ 로 나타낸 편차를 측정함으로써 검증할 수 있습니다.

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

여기서 Δx 는 X 의 부공간을 나타내고, Δt 는 임시 창을 나타냅니다. 따라서 이 편차는 X 의 각 부공간 별로 또한 시간 기간 별로 계산할 수 있고, 주어진 기간 동안 그 부공간이 관련 법칙을 얼마나 많이 위반했는지 측정치를 생성합니다.

이 측정치를 사전 정의 임계치 $\xi_{\text{임계치}}$ 와 비교함으로써, 법칙을 위반한 하부공간을 검출할 수 있습니다.

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

위반 임계치 $\xi_{\text{임계치}}$ 는 그 임계치보다 훨씬 더 법칙을 회피할 수 있는 신호 의 동시적 부상이 거의 일어나지 않게 선택합니다. 이를 통해, 특정 데이터 서브셋이 법칙 위반인지 높은 충분한 통계적 중요성으로, 데이터의 의미 자체에 대한 사전 지식 없이 자동으로 검증할 수 있습니다.

신호가 시간과 공간 모두에서 변하기 때문에, 다른 임시 창 은 법칙 위반으로서 검출된 다른 부공간을 생성할 수 있습니다. Endor 는 (데이터가 아닌 법칙에서 도출한) 사전 정의된 고정된 임시 창 세트를 사용합니다. $\Delta t = 1$ 일, 7-일, 30 일, 90 일.

데이터가 매우 역동적이면, 임시 창이 길수록 편차 그룹을 생성할 가능성이 없고 데이터가 정적이면, 임시 창이 짧을 수록 편차 그룹을 생성할 가능성이 없습니다. 그럼에도 불구하고, 정의 상 노이즈는 일관된 법칙 위반을 생성할 수 없기 때문에 어느 창도 “정크 그룹”을 생성하지 않습니다(또는 임계치 $\xi_{\text{임계치}}$ 를 선택하는 방식 그렇기 때문에 더 공식적인 용어로, 노이즈가 큰 충분한 법칙 위반을 생성한 확률은 0 에 가깝습니다).

지식 영역은 모든 관련 임시 창에 대해 모든 법칙과의 모든 그룹 편차의 총합입니다.

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

과정은 질문할 쿼리에 의해 영향을 받지 않고 원데이터에 기원을 둔 내부 행동 구조에 영향을 받기 때문에 이 지식 영역은 일단 데이터 세트에 의거하여 계산합니다. 추상적 관점에서 Endor 의 사회 물리학 엔진은 익명의 가공 전 데이터를 행동 관련 규범적 표상으로 "압축합니다."

2.5.3 예

다음은 데이터 내 위반 패턴을 검출하기 위해 법칙을 어떻게 사용하는지 보여주는 두 가지 예입니다. 묘사 방식에 의해, 우리는 이 메커니즘을 입증하기 위해 알려진 수학적 현상을 사용합니다.

예 1: 비전

$d(x,t)$ 는 특정 프레임에서 특정 픽셀 x 의 색상을 나타낸다고 가정합니다. 이제 동일 색상 법칙을 정의할 수 있으며 소구역의 각색상은 일부 사전 정의 색상 C 이어야 할 것을 요구합니다.

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

이 연산자는 구역 x 와 시간 창 t 로 이 입력 데이터를 위한 평균 색상을 계산합니다.

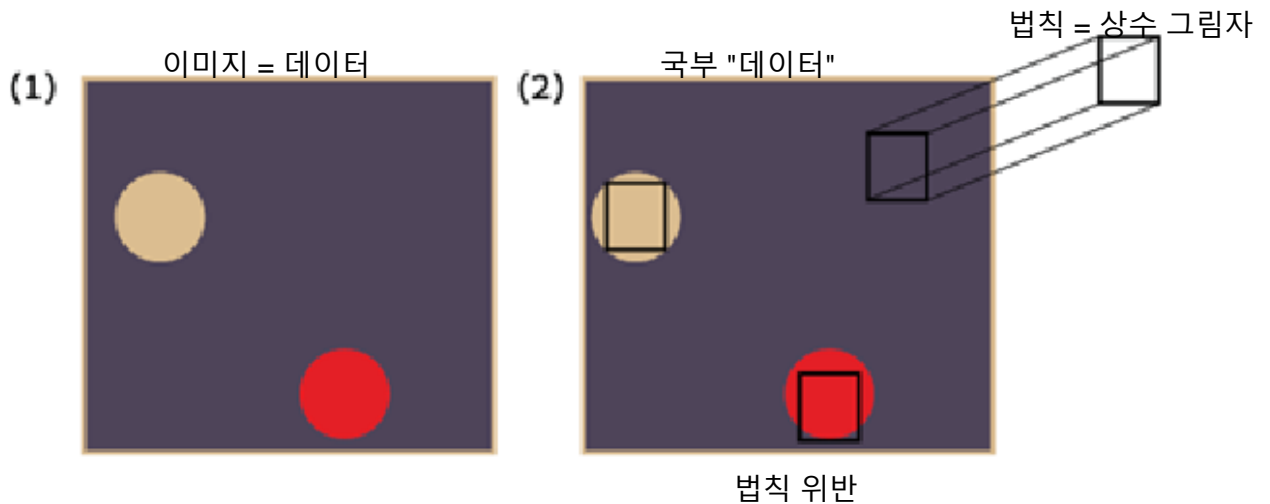
이 예는 네비게이션에 의한 법칙 개념을 설명합니다. 동영상에서 픽셀에 관한 고유의 연역적 법칙은 알려져 있는 것이 없습니다 (확실히, 각 소구역이 동일한 평균 색상이라고 가정하는 법칙이 아닌...). 이 저 연산자와 상수 $C =$ 하늘색을 고려하여 이제 Δx 를 크기 $N \times N$ 픽셀의 제곱 창으로 정의하고 단일 프레임 $\Delta t = 1$ 을 사용할 수 있습니다.

이로써 하늘색과의 $N \times N$ 창의 국부 편차로서 다음과 같은 값이 나옵니다.

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

2 차원 이미지에서, 이것은 쉽게 적용하여 하늘색과의 패치 편차를 측정하고, 이 편차를 사전 정의 임계치와 비교할 수 있습니다. 편차가 이 임계치를 초과하는 패치는 "유사한 성질을 가진 픽셀 클러스터"로 분류되고 "하늘색 법칙"을 위반한 정사각형 안에 있는 모든 픽셀을 나타냅니다.

우리는 이를 약간 더 포괄적인 가설 법칙을 사용하여 묘사할 수 있습니다. 이미지가 항상 “부드럽다” (즉, 작은 규모로 단조로우며 또는 “피크” 또는 국부적 최고/최소가 부족합니다)고 가정합니다. 이 가정 하에서 다음과 같은 “데이터”를 관찰할 수 있습니다. 부드러운 색상의 "합법적인 배경과 두 개의 패치 (하나는 파란색, 다른 것은 빨간색)가 들어 있는 이미지. 많은 수의 작은 정사각형 모양의 무작위 구역의 샘플링을 통해 이 두 가지 “위반”을 쉽게 특정할 수 있습니다.



그 다음, “쿼리”는 빨간색 픽셀 형식으로 물을 수 있습니다. 그 같은 픽셀은 빨간색 위반 안에 속해 있는 것으로 식별할 것이며 그 결과로 빨간 색 패치가 되돌아 옵니다.

이러한 예의 의도는 사회 물리학의 수학 및 역학을 설명하려는 것이지 간단한 비전 기반 애플리케이션에서 그것의 이점을 제안하려는 것은 아니라는 점을 다시 한번 유의하십시오. 그 같은 애플리케이션은 전통적 산출 비전 또는 딥 러닝 방법으로 더 잘 다룰 수 있습니다.

예 2: 척도 없는 네트워크

이 예에서 $\mathbf{d}(\mathbf{x}, \mathbf{t})$ 는 추상적으로 그래프 노드인 x 로 구성된 그래프를 나타냅니다. 법칙 연산자는 각도 분포 연산자이며 다음과 같이 형성됩니다.

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

이 벡터 연산자는 각 노드의 각도로 1 을 생성합니다. 모든 그래프의 노드 에서 이 연산자 부록 A: 사회 물리학 설명 결과값의 총합은 그래프의 각도에 대한 집합원 벡터 (노드 수로 나뉘었을 때 각도 분포와 같음) 입니다. 25

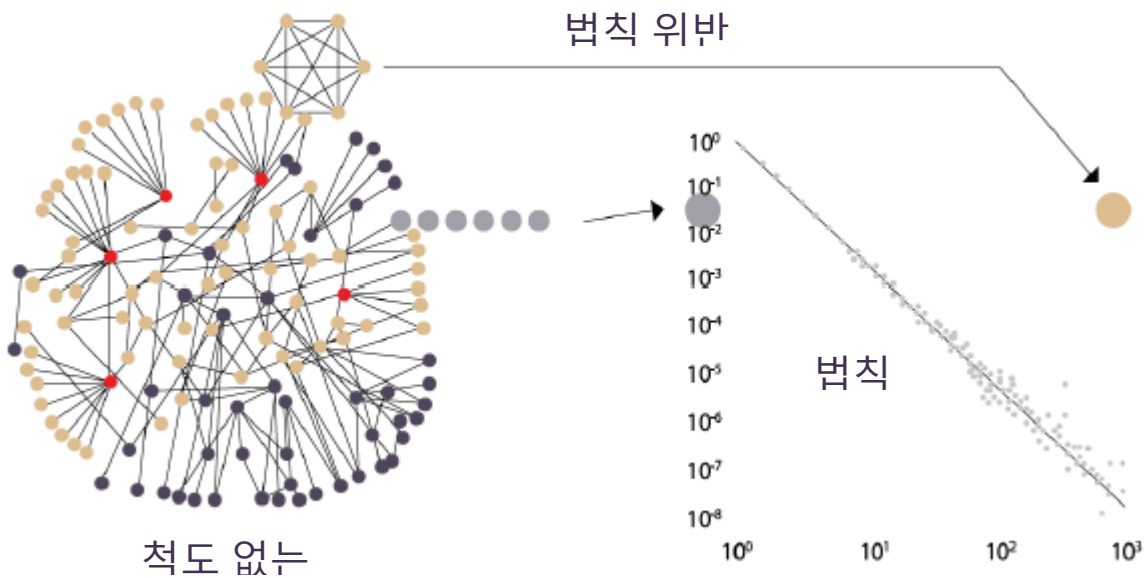
이 예에서 우리는 그래프는 척도 없는 네트워크이라고 가정해야 합니다. 그러므로 멱 법칙 각도 분포를 가정하는 법칙 상수를 적용할 수 있습니다 (일부 정규화 상수 α 를 위해):

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

이 법칙 상수는 그 다음 다음과 같이 표현할 수 있습니다.

