

답변 요약서

6조 T-SAN(티끌모아 태산)

1. labeling 관련한 기존의 기술대비 (Amazon SageMaker GroundTruth 등) 핵심 기능이 어떤 것인지 생각해보면 좋을것 같습니다. (이경용 교수님)

Amazon SageMaker GroundTruth 의 경우 의뢰자가 가공되지 않은 데이터셋을 제공하면 해당 레이블링 주제에 대한 모델을 구축합니다. 이후 데이터를 학습된 모델과 레이블링 참여자를 통해 레이블링 작업을 해서 의뢰자에게 제공하는 방식입니다.

1) 공공데이터로 공개

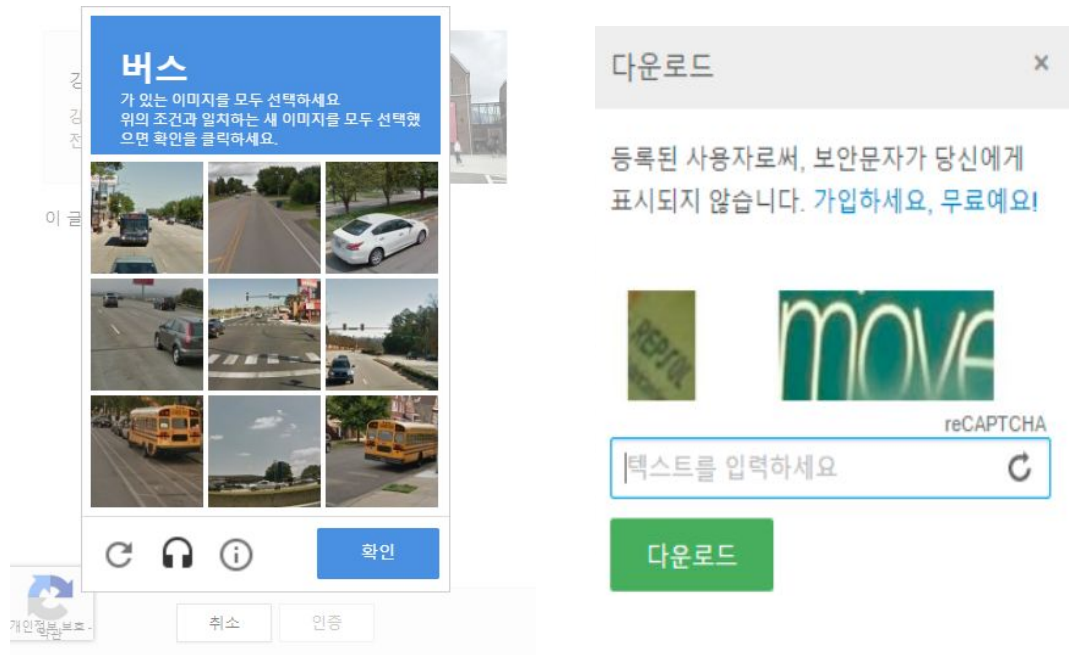
Amazon 플랫폼과 **차별성**을 말씀드리자면, 저희 플랫폼은 의뢰자가 제공하는 데이터셋을 제공하는 것뿐만 아니라, **자체적으로 크롤링 하여 모은 데이터를 레이블링** 할 예정입니다. 크롤링한 데이터는 레이블링 이후 **공공 데이터로 공개**할 예정입니다. (데이터는 모두 저작권 관련한 문제가 없는 데이터로 수집할 예정입니다.)

2) CAPTCHA 인증 서비스 제공

또한, 이미지 선택 레이블링과 텍스트 단답형 레이블링을 이용하여 **CAPTCHA 인증방법¹**을 고안하였습니다. CAPTCHA 인증방식은 두 번의 인증을 받습니다. 첫번째 인증은 사용자가 봇인지 판단하기 위한 것으로 이것을 제대로 입력하면 사람으로 인정합니다. 그리고 두번째 레이블링 받고 싶은 데이터에 대한 입력을 받으며 이는 특별한 인증 절차없이 올바르게 해석한것으로 간주합니다.

