블록체인 기술을 활용한 보상형 공동 번역 모델

구본철, 남영우, 김승수, 이건명
충북대학교 소프트웨어학과
e-mail:boncheolb@gmail.com, nywoo94@gmail.com, nablekim94@gmail.com, kmlee@cbnu.ac.kr

A Blockchain Based Reward Model for Co-translation

Bon Cheol Ku, Young Woo Nam, Seung Soo Kim, Keon Myung Lee
Dept of Computer Science, Chung-Buk National University

요 약

최근 딥러닝과 신경망 기술을 바탕으로 한 다양한 번역 시스템이 상용화되고 있다. 그러나 문맥을 고려해야하는 경우나 의료, 법학 등 전문적인 지식을 요구하는 경우, 기존의 번역 시스템으로는 원문 작성자가 의도한 대로 자연스러운 번역 결과물을 얻기 어려운 문제점이 있다. 이로 인해 전문적인 지식을 요구하는 문서는 전문 번역 기관에 의뢰하여 번역이 이루어지고 있으며 소비자의 입장에서는 업체로부터 받은 번역물을 평가 과정 없이 사용할 수밖에 없는 상황이다. 따라서 본 논문에서는 사람이 직접 번역한 결과물을 사용자가 가진 번역 신뢰 점수를 반영하여 평가함으로써 양질의 결과물을 도출하고 번역 활동 기여도에 따라 암호화폐로 보상금을 지급하는 블록체인 기반 보상형 공동 번역 모델을 제안한다. 다수의 번역 평가자들로부터 높은 점수를 받은 우수한 번역 결과물을 도출하고, 번역글 평가에 사용되는 사용자별 신뢰도 점수를 블록체인에 저장하여 조작을 방지함으로써 믿을 수 있는 번역 생태계를 조성할 것으로 기대한다.

1. 서론

최근 인공지능 기술을 활용한 번역 프로그램이 우리의 생활 속에서 널리 사용되고 있다. 이들은 딥러닝(Deep Learning)을 기반으로 개발된 인공신경망 기계번역 기술을 사용하여 특정한 번역자 없이 간편하게 사용할 수 있는 것이 큰 장점이다. 대표적으로는 Google의 'Google Translator'[1], 네이버의 '파파고' 등이 있다. 그러나 문맥을 고려해야하는 경우나 의료, 법학 등 전문적인 지식을 요구하는 경우, 기존의 번역 시스템으로는 사용자의 기대를 충족시키기 어려운 실정이다[2].

본 논문에서는 사람의 집단 지성을 이용한 블록체인 기반 보상형 공동 번역 모델을 제안한다. 사람이 직접 번역한 결과물을 사용자가 가진 번역 신뢰도 점수를 반영하여 평가함으로써 양질의 번역 결과물을 도출하고 번역 활동 기여도에 따라 암호화폐로 보상한다. 이때, 번역글 평가에 영향을 미치는 사용자별 번역 신뢰도 점수를 블록체인[3]에 저장하여 유지 및 관리함으로써 신뢰성을 보장한다. 블록체인의 분산 원장 기술[4]을 활용하여 믿을 수 있는 번역 생태계를 조성하고 사람들의 직접 참여로 양질의 번역 결과물을 얻을 수 있다.

2. 관련 연구

2.1 기계 번역

최근 학술계, 산업계에서 큰 주목을 받고 있는 딥러닝 기술이 기계 번역에 활용되는 다양한 연구가 진행되어 왔다. 지난 수년간, 뉴럴 기계번역과 기타 심층학습을 이용한 기계 번역 모델 등 광범위한 연구가 진행되며 주목할 만한 성과를 이루었고 번역 성능을 높였다[5]. 하지만, 딥러닝을 이용한 기계 번역 모델들은 방대한 모델 파라미터로 인한 학습 및 번역 속도의 문제, 제한된 출력 어휘 사이즈(Output Vocabulary Size)로 인한 번역 성능의 저하 문제 등 여전히 많은 문제점을 갖고 있었다. 이로 인해, 번역은 정확한 결과물을 얻기 위해 아직까지 사람의 개입이 필요한 영역으로 남아있다.