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

이 법칙이 암시하는 바는 전체 그래프는 모든 노드의 각도의 멱 법칙 분포를 따라야 한다고 암시합니다. 하지만, 많은 큰 실제 세상의 척도 없는 그래프에서, 그 같은 분포와의 상당히 큰 국부 편차가 있을 수 있습니다. 예를 들어 이는 파벌 (즉, 완전히 연결된 하부 그래프) 또는 체인 (즉, 노드 없이 2개의 이상의 이웃이 있는 연결된 트리를 형성하는 노드 서브셋) 주위에서 발생할 수 있습니다. 그같은 편차 또는 법칙 위반은 다음과 같은 차트에서 나타내며 구조적 표상(좌)과 인접 매트릭스 (우) 모두에서 그것들이 발현되어 있습니다.



법칙이 주어진 상태에서, 그같은 위반은 다음과 같은 다양한 측정치에 의해 쉽게 검증할 수 있습니다.

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

□

이 편차는 차이의 누적 제곱을 측정합니다 (그 같은 측정치의 또 다른 예는 두 확률 분포의 *KL-발산* 이 될 수 있습니다). 여기서, Δx 는 그래프의 모든 가능한 하부 그래프를 나타냅니다. 분명히, 어느 한 입력 그래프 내 모든 가능한 하부 그래프를 스캐닝하는 것은 그것이 “*비 다항식 어려운 문제*”로서 알려진 "어려운 문제" 등급의 한 멤버이기 때문에 실행 가능하지 않습니다. 이런 의미에서, _ 법칙 위반 검증 (법칙의 세부사항을 알아야 함)과 법칙 위반 검출 (각 법칙 별로 특별히 개발된 독점 기법 한 세트를 요구함)을 구분하는 것이 중요합니다. 그같은 검출 기법은 Endor 의 독점 기술의 일부이며 사회 물리학 수학 법칙에 맞게 맞춤형으로 만들어졌습니다.

우리의 척도 없는 예로 되돌아와 우리가 그래프에서 그 같은 국부 간섭을 찾기 위한 효율적인 기법을 소유하고 있다고 가정할 때, 그러한 기법은 아래에서와 같이 생성될 수 있는 하부 그래프의 콜렉션을 만들어 낼 수 있습니다. :

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

이 법칙이 암시하는 지식 영역은 다음과 같이 정의됩니다.

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

일단 사회 물리학 법칙 위반 콜렉션이 들어 있는 지식 영역이 사용 가능해지면, 주어진 레벨 부착 전형과 "유사한 것"을 검출하는 데 사용할 수 있고, 동일 도메인에서 나온 객체들의 목록으로서 정의될 수 있습니다. 예에서, 그래프 노드 목록이 주어졌을 때, 다른 그래프 노드 모두는 그것들이 라벨 표시 전형과 공유한 클러스트의 수에 따라 점수를 매길 수 있습니다. 또는 대신, 그것들이 단지 입력으로서 검출된 클러스터와 라벨 표시 목록에만 의존하는 한, 인구 점수 형태로 출력을 생성하기 위해 다른 점수 매트릭스를 사용할 수 있습니다. 우리는 그 같은 매트릭스 또는 점수 메커니즘을 “*득점자*”로서 지칭하며 다음 섹션 이를 더 자세히 살펴 볼 것입니다.

다른 임시 창은 다른 지식 영역을 생성할 수 있으며, 그래프 노드 사이에서 매우 다른 유대를 나타낼 수 있다는 점에 유의합니다. 또한 그러한 클러스터는 소셜 미디어 또는 소셜 네트워크와 같은 외부 데이터 소스에서는 생성되지 않지만 차라리 고객 자체의 내부 거래 데이터 소스에서는 생성되는 행동 연관성을 나타냅니다. 이를 통해 Endor 는 다른 사용 가능한 데이터에서는 명시적으로 나타나지 않는 암시적 행동 클러스터를 검출할 수

있습니다.

2.5.4 질문에 대답: 간단한 “득점자”

일단 지식 영역을 사용할 수 있으면, 사용자는 의미적 의미가 있는 라벨 표시 데이터의 목록 (가능한 매우 적은)인 쿼리를 제공함으로써 사업 관련 질문을 묻기 시작할 수 있습니다.

$$y. \in X.$$

이 섹션에서는 출력으로서 순위가 정해진 인구 목록을 생성하기 위해 행동 클러스터 + 쿼리를 사용하는 함수인 “득점자”의 두 가지 예를 제시합니다.

논의한 첫 번째 득점자는 간단한 공통 클러스터 종합 득점자로 각 후보 별로 $\tilde{y} \in X$ 다음과 같은 점수를 계산합니다.

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

각 “후보”는 라벨 표시 데이터의 구성원이 함께 있는 지식 영역 내 클러스터의 수에 의거하여 점수가 매겨집니다. 이 간단한 득점자는 입력 라벨 표시 목록이 있는 X의 구성원 행동 유사성 수를 세는 것을 목표로 합니다. 라벨 표시 입력 목록 내 객체와 클러스터를 공유하는 객체 목록은 다음으로 나타내 집니다.

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

예를 들어, 색상 포인트와 예 1에서의 “동일 색상 법칙”이 주어질 때, 한 점 주위의 색상 영역은 입력으로서 주어집니다. “척도 없는 그래프 법칙”에서 노드가 주어질 때 출력은 각도 분포를 국소적으로 위반한 노드가 들어 있는 모든 하부 그래프 내 노드가 됩니다.

2.5.5 질문에 대답: 딥 러닝 기록원

이전 섹션에서 논한 득점자는 쿼리의 성격을 미리 알지 않고도 어떠한 "사례 입력" (즉, 쿼리)에 대해서도 고품질 "답은 꿀 예측"을 생성하기 위해 사용될 수 있는 사회 물리학 지식 영역으로 구성된 클러스터의 방식에 관한 직관적인 예입니다. 이 간단한 예는 사회 물리학의 규범적 데이터 표상의 이점을 설명합니다. 하지만, 정확한 예측을 생성하기 위해 Endor는 첨단 딥 러닝 알고리즘에 기초한 더 단단한 득점자를 개발했습니다.

위에서 언급했듯, 인간 행동과 관련하여 딥 러닝을 효율적으로 사용하려면 교육 단계에서 방대한 양의 데이터가 필요합니다. 각 라벨 표시 쿼리에는 일반적으로 적은 수의 라벨이 들어 있고, 전형적으로 쿼리 당 단지 몇 개의 예만 있고 그래서 고객 당 상대적으로 적은 수의 여러 쿼리가 있기 때문에 원 데이터를 사용하여 이를 이루기는 매우 어렵습니다. 이러한 유형의 복잡하고 역동적이며 적은 크기의 라벨 데이터는 딥 러닝에 잘 맞지 않습니다. Endor의 사회 물리학 엔진은 모든 고객에게서 받은 모든 데이터 세트와 모든 쿼리의 예를 단일 기본형의 큰 컬렉션으로 변환함으로써 이러한 제한을 극복합니다. 사람 클러스터로 구성된 지식 영역과, 그러한 클러스터를 지칭하는 예 모든 이러한 출처에서 나온 모든 데이터를 결합함으로써, Endor는 라벨과 지식 영역이 주어질 때, 각 사람 별로 점수를 매기는 딥 러닝 네트워크를 교육하기에 충분한 큰 라벨 표시 교육세트를 만들 수 있습니다. 이 과정이 일단 완료되면 그 결과로 교육된 높은 수준의 딥 러닝 모델이 나오고, 그 모델은 새로운 데이터 소스와 새로운 쿼리에서 생성된 새로운 클러스터를 효율적으로 처리하는데 사용할 수 있습니다.

이 교육된 모델은 클러스터 컬렉션 (즉, 사회 물리학 규범적 표상 형식)으로서 대표되는 많은 데이터 예에 관해 교육되었기 때문에 새로운 쿼리 또는 새로운 유형의 데이터 세트에 대해서 정기적으로 재교육할 필요가 없다는 점에 유의하십시오. 이 방법을 사용하여, Endor는 사회 물리학과 딥 러닝의 장점을 결합합니다. 사회 물리학은 익명의 원 데이터를 기본형으로 사회 물리학 법칙 위반에 기초하여 가 변환하며 딥 러닝 알고리즘은 그 다음 후보의 점수를 생성된 지식 영역과 (극히 적은) 라벨 표시 데이터에 기초하여 매깁니다.

2.6 노이즈에 강함

Endor 예측 플랫폼의 주요 강점 중 하나는 데이터 격차와 노이즈 데이터에 대한 높은 탄력성입니다. 전통적으로, 각 데이터-분석 프로젝트는 “데이터 위생” 단계로 시작하며, 다음과 같은 원하지 않는 데이터 세그먼트 검출 및 원하는 완화로 구성됩니다.

