

FABRK

기술 제안서 v1.0

위임 가치 증명(DPoV)

네트워크 상의 평판을 합의 알고리즘에 결합하는 FABRK의 접근 방식은 DPoS와 PoI를 결합한 하이브리드 방식입니다. 지분과 사용자가 FABRK 소셜 네트워킹 플랫폼에서 생성한 가치 모두가 위임 노드 선출에 있어서 사용자 노드의 가치에 영향을 미칩니다. 위임 가치 증명(DPoV) 알고리즘은, 커뮤니티 구성원들이 블록체인의 보안을 유지하고 관리할 위임자를 선출한다는 점에서 위임 지분 증명(DPoS)의 접근 방식과 유사합니다. 하지만, 절대적으로 토큰 보유량이 투표권을 결정하는 위임 지분 증명(DPoS)와는 다르게, 위임 가치 증명(DPoV)는 투표권 v_i 를 각 구성원 i 에게 할당합니다. 이는 즉, 자신들이 창출해 낸 것을 소비하는 사용자들의 투표권이 파트너 간의 네트워크 거리에 따라 변화하며, 거리가 먼 구독자들로부터 요청을 받는 노드들이 중점적으로 투표권을 선점하는 투표권 결정 함수를 의미합니다. 예를 들면, 사진을 게시하여 사용자가 획득하는 가치는 다음에 따라 결정됩니다.

1. 사진을 요청한 사용자의 수 (예. 조회 횟수)
2. 사진을 요청하는 사용자의 가치 (예: 유명인 vs. 스팸 봇)
3. 해당 사용자 간의 네트워크 거리(예: 동일 파벌 혹은 집단에 속해 있는지)

플랫폼 상에 존재하는 모든 읽기 활동(예: 사용자 노드의 로컬 저장소에 있는 정보, 혹은 사용자 노드가 다른 곳에 저장한 데이터에 대한 권한을 보장하는 서명된 일회성의 URI를 획득)은 기본적으로 일정량의 FAB를 소모합니다. 읽기 작업에 대한 기초비용은 사용자들이 데이터 저장 미 서비스에 대한 기초 비용을 회수할 수 있도록 하며, 이 기초 비용은 위임자 노드에 의하여 주기적으로 계산 및 갱신됩니다. DPoV는 데이터 교환 기록을 사용하여, 단순하게 최대의 지분만 가진 사람이 아닌, 본인들의 콘텐츠(예: 접근가능한 데이터)의 구독자 수를 확보하여 플랫폼에 긍정적인 영향을 끼친 이들을 보상합니다.

사용자 i 에 귀속된 콘텐츠에 접근할 때, 그 사용자의 PoV, v_i 가 갱신되고, 그 다음에는 분산된 장부에 대한 업데이트가 진행된다. 분산 장부 내의 노드간 상호작용 네트워크는 사용자가 끝 값 $V(A)$ 을 형성하는 가중치가 부여된 유도된 그래프라 여겨진다. 유도된 엣지 $E(A)$ 내의 e_{ij} 는 사용자 j 가 사용자 i 의 내용에 접근한 횟수에 상응하는 정수 값의 가중치를 갖게 되며, j 가 처음으로 i 를 읽을 때 (예: 업데이트 시에 $e_{ij}=0$) PoV 업데이트는 j 와 i 간의 거리를 최단거리 in 으로 인식한다. 만약 사용자 j 가 사용자 i 의 콘텐츠에 수시로 접근할 경우, e_{ij} 의 가중치가 늘어나며, j 와 i 사이의 비대칭 거리는 줄어들게 된다. 다중 선행 읽기의 경우, $e_{ij} > 0$ 이며, 거리는 엣지 가중치의 역수($1/e_{ij}$)이다.