2.2 보상 모델

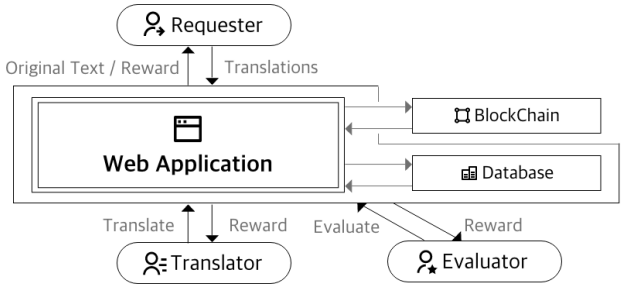
보상 모델은 마르코프 보상 모델[6], 대가성 보상 모델[7] 등 확률, 비즈니스 분야를 포함한 여러 분야에서 사용되고 있다. 대가성 보상 모델의 사례로 사용자 제작 컨텐츠 플랫폼인 유튜브가 있다. 영상 제작자가 올린 컨텐츠의 조회 수에 따라 보상금을 지급하기 때문에 제작자는 더 좋은 영상을 만들고, 사용자는 원하는 영상을 골라볼 수 있는 선순환 구조를 형성할 수 있다. 하지만, 조회 수와 같이 보상금 지급의 기준이 되는 데이터는 특정 기업의 중앙화된 저장소에 저장되므로 조작의 위험으로부터 데이터의 신뢰성을 보장할 수 있는 수단이 필요한 상황이다.

2.3 블록체인

블록체인은 데이터를 집중형 서버에 보관하지 않고, 거래에 참여하는 모든 노드에 저장하는 방식으로 공공거래 장부라고도 불린다. 한번 기록된 사실에 대해 수정이나 삭제가 불가능하다는 특징을 가지므로 데이터 신뢰성 및 안정성을 제공할 수 있는 기술로 대두되고 있다[8]. 금융, 물류, 유통, 제조, 공공 서비스 등 신뢰성과 안정성을 요구하는 다양한 산업 부문에서 접목되어 사용되기 위해 관련 연구가 활발히 진행되고 있다[9-11].

3. 블록체인 기반 보상형 공동 번역 모델

3.1 시스템 구성도

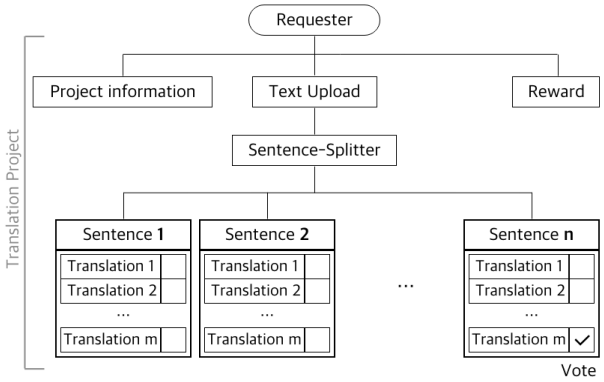


(그림 1) 시스템 내 글 등록, 번역, 평가 절차

본 논문에서 제안하는 블록체인 기술을 활용한 보상형 공동 모델은 그림 1과 같이 구성된다. 시스템은 웹 어플리케이션으로 구현하여 사용자의 접근성을 높이고, 비동기 방식을 채택하여 다수의 사용자가 공동으로 번역하기에 적합한 환경을 구축한다. 사용자는 번역을 요청하기 위해 글을 등록하는 원문 등록자(Requester), 원문을 번역하는 번역자(Translator), 번역한 문장을 평가하는 평가자(Evaluator)로 구성된다. 번역자에게는 번역 기여도에 따른 보상을 제공, 원문등록자에게는 평가된 양질의 번역물을 제공, 평가자에게는 평가 참여에 대한 소량의 보상을 제공한다.

3.2 시스템 활용 절차

제안 모델에서의 전체적인 활동은 그림 2와 같이 일어난다. 원문 등록자는 원문 내용, 마감 기한과 보상금 입력 후 글을 등록한다. 시스템은 입력한 원문 내용을 문장 단위로 파싱하고 등록된 글은 번역 프로젝트(Project)로 지칭하여 관련 정보를 저장한다. 번역자가 각각의 문장에 대해 번역을 작성하면 평가자는 각 문장에 대해 다수의 번역을 확인하고 가장 적절한 번역 하나만 선택하여 투표(Vote)할 수 있다. 마감 기한이 지나면 해당 번역 프로젝트는 마감 상태로 전환되며, 가장 높은 점수를 받은 번역 문장을 조합하여 최종 번역본을 형성한다. 이후 번역자의 번역 기록과 평가자의 평가 기록, 최종 번역 결과를 기준으로 조정된 사용자별 번역 신뢰 점수를 블록체인 네트워크에 저장하여 투명하게 관리한다.