- 데이터 격차 (즉, 전체적 또는 부분적으로 데이터 손실 기간)
- 가공 전 데이터 내 횡설수설 삽입
- 한 개 이상의 형태로 데이터에서 나오는 카테고리 이름과 같이 의미상의 애매모호함
- 수치 값의 정상화 이슈
- 수치 값의 요구되는 비닝
- ...

머신 러닝 알고리즘은 일반적으로 노이즈에 매우 민감하고 데이터는 정상적으로 노이즈하다는 사실을 고려하면, 데이터 문제를 완화하는 단계는 전형적으로 비용이 많이 들어가고 오래 걸립니다. 이유는 전통적 머신 러닝 기법으로 분석하기 위해 실제 세상에서는 데이터는 원래의 형태 (금융 거래, 전화 통화, 등)로 생성되는 반면, 이것은 “특징” 또는 “성질”로 지칭되는 하나의 종합적인 형태로 전환되어야 합니다. 이러한 종합은 다시 적은 수의 틀린 데이터 값 또는 다른 사용자에게 대한 다른 수량 값으로 크게 영향을 받을 수 있습니다.

Endor 는 이 요건을 원데이터 자체를 분석함으로써 (이전 섹션에서 자세히 설명) 극복합니다. 또한 Endor 의 엔진은 예측을 위해 사용할 수 있는 패턴 검색 시 데이터의 통계적 분석은 하지 않고 대신 데이터 밖에 있고 그것에 의해 영향을 받지 않는 사회 물리학의 법칙인 수학 불변성을 활용합니다. 이 접근방식에는 다음과 같은 기본 개념에서 유래한 서너 개의 중요한 이점이 있으며 있습니다.

노이즈는 "임의로 나오게 할 수 없는" 데이터 패턴을 생성할 수 없습니다.

(여기서 후자는 무시할 수 있는 확률로 존재하기 위해 분석적으로 보여줄 수 있는 데이터-패턴 으로 정의됨)

이러한 관찰이 사전 클리닝 없이도 어떻게 인간-데이터에서 데이터 통찰력을 자동으로 추출하는지 이해하기 위해 추출 과정은 사회 물리학 법칙 중 하나를 위반한 데이터 내에서

부록 A: 사회 물리학 설명 32

객체 그룹을 찾는 것임을 상기합니다. 즉, 우리가 증명할 수 있는 데이터 패턴을 표시하는

그룹은 데이터 안에서 자연적으로 부상할 수 없습니다 (사회 물리학으로 가능한 수학적 분석 사용).

이는 노이즈가 우리의 엔진으로부터 나온 통찰력을 정말로 숨길 수 있는 반면, 정의 상, 법칙 위반으로서 검출될 수 있는 데이터 패턴은 (거의) 절대로 생성할 수 없음을 의미합니다. 노이즈 데이터는 사회 물리학 법칙을 위반할 수 없으며, 인간주도 신호 데이터만 할 수 있습니다.

3. 결과

이 섹션은 Endor 예측 시스템의 활용 방식을 나타내는 다양한 이용 사례를 제시합니다. 4 개의 예측 질문에 정확히 답하기 위해 7 일 간의 대형 금융 투자 플랫폼의 활동으로 구성된 데이터를 사용하여, 전반적 예측 과정에 관한 세부적 설명 (Google 의 텐서 플로우 딥 러닝 플랫폼과 비교 분석 포함)

1. Kaggle 도전을 무력화하기 위해 완전 자동 방식으로 Endor 를 사용한 예:
2. Endor 가 다수의 사업 질문에 대해 POS 데이터를 사용하여 24 시간 이내에 정확한 예측을 제공한 코카콜라 사례연구.

3.1. 정확하고 자동으로 4 가지 예측 질문에 답하기 위해 7 일 이상 금융 활동 사용

이 섹션에서는 전반적 예측 과정을 Endor 시스템을 사용하여 보여줍니다.

- 사용한 원 데이터에 관한 설명
- 데이터를 사회 물리학 표상으로 변환 (즉, 행동 클러스터 세트로 구성된 지식 영역)
- 4 개 예측 질문의 정의
- 사회 물리학 형식으로 쿼리 발현
- 전반적 예측 정확성
- Google 의 Tensor-Flow 딥 러닝 플랫폼과 비교.

3.1.1. 데이터:

이 예에서 사용한 데이터는 소매 금융 투자 플랫폼에서 나왔으며 투자 커뮤니티 구성원의 전체 투자 거래가 들어 있습니다. 데이터는 익명화 했으며 MIT 에서 연구 목적으로 공개되었습니다 (데이터는 요청 시 공유할 수 있습니다).

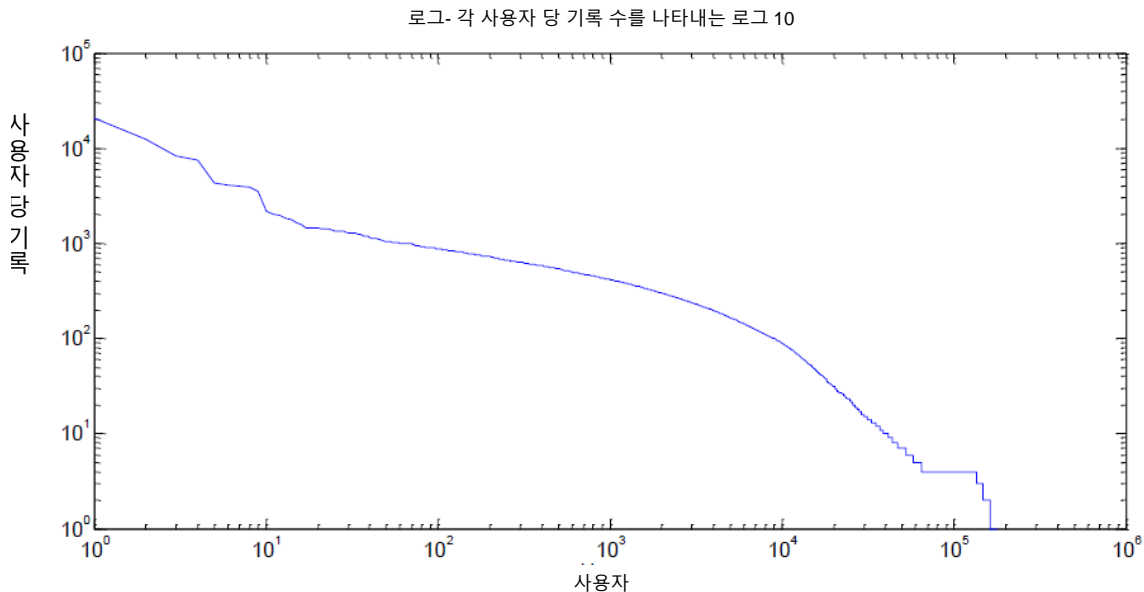
다음은 데이터 세트에 관한 전반적인 요약입니다.

- 7 일 간의 데이터
- 3,719,023 개 열
- 178,266 명의 고유 사용자
- 데이터의 맥락 또는 의미적 해석을 제공하지 않았습니다.
- 12 개의 데이터 필드:

필드명	유형	# 고유값
시간 1	시간	501573 고유
시간 2	시간	4 개의 고유
각 사용자 ID	INT32 단정적	178,266 명의 고유
기록 ID	INT32 단정적	1,053,574 명의 고유
성질 1	INT32 단정적	24 명의 고유 사용자
성질 2	INT32 단정적	7 명의 고유 사용자
성질 3	INT32 단정적	134 명의 고유
성질 4	INT32 단정적	77,527 명의 고유
성질 5	INT32 단정적	10 명의 고유 사용자
성질 6	INT32 단정적	27 명의 고유 사용자
성질 7	INT32 단정적	9 명의 고유 사용자
성질 8	이중 숫자	3,772 명의 고유

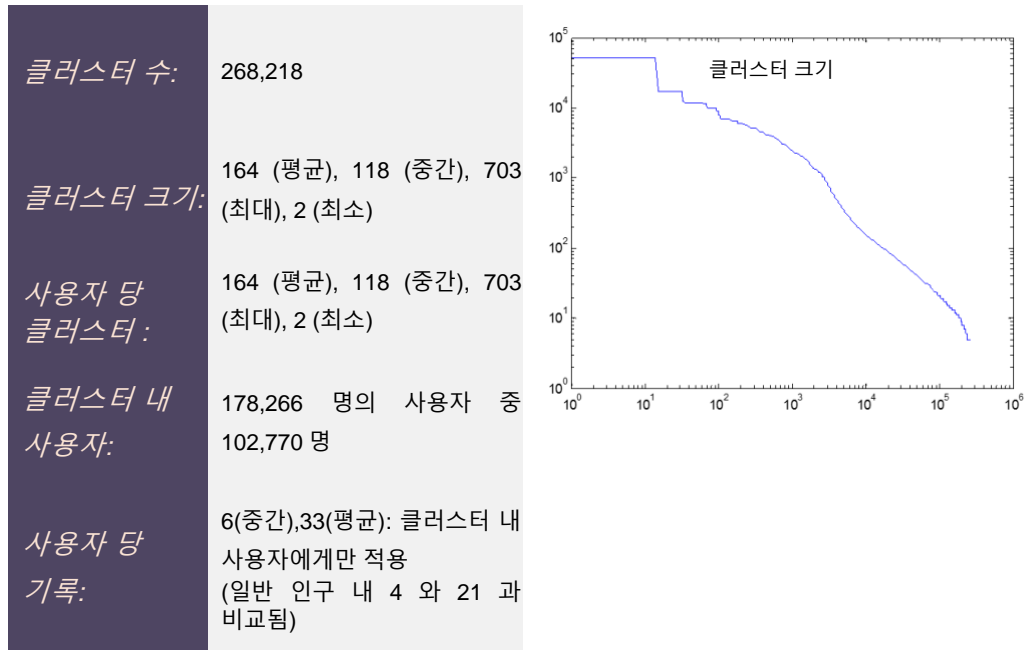
데이터 세트의 다음과 같은 주요 측면에 유의해야 합니다.