높은 수준에서 구현시에 중요도의 순서대로 다음 요소들을 고려합니다:

1. 읽기 작업 이전의 PoV, $(V_{t-1})^i$
2. 읽기 작업 중의 독자의 PoV, $(V_{t-1})^j$
3. 장부 그래프
 - a. 장부 그래프 상의 j 에서 i 까지의 거리.

b. 향후 기술 메모에서 자세히 설명될 장부내의 i의 로컬 토폴로지

분산 장부 내의 노드간 상호작용 네트워크는 사용자가 끝 값을 형성하는 가중치가 부여된 유도된 그래프라 여겨진다. 사용자 i의 콘텐츠가 접근되었을 때, 그 사용자의 PoV 값, v_i^i 가 업데이트 되는데, 이 v_i^i 는 극 값 i의 가중치를 의미합니다. 유도된 엣지 E(A)내의 e_{ji} 는 사용자 j가 사용자 i의 내용에 접근한 횟수에 상응하는 정수 값의 가중치를 갖게 된다. 만약 사용자 j가 사용자 i의 콘텐츠에 수시로 접근할 경우, e_{ji} 의 가중치가 늘어나며, j와i사이의 비대칭 거리는 줄어들게 된다. 일반적으로, 이전의 PoV 값과 마찬가지로, 장부 연결 구조는 업데이트 작업의 입력 자료입니다.

$$v_i^i = f(v_{t-1}^i, v_{t-1}^j, E(A_{t-1}))$$

노드 i의 PoV, v_i^i 는 i의 콘텐츠의 내용이 읽힐 때 마다 갱신됩니다. 거래가 검증되면, 엣지 e_{ji} 의 강화를 반영하기 위해 상태가 갱신되며, 이 업데이트 과정은 가장 최근의 가치 증명 상태가 과거 역학을 인코딩 한다는 점에서, Markov 프로세스이며, 잊어버림 계수는 각 업데이트의 영향력을 제어합니다.

$$v_{t-1}^i = v_{t-1}^i + (1 - \alpha)\epsilon$$

ϵ 는 노드의 콘텐츠를 읽는 행위의 기본 가치를 커뮤니티에 반영하는 작은 글로벌 상수입니다. v_i^i 의 업데이트 값은, 독자와 콘텐츠 소유자의 거리가 가까워지면 A_{t-1} 만큼 불이익을 받으며, 읽기 행위가 수시로 있는 경우에, 새로운 업데이트 값들을 기하 급수적으로 감소합니다.

$$\exp(-\lambda / d)$$

λ 는 기하급수적인 감소의 정도를 결정하는 감소 계수이고, d는 j에서 i까지의 거리를 의미한다. j가 i의 콘텐츠를 몇 회 읽었는지는 e_{ji} 에 저장되며, e_{ji} 가 0일 때, j에서 i까지의 최단 경로가 거리 d로 사용된다. 하지만, $e_{ji} > 0$ 일 경우에는, 거리는 $1/e_{ji}$ 로 지속적으로 감소하게 된다.

$$d = \begin{cases} \frac{1}{e_{ji}} & e_{ji} > 0 \\ d_{\Lambda}(j, i) & e_{ji} = 0 \end{cases}$$

대부분의 감소 계수 λ 경우, $\exp(-\lambda e_{ji})$ 는 5번의 읽기 작업 후 0에 가까워진다. 최종적으로, 업데이트 값은 독자의 PoV, v_i^i 와 선형으로 비례하며, 관계식은 다음과 같습니다.

$$v_i^i = \alpha v_{t-1}^i + (1 - \alpha)\exp(-\lambda / d) v_{t-1}^j$$

복합적으로, 이 업데이트 식은 네트워크에서 새로운 노드를 자동으로 생성하거나 노드 클러스터 간의 연속적이 읽기 작업을 통해, 일련의 간단한 공격 경로를 완화합니다. 이 업데이트 값은 i의 콘텐츠가 멀리 위치한 새로운 독자들에게 의해 접근되거나, 이전의 독자들의 특별히 높은 PoV 값을 지닐 때 의미가 있는데, 이는 많은 읽기 작업들이 멀리 있는 새로운 독자들에게 의해 수행되지 않으면 쉽게 달성할 수 없습니다. 또한, 이는 다양한 참가자들을 위한 매력적인 콘텐츠를 만들도록 권장합니다.