(그림 2) 시스템 내 글 등록, 번역, 평가 절차

3.3 번역 문장 평가

본 절에서는 제안 모델의 핵심 기능 중 하나인 평가 부분을 설명한다. 각 문장별 평가를 종합하여 최종 번역본을 도출하는 과정에 대해 설명한다.

$$SES_{ij} = \sum_{k}^{n} UR_k \tag{1}$$

원문 내용 기준 i번째 문장에 대한 j번째 번역의 평가 점수(SES_{ij})는 식 (1)을 따른다. n는 해당 번역 문장을 평가한 평가자 수를 의미하며, UR_k은 사용자별 번역 신뢰 점수를 말한다. 따라서 번역 문장의 평가 점수는 평가자들의 신뢰 점수들의 누적 합이다.

$$\begin{aligned} S_i &= \{s_{i1}, s_{i2}, \ldots, s_{ij}\} \\ FS_i &= \{s_{ij} | Max(SES_{ij}) \text{ and } s_{ij} \in S_i\} \end{aligned} \tag{2}$$

원문 내용의 각 문장(S_i)은 여러 사람에 의해 번역될 수 있다. 이로 인해 같은 원문 내용에 대해서도 여러 개의 번역 문장(s_{ij})이 존재하게 된다. 원문 등록자가 정한 마감 기한이 지날 경우, 각 문장별로 가장 높은 번역 평가 점수를 받은 문장이 그 문장의 최종 번역 문장(FS_i)이 된다. 식 (2)를 통해 최종 번역 문장들을 모으면 원문에 대한 최종 번역본을 얻을 수 있다.

3.4 보상 지분 계산

최종 번역 문장이 결정되면 번역 및 평가 활동에 따른 보상이 진행된다. 원문 등록자가 앞서 설정한 보상금은 표 1과 같이 8:1:1 비율로 번역자 활동 보상금, 평가자 활동 보상금, 수수료(블록체인 트랜잭션 발생비용 포함)로 이용된다. 보상 부분에서는 최종 번역 문장으로 선정된 문장의 번역자와 평가자에게 기여도에 따른 보상금을 지급한다.

번역 또는 평가한 각 문장이 전체 문장에서 차지하는 비중에 비례하여 보상금을 지급하기 위해 먼저 원문 전체 문장 중에 각 문장이 차지하고 있는 문장별 작업량(W_i)을 구한다. 식 (3)에서 U_i은 Unicode 기준 원문 전체 문장의 Bytes, P_i은 해당 문장의 Bytes를 의미한다.

$$W_i = \frac{P_i}{U_i} \tag{3}$$

번역자 A가 하나의 번역 프로젝트(project)에서 갖는 전체 번역 기여도(TTC_A)는 식 (4)를 통해 구한다. 이는 A가 전체 원문 중에 번역한 문장들의 작업량을 합한 값이다. 여기서 k은 A가 번역한 문장 번호를 의미한다.

$$TTC_A = \sum_{k} W_k, \quad k \in \{k_1, k_2, \ldots, k_m\} \tag{4}$$

평가자 B가 평가한 한 문장에 대해 갖는 평가 기여도(EC_{iB})는 식 (5)과 같다. M_i은 해당 문장을 평가한 사람 수, W_i은 평가한 문장이 전체 문장에서 차지하는 문장 작업량을 의미한다.

$$EC_{iB} = \frac{W_i}{M_i} \tag{5}$$

평가자 B가 하나의 번역 프로젝트에서 갖는 전체 평가 기여도(TEC_B)는 식 (6)을 통해 구한다. 이는 B가 평가한 문장들의 평가 기여도(EC_{kB}) 총합이다. 여기서 k은 B가 평가한 문장 번호를 의미한다.

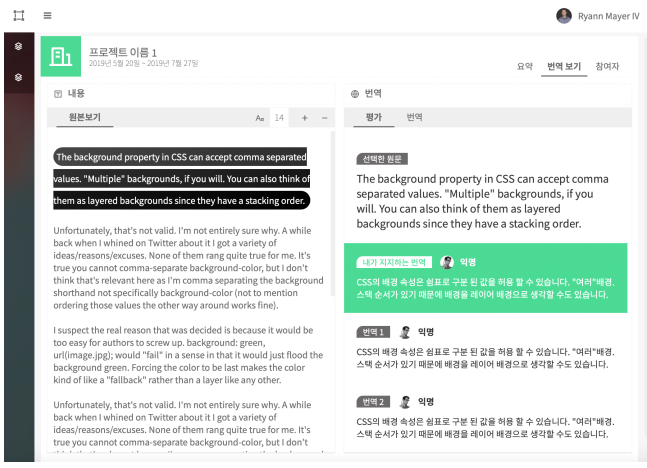
$$TEC_B = \sum_k EC_{kB}, \quad k \in \{k_1, k_2, ..., k_m\} \tag{6}$$