- 데이터는 가공 전 합계를 구하지 않은 형태로 주어졌고, 그 안에는 각 사용자 차원에서의 이벤트가 들어 있었습니다.
- 데이터의 맥락 또는 의미적 해석을 제공하지 않았습니다.
- 데이터 위생 미실시. 데이터에 노이즈와 격차, 이중 기록 등이 들어 있었습니다.
- 데이터에 사용자 당 기록이 극히 불균일한 분포로 들어 있었습니다. (사용자의 78%가 10 건 이하로 거래했지만 200 개 이상의 기록을 가진 사용자는 4,000 명 입니다. 중간값은 사용자당 4 개 기록):



3.1.2. 자동 클러스터 추출

첫 번째 데이터 분석 시, Endor 시스템은 데이터 역동성으로 사회 물리학의 수학 불변성을 위반하는 사용자 그룹인 “행동 클러스터”를 검출 및 추출합니다. 데이터는 시스템에 입력으로서 제공되었기 때문에, 이러한 클러스터는 그 데이터의 모든 행에 기초하지만 지난 7일로 제한됩니다. 다음은 시스템에 의해 검출된 데이터 세트에서 나온 행동 클러스터에 관한 요약입니다.



3.1.3. 예측 쿼리

다음 예측 쿼리가 정의되었습니다:

- **“고래”가 될 새 사용자** : 지난 2 주에 합류한 사용자는 다음 90 일 이내에 최소 \$500 를 커미션으로 생성합니다.
- **활동 감소** : 지난 주 적극적이었지만 다음 30 일 내 활동을 50%로 줄일 사용자 (서비스 제공자를 바꾸지 않고 여전히 거래를 계속함)
- **“고래” 전환 가입** : 지난 주에 적극적이었고 다음 30 일 내 비 적극적이 될 현재 적극적인 “고래” (지난 90 일 동안 활동으로 정의되는)
- **최초로 Apple 거래를 할 것인가** : Apple 지분을 한 번도 산 적이 없지만 향후 30 일 이내에 처음으로 지분을 살 사용자

알 수 있듯이 위의 질문 모두는 7 일 데이터 세트를 넘어 연장된 데이터를 지칭합니다 – 과거

데이터 (“검색 인구” 또는 “예” 생성에 사용)와 미래 데이터 (예측 검증에 사용) 모두. 시스템에 7 일 기간을 벗어난 정보를 제공하지 않기 위해, 쿼리는 사용자_ID 값이 들어 있는 목록 형태로 만들었습니다. 예를 들어:

- **쿼리명:** “고래”가 될 “새 사용자
- **검색 인구:** 지난 2 주 동안 가입한 사용자가 포함된 user_IDs 목록 (*End_of_data* 전).
- **“예”:** “고래”로 알려진 사용자가 포함된 user_IDs 목록(즉, *End_of_data* 전 90 일 내 커미션으로 \$500 이상을 생성한 사용자).
- **찾아야 할 타겟 목록:** user_IDs 목록, “검색 인구”의 서브셋으로 *End_of_data* 후 90 일 내 커미션으로 \$500 이상을 생성한 사용자가 들어 있음. 이 목록은 검증 목적으로만 사용되었으며 시스템에는 제공하지 않았습니다.

3.1.4. 쿼리의 지식 영역 발현

검색 쿼리 정의는 행동 클러스터 추출 및 쿼리 정의와는 독립적으로 이루어 지는 지식 영역 생성에 완전히 직각임을 명심하는 것이 중요합니다. 그러므로, 시스템으로 검출한 클러스터 내 쿼리 발현을 분석하는 것은 흥미로우며, 다음과 같은 사실에도 불구하고 클러스터에는 쿼리 정의와 관련된 정보가 들어 있습니다.

- 클러스터는 완전 자동으로 데이터에 관한 의미 정보를 사용하지 않고 추출했습니다.
- 쿼리는 클러스터 추출 후 정의했으며, 이 과정에 영향을 미치지 않았습니다.

이 분석은 매우 높은 밀도로 “샘플”이 들어 있는 클러스터 수를 측정하여, 다시 말해, 통계적으로 기대했던 것 보다 더 많은 예가 포함된 클러스터를 찾음으로써 실시했습니다.” 높은 수의 그 같은 클러스터 (동일 인구를 무작위로 샘플링 했을 때 받는 양 보다 훨씬 더 많은 경우)는 이 과정이 가치 있는 관련 의미 통찰력을 완전 자동으로 추출할 수 있음을 입증합니다.

아래 표는 이 관찰내용을 보여주고 동일한 샘플의 양이 들어 있는 “무작위 클러스터”의 수와 비교되는 일정량의 “샘플”이 들어있는 행동 클러스터 수를 비교합니다. 무작위 클러스터는 고객 인구에서 무작위로 샘플링한 N 개 사용자 그룹 세트를 지칭하며 그 N 개는 시스템에 의해 검출된 행동 클러스터의 수와 같으며 무작위로 샘플링된 그룹의 크기는 시스템에 의해 검출된 클러스터의 크기와 같습니다. 이것은 그 정보에는 무작위 샘플과 비교하여 높은 일관성을 지닌 “타겟 사용자” 클러스터가 상당히 크게 더 많이 포함되어 있기 때문에 (또한 그것들이 “질문” 정의 전 시스템에 의해 검출되었음을 상기할 때) 그 정보가 행동 클러스터에 의해 밀폐되었음을 입증하는 데 사용됩니다.

요청된 “샘플” 수는 “기준선” 단위로 주어집니다. 즉 – “활동 감소” 쿼리“에 대한 기준선의 X5” (기준선이 약 11%)는 클러스터가 “샘플”과 55% 중복되었다는 의미입니다 (즉, 그들의 구성원 55% 역시 “활동을 줄였던 이전 사용자” 목록에 나옵니다).

<i>많은 타겟 고객이 포함되어 있는 클러스터</i>		<i>#무작위 클러스터</i>	<i>#행동 클러스터</i>
활동 감소	기준선에서 X2	0	98
	기준선에서 X2	0	11
“고래” 전환 가입	기준선에서 X2	212	1678
	기준선에서 X2	67	525
	기준선에서 X2	21	114
구매한 적 없음	기준선에서 X2	3962	60044
	기준선에서 X10	1332	25542
	기준선에서 X20	415	9090
새 고래	기준선에서 X2	38	1898
	기준선에서 X5	0	65

3.1.5. 예측 결과

다음의 표는 4 개 쿼리에 대한 예측의 정확성을 나타냅니다.

- 기준선: 무작위 인구 샘플 내 요청된 타겟 고객의 평균 부분, 무작위 추측의 정확성을 나타냄.
- 후보: 검색 인구 크기. 예를 들어, “고래가 된 신규 사용자”에서 후보 수는 신규 사용자 수를 지칭합니다.
- 상위 100: 예측 보고의 상위 100 개 구성원 내 요청된 타겟 고객 부분 (마찬가지로, 상위-250 및 상위-500).

	<i>기준선</i>	<i>상위 100</i>	<i>상위 250</i>	<i>상위 500</i>
“고래”가 될 “새 사용자 지난 2 주에 합류한 사용자는 다음 90 일 이내에 최소 \$500 를 커미션으로 생성합니다.	7.5% (2270 명 후보 중 170 명)	37%	28.8%	21%
활동 감소 현재 적극적 사용자 중 다음 30 일 이내에 활동을 50% 줄이는 자 (하지만 해약하지는 않음)	11.4% (2233 명 후보 중 255 명)	21%	23%	20.2%
“고래” 전환 가입 현재 적극적인 “고래” 중에서 누가 지난 주에 적극적이었고 다음 30 일 내 비 적극적이 될 것인가?	1.66% (4141 명 후보 중 69 명)	10%	10.8%	9.2%
최초로 Apple 거래를 할 것인가 Apple 지분을 한 번도 산 적이 없는 자 중 누가 향후 30 일 이내에 처음으로 지분을 산 것인가?	0.5% (161382 명 후보 중 839 명)	14%	12%	10%

볼 수 있고 기대했듯이, 정확성은 예측 목록을 더 심도있게 살펴 봄에 따라 감소합니다.

3.1.6. 텐서 플로우와 비교

이 섹션에서는 Endor 시스템으로 획득한 예측 결과와 Google 의 텐서 플로우를 비교해 보도록 하겠습니다. 다른 딥 러닝 라이브러리와 마찬가지로 텐서 플로우는 논의 중에 있는 것과 유사한 데이터를 다룰 때 몇 가지 어려움을 맞이 합니다.

- 사용자 당 기록 수의 극히 불균일한 분포로 데이터의 정전화가 필요하며 또한 다시 다음을 요구합니다.
 - 최소 몇 가지 데이터 과학에 관한 이해가 있는 개인에 의해 실시된 일부 수작업.
 - 시간 투자는 물론 데이터 소유자 또는 제공자로의 액세스를 요구하는 데이터 의미에 관한 일부 이해
- 극히 균일하지 않은 포지티브 대 네거티브 샘플 분포를 사용하는 단일 클래스 분류는 결과의 과대 적합을 낳고 일부 사소하지 않은 조작을 요구하는 경향이 있습니다. 다시, 딥 러닝 전문가의 개입이 필요합니다. (사업 또는 제품, 마케팅 전문가가 사용할 수 있는 Endor 시스템과 다르며 선행 조건으로 머신 러닝 또는 데이터 과학을 요구하지 않음).

딥 러닝 전문가에게 2 주 동안 텐서 플로우에 기초한 솔루션을 제작하도록 요청했으며 그는 데이터 취급을 위한 충분한 전문지식을 갖고 있습니다. 생성된 솔루션은 다음과 같은 보조 기법을 사용합니다.

