

MultiSAGE 모델과 ESG 지표를 적용한 상품 추천 시스템 개발[☆]

Development of Product Recommendation System Using MultiSAGE Model and ESG Indicators

김 현 우¹ 김 용 준² 유 길 상^{3*}
Hyeon-woo Kim Yong-jun Kim Gil-sang Yoo

요 약

최근 소비자들은 환경, 사회, 지배구조 관련 정보를 확인하고 더 나은 사회적 가치와 환경 친화적인 제품을 선택하려는 경향이 증가되고 있다. 본 논문에서는 GraphSAGE와 GAT를 결합한 모델인 MultiSAGE를 활용하여 최근 소비 트렌드인 가치소비에 맞추어 ESG 지표를 적용한 상품 추천 시스템을 제안하였다. 이를 위하여 한국 ESG 기준원에서 수집한 2022년 1,033개 기업의 ESG 등급 데이터와 실제 N기업의 쇼핑의 상품 데이터를 Heterogeneous Graph 형식의 데이터로 바꾸는 데이터 처리 과정과 MultiSAGE를 적용하여 머신러닝에 적용하고, 특정 상품을 입력하면 그 상품의 친환경 대체제를 추천해주는 추천 시스템을 구현하였다. 구현결과, 소비자들은 기업의 ESG지표를 적용한 제품을 쉽게 비교하여 구매할 수 있고, 이를 통해 사회적 가치와 환경친화적인 제품을 추천하는 시스템에 활용될 것으로 기대한다.

☞ 주제어 : MultiSAGE, 추천 시스템, ESG, GraphSAGE

ABSTRACT

Recently, consumers have shown an increasing tendency to seek information related to environmental, social, and governance (ESG) aspects in order to choose products with higher social value and environmental friendliness. In this paper, we proposes a product recommendation system applying ESG indicators tailored to the recent consumer trend of value-based consumption, utilizing a model called MultiSAGE that combines GraphSAGE and GAT. To achieve this, ESG rating data for 1,033 companies in 2022 collected from the Korea ESG Standard Institute and actual product data from N companies were transformed into a Heterogeneous Graph format through a data processing pipeline. The MultiSAGE model was then applied in machine learning to implement a recommendation system that, given a specific product, suggests eco-friendly alternatives. The implementation results indicate that consumers can easily compare and purchase products with ESG indicators applied, and it is anticipated that this system will be utilized in recommending products with social value and environmental friendliness.

☞ keyword : MultiSAGE, Recommend System, ESG, GraphSAGE

1. 서 론

최근 전 세계적으로 지구 온난화나 환경오염, 이상 기

후 등 여러 환경 문제가 증가하고 있으며, 이러한 문제들을 해결해야 한다는 목소리가 커지고 있다[1]. 이러한 맥락에서 기업들은 지속 가능한 성장을 위해 기업의 사회적 책임을 강조하고, 기업의 지배구조의 투명성을 요구하는 시민들과 투자자들의 관심이 높아지고 있다[2]. ESG (Environmental, Social, Governance) 경영이란 환경 문제, 사회적 책임, 기업의 지배구조의 투명성 등 과거에는 크게 강조되지 않았던 기업의 비재무적인 측면이 중요시되고 있는 경영 방식을 의미한다[3]. 실제 사례로 대한민국의 모 기업은 잘못된 ESG 경영으로 소비자와 투자자들의 신뢰를 상실하고, 매출 하락과 치명적인 제품 인지도 손상, 기업 투자 감소 등의 결과를 초래하였다.

¹ Industrial and Information Systems Engineering, Seoul National University of Science and Technology, Seoul, 01811, Korea

² Department of Applied Statistics, Gachon University, Seongnam, 13120, Korea.

³ College of Informatics, Korea University, Seoul, 02841, Korea.

* Corresponding author (ksyoo@korea.ac.kr)

[Received 25 September 2023, Reviewed 1 October 2023(R2 6 December 2023), Accepted 28 December 2023]

☆ This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(RS-2023-00246191)