만약 기업이 T-SAN의 인증방식을 사용한다면, 이용료를 지불하지 않습니다. 오히려 기업이 데이터 레이블링에 대한 보상을 토큰으로 받을 수 있습니다. 그리고 저희는 T-SAN 뿐만 아니라 외부에서도 레이블링 된 데이터를 받을 수 있게 됩니다.

¹ 출처: 한국전화결제산업협회, <https://m.blog.naver.com/ezpbill/221003296789>

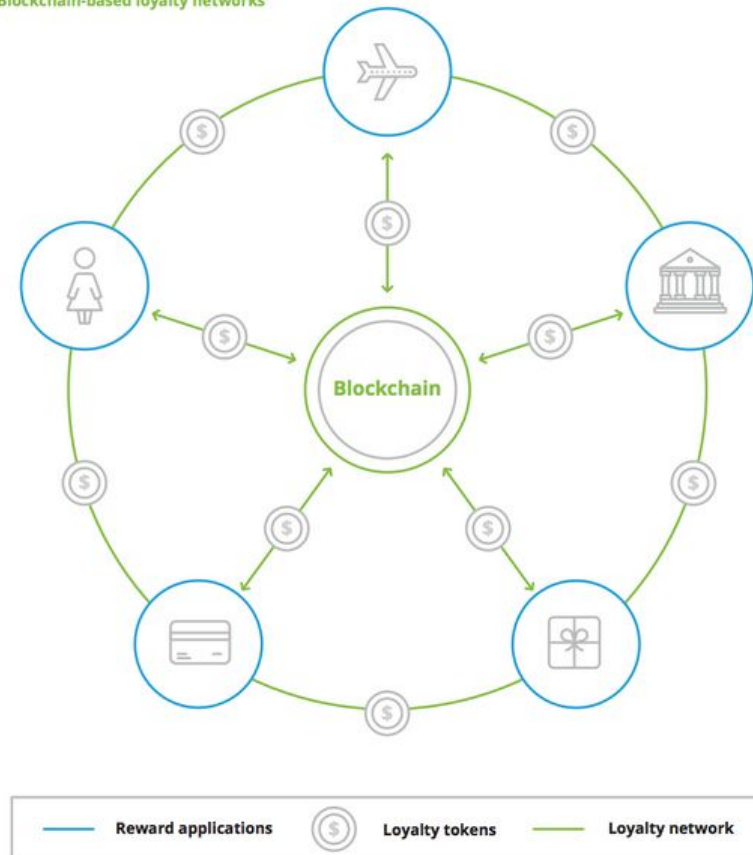


[그림1. CAPTCHA예시 이미지]

2. 블록체인 포인트? 외에 다른 방식의 보상 시스템은 없을지 고민해보면 좋겠습니다.
블록체인이 구체화되어 리워딩으로 사용하면 좋겠습니다.
(김상철 교수님, 주용수 교수님)

- 사용자는 단순히 포인트로 느끼고 사용하지만 저희 T-SAN이 고객의 포인트를 블록체인으로 투명하게 관리할 것입니다. 그래서 포인트를 블록체인으로 관리하는 방법을 채택하게 되었습니다.

Figure 2: Blockchain-based loyalty networks



- 블록체인을 포인트로 관리한다면, 추후 다른 기업들과 제휴를 맺으며 블록체인을 통해 거래가 가능해지고 사용자는 포인트를 이용하여 다른 업체에서 구매하는 것이 가능해집니다. 이를 통해 ***로열티 네트워크 플랫폼** 형성이 가능해집니다.
- ***로열티 네트워크 플랫폼²**
- 로열티 네트워크를 수용하는 플랫폼은 서로 다른 여러 조직 및 로열티 프로그램을 수용할 수 있으며, 특히 그들의 포인트의 교환이라는 측면에서 상호 작용을 촉진할 수 있다. 이 플랫폼은 중개인이나 정보 센터가 필요없이 거래에 대한 합의에 이르는 모든 대리인을 통해 이들을 연결한다.