집계 투표권은 투표 거래가 검증될 때 계산됩니다. 집계 투표권은 가치 증명, 지분, 그리고 네트워크 토폴로지에 의해 결정됩니다. 명확한 표기를 위하여, 이전의 PoV 값, v_i^i 를 V, 그리고 코인 지분을 S라고 표기하면, 위임 노드 선출전의 사용자의 투표권을 다음과 같이 표기할 수 있습니다.

$$g(V, S) = \ln(\beta(S) + 1) V$$

상기의 식에서 $\ln()$ 는 자연 로그를 의미합니다. 사용자의 투표권은 지분과는 대수적인, 가치증명과는 선형적인 관계로 확장됩니다. 큰 지분의 경우, 자연 로그 값을 이용하기에, 이 큰 지분의 투표권이 극단적으로 줄어들게 됩니다. 이렇게 함으로써, 투표 행사력에 지분의 양이 미치는 영향을 조절할 수 있게 됩니다. 또한, 2019년 4분기에는 FABRK 테스트 네트워크에서 시뮬레이션 된 활동에 대한 이 매개 변수의 다양한 영향에 대한 영향 선언문을 발표할 것입니다.

투표권에서의 사회 공헌의 역할

V는 주어진 노드에 의해 게시된 데이터에 대한 시간의 흐름에 따른 커뮤니티의 총체적인 관심에 대한 통찰력을 제공하지만, 콘텐츠의 소비가, 단순하거나, 순수한 것이 아니라는 사실을 설명하지 못합니다. 네트워크 과학 분야의 기초 연구와 최근 연구에서, 거짓된 루머와 같은, 네트워크의 안정성에 해를 끼칠 수 있는 것으로 일반적으로 인식되는 특정 콘텐츠의 저자와 유포자를 네트워크 상의 실시간 로컬 토폴로지로 식별¹할 수 있음이 밝혀졌습니다. 가치 증명 매개 변수 V 외에 추가적으로, 우리는 네트워크에 대한 노드의 사회적 기여도를 나타내는 네트워크 매개 변수를 소개하며, 이는 네트워크 내의 노드의 로컬 토폴로지와 상호작용에 기반하여 계산됩니다.

FABRK 커뮤니티는 모듈식 빌딩 블록을 광범위하게 제어하여 노드의 중요도나 평판을 결정하는 낮은 레벨의 메커니즘부터, 참된, 거짓된 혹은 바이럴 콘텐츠의 전파자를 식별하는 높은 수준의 메커니즘에 이르기까지의 사회적 공헌도를 수치화 하여 나타냅니다. 다른 커뮤니티들은 투표권에 관한 사회적 요소에 관한 특정한 합의에 도달하기 위하여 장부내에서 포크 할 수 있다.

장부에서 노드의 인지도와 상호 작용을 통해 위임자는, 대리인은 바이럴 콘텐츠가 몇 개의 벡터 노드로 전파되는 것을 추적하는 것부터, 장부 내의 특정 상호작용 패턴이 사회적으로 긍정적인지 부정적인지 분류하는 것까지, 노드의 사회적 특성을 짐작하는 알고리즘을 제공할 수 있습니다. 그런 다음, 커뮤니티는 커뮤니티의 규칙을 이용하여 노드의 사회적인 영향을 긍정적 혹은 부정적으로 수치화 할 수 있습니다. 전반적으로, 네트워크 분석 및 머신 러닝 기술이 활용될 수 있으며, FABRK 재단과 커뮤니티가 개발에 앞장설 것입니다.

네트워크가 사회 공헌의 합의된 정의에 도달하게 되면, 우리는 투표권 계산의 향후 개정에 구조적이고 근본적인 영향을 끼칠 것을 상상할 수 있습니다. 예를 들면, 스칼라의 경우, 스칼라 값을 직접 지분의 영향력에 연관 지을 수 있는데, 이를 통해 커뮤니티가 다음의 방식을 이용하여, 불안정성을 야기하는 네트워크 토폴로지나 방대한 코인 지분을 지닌 사람들의 투표권을 축소할 수 있습니다.

$$g(\tau, V, S) = \beta (u_s - S)^2 (u_\tau - \tau)^5 V$$

u_τ 가 A에 있는 모든 노드의 네트워크 참여도의 평균값인 경우, u_s 는 상대적인 네트워크 지분의 평균값(전체 네트워크에 대한)이며, β 는 지분과 네트워크 점유율에 변동성이 있는 참여자들이 생산적인 활동을 할 수 있도록 장려하는 증폭 계수입니다. 우리는 먼저, 이 표현의 형식을 설명하고 τ 의 계산에 대해 논의합니다.