2025년을 시작으로 2030년까지 상장 기업들은 ESG 관련 정보를 의무적으로 공시해야 하는 등 ESG에 대한 개인, 기업, 정부의 관심과 이에 따른 관련 연구가 활발하게 이루어지고 있다[4][5][6]. 이에 따라 소비자 중에서는 사회적 가치를 중시하는 가치소비를 실천하려는 사람들이 늘어나고 있다[7]. 이러한 가치관을 가진 소비자들은 제품을 구매할 때 높은 가격을 내더라도 환경친화적인 제품을 선호하며, 또한 제품을 생산한 기업의 투명성과 사회적 책임을 확인하는 경우가 많아지고 있다. 그러나 현재 쇼핑몰과 온라인 플랫폼에서는 상품 추천 시스템이 주로 가격, 인기, 관련성 등을 기반으로 작동되고 있으며, ESG 요소를 전혀 고려하지 않고 있다. 따라서 소비자들이 손쉽게 가치 있는 소비를 실천하고 제품의 환경, 사회, 지배구조 측면에서 정보를 확인할 수 있는 추천 시스템이 요구되고 있다. 이에 따라 본 연구에서는 기존 추천 시스템을 보완하여 ESG 지표를 추가로 고려한 상품 추천 시스템을 연구하고 구현하였다. 이를 통해 소비자들이 구매 결정을 내릴 때 더 나은 사회적 가치와 환경친화적인 제품을 선택할 수 있도록 하였다.

본 논문의 구성은 다음과 같다. 2장에서는 추천 시스템과 그래프에 대한 기본 개념을 간략히 설명하고 그래프를 기반으로 한 추천 모델과 그 중 MultiSAGE의 기초가 되는 GraphSAGE와 GAT에 대해 설명한 후 이 두 모델이 MultiSAGE에 어떻게 적용되었는지에 대해 기술하였다. 3장에서는 제안한 추천 시스템의 모델링 과정을 설명하고, 4장에 실험 결과를 기술하였다. 마지막 5장에서는 연구 결과에 대한 결론과 향후 연구과제에 대하여 제시하였다.

2. 관련 연구

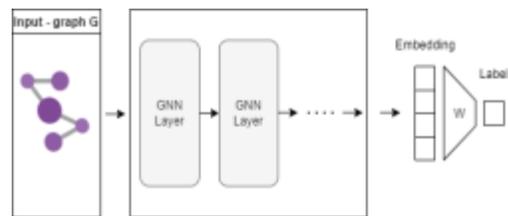
2.1 추천 시스템

추천 시스템이란 사용자에게 상품, 서비스, 콘텐츠 등을 추천해주는 알고리즘으로 오래전부터 사용되었다[8]. 이러한 추천 시스템은 다양한 종류가 존재하는데 전통적으로 사용자가 이전에 평가한 항목과 비슷한 특징을 가진 항목을 추천해주는 콘텐츠 기반 추천 시스템, 사용자의 과거 행동 데이터를 기반으로 추천해주는 협업 필터링, 그리고 이러한 시스템들의 강점을 활용한 하이브리드 추천 시스템 등이 있다. 이외에도 딥러닝의 발전에 따른 딥러닝 기반 추천 시스템, 사용자의 현재 위치나 시간을 고려하는 컨텍스트 기반 추천 시스템 등이 있다.

2.2 그래프

그래프는 노드들과 그 노드를 잇는 엣지들을 모아 구성한 자료구조의 일종이다. 그래프는 entity의 개수에 따라 두 가지 종류로 나눌 수 있는데 1개의 entity로만 이루어진 그래프이면 Homogeneous Graph라고 하고, 2개 이상의 entity로 이루어진 그래프이면 Heterogeneous Graph라고 한다. 전자는 노드들 사이에서 유사성을 정의하기 쉬우므로 협력 필터링 추천 시스템과 같은 일반적 추천 시스템에서 사용된다. 이 그래프는 나와 비슷한 사람이 구매한 상품은 알 수 있지만, 이 상품과 비슷한 다른 상품을 추천받거나 내가 원하는 특징에 맞춘 추천을 받기에는 한계점이 있다. 후자의 경우 노드와 엣지의 다양한 유형, 노드 간의 관계를 통해 여러 정보를 표현할 수 있다. 또한 특정 특징에 맞춘 추천을 제공할 수도 있고, 각 특징에 가중치를 부여하여 여러 상품에 대한 정보를 통합해 종합적인 추천을 제공할 수도 있다.