- 데이터 시퀀스를 고객 당 200 개 기록으로 자르고 기록이 200 개 미만인 사용자를 위해 중립 기록으로 스트림을 채웁니다.
- 200 개의 교육 세트를 만들어 각각에 1,000 명의 고객 (50% 알려진 포지티브 라벨, 50% 알려지지 않은 라벨)이 속하게 하고 그 다음 그러한 교육 세트를 모델 교육에 사용합니다.
- 시퀀스 분류 (128 개 LSTM 과 함께 RNN)를 2 개의 출력 뉴런 (포지티브, 네거티브)과 함께 사용하며, 전체적인 결과로 두 개의 점수 간에 점수 차가 있습니다.

아래 표는 이러한 기법을 사용하여 얻은 결과 (빨간색)를 Endor 의 예측 (파란색)과 비교한 것입니다.

	기준선	상위 100	상위 250	상위 500
“고래”가 될 “새 사용자 지난 2 주에 합류한 사용자는 다음 90 일 이내에 최소 \$500 를 커미션으로 생성합니다.	7.5% (2270 명의 후보 중 170 명) (2135 개 예)	37% 21%	28.8% 27.2%	21% 19.6%
활동 감소 현재 적극적 사용자 중 다음 30 일 이내에 활동을 50% 줄이는 자 (하지만 해약하지는 않음)	11.4% (2233 명의 후보 중 255 명) (366 개 예)	21% 8%	23% 18.8%	20.2% 19.4%
“고래” 전환 가입 현재 적극적인 “고래” 중에서 누가 지난 주에 적극적이었고 다음 30 일 내 비 적극적이 될 것인가?	1.66% (4141 명 후보 중 69 명) (21156 개의 예)	10% 11%	10.8% 12.4%	9.2% 8.4%
최초로 Apple 거래를 할 것인가 Apple 지분을 한 번도 산 적이 없는 자 중 누가 향후 30 일 이내에 처음으로 지분을 살 것인가	0.5% (161382 명 후보 중 839 명)	14% 1%	12% 0.8%	10% 1%

관찰내용:

- Endor 는 4 개 쿼리 중 3 개에서 텐서 플로우보다 우수하여 , 그 결과 4 번째에서 동일한 정확성이 나옵니다.
- Endor 의 우수성은 상위 500 개 보다는 상위 100 개를 중심으로, 작업이 “더 어려워”질수록 더욱 뚜렷해집니다.
- 누가 Apple 에서 최초로 거래할 것인가”라는 쿼리 보다 “덜 역동적인 쿼리” (고래가 되고, churn, 활동이 감소” – 그것을 위한 정적인 신호가 더 쉽게 검출될 가능성이 큰) 간에 더 분명한 차이가 있으며, (a) 더 역동적이고 (b) 기준선이 매우 낮으며, 후자에 대해서 Endor 는 10 배 더 정확합니다!
- 이전에 언급했듯이 – 여기서 나타낸 텐서 플로우 결과는 딥 러닝 전문가에 의해 완료된 한 2 주 간의 수작업 개선을 사용한 반면 Endor 결과는 100% 자동으로 생성되었습니다.

3.2. KAGGLE 도전을 3 시간 내 자동으로 무력화

또 다른 예에서, 우리는 Endor 시스템을“가치 있는 쇼핑객 도전 획득”으로 알려진 Kaggle 경연대회에서 나온 일반에게 공개되는 데이터로 시험했습니다. 그 데이터에는 수 십만 고객을 지칭하는, 거의 3억 개의 POS 기록이 담겨 있는 한편 도전은 특정 프로모션 쿠폰을 받은 사용자 중 누가 되돌아온 고객이 될지에 관한 예측을 수반했습니다.

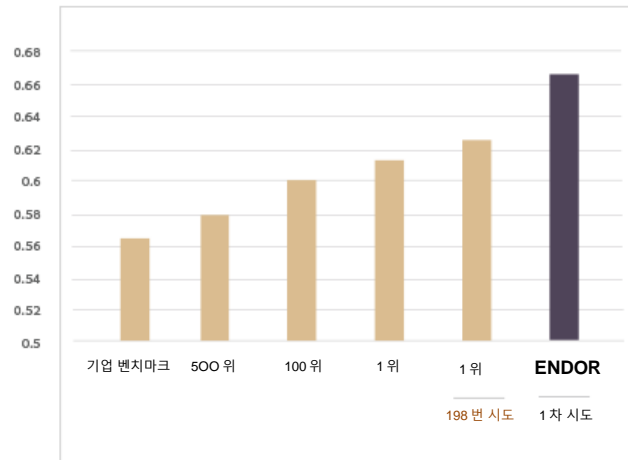
원래 도전은 952 개 팀을 포괄하고 3 개월 간 지속되었습니다.

도전 참가자에게 주어진 원 데이터 상에서 완전 자동으로 가동하는 Endor 기계는 원래 1 위 였던 팀보다 더 나은 예측을 생성할 수 있었습니다.



- 952 개 경쟁 데이터 과학팀
- 3 개월 도전

ENDOR:
몇 번의 클릭, 1 위



부록 B – Endor 의 기업용 상용 사례

퍼스널 Oracle 을 가지면 어떨것 같습니까? Google 및 Facebook 등 만 이용하는 강력한 데이터 엔진에 액세스 할 수 있으면 어떨것 같습니까? 지금까지는 Endor 인 가장 강력한 그 같은 엔진을 사용하여 기술력이 뛰어난 자만이 주는 가장 신뢰할 수 있는 답을 얻을 수 있으면 어떨것 같습니까? 선진 머신 러닝을 민주화함으로써 블록체인 예측을 모두가 접근 가능하고 모두를 위해 유용하게 만들겠다는 비전으로, *Endor.coin* 은 지금 까지는 내부적으로 전문 데이터 전문가 팀을 갖춘 기술 자이언트에게만 가질 수 있었던 능력을 누구나 가질 수 있게 합니다.

Endor 의 기업 고객이 요구하는 예측 상용 사례는 아래에서와 같이 요약됩니다. 이용 사례는 Endor 가 현재 서비스를 제공하는 주요 세그먼트 별로 분류됩니다.

- 소매 은행
- 보험사
- 소매 및 e-커머스

하지만, *사회/물리학*를 활용한 덕분에, 새로운 이용 사례는 쉽게 지지될 수 있으며 – 평균적으로 Endor 의 판매 엔지니어에게 서너 시간을 요구합니다.



ENDOR



이용 사례 : 소매 은행

ENDOR

상용 사례

더 많이 판매



- 구매 의사
- 교차 판매
- 상향 판매
- 씬씬이가 큰 사람
- 신규 서비스 - 얼리 어답터

문제의 예

- 구매 의사: 지난 3 개월 동안 모든 자격이 되는 적극적 고객 중, 다음 주에 전화를 받았을 때 누가 \$X 이상의 대출을 받을 것인가? 누가 신용 카드 없이 대출을 받을 것인가? 단기 대출을 받을 것인가? 장기 대출을 받을 것인가?
- 교차 판매: 지난 6 개월 내에 제품 X 를 사용했던 고객 중 다음 달에 프로모션 Y 를 하면 누가 제품 Z 를 사용하기 시작할 것인가
- 씬씬이가 큰 사람: 지난 30 일 동안 가입한 신규 고객 중 누가 향후 3 개월 동안 매달 \$X 이상을 지출할 가능성이 있는가?
- 신규 서비스: 지난 3 개월 동안 적극적 신용 카드 고객 중에서, 누가 당사가 2 주전 출시했던 신규 서비스에 등록할 것인가?

상용 사례

더 많이 판매



추가적인 예

- 지난 90 일 동안 적극적 신용 카드 고객 중에서, 누가 두 번째 카드를 받으면 지출을 \$/%X 이상 증가시킬 것인가?
- 지난 90 일 동안 비 적극적 신용 카드 고객 중에서, 당사가 제품 A 를 B 로 변경하면 누가 카드를 사용하기 시작할 것인가?
- 지난 6 개월 동안 적극적 신용 카드 고객 중에서, 누가 다음 3 개월 동안 잔고를 유지할 것인가?
- 지난 6 개월 동안 적극적 신용 카드 고객 중에서, 누가 다음 0 개월 동안 기꺼이 Tasa 를 가져갈 것인가?
- 모든 적극적 골드 및 플래티늄 계정 고객 중 누가 다음 달에 파일 경비에 카드를 기꺼이 포함시킬 것인가?
- 지난 6 개월 동안 관심을 갖게 된 고객 중에서, 다음 3 개월 내 그들의 잔고를 청산할

것인가?

- 당사의 여행 웹사이트에서 비행기 표를 구입한 사람 중, 누가 호텔 숙박권까지 구입하는 데 관심을 가질 것인가?
- 어느 플래티늄과 골드 포트폴리오 고객이 다음 2 개월 내에 높은 비용 범주에서 포인트를 만회할 것인가?
- 플래티늄 카드 높은 가치 고객 중 누가 다음 달에 카드를 취소할 것인가?
- 지난 12 개월 동안 적극적 고객 중에서, 누가 다음 2 개월 동안 제품 모두에서 \$Y 이상 소비할 가능성이 큰가?
- 지난 6 개월 동안 적극적 제품 Y 고객 중에서, 누가 다음 2 개월 동안 제품 Y 에서 \$Z 이상 소비할 가능성이 큰가?
- 작년 동안 카드가 없었던 고객 중 누가 다음 주 연락을 받으면 다음 달에 새 신용카드를 발급받겠는가

상용 사례



마케팅 효율성

- 프로모션 효율성
- 채널 전체에서의 연결 유지
- 고객 충성도

문제의 예

- 프로모션 효율성: 적극적인 고객 중 누가 문자 메시지? 전화? 프로모션 γ 를 받았을 때 프리미엄 서비스 x 에 가입하겠는가?
- 디지털 카드: 목록에 있는 웹사이트 상에서 최소 γ 번 구매했던 고객 중 누가 다음 달페이백 서비스를 사용하여 구매할 것인가
- 지난 x 개월 동안 면세점에서 최소 1 회 구매했지만 탑승해서는 아무것도 구매하지 않은 고객 중 누가 앞으로 3 개월 내 탑승해서 신용 카드로 구매할 것인가?
- 적극적인 모든 고객 중 누가 다음 30 일 내에 뉴욕 또는 영국 목적지 X 로 여행하겠는가?
- 지난 90 일 동안 가입한 고객 중 누가 카드를 활성화하지 않았으며 누가 \$10 보너스를 받으면 활성화할 것인가? \$20?
- 지난 6 개월 동안 모든 적극적인 카드 보유자 중 누가 다음 달 카드 사용을 중단할 가능성이 있는가?