The majority of nodes will not be affected by the polynomial product $(u_s - S)^2 (u_\tau - \tau)^5$, which evaluates to one except when a node has extreme stake or network participation values. When a node appears in either tails of the stake/participation distributions, $(u_s - S)^2 (u_\tau - \tau)^5$ modifies the proof of value V according to the following principle. If a node's network participation is in either tail of the population distribution, the node's voting power is amplified or diminished accordingly. However, a node's stake relative to the population stake distribution leads to a refinement. When the node's participation is in the low tail of the population but its stake is low, this effect is lenient. When the node's participation is in the high tail yet its stake is high, this effect is less rewarding. This interaction between stake and network participation is controlled by:

대부분의 노드는 특정 노드가 지나친 네트워크 참여도나 지분율을 갖고 있는 것을 평가하는 다항식, $(u_s - S)^2 (u_\tau - \tau)^5$ 의 영향을 받지 않습니다. 한 노드가 지분/참여 분포 중 어느 꼬리값 에라도 나타나면, $(u_s - S)^2 (u_\tau - \tau)^5$ 는 다음의 원리에 따라 가치증명 V를 수정합니다. 노드의 네트워크 참여가 인구 분포의 어느 한 쪽의 꼬리 부분에 있는 경우, 노드의 투표 권한이 알맞게 증폭되거나, 줄어듭니다. 하지만, 인구 지분 분포에 따른 지분 분포는 세분화됩니다. 노드의 참여도가 인구 분포의 낮은 꼬리에 위치하지만, 지분은 낮을 때, 이러한 효과는 관대합니다. 노드의 참여도가 높은 꼬리에 있는데, 지분이 높을 때, 보상이 적습니다. 지분과 네트워크 참여도의 이러한 관계는 다음의 식으로 조절됩니다.

$$\beta = \begin{cases} 1.0, & S > \mu_s + \sigma_s, \tau > \mu_\tau + \sigma_\tau \text{ or } S > \mu_s - \sigma_s, \tau > \mu_\tau - \sigma_\tau \text{ 일 때} \\ \text{...} \end{cases}$$

1 나머지

우리는 소셜 네트워크의 분석 및 소셜 미디어의 광범위한 문헌 모음²³을 통해 노드의 네트워크 값(가치 증명 계산에서 추정되는 콘텐츠의 가치와 반대되는)을 확률적으로 추론할 수 있게 됩니다. 머신 러닝 기술은 이러한 분석을 노드의 주변, 로컬 이웃에 대한 내장된 표현으로 더욱 향상시킵니다.

FABRK 팀은, 게임 이론가, 머신러닝 전문가, 소셜 네트워크 과학자 및 UX 디자이너 등으로 구성된 팀으로써, 현대의 디지털 소셜 네트워크의 콘텐츠와 품질 간의 관계에 대한 혁신적인 연구를 제공하고, 탈중앙화된 시스템의 응용프로그램들을 탐색할 준비가 되었습니다

우리는 투표권 계산에서 사용자의 행동의 가치에 효과적으로 가중치를 부여합니다. 한 구성은 다른 구성보다 더 조작성이 뛰어납니다. 그렇기 때문에, 성급하게 정의하지 않고, 미래에 공개하는 것이 매우 중요합니다. 당사의 구체적인 접근 방식은 추후의 기술 메모에 자세히 설명되어 있습니다.

우리는 현재 비공개 테스트 넷에서, 시뮬레이션을 통해 이러한 매개 변수를 테스트하고 연구하고 있으며, 차후 2019년 4분기에 예정된 백서에서 특정 학습방법들을 발표할 것입니다. FABRK 기여자 커뮤니티는 이러한 요소들과 이들의 거버넌스 영향 범위를 개선하는 데 도움을 줄 것입니다. 추가적인 수정은 FABRK 재단에서 수행하며, 네트워크 업데이트의 형태로 배포됩니다.

노렌(Noren): 연합 학습 요약 및 평가 프로토콜

고품질의 훈련 데이터를 위한 시장을 활성화시키기 위해서, 개발자는 확신을 갖고 값을 지불하고 사용자의 데이터에 기반하여 훈련할 기회를 갖도록 해야 한다. 마찬가지로, 사용자들은 그들의 개인 정보가 적절하게 보호된다는 확신을 가져야 한다. FABRK의 노렌(Noren) 프로토콜은 로컬 교육 모델의 안전한 집계 프로토콜뿐만 아니라 익명의 요약 및 감정 프로세스를 성립한다.