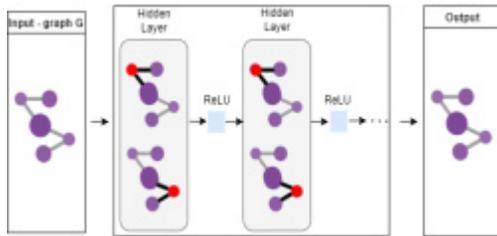
이러한 그래프를 활용한 것이 GNN(Graph Neural Network)[9][10]으로 GNN은 그림 1과 같이 그래프 구조를 유지하면서, 노드 간의 메시지 패싱을 통해 노드 또는 엣지의 표현을 학습하는 신경망의 한 종류이다. GNN은 각 노드를 잘 표현할 수 있는 임베딩 추출을 기대할 수 있다. 즉, 그래프 구조를 활용하여 로스를 최적화시키는 과정이라고 할 수 있다.



(그림 1) GNN 구조
(Figure 1) GNN Structure

GNN은 크게 Spectral한 방법과 Spatial한 방법으로 나눌 수 있는데 전자는 eigen-decomposition을 적용하여 더 많은 정보를 반영하지만 새로운 그래프에 대해서는 일반화할 수 없고 후자는 이웃 집계로 간단히 표현하는 방법이다. 즉, 어떠한 노드의 주변 정보를 가지고 자기 자신을 업데이트하는 방식을 의미한다. 그래프를 활용한 기술이 발전함에 따라 초기에는 Spectral한 방법으로 접근하였지만, 현재에는 대부분 Spatial한 방법론들이 사용되고 있다. 딥러닝 모델에서 흔히 다루는 이미지, 텍스트, 정형 데

이터는 격자 형태로 표현이 가능한 데이터이다. 이는 유클리디언 공간상의 격자 형태로 표현할 수 있다. 하지만, 소셜 네트워크와 분자 데이터 등은 유클리디언 공간으로 표현할 수 없다. 즉, 유클리디언 공간이 아니므로 거리가 중요하지 않으며, 연결 여부와 연결 강도가 중요하다. 또한, 이러한 데이터들은 각 데이터 포인트 간의 독립성이라는 기본 가정이 성립하지 않기 때문에 만약 이미지를 인접한 픽셀끼리 연결된 그래프인 경우, 이 또한 그래프 컨볼루션의 한 종류라 볼 수 있다. 즉, 그림 2와 같이 그래프 컨볼루션은 노드와 연결된 이웃들의 정보를 가중 평균함으로써 컨볼루션 효과를 만들어 낸다. 이러한 개념에서 만들어진 모델이 바로 GCN(Graph Convolutional Network)[11]이다.



(그림 2) GCN 구조
(Figure 2) GCN Structure

딥러닝 모델이 지속적으로 발전함에 따라 추천 시스템도 같이 발전하였는데 이러한 기술을 활용한 추천 시스템이 여러 구조화된 그래프 태스크에서 SOTA(State-Of-The-Art)를 달성하며 인정받고 있다. PinSage[12]는 웹 규모에서 대용량의 데이터를 활용하여 사용자에게 추천을 제공함으로써 GCN이 Large Graph에서 저조한 성능을 보인다는 문제점을 해결한 추천 시스템이다. 실제 이 모델은 Pinterest에서 사용되는 모델로서 노드 간의 확률적인 이동을 모델링한 효율적인 랜덤 워크와 그래프 구조 그리고 노드 특징 정보를 임베딩하는 것을 결합한 모델이다. 이 모델은 다른 딥러닝과 그래프 기반 모델과 비교해서 높은 성능을 보이는 것을 알 수 있다.

2.3 MultiSAGE

GraphSAGE(Graph Sample and Aggregate)[13][14]의 경우 크게 Sample과 Aggregate라는 두 가지 특징이 있다. 먼저 Sample의 경우 Large Graph는 계산량이 많아 Sampling으로 sub-graph를 작성하는데, 이때 노드별로 연결된 노드

중 Uniform 분포 등으로 무작위로 이웃을 선택한다. 선택된 이웃들은 유사하게, Negative sample은 멀어지게 학습하게 되는데 이를 통해 GraphSAGE는 Random Walk를 통한 Sampling으로 이웃을 정의함을 알 수 있다. Aggregate의 경우 Learnable parameter를 이용해 이웃에 대한 정보를 통합하는 과정을 의미한다. 이때 단순 평균, LSTM, Pooling, GAT 같은 방법으로 이웃에 대한 정보를 통합할 수 있다. 순서와 무관하게 얻은 노드별 이웃 정보를 통합해 자신 노드에 저장하는데, 이 때문에 학습되지 않은 새로운 노드를 통해서도 임베딩 결과를 도출할 수 있다.