ENDOR

상용 사례

마케팅 효율성



추가적인 예

- 매월 소비자 신용 프로모션 내용이 담긴 이메일을 열어보지 않은 고객 중 당사가 다음 주 (1) 전화를 하면 또는 (2) SMS 를 전송하면 누가 다음 3 주 내 새 신용을 얻을 것인가
- 3.000 ~ 24.000 포인트가 있는 기혼 고객 중 당사가 내일 이메일을 보내면 누가 다음 4 일 내 포인트를 향수로 교환할 것인가?
- 식당에서 카드를 사용하고 카드 한도가 \$1 백 만 달러 이상이며 지역 x 에서 사는 고객 중 누가 이번 주에 이메일을 받으면 다음 주에 제품 x 에서 $\gamma\%$ 이상 소비를 늘릴 것인가
- 다음 주 특별 주유 할인권을 전송하면 어느 고객이 다음 달에 주유소에서 카드를 사용할 것인가?
- 지난 6 개월 동안 미불 채무로 재상담 했던 고객 중 잔액 납부 안내 SMS 를 내일 받으면 누가 다음 달에 연체자가 될 것인가?
- 지난 2 개월 동안 카드를 사용하지 않은 고객 중 다음 주 주유 특별 할인권을 받으면 누가

카드를 다시 사용하기 시작할 것인가?

ENDOR

상용 사례

디지털 변환



- 온라인 활동 증가
- 지점에서 온라인으로 이동
- 웹사이트 방문자 리타겟팅
- 최초 구매자, 반복 구매자, 유지
- 씬스름이가 큰 사람

문제의 예

- 2개월 동안 비 적극적으로있던 온라인 고객 중 누가 전화를 받았을 때 다음 7일 내 적극적으로 될 것인가? 인센티브 X?
- 전통적 거래 (지점, 콜 센터)를 하는 모든 고객 중 누가 프로모션 x 를 하면 다음 주에 디지털을 사용할 것인가?
- 지난 7일 동안 신규 온라인 사용자 중 누가 첫 번째 3개월 동안 누적액 \$x 거래를 할 가능성이 있는가?
- 지난 3개월 중 디지털 채널을 통해 신 상품에 가입한 고객 중 프로모션 x 를 받으면 다음 달 내에 두 번째 상품에 가입하겠는가?
- 지난 3개월 동안 웹사이트 방문자 중 우리가 리타겟팅하는 경우 어느 방문자가 당사 온라인 बैं킹에 등록할 가능성이 있는가?

ENDOR

상용 사례

디지털 변환



추가적인 예

- 새 온라인 사용자 중 누가 지난 달 등록했으며, 누가 향후 3개월 동안 계속해서 온라인을 사용할 것인가?
- 당사의 모든 온라인 बैं킹고객 중 누가 다음 달 온라인 बैं킹 사용을 중지하겠는가?
- 지난 달 등록한 모든 신규 온라인 사용자 중 다음 주에 인센티브 x 를 받으면 누가 반복 구매 온라인 사용자가 될 것인가
- 기능 x 는 사용하지만 y 는 사용하지 않는 모든 온라인 고객 중 연락을 받으면 누가 다음 달에 기능 y 를 사용할 것인가
- 지난 3개월 중 새로운 디지털 상품(X,Y,Z) 에 가입한 모든 고객 중 프로모션 X 를 받으면 누가 다음 달 내에 두 번째 상품에 가입할 것인가
- 지난 3개월 중 디지털 채널을 통해 상품에 구매했던 고객 중 누가 1개월 내 떠날

것인가

- 지난 3 개월 동안 모든 웹사이트 방문자 (쿠키 ID) 중 리타겟팅되면 누가 당사 온라인 뱅킹에 등록할 가능성이 있는가
- 한 가지 은행 상품만 갖고 있는 고객 중 누가 다음 달에 두 번째 상품을 구매할 가능성이 있는가?
- 당사 웹사이트에 한 번도 로그인 한적 없는 고객 중 내일 이메일을 보내면 누가 다음 주 중에 로그인 하겠는가?



ENDOR

이용 사례 : 보험

ENDOR

상용 사례

매일 어떻게 예측을 하십니까? 결정? 실적 증진?

다음과 같은 질문에 대답해야 합니다.



판매

- 누가 보조 커버리지 X를 추가하는가
- 다음 주 접근하면 누가 상향 판매/교차 판매할 가능성이 있는가
- 사전 유효 일자를 취소했던 사람 중에 누가 재구매할 것인가
- 일주일 전에 출시된 신제품/신규 서비스를 누가 구매할 것인가
- 누가 다음 주에 새 계획을 신청할 것인가
- 고객 중 누가 씬셈이가 큰 사람 / 프리미엄 고객이 되겠는가



마케팅

- 누가 프로모션 X에 반응할 것인가
- 누가 처음 90일 이내에 계약을 해지할 것인가
- 누가 당사의 모바일 서비스 (app)를 사용할 것인가
- 누가 두 달 내에 뉴욕과 멕시코를 여행할 것인가
- 누가 친구를 추천할 가능성이 높은가?



디지털

- 누가 \$/%X 이상 활동을 증가시킬 것인가
- 누가 새 디지털 서비스를 이용할 것인가
- 어느 비 적극적인 고객이 적극적인 고객이 될 것인가
- 누가 지점에서 온라인으로 이동할 것인가
- 누가 프로모션 X에 반응할 것인가
- 새로운 구매층에게 맞추면 어느 웹사이트 사용자가 이전하겠는가

상용 사례

더 많이 판매



- 구매 의사
- 교차 판매
- 상향 판매
- 씬셈이가 큰 사람
- 신규 서비스 - 얼리 아답터

문제의 예

- 구매 의사: 모든 잠재적 고객 중 누가 방침 X를 추가할 가능성이 있는가? 모든 기존의 고객 중 누가 보조 커버리지를 추가할 가능성이 있는가?
- 교차 판매: 플랜 x 고객 중, 프로모션 z를 하면 누가 제품/서비스 y를 구매할 가능성이 있는가?
- 씬셈이가 큰 사람: 지난 30일 동안 가입한 신규 고객 중 누가 향후 3개월 동안

매달 \$X 이상을 지출할 가능성이 있는가?

- 신규 서비스: 지난 3 개월 동안 적극적 신용 카드 고객 중에서, 누가 당사가 2 주전 출시했던 신규 서비스에 등록할 것인가?

상용 사례

디지털화



- 온라인 활동 증가
- 지점에서 온라인으로 이동
- 웹사이트 방문자 리타겟팅
- 최초 구매자, 반복 구매자, 유지
- 씬씬이가 큰 사람

문제의 예

- 2개월 동안 비 적극적으로었던 온라인 고객 중 누가 전화를 받았을 때 다음 7일 내 적극적으로 될 것인가? 인센티브 X?
- 오프라인 (지점, 콜 센터)으로 연락했던 모든 고객 중 누가 프로모션 X를 하면 다음 주에 디지털을 사용할 것인가?
- 지난 7일 동안 신규 온라인 사용자 중 누가 첫 번째 3개월 동안 \$x을 지출할 가능성이 있는가?
- 지난 3개월 중 디지털 채널을 통해 신 상품에 가입한 고객 중 프로모션 x를 받으면 다음 달 내에 두 번째 상품에 가입하겠는가?
- 지난 3개월 동안 웹사이트 방문자 중 우리가 리타겟팅하는 경우 어느 방문자가 온라인으로 커버리지를 추가할 가능성이 있는가?

ENDOR

상용 사례

마케팅 효율성



- 프로모션 효율성
- 신제품
- 고객 충성도 및 추천
- 타겟팅된 프로모션

문제의 예

- 프로모션 효율성: 적극적인 고객 중 누가 문자 메시지? 전화? 프로모션 Y를 받았을 때 프리미엄 서비스 X에 가입하겠는가?
- 타겟팅된 프로모션 적극적인 모든 고객 중 누가 다음 30일 내에 목적지 X로 여행하겠는가?
- 추천: 적극적인 모든 고객 중 누가 인센티브 X를 받으면 친구를 추천할 가능성이 있는가?

- 신제품: 적극적인 모든 고객 중 누가 다음 주에 접촉하면 신규 디지털 서비스 X를 사용할 가능성이 있는가?

상용 사례



유지

- 증권 갱신
- 해지
- 다시 돌아온 고객

문제의 예

- 해지: 증권 X의 모든 신규 고객 중 누가 처음 90일 이내에 계약을 해지할 가능성이 있는가?
- 다시 돌아온 고객: 발효일 전 취소했던 고객 중 누가 다시 계약을 할 것인가?
- 갱신: 지난 60일에 새 보험에 가입한 모든 고객 중 누가 발효일 전에 취소할 것 같은가?

ENDOR



이용 사례 : 소매 및 e-커머스

ENDOR

상용 사례

매일 어떻게 예측을 하십니까? 결정? 실적 증진?

다음과 같은 질문에 대답해야 합니다.