FABRK는 분산된 장부를 위한 신경 익명화 프로토콜에 의해 더욱 강력해진다. 노렌(Noren) 프로토콜은 안전한 익명의 데이터 거래를 가속화하기 위해서 노드 사용자와 장부 서비스 사이의 인터페이스 역할을 한다. 노렌(Noren)의 틀 안에서, FABRK 사용자는 콘텐츠로 수익을 창출하고 광고 기회를 얻기 위해 암호화된 익명의 장부 요약본을 생성할 수 있으며, 이러한 방식으로 개인 정보의 보호를 염두에 둔 데이터 공유 기반이 형성된다. 데이터를 중점에 둔 목표를 가진 FABRK 개발자들은 자신들의 콘텐츠에 대한 공공 AI 감정에 참여하는 사용자들로부터 정보를 얻을 수 있다.

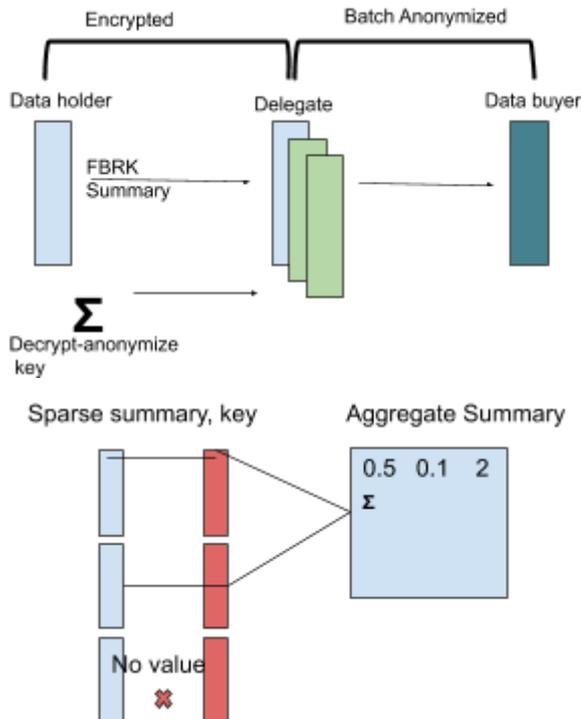
노드의 내부의 기본적인 컨테이너는 외부 서비스에 대한 접근 없이 암호화로 인해 안전하다. 노렌(Noren)을 선택함으로써 인해 컨테이너들은 1) 데이터 거래를 위한 컨테이너 상태에 대한 익명의 통계적 요약을 생성하고, 2) 다양한 상업적 목표를 향한 컨테이너의 활용성을 집착할 수 있다. 또한, FABRK는 제3의 개발자가 이러한 가능성과 대체적인 데이터 프로토콜을 창출할 수 있도록 대체 도구를 만들어내는 것을 권장한다. FABRK의 기술 제안서 V1에서 노렌(Noren)의 감정법과 요약문이 좀 더 자세하게 기술되어 있다.

FABRK 요약문

FABRK의 요약문은 익명의 통계적 컨테이너 요약문입니다. 컨테이너 상태는 통계적으로 고차원의 내장 형태로 인코딩되며, 요약문의 해독된 상태에서 내장 형태 내의 컨테이너의 좌표를 산출합니다. 이 내장형 표현은 응용프로그램들이 활용할 수 있는 관련 정보들을 포함하며, 컨테이너의 콘텐츠를 FABRK 요약문에 인코딩한 것은 오픈 소스 프레임워크인 노렌 네트워크를 통해 완료됩니다. 프레임워크의 배포는, 블랙박스의 형태로 컨테이너 내에서 발생하며, 이는 접근 불가능한 알고리즘 매개 변수를 수반합니다. 계산이 완료가 되면, 암호화된 요약문은 컨테이너 내에 제한적/공개적 권한으로 저장되어, 위임자들, 개발자들 혹은 제3의 서비스 업체들이 접근하며 소비할 수도 있고, 익명화 될 수 있다.