GraphSAGE의 특징으로는 크게 3가지가 있다. 첫 번째, 노드 자신의 feature와 자신의 이웃 정보가 들어오면 학습되지 않은 노드도 임베딩이 가능하다. 두 번째, 내 노드의 거리에 따라 연결된 노드 중 일부 노드만 이웃으로 정의하면 되기 때문에 전체 degree 정보를 볼 필요가 없다. 세 번째로 자신의 주변을 통해 그래프의 구조적 정보 없이 임베딩을 하기 때문에 샘플의 분산 효과를 통해 전체 그래프를 보는 GCN과 유사한 효과를 얻을 수 있다.

GAT(Graph Attention Network)[15]의 경우 크게 Attention과 Aggregate라는 두 가지 특징이 있다. 먼저 Attention의 경우 Coefficient matrix를 활용해 이웃 별 Attention을 계산하고 FC Layer에서 LeakyReLU, Masking 등을 이용해 Score를 계산한다. 이때 Attention Score는 노드 데이터 간 상대적 중요성을 의미한다. 즉 Attention Score로 정확한 이웃 별 가중치 부여가 가능하다. Aggregate의 경우 기존 모델의 Aggregate 방식이 그래프 구조적 정보 기반 유사도만 반영하는 것에 비해 GAT Layer를 사용해 구조적 정보와 노드 안의 데이터도 반영한 노드 간 Attention Score를 계산할 수 있게 된다. 즉 GAT Layer 사용 시 더 정확한 이웃 별 가중치를 부여할 수 있게 된다.

GAT의 특징으로는 크게 3가지가 있는데 먼저 Aggregate 과정에서 이웃 별 가중치를 구할 때 GAT Layer를 사용하면 Trainable attention weight를 구할 수 있다. 그리고 어떤 이웃이 나랑 가장 유사한지, 어떤 이웃이 나에게 유익한 정보를 주는지 판단 가능해진다. 또한 GAT Layer는 Aggregator 역할을 수행할 수 있으므로 SAGE 구조에 적용할 수 있는데, 이는 GAT가 Heterogeneous Graph를 잘 분석할 수 있는 알고리즘임을 보여준다.

MultiSAGE(Multi Sample and Aggregate)[16]는 GraphSAGE와 GAT를 모두 합한 알고리즘으로 전자는 Large Scale Graph에서 좋은 성능을 발휘하고, 후자는 Heterogeneous Graph를 잘 분석할 수 있는 모델이기 때문에 본 연구의

학습 모델로 MultiSAGE를 적용하였다. MultiSAGE의 특징으로는 크게 3가지가 있는데 먼저 상품 - 특징 - 상품 구조에서 상품별로 다양한 특징들을 연결함으로써 상품별로 다양한 연결 relation을 설정할 수 있다. 이를 통해 다양한 특징들을 이용한 Multi-relation을 표현할 수 있다. 다음 특징으로 노드별로 Positive/Negative Node를 정의해 Max Margin Ranking Loss를 활용해 Query 노드와 Positive 노드는 유사하도록, Query 노드와 Negative 노드는 멀어지도록 학습한다. 마지막으로 MultiSAGE는 GraphSAGE 처럼 Aggregate된 통합 이웃 노드와 자신을 통해 최종 노드를 표현하는데, 통합 이웃 노드는 연결된 특징 노드들의 가중평균으로 임베딩하며 이를 Contextual Embedding이라 한다. 기존 연구에서 Contextual Embedding을 만드는 다양한 방법 중 GAT Layer를 쓴 방식이 가장 우수한 성능을 보여 주었기 때문에[17] 본 연구에서는 GraphSAGE와 GAT를 모두 적용하였다.