판매

- 누가 다음 달에 X를 구입할 것인가
- 누가 1주일 전에 출시한 제품을 구입할 가능성이 높은가
- 누가 프리미엄으로 이동할 것인가
- 누가 큰 소비자가 될 것인가
- 어느 X 고객이 Y를 구입할 것인가



마케팅

- 누가 프로모션 X에 반응할 것인가
- 어느 캠페인을 유지할 것인가
- 어떤 고객이 온라인으로 구입할 것인가
- 누가 당사의 모바일 앱을 사용할 것인가
- 누가 \$X 이상 활동을 증가시킬 것인가



E-커머스

- 어느 쿠키 ID를 전환할 것인가
- 누가 프로모션 X에 반응할 것인가
- 누가 첫 번째 구매자가 될 것인가
- 누가 활동을 증가시킬 것인가
- 어떤 비 적극적인 고객이 프로모션 X를 실행하면 적극적인 고객이 될 것인가



스토어/제품

- 어느 매장이 제품 Y을 \$X 판매량으로 올릴 것인가
- 어느 매장의 실적이 기준에 미달할 것인가
- 어느 제품이 다음 달 20% 이상의 증가를 경험할 것인가



전략

- 조기 경고 시스템 어디서 결함이 발생할 것인가
- 어디서 제품 X을 위한 순풍을 얻을 것인가

상용 사례

더 많이 판매



- 구매 의사
- 교차 판매
- 상향 판매
- 향후 씬셈이가 큰 사람
- 신제품-얼리 아답터

문제의 예

- 지난 3 개월 동안 모든 적극적 고객 중에 누가 향후 3 개월 내 352553 을 구입할 가능성이 있는가?
- 지난 30 일 내 제품 219 를 구입했던 모든 고객 중에 누가 향후 90 일 내 제품 890 을 구입할 가능성이 있는가?
- 지난 3 개월 동안 적극적 제품 32236 고객 중에 누가 다음 7 일 동안 프리미엄으로 업그레이드할 가능성이 있는가?
- 지난 6 개월 동안 모든 적극적 고객 중에 누가 당사가 3 주 전에 출시한 신제품을 구입할 가능성이 있는가?
- 지난 30 일 내 모든 신규 고객 중 누가 처음 3 개월 내 누적액 \$10K 이상 구매할 것인가?

ENDOR

상용 사례

마케팅 효율성



- 더 나은 분할
- 프로모션 효율성
- 채널 전체에서의 연결 유지
- 고객 충성도

문제의 예

- 지난 6 개월 동안 상위 10%를 소비하는 고객 중 누가 2 개월 내 소비를 최소 20% 줄일 것인가?
- 지난 6 개월 동안 하위 20%를 소비하는 고객 중 누가 3 개월 내 소비를 최소 30% 적극적으로 늘릴 것인가?
- 모든 고객 중 다음 주 20% 할인 쿠폰을 받으면 누가 제품 6347 을 구매할 것인가?
- 모든 고객 중 당사자가 다음 주 텍스트 메시지를 보내면 누가 제품 6347 을 구매할

것인가?

- 지난 90 일 동안 웹사이트를 통해 구매했던 모든 고객 중 누가 다음 달에 모바일 앱을 사용할 가능성이 있는가?

ENDOR?

상용 사례

증가된 활동



- 효율적인 사용자 취득 (리타겟팅)
- 첫 번째 구매자
- 반복 구매자
- 증가된 활동

문제의 예

- 모든 웹사이트 방문자 (쿠키) 중 누가 다음 주 Google 을 통해 리타겟팅되면 고객이 될 가능성이 있는가
- 모든 웹 사이트 방문자 (쿠키) 중 누가 처음 3 개월 내 누적액 \$10K 이상을 구매할 것인가?
- 지난 6 개월 간 비적극적인 모든 고객 중 누가 다음 달에 FTB 가 될 것인가?
- 지난 7 일 동안 처음 식료품 주문을 했던 모든 고객 중 누가 60 일 이내 2 건의 새 주문을 할 것인가?

많은 다른것...



제품 / 매장

- 어느 서비스가 다음 달 20% 이상의 증가를 경험할 것인가
- 어느 서비스가 다음 3 개월 내 50% 이상 감소를 경험할 가능성이 있는가
- 어느 매장이 다음 3 개월 내 카테고리 판매가 50% 이상에 도달할 것인가



운영

- 누가 내일 X 때문에 콜 센터/기술 지원팀에 연락할 것인가
- 누가 내일 지점에서 거래를 할 것인가



전략

- 다음 달 브랜드 X 결함이 어디에서 나올 것인가 (역풍)
- 어디서 브랜드 Y 를 위한 순풍을 얻을 것인가

”사전 정의 메뉴”는 없습니다

비즈니스 사용자는 사업 성장에 필요할 때 마다 묻습니다.

부록 C – 사전 정의 예측의 Endor.coin 예

Endor.coin 의 웅대한 비전은 보험 및 은행, e 커머스, 건강 등과 같은 블록체인에 의해 변환될 섹터에 겨냥하고 있는 한편, 우리는 거래에서 주도권을 추구하는 암호 보유자에게 암호화 화폐 통찰력을 제공하기 위한 이전에 없었던 예측 플랫폼을 제공하고 있습니다. 블록체인의 신뢰 입증 분산화된 기반시설 위에서, 누구에게나 어떤 데이터 소스에서든 그들의 실제 거래 전략을 공개하지 않고도 그들의 가설을 시험할 수 있게 허용하며, *Endor.coin* 은 다른 기존 기술의 렌즈를 통해 관찰할 수 있기 이전에 미래를 보는 힘을 부여합니다.

다음은 출시 시 *Endor.coin* 플랫폼에 의해 지지될 첫 번째 사전 정의 블록체인 예측 일부의 예입니다. 새로운 이용 사례는 *RFP* (예측 요청) 메커니즘을 사용한 커뮤니티의 요청에 따라 점차적으로 추가됩니다.

암호화 화폐 주소 예측: 이러한 예측은 사전 정의 주소 목록 (즉, "지난 달 최소 한 번의 발신 거래가 있었던 비트코인 주소")을 수신하고 최근의 사전 정의 행동과 그것들의 행동 유사성에 의거하여 순위를 매깁니다 (즉, "지난 주 최소 0.1 비트코인을 받았던 주소"). 결과 목록에는 사전 정의 행동을 가장 유사한 주소가 상부에 포함될 것이며 (그러므로 미래 동일한 행동을 표시할 가능성이 통계적으로 더 높은), 한편 결과 목록 하단에는 사전 정의 행동과 가장 유사하지 않은 주소가 나와 있을 것입니다.

- **활성 주소:** 지난 달 최소 1 회 활성화 되었던 모든 주소 중, 최근 거래수가 크게 증가한 주소와 가장 유사한 주소는?
- **거래가 활발한 주소:** 지난 달 최소 1 회 활성화 되었던 모든 주소 중, 지난 달 전체 거래량이 10BTC 를 넘는 주소와 가장 유사한 주소는?
- **비 활성화 주소 되기:** 지난 달 최소 1 회 활성화 되었던 모든 주소 중, 최근 전체 거래량이 50% 감소한 주소와 가장 유사한 주소는?

토큰 예측: 이러한 예측은 사전 정의 토큰 목록 (즉, "지난 달 거래량이 최소 1 천만 달러인 토큰")을 수신하고 최근의 사전 정의 행동과 그것들의 행동 유사성에 의거하여 순위를 매깁니다 (즉, "평균 월 거래량이 2 증가했던 토큰"). 결과 목록 상단에는 사전 정의 행동과 가장 유사한 (또한 그러므로 미래에 동일한 행동을 나타낼 가능성이 통계적으로 더 큼) 토큰이 올 것이며 한편 사전 정의 행동과 가장 유사하지 않은 토큰은 목록 하단에 배치됩니다.

- **수익성 토큰:** 지난 달 거래량이 최소 1 백만 달러였던 모든 토큰 중에서 어느 것이 거래 지난 달 BTC 와 관련하여 가격이 50% 이상 인상된 토큰과 가장 유사한가?
- **비수익성 토큰:** 지난 달 거래량이 최소 1 백만 달러였던 모든 토큰 중에서 어느 것이 거래 지난 달 BTC 와 관련하여 가격이 50% 이상 하락된 토큰과 가장 유사한가?
- **변동성 토큰:** 지난 달 거래량이 최소 5 백만 달러였던 모든 토큰 중에서 어느 것이 지난 달 가장 변동성이 큰 10 개 토큰과 가장 유사한가?
- **안정적 토큰:** 지난 달 거래량이 최소 5 백만 달러였던 모든 토큰 중에서 어느 것이 지난 달 가장 변동성이 적은 10 개 토큰과 가장 유사한가?
- **토큰 양 증가:** 지난 달 거래량이 최소 2 천만 달러였던 모든 토큰 중에서 어느 것이 거래 지난 달 월 거래량을 배가시킨 토큰과 가장 유사한가?
- **토큰 양 감소:** 지난 달 거래량이 최소 2 천만 달러였던 모든 토큰 중에서 어느 것이 거래 지난 달 월 거래량이 50% 감소된 토큰과 가장 유사한가?

부록 D – 지식 영역 클래스: API 접근

이 부에는 – *Endor.coin* 프로토콜의 기초 건축 블록인 “지식 영역” 데이터 구조의 사양은 물론 액세스를 제공하고, 미래 분석 엔진의 구현을 위해 , to be plugged to the *Endor.coin* 네트워크에 연결되는 등급에 대한 전체 설명이 수록되어 있습니다.

이용 설명: 클러스터 객체는 3 개의 객체로 구성됩니다.