² 출처: <https://steemit.com/kr/@jayden.yoo/3poond>

- 1번에서 언급하였듯이, 저희가 기존에 개발한 레이블링 UI 를 리캡차와 같은 인증방법으로 사용하는 것도 고안하였습니다. 저희 플랫폼에서 제공하는 인증방식을 이용하면 해당 기업은 사용료를 지불하는 것이 아닌 오히려 저희에게 데이터 레이블링에 대한 일정부분 보상을 받게 됩니다. 그리고 이 보상을 포인트로 지급하여 블록체인을 리워딩으로 사용할 수 있게 됩니다.

3. 실용성이 목적인지, 기술적인 문제 해결이 먼저인지 명확히 하고, 실용성이 목적이라면 시제품까지 구현하고 시연을 할 수 있도록 하면 좋겠습니다. (강승식 교수님)

- 저희는 **실용성 목적을 우선**으로 하여 시제품까지 구현할 예정입니다.
- 실용성 측면에서는, 기존의 레이블링 플랫폼은 데이터 레이블링하는 시간을 줄이고자 더 빠르고 효과적으로 레이블링할 수 있도록
 - 1) 이미지 데이터의 경우: 이미지 캡처, 이미지 선택 방식
 - 2) 텍스트 데이터 경우: 주어진 보기를 선택하는 방식, 단답형 방식으로 레이블링이 가능합니다.
 이 방식을 통해 쉽고 빠르게 레이블링 할 수 있도록 하며 더 많은 레이블링 된 데이터를 얻을 수 있을 것으로 기대됩니다.
- 기존의 사람이 직접 레이블링된 데이터를 채점하는 방식을 머신러닝을 이용하여 인공지능으로 레이블링된 데이터를 검증하는 방식으로 연구 및 개발할 것입니다.

4. 프로젝트 결과의 유효성 검증을 위한 방법을 보다 명확히 설계하는 편이 좋겠음 (대규모의 사용자 참여가 요구될 가능성이 있으므로) (이시윤 교수님)

- 계획서 발표 때, 저희는 사용자가 하루에 참여할 수 있는 프로젝트 개수인 Cycle을 제한하는 방식으로 프로젝트 수와 레이블링 참여자 수의 불균형을 해소하고자 하였습니다. 하지만 그렇게 하면 숙련된 사용자의 레이블링 참여가 제한되므로 프로젝트 결과에 대한 유효성이 떨어질 수 있다는 피드백을 주셨습니다.
- **‘의뢰 금액 * Softmax(신뢰도) = 1인당 보상금액’**
 교수님의 조언을 토대로 신뢰도 기반으로 사용자의 프로젝트 참여 제한(Cycle)을 없앴습니다.
 그리고 보상 방식을 더 구체화해보았습니다. **의뢰 금액 * Softmax(신뢰도)로 1인당 보상 금액을 계산**하여 더 많은 참여자들이 참여할 수 있도록 동기부여하였습니다.
 이를 통해 이전의 대규모의 사용자 참여를 제한할 수 있다는 한계를 극복할 수 있게 되었습니다.

- **데이터에서 100개는 이미 답을 알고 있는 데이터로 구성**

등록된 주제의 데이터에서 100개의 데이터를 이미 답을 알고 있는 데이터로 구성합니다. 이는 신규 회원의 신뢰도 측정에도 큰 도움이 될 것이며, 프로젝트 결과의 유효성을 검증하기 위한 좋은 방법이 될 것이라고 생각합니다.

- **2차 검증 도입**

레이블 검증 단계에서 레이블 확정이 안된 소수의 데이터들이 발생하였을 때, 그 데이터들을 신뢰도가 높은 참여자들이 참여할 수 있도록 긴급 프로젝트를 개설하여 레이블링을 다시 받도록 합니다. 그래서 프로젝트 결과의 유효성을 높이도록 하겠습니다.