3. 제안한 추천 시스템

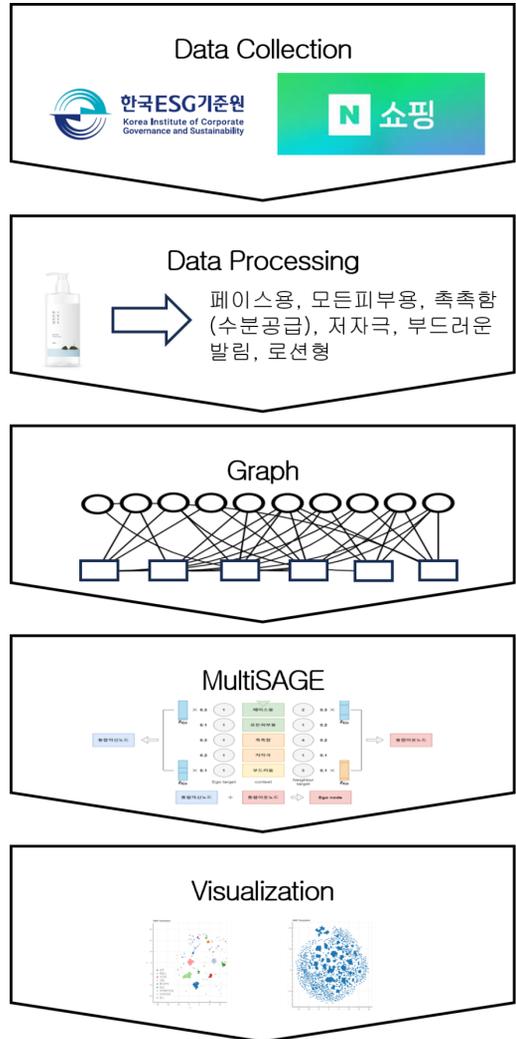
그림 3은 데이터 수집부터 MultiSAGE 기반 모델링 후 시각화까지의 과정을 보여주고 있다. 먼저 ESG Score 데이터와 상품 데이터를 수집 후 데이터를 각각 다른 방식으로 처리해 MultiSAGE 모델이 잘 학습할 수 있도록 데이터 형태를 변형한다. 마지막으로 MultiSAGE 기반 모델링을 진행해 추천 시스템을 완성한다.

3.1 데이터 수집

ESG Score는 한국 ESG 기준원에서 평가한 2022년 1,033개의 기업 등급 데이터를 활용하였다. 이 데이터는 S등급부터 D등급까지 기업의 환경, 사회, 지배구조 요소를 평가한다.

평가 기준으로 환경 요소는 리더십과 거버넌스, 위험 관리, 운영 및 성과, 이해관계자 소통 등을 사용하고, 사회 요소는 리더십과 거버넌스, 노동 관행, 직장 내 안전보건, 인권 등을, 지배구조 요소는 이사회 리더십, 주주권 보호, 감사, 이해관계자 소통 등을 사용한다. 상품 데이터는 N 기업의 쇼핑 화장품/미용 카테고리 ‘가격 비교’ 탭의 총 221,700개 중에서 119,536개를 수집하였다. 수집 시 상품 고유 ID가 같은 제품이거나 상품명과 제조사가 동시에 같은 제품, 상품명과 브랜드가 동시에 같은 제품의 경우 같은 제품으로 간주하고 제거하였다. ‘전체’ 탭의 경우 36,027,459개의 상품이 있었지만, 쇼핑물별로 중복 상

품이 너무 많아 ‘가격 비교’ 탭에 있는 상품 데이터만 활용하였다.



(그림 3) 제안한 전체 시스템 흐름도
(Figure 3) Proposed Full System Flowchart

3.2 데이터 처리

기업별 ESG 등급 데이터의 경우 각 등급을 기준으로 S등급부터 D등급까지의 총 7개 등급을 10점 만점을 기준으로 S등급이면 10점, 그 밑에 등급인 A+은 8.5점을 가지는 식으로 7등분하여 설정한 점수 환산표에 따라 ESG 등급을 점수로 환산한 후 상품 생산 기업이 ESG 실천 기업

에 해당하면 등급별 가중치와 등급별 점수를 곱하여 최종 점수를 계산하는 방식으로 상품별 ESG Score를 계산한다. 이때 등급별 가중치는 기본값을 환경 50%, 사회 30%, 지배구조 20%로 설정하였는데 해당 가중치는 사용자가 ESG 요소들을 중요하게 생각하는 정도에 따라 조정할 수 있도록 설정했다. 또한 친환경 제품일 경우 생산 기업의 ESG Score에 환경 부분 최고 점수인 5점을 가산하였다. 가령 A 기업이 환경 등급 A, 사회 등급 A+, 지배구조 등급 A를 받았다면, A 기업의 ESG Score는 각 등급별 점수와 가중치를 곱한(0.5*7 + 0.3*8.5 + 0.2*7) 값인 7.45점이 된다. 만일 A 기업이 생산한 어떤 제품이 친환경 제품이면 환경 부분 점수에 5점을 가산하여 이 제품의 ESG Score는 12.45점이 된다.