- 희소 행렬 M , 차원 ($|$ 검색가능 객체 $| \times |$ 행동 클러스터 $|$), 검색가능 객체와 행동 클러스터 간 연결성을 나타내며 인덱스 $i \in$ 클러스터 j 의 $M_{i,j} = 1$ iff 검색가능 객체를 갖습니다 . 검색가능 객체는 토큰과 지갑 주소, 위치, 전화 번호 또는 기타 유형의 객체를 지칭하며 즉, 데이터 내 포함되고 예측의 기초 역할을 합니다.
 - 배열 A_M 각 검색가능 객체 SO 를 희소 행렬 M 내 인덱스로 매핑.
 - 데이터프레임 D_M 기타 클러스터 성질이 들어 있으며 각 예측 엔진이 정의 및 계산한 바와 같습니다. 그 같은 성질은 예를 들어 클러스터 크기와 내부 커넥티비티 대 외부 커넥티비티의 비율, 예측 엔진이이 클러스터 등을 생성하기 위해 사용하는 내부 모듈이 될 수 있습니다.

사용할 클러스터 객체를 추출 및 구축하기 위해 , 다음과 같은 파일이 필요합니다.

- 매트릭스 M 구축을 위해:
- 매핑 배열 A_M 구축을 위해:
- 클러스터 성질 데이터프레임 D_M 구축을 위해:

위에 언급한 파일은 특정 경로에 놓일 것이며 클러스터에 의해 아래 나타낸 sExtractor 등급으로 지칭됩니다.

<cluster_files_path>

사용예: 관련 5 개 클러스터 파일을 i 클러스터 파일 경로 i 에 놓은 후, 아래 코드를 사용하여 클러스터를 추출할 수 있습니다.

이 예에서 “클러스터 맵으로의 팝”은 희소 행렬 M 를 가리킵니다. “클러스터 프롭”은 클러스터 성질이 들어 있는 데이터프레임 D_M 을 가리킵니다. “해석 팝”은 각 검색가능 객체를 희소 행렬, A_M 내 인덱스에 매핑하는 배열을 가리킵니다.

클러스터 추출기 등급: 다음은 지식 영역 API 에 관한 설명 전문입니다.

i

용어집

- [1] Wikipedia – *사회 물리학* (2017).
URL https://en.wikipedia.org/wiki/Social_physics
- [2] W. Pan, Y. Altshuler, A. Pentland, 금융 거래 네트워크에서 사회적 영향력과 집단 지혜 디코딩: 프라이버시, 보안, 리스크, 신뢰 (PASSAT), 2012 년 소셜 컴퓨팅 국제 회의 (SocialCom), IEEE, 2012, pp. 203–209.
- [3] Y.-Y. Liu, J. C. Nacher, T. Ochiai, M. Martino, Y. Altshuler, 온라인 금융 거래에 관한 전망 이론, PloS one 9 (10) (2014) e109458.
- [4] Y. Altshuler, W. Pan, A. Pentland, 사회 확산 모델을 사용한 트렌드 예측, arXiv.org, 2011.
- [5] P. M. Krafft, J. Zheng, W. Pan, N. Della Penna, Y. Altshuler, E. Shmueli, J. B. Tenenbaum, A. Pentland, 분산 베이지안 추론으로서 인간 집단지성, arXiv preprint arXiv:1608.01987.
- [6] Y. Altshuler, A. S. Pentland, G. Gordon, 사회적 행동 편향 및 지식 관리 극대화, in: 소셜 컴퓨팅 및 행동 문화 모델링, 예측, Springer, 2015, pp. 258–263.
- [7] Y. Altshuler, A. Pentland, 최적의 성과를 얻기 위한 네트워크 조율 방법과 장치, 미국 특허 8,914,505 (Dec. 16 2014).
URL <https://www.google.com/patents/US8914505>
- [8] W. Pan, Y. Altshuler, A. Pentland, N. Aharony, 네트워크 내 행동 예측 및 수정을 위한 방법 및 장치, 미국 특허 9,098,798 (Aug. 4 2015).
URL <https://www.google.com/patents/US9098798>
- [9] 집단 지성을 얻기 위해 소셜 네트워크 튜닝(MIT Media Lab 웹사이트) (2017).
URL <https://www.media.mit.edu/research/highlights/tuning-social-networks-gain-wisdom-crowd>
- [10] 시장 통찰력: 시장에서의 트위터 영향을 깨닫게 되다 (*파이낸셜 타임즈*) (2017).
URL http://web.media.mit.edu/~yanival/Markets_Insight.htm
- [11] 에코 챔버를 너머 (하버드 비즈니스 리뷰) (2017).
URL <https://hbr.org/2013/11/beyond-the-echo-chamber>
- [12] 예측 분석 재사고 (*FirstMark 의 데이터 주도*) (2017).
URL <http://firstmarkcap.com/insights/rethinking-predictive-analytics/>
- [13] 소셜 미디어가 투자자에게 돈을 벌 수 있게 해주는지 보기 위한 MIT의 1 백만 달러 시험 (2013).
URL <https://tinyurl.com/MIT-1M-USD>
- [14] Endor – “예측 분석을 위한 Google” 발명 (2017).
URL <http://news.mit.edu/2017/endor-inventing-google-predictive-analytics-1220>
- [15] Endor 선도적인 투자자 – *Innovation Endeavors* (2014).
URL <http://www.innovationendeavors.com>

- [16] A. Boehme, Y. Altshuler, 소비자 행동을 예측하기 위해 사회 물리학 활용, in: 네트워크 사이언스 (NetSci), 2017.
- [17] 마스터카드는 5 개의 새로운 스타트업을 스타트 경로 가속 프로그램에 참여하게 합니다 (2016).
URL <https://tinyurl.com/MasterCard-Endor>
- [18] Endor – *Finnovate Fall 2017* (2017).
URL <https://www.youtube.com/watch?v=69rUQloq-qA>
- [19] Endor – *가트너 선정 주목할 만한 기업*(2017).
URL <https://www.gartner.com/doc/3727117>
- [20] Endor – *세계 경제 포럼의 기술 파ioni어 인정*(2017).
URL <http://widgets.weforum.org/techpioneers-2017/>
- [21] *DARPA Network Challenge* (2011).
URL <http://archive.darpa.mil/networkchallenge/>
- [22] 2012 년 mckinsey 수상자 (2012).
URL <https://hbr.org/2013/04/the-2012-mckinsey-award-winners>
- [23] *Google Scholars – Professor Alex “Sandy” Pentland* (2017).
URL <https://scholar.google.com/citations?user=P4nfoKYAAAAAJ&hl=en>
- [24] *Tim O’Reilly: 세계 가장 강력한 7 명의 데이터 과학자* (2017).
URL <http://www.forbes.com/pictures/lmm45emkh/6-alex-sandy-pentland-professor-mit/>
- [25] Y. Altshuler, A. Pentland, A. M. Bruckstein, *검색에서 스웸과 정보망*, Springer, 2017.
- [26] Y. Altshuler, Y. Elovici, A. B. Cremers, N. Aharony, A. Pentland, *소셜 네트워크 내 보안과 프라이버시*, Springer Science & Business Media, 2012.
- [27] H. Shrobe, D. L. Shrier, A. Pentland, *사이버 보안을 위한 새로운 솔루션*, MIT Press, 2018.
- [28] J. Clippinger, D. Bollier, *비트코인부터 버닝맨 그 너머 까지: 신원 요청과 디지털 사회에서 자율성, ID3 and Off The Common Books*, 2014.
- [29] T. Hardjono, D. Shrier, A. Pentland, *TRUST:: 데이터: 아이덴티티 및 데이터 공유를 위한 새로운 프레임워크*, 2016.
- [30] A. Pentland, T. Heibeck, *슬직한 신호: 그것이 우리 세상의 모습을 만드는 방식*, MIT press, 2010.
- [31] A. Pentland, *사회 물리학: 좋은 아이디어가 확산되는 방식- 새로운 과학에서 배우는 교훈*, Penguin, 2014.
- [32] D. Shrier, *금융 기술의 최전선: 미래 커머스 탐험, 블록체인과 디지털 banking에서 예측 시장 그 이상까지*, 출판사: Visionary Future, 2016.
- [33] *Endor.coin 프로토콜 GIT* (2017).
URL <https://github.com/orgs/EndorCoin>
- [34] 아직 블록체인 기반 AI 를 잘 모르면서 그것을 원하는 이유는 (2017).
URL <https://tinyurl.com/blockchain-based-AI>
- [35] Y. Altshuler, N. Aharony, A. Pentland, Y. Elovici, M. Cebrian, 현실 절도: 범죄자가 데이터 과학자 (또는 그 반대)가 될 때, *지능 시스템*, IEEE 26 (6) (2011) 22–30. doi:10.1109/MIS.2011.78.
- [36] M. Ulieru, *블록체인: 이것은 무엇이며 이것이 정말로 세상을 바꿀 수 있을까*, 세계 경제 포럼.
- [37] *망가진 금융 시스템을 고치는 데 기술이 기여하는 방식* (2017).
URL <https://tinyurl.com/technology-fixing-our-financia>
- [38] J. Pieprzyk, T. Hardjono, J. Seberry, *컴퓨터 보안의 펀드멘탈*, Springer 과학 & 비즈니스 미디어, 2013.
- [39] T. Hardjono, L. R. Dondeti, *멀티 캐스트 및 그룹 보안*, Artech House, 2003.
- [40] T. Hardjono, L. R. Dondeti, *무선 LANS 및 MANS 내 보안 (Artech House 컴퓨터 보안)*, Artech House, Inc., 2005.
- [41] J. Seberry, T. Hardjono, *인도네시아와 말레이시아어 암호해독을 향하여*, 1989.
- [42] S. G. Ong, J. Seberry, T. Hardjono, A. D. F. Academy., *중국어 암호해독을 향하여 (Pinyin)*, 1991.

면책고지: 이 백서는 논의용으로만 사용됩니다.
Endor.coin 은 이 백서에서 도달한 결론의 정확성을 보증하지 않습니다.

Copyright ©2018 Endor.coin.