사용자들은 종종 이미지, 비디오, 오디오, 텍스트 파일과 같은 고차원의 복잡한 객체들을 저장합니다. 노렌 프레임워크는 컨테이너 내의 객체들을 인식 가능한 포맷과 적절한 권한 설정을 통해 구성합니다. 인코딩을 한다면, 이러한 객체에 시도됩니다. 노렌은 이후 각각의 객체를, 형식에 따른 머신 러닝 모델 묶음을 통해 인코딩하려고 시도합니다. 최종적으로, 네트워크는 모든 실행가능한 객체들을 단일 벡터 표현으로 출력합니다. 우리는 다중 모달 학습⁴⁵⁶이나 정보 융합에 대한 추가적인 작업을 수행하여, 광범위한 데이터 중심 프로젝트에 근접하도록 이러한 표현이 가능한 한 다양하고 유용하도록 만드는 것을 목표로 합니다.

암호화와 익명화를 분리함으로써, 익명화 및 익명화의 실행 사양은 계산상의 이유로 혹은 검증상의 이유로 위임할 수 있게 됩니다. 대부분의 머신 러닝 어플리케이션에서, 분산형 머신러닝과 연합 학습⁷⁸에 관한 최근의 연구는 머신러닝에 대한 입력으로써, 고차원의 집계 데이터의 견고함을 시사한 바 있습니다. FABRK과 같은 분산된 장부에서는, 집계 및 위임된 방식으로 익명화를 달성합니다.



대부분의 경우에, 프로토콜은, 요약문이 이미 예정된 변환에 의해 익명화 되거나 암호 해독키 스스로가 암호 해독-익명화 절차인 경우에만 적용가능한 암호 해독키를 생성합니다. 우리는 위임된 익명화와, 집계 익명화 기술인, Google의 보안 집계 프로토콜⁹을 활용하는 예시 체계를 제시합니다. 절차 전체에 대한 컨테이너 요약문의 상태는 왼쪽 그림에 자세히 나와있습니다.

계산이 끝난 후에, 임의의 시드를 사용하여 이 요약문의 임의의 위치를 영점으로 만듭니다. 결과적 희소 행렬은 0 아닌 값을 정렬되지 않은 배열로 저장하는 반면, 무작위 시드는 암호 해독 키 또는 위치에서 위치로의 맵 역할을 하여 행렬의 구조를 복구합니다. 위임된 암호화 및 익명화 휴리스틱은 오른쪽 그림에 나타나 있습니다. 암호 해독 키는 일괄적으로 활성화되며, 집계 행렬의 위치에 대한 반복 작업은 그 위치에 non-zero sum 값을 생성하고, 이는 다음 위치로 옮기기 전에 평균화 됩니다. 그런 다음, 집계 요약이 데이터를 요청한 사람에게 전달됩니다.

FABRK 컨테이너

FABRK 장부는 데이터 콘텐츠 평가를 통해 데이터 트랜잭션 절차를 가속화합니다. 사용자와 비즈니스들에게 부여된 인지도 외에도, 작업에 적합한 데이터를 FABRK 데이터 공유점으로부터 쿼리하는 사용자 방송 평가를 통해서 더 많은 정보를 얻을 수 있습니다.

사용자들은 평가 프로세스를 통해 FABRK 활용 등급(UR) 혹은 유사한 타사의 등급제도를 선택할 수 있습니다. UR은 다양한 상업적 목표를 향한 컨테이너의 활용도의 근사치를 찾는 속성값의 집합입니다. 이러한 목표에는 광고를 위한 인구 통계학적 스케치와 머신러닝을 위한 데이터 수집이 포함되며, 작업에 대한 활용도 등급(UR)은 I/O 테스트와 머신러닝 평가를 모은 것에 의해 결정됩니다. I/O 테스트 모음은 컨테이너 내의 객체들의 읽기/접근 상태 뿐만 아니라 파일 형식, 파일 손상, 잘못 정의된 필드 혹은 무효 값(NAN)의 존재 등의 객체들의 기본 속성까지도 확인합니다. 일반적인 머신 러닝 응용 프로그램과 I/O테스트를 통과한 컨테이너 콘텐츠들의 경우, 노렌 프레임워크가 머신러닝 평가를 실시하여 특정 어플리케이션에 대한 작은 데이터 집단의 유용성을 짐작할 수 있습니다.