상품 데이터의 경우 리스트 형태로 저장되어 있는 상품별 특징들을 추출해 이를 bool 형태로 변형해서 데이터 형식을 Tensor 형태로 전처리 하는 과정을 거친다. 상품 노드에는 해당 상품별로 가지는 특징들과 느낌들을 저장하고, 특징 노드에는 특징별로 지정한 ID를 저장한다. 옛지에는 상품과 그 상품이 가지는 특징을 연결하고 연결된 상품의 상품별 평점을 저장한다. 이때 생성된 Heterogeneous Graph는 ‘상품’, ‘특징’이라는 두 개의 노드 타입에 ‘Define’, ‘Define-by’라는 두 개의 옛지 타입으로 이루어진 그래프이다. 이 그래프에는 상품 노드 119,536개, 특징 노드 8,540개 총 128,076개의 노드가 있으며, ‘Define’과 ‘Define-by’ 모두 874,830개씩 총 1,749,660개의 옛지를 가진다.

3.3 MultiSAGE 기반 모델링

본 연구에서는 Python 3.11을 이용하여 MultiSAGE 기반의 ESG 지표를 적용한 상품 추천 시스템을 구현하고자 했다. 그림 4는 MultiSAGE기반의 학습 과정을 보여주고 있다. 상품의 특징 데이터를 가진 상품 노드를 추가하면 MultiSAGE는 아래와 같이 모델링을 진행한다.

첫 번째로 각종 노드를 정의한다. 각 Query 노드별로 Positive/Negative 노드와 이에 대응되는 Context node, 이웃 노드(Neighbor target)를 정의한다.

두 번째로 노드별로 Contextual Embedding을 진행한다. Target 노드에 Contextual Embedding을 진행한 후 Ego target과 Neighbor target에 모두 Contextual 정보를 반영한다.

세 번째로 노드별 Attention Score를 계산한다. Contextual Embedding에서 나온 결과로 Attention Score를 계산하고 GAT Layer를 이용해 이웃들의 정보를 통합한다.

네 번째로 최종 노드의 Loss를 계산한다. GAT Layer를 이용해 통합자신노드와 통합이웃노드를 계산한 후 최종 노드를 계산해 Loss를 산출한다.



(그림 4) MultiSAGE 학습 과정
(Figure 4) Visualization of MultiSAGE learning process

4. 실험 결과

4.1 하이퍼파라미터 조정

표 1은 MultiSAGE 모델을 이용한 학습에서 사용한 최적의 하이퍼파라미터 값을 나타낸 것이다. 표 1의 하이퍼파라미터 값은 선행 연구[18]가 사용한 PinSAGE 알고리즘의 하이퍼파라미터의 값을 참고하였다.

Random Walk Length는 상품 - 특징 - 상품 그래프 구조에서 Target 노드가 상품이기에 2로 설정한다.

Random Walk Restart Probability는 연결된 이웃 검색 시 특정 특징에 치우치지 않도록 하기 위해 0.5의 확률로 epoch 진행 시 재실행하도록 설정한다. Number of Random Walk는 10으로 설정해 총 10가지의 상품 - 특징 - 상품의 sequence들을 임베딩에 활용했으며, 모든 이웃에 대한 정보를 저장할 수 없으므로 이 중 5개의 High order 이웃 정보만 활용했다.

(표 1) 학습에 설정한 하이퍼파라미터 값
(Table 1) Hyperparameter values set for learning

변수명	값
Random Walk Length	2
Random Walk Restart probability	0.5
Number of Random Walk	10
Number of Neighbors	5
Number of Layers	2
Number of GAT Layers Headers	3
Hidden Dimension	512
Batch Size	256
Epoch	200
Batches per Epoch	50
Learning rate	1e-4

Batch Size를 256으로 설정해 16X16 행렬로 Attention Score를 계산하도록 설정한다.