번역자와 평가자들의 기여도 계산이 끝나면 식 (7)과 표 1에 따라 보상금을 분배하고 블록체인에 송금 트랜잭션을 발생시켜 사용자에게 암호화폐로 보상금을 지급한다. 하나의 번역 프로젝트를 기준으로 생각했을 때, TR_A는 번역자 A가 받는 번역 활동에 대한 총 보상금, ER_B는 평가자 B가 받는 평가 활동에 대한 총 보상금, R은 전체 보상금을 의미한다.

$$\begin{aligned} TR_A &= R \times 0.8 \times TTC_A \\ ER_B &= R \times 0.1 \times TEC_B \end{aligned} \tag{7}$$

3.5 사용자 인터페이스

그림 3은 번역이 진행 중인 번역 프로젝트를 볼 수 있는 사용자 인터페이스이다. 좌측 내용에 원문의 특정 문장을 선택하면 그 문장에 대한 번역본들이 우측에 나타난다. 다수의 번역본들 중에 가장 번역이 잘 되었다고 판단하는 번역 문장을 선택하여 투표할 수 있다. 만약 새로운 번역 문장을 등록하고 싶은 경우, '새 번역' 버튼을 클릭하여 번역에 참여할 수 있다.



(그림 3) 번역 프로젝트 화면에서의 사용자 인터페이스

4. 사례 적용

본 논문에서 제안한 블록체인 기반 보상형 공동 번역 모델을 적용하기 위해 'Robert Frost의 Fire and Ice' 시 본문을 활용하였다. 원문은 표 1과 같이 4 문장으로 분류되었고 각 문장이 전체 문장에서 차지하는 바이트 수를 Unicode UTF-8 방식 기준으로 계산했다. 또한, 4명의 번역자(Ta, Tb, Tc, Td)와 3명의 평가자(Ea, Eb, Ec)가 번역 및 평가 활동에 참여하여 표 2와 같은 결과를 얻었다. 평가자들의 번역 신뢰 점수는 각각 8점, 5점, 2점을 가지며 총 보상금은 1000peb[1]라 간주한다.

<표 1> 원문 예시 및 문장별 바이트 수

문장 번호	원문 내용	Bytes	문장별 작업량
1	some say the world will end in fire, some say in ice.	53	53/244
2	From what I've tasted of desire, I hold with those who favor fire.	66	66/244
3	But if it had to perish twice, I think I know enough of hate	60	60/244
4	To know that for destruction ice is also great and would suffice.	65	65/244
합계		244	1

평가자들은 가장 잘 된 번역 문장을 선택하여 자신의 번역 신뢰 점수를 평가에 반영한다. 마감 기한 종료 시, 가장 높은 평가 점수를 받은 문장들을 모아 최종 번역본을 만든다. 이후, 번역자와 평가자에게 보상을 주기 위해 표 1, 2의 데이터와 3절의 식 (1)-식 (7)을 활용하여 기여도에 따른 보상금을 계산한다. 위 사례의 경우 사용자별 암호화폐로 보상받을 금액은 다음과 같다.

$$\begin{cases} TR_{Ta} = 1000 \times 0.8 \times \left(\frac{53}{244} + \frac{60}{244}\right) \fallingdotseq 370\,peb \\ TR_{Tb} = 1000 \times 0.8 \times \frac{65}{244} \fallingdotseq 213\,peb \\ TR_{Tc} = 0\,peb \\ TR_{Td} = 1000 \times 0.8 \times \frac{66}{244} \fallingdotseq 216\,peb \end{cases}$$

$$\begin{cases} ER_{Ea} = 1000 \times 0.1 \times \left\{\left(\frac{53}{244} \times \frac{1}{2}\right) + \left(\frac{66}{244} \times \frac{1}{2}\right) + \left(\frac{60}{244} \times \frac{1}{3}\right)\right\} \fallingdotseq 32\,peb \\ ER_{Eb} = 1000 \times 0.1 \times \left\{\left(\frac{53}{244} \times \frac{1}{2}\right) + \left(\frac{60}{244} \times \frac{1}{3}\right)\right\} \fallingdotseq 19\,peb \\ ER_{Ec} = 1000 \times 0.1 \times \left\{\left(\frac{66}{244} \times \frac{1}{2}\right) + \left(\frac{60}{244} \times \frac{1}{3}\right) + \left(\frac{65}{244}\right)\right\} \fallingdotseq 48\,peb \end{cases}$$