사용자는 컨테이너에서 노렌 프레임워크가 지원하는 많은 머신러닝 작업에 대한 평가를 수행하고, 방송할 활용도 등급의 일부분을 선택하고, 관심을 갖는 집단과 데이터 트랜잭션에 참여할 수 있습니다. FABRK 요약문을 사용하여 계층화되면, 상기의 트랜잭션은 보호되고 익명화 될 수 있습니다.

이제 일반적인 머신러닝 평가의 예를 들어봅시다. 컨테이너 데이터 집합과 예제의 데이터 집합이 주어졌다는 가정 하에 우리는, 컨테이너 데이터 집합과 예제의 데이터 집합을 생성해낸 소스의 분포, 순서대로 X와 X'가 서로 얼마나 비슷합니까? 와 같은 유용성에 관한 광범위한 여러 질문을 대변합니다. 만약 두 데이터 집합 모두 같은 분포에서 샘플링 되었다면, X'에서 추출된 샘플에 대한 접근 권한을 구매한 개발자는 그들의 데이터 집합(X'=X) 절대적으로 증가시킨 것입니다. 통계적 추론 및 정보 이론을 활용하여 컨테이너의 콘텐츠와 예제의 콘텐츠가 얼마나 유사한지 수치화한 정보를 얻습니다. 이미지에 대한 데이터 및 이미지에 대한 머신러닝에 대해서 평가할 때 가장 중요한 점은, 이미지들이 자연적으로 발생하는 것 인지입니다. 이것은 아래의 기술에 적합한 응용/어플리케이션입니다

두 소스 분포 사이의 편차를 실증적으로 추정하기 위해서, 많은 측정 값이 사용될 수 있습니다 예를 면, Kullback-Leibler 발산¹⁰¹¹ $DKL(P \parallel Q)$ 는 근사값 분포도 Q가 질문의 '참' 분포도 P 대신에 사용될 경우의 정보의 손실을 나타냅니다. 낮은 발산, $DKL(X \parallel X') < \epsilon$, 이 의미 하는 바는 두 데이터 집합에서 샘플링 된 분포가 통계적으로 유사하다는 점 입니다.

그러나 저장된 데이터는 종종 이미지, 사운드 파일 및, 텍스트 파일과 같은 고차원의 객체 형태를 취합니다. 예시의 객체를 사용하여 직접 이 객체간의 측정값은 계산하는 대신, 통계적 추론의 기술을 활용하여 순서대로 Z와 Z'으로 표시되는 X와 X' 사이의 가장 핵심적인 변동 요소를 추출합니다. 실제로, 노렌 네트워크는 미리 훈련된 매개 변수를 이용하여 작업에 부합하는 인코더 F를 선택합니다. F는 항상은 아니지만, 대개 신경 네트워크입니다. 이 인코더는 두 데이터 집합을 잠재적으로 Z와 Z'로 나타내며, 공통 잠재 공간에 매핑합니다.

$$F(X) = Z \quad F(X') = Z'$$

이러한 잠재적인 표현은 주로 작업과 관련된 정보 위주이며, 낮은 차원의 표현입니다. Kullback-Leibler 발산은 누적 분포 함수나 밀도 추정¹⁰을 계산하여 그러한 표현을 추정합니다. 이러한 발산 정보를 사용자에게 제시한 후, 사용자는 이러한 작업에 특정한 검색 결과를 방송할지 선택할 수 있다.

장부의 데이터를 평가할 때 부딪히는 몇 가지 난관들 중 하나는, 컨테이너 내에서 몇가지의 관측 값을 찾는 것입니다. 이 경우에, 소스의 분포 간의 측정의 신뢰도가 떨어집니다. 이와 같은 데이터 제약의 문제를 해결하기 위해서, 수 많은 논문이 작성되었으며, 테스트 네트워크 기간동안 수행될 많은 작업은 견고한 해결책을 찾아낼 수 있을 것입니다. 또다른 난관은 가설 분포 사이의 측정 값이 매우 낮지만, 명시된 작업에 대한 정보가 없는 데이터의 진위를 판별하는 것(복제나 적대적으로 생성된 데이터와는 반대로)입니다. 이를 해결하기 위한 잠재적인 방안은, 보다 구체적인 가정에 기반한 세번째 평가법을 고안하는 것입니다.