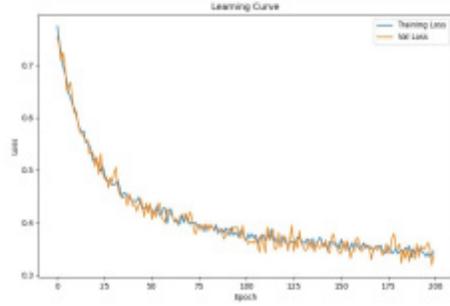
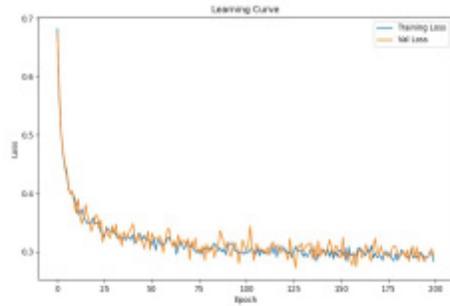
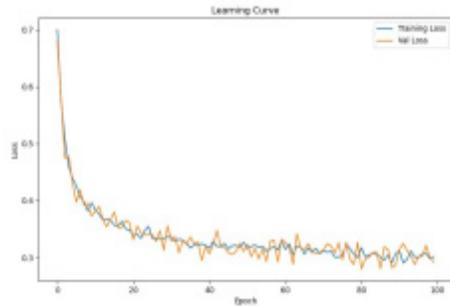
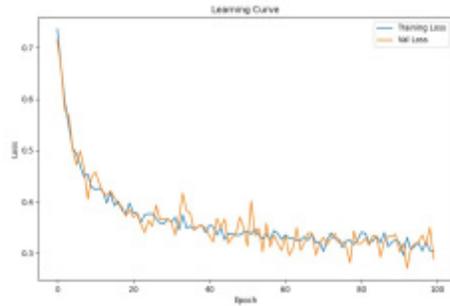
이 실험에서는 최적의 하이퍼파라미터를 찾기 위해 학습률이 1e-4일 때와 1e-5일 때, Batches per epoch가 25, 50, 100일 때의 Learning Curve를 비교하였다. 모든 실험은 MultiSAGE 모델을 이용해 200 에포크 동안 진행하였다.

그림 5는 에포크 당 배치 크기와 학습률에 따른 Learning Curve를 나타낸 것이다. 첫 번째 결과는 에포크 당 배치 크기가 25일 때, 두 번째 결과는 50일 때의 Learning Curve이며, 세 번째 결과는 학습률이 1e-4, 네 번째 결과는 1e-5일 때의 Learning Curve이다.

에포크 당 배치 크기가 25일 때 Learning Curve를 관찰했을 때, 에포크가 0에서 20일 때 학습 속도가 빠르게 상승하였으나, 20 이후부터 Val Loss가 불안정한 양상을 보여주고 있다. 이는 에포크 당 배치 크기가 작을수록 노이즈에 민감하기 때문이다. 반면 에포크 당 배치 크기가 50일 때 Learning Curve는 노이즈에 비교적 덜 민감하기 때문에 이전보다 안정적으로 수렴하는 양상을 보여주고 있다. 이는 학습이 안정적으로 이루어지고 있음을 의미한다.

학습률이 1e-4일 때 Learning Curve는 에포크가 100이 넘어가는 시점부터 Loss가 0.3으로 수렴하며 안정적인 학습이 진행되고 있음을 보여주고 있다. 반면 학습률이 1e-5일 때 Learning Curve는 에포크가 200이 넘었음에도 Loss가 0.3보다 크며 수렴하는 양상을 보이지 않는다. 즉, 모델이 데이터의 패턴을 빠르게 학습하지 못하고, 학습과정이 지연되고 있음을 의미한다.

위와 같이 데이터 세트의 구조와 하이퍼파라미터에 따라 모델 성능 차이를 알아보는 실험 결과, 1e-4의 학습률과 에포크 당 50의 배치 크기가 가장 안정적인 학습 속도와 안정적인 수렴을 제공하여 제일 효과적인 하이퍼파라



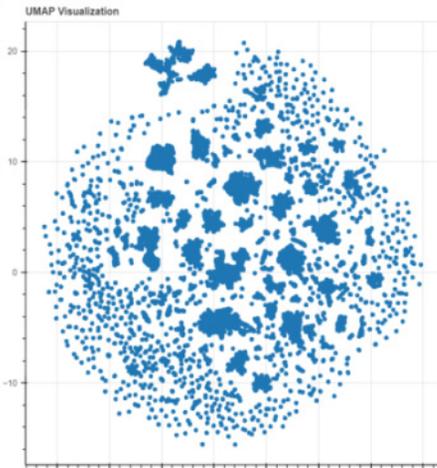
(그림 5) 에포크 당 배치 크기와 학습률에 따른 Learning Curve

(Figure 5) Learning Curve by Batches per Epoch and Learning Rate

미터로 확인되었다. heterogeneous graph 기반 모델은 다른 종류 노드들이 그래프 상에서 상호 작용을 학습하는 데 특화되어 있는데 하이퍼파라미터가 변화하면서 Learning Curve가 변화하고 있기 때문에 이번 실험은 heterogeneous graph 기반 추천 시스템에서 그래프의 구조 설정과 최적의 하이퍼파라미터 조정이 매우 중요함을 잘 보여주고 있다.