1) https://docs.klaytn.com/klaytn/design/klaytn-native-coin-klay

<표 2> 번역 및 평가 결과

문장번호	번역 문장	번역자	번역 문장 선택	평가점수
1	어떤 사람은 세상이 불로 끝날 거라 말하고, 어떤 이는 얼음이라고 말한다.	Ta	Ea, Eb	13
	세상이 불로 끝장난다는 사람도 있고, 얼음으로 끝장난다는 사람도 있다	Tb	Ec	2
	어떤 사람들은 세상이 불로 끝날 것이라고 말하고, 어떤 사람들은 얼음으로 말을 합니다.	Tc	-	0
2	내가 욕망을 맛본 바로는, 불을 택한 사람들에 공감이 간다.	Ta	Eb	5
	내가 맛본 욕망에 의하면, 나는 불로 끝날 거라 말하는 사람들에 동의한다.	Td	Ea, Ec	10
3	허나 세상이 두 번 망해야 한다면, 나는 증오의 감정도 충분히 안다고 생각하기에	Ta	Ea, Eb, Ec	15
4	파괴에는 얼음도 충분히 위력적이라고 생각한다.	Tb	Ec	2

5. 결론 및 향후 연구

본 논문에서는 기존 번역 프로그램의 부정확한 번역 문제를 해결하기 위해 블록체인 기술을 활용한 보상형 공동 번역 모델을 제안했다. 원문 등록부터 활동 기여도 계산을 통한 보상금 분배까지 전체적인 과정을 살펴보았고, 모델의 핵심 기능인 번역 문장 평가와 보상 지분 계산을 중점으로 설명하였다. 또한, 예시 사례를 통해 번역자와 평가자에게 보상금이 어떻게 배분되는지 살펴보았다. 보상형 모델을 통해 사용자들의 자발적인 참여와 양질의 데이터를 유도할 수 있을 것이며, 번역 평가에 영향을 미치는 데이터를 블록체인 저장함으로써 신뢰성 있는 번역 생태계를 구축할 수 있을 것으로 기대한다. 향후에는 중앙화된 기업을 중심으로 보상이 일어나는 다양한 서비스에 블록체인 기반 보상형 모델을 적용하기 위한 연구를 진행할 계획이다.

감사의 글

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학 사업의 지원으로 수행되었음. (2019-0-01183)

참고문헌

[1] S. Lim and M. Han, “A Case Study on “Google Translate”: The Correlation between Translation Quality and Units of Translation in Machine Translation”, 번역학연구, Vol. 15, No. 1, pp. 177-209, 2014

[2] H. Kim and S. Yoon, "A Study on Limits and Prospects of Machine Translation Through Back Translation: Focusing on The Vegetarian", Folktale and Translation, Vol. 37, No. 0, pp. 13-37, 2019

[3] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System", 2008, Available: https://bitcoin.org/bitcoin.pdf (accessed on Sept. 21, 2019)

[4] H. Youm, “분산원장기술 표준기술”, 한국통신학회지, Vol. 36, No. 7, pp. 15-19, 2019

[5] J. Li and J. Lee, "RA Survey on Neural Machine Translation", 정보과학회지, vol. 33, No. 10, pp. 40-47, 2015

[6] A. Gouberman and M. Siegle, “Markov Reward Models and Markov Decision Processes in Discrete and Continuous Time: Performance Evaluation and Optimization", In: International Autumn School on Rigorous Dependability Analysis Using Model Checking Techniques for Stochastic Systems. Springer Berlin-Heidelberg, p. 156-241, 2012

[7] 이기훈, “UCC 기반의 온라인 동영상 사업자의 수익 및 보상 모델”, 정보통신방송정책, Vol. 23, No. 18, pp. 82-95, 2011

[8] 길성원, “블록체인과 가상화폐 시장의 동향 및 시사점”, 정보통신산업진흥원 이슈리포트, Vol.19, pp.2, 2017

[9] S. Oh and C. Lee, “Block Chain Application Technology to Improve Reliability of Real Estate Market”, 한국전자거래학회지, Vol 22, No. 1, pp.51-64, 2017

[10]Z. Yang, “Blockchain-Based Decentralized Trust Management in Vehicular Networks”, IEEE Internet of Things Journal, Vol. 6, No. 2, pp. 1495-1505, 2019

[11]J. Choi, J. Park, M. Kim and H. Kim, “블록체인 기반 탈중앙화 사물인터넷 플랫폼 연구”, 정보보호학회지, Vol. 27, No. 6, pp. 5-14, 2017