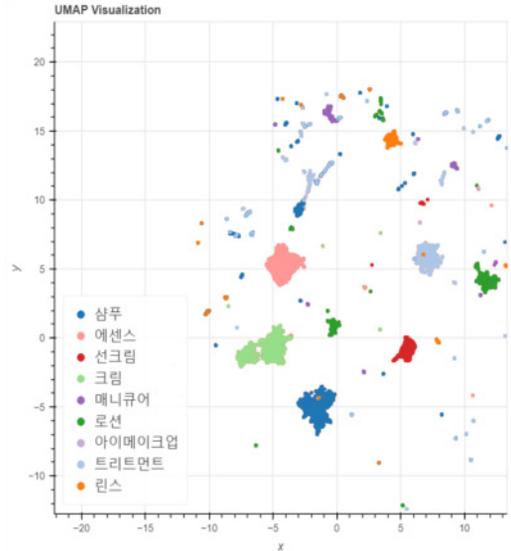
4.2 시각화 분석

그림 6은 각 상품별로 가지는 각각의 특징들의 위치들을 모두 2차원에 시각화한 것이며, 그림 7은 8,540개의 전체 특징 중에서 가장 많은 제품을 가지고 있는 상위 9개의 특징을 특징별로 묶어서 특징들이 어떻게 분포되어 있는지 좌표평면에 나타낸 것이다.



(그림 6) 전체 특징에 대한 그래프 시각화
(Figure 6) Graph visualization for full features

이 연구에서는 상품별 실험 결과를 시각화하기 위해 UMAP(Uniform Manifold Approximation and Projection)[19]이라는 모델을 사용했다. UMAP은 데이터 시각화 및 차원 축소 기술 중 하나로, 리만 기하학과 위상수학에 기반해 고차원 데이터를 저차원 공간으로 매핑하는 데 사용된다. 특히 데이터의 복잡한 구조를 보다 효과적으로 보존하면서 데이터를 저차원으로 투영하여 시각적으로 해석하기 쉽게 만들기 때문에 클러스터링, 패턴 인식 및 데이터 탐색을 위한 데이터 시각화 분석에 많이 활용되고 있다. 시각화된 결과를 보면 클러스터링 된 형태로 특정

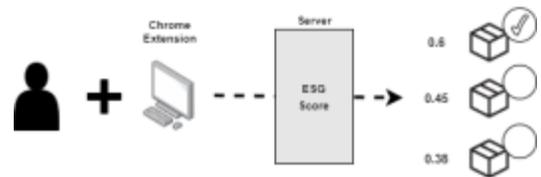


(그림 7) 9개 특징에 대한 그래프 시각화
(Figure 7) Graph visualization for 9 features

지점에 형태로 특정 지점에 점들이 모여 있는 것을 볼 수 있는데, 약 8,000개의 특징 중 10개만 뽑아서 좀 더 자세히 시각화해 보면 같은 특징을 가지는 제품이 특정 지점에 모여 있는 것을 확인할 수 있다.

4.3 ESG를 적용한 추천 시스템 구현

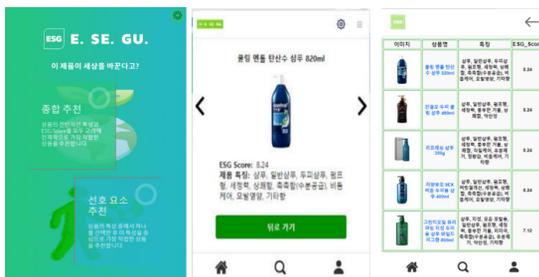
본 연구에서 제안한 ESG지표를 적용한 추천 시스템은 N 포털 기업의 쇼핑 화면에서 크롬 확장 프로그램을 적용하고 모델의 성능을 확인하였다.



(그림 8) 제안한 추천 시스템 구현 및 동작 과정
(Figure 8) The Implementation and Operation Process of the Proposed Recommendation System

제안한 시스템은 그림 8과 같이 크롬 확장 프로그램(Chrome extension)으로 구현되었다. 사용자가 확장 프로그램을 통해 서버에 요청을 보내면 해당 제품의 제조사

와 친환경 제품 여부를 확인 후 DB에 있는 그래프와 모델, 학습 후 나온 노드별 가중치 데이터 등을 이용해 해당 상품의 ESG Score를 계산해 추천 상품 100개 중 ESG Score가 높은 5개의 상품이 추천된다. 여기서 등급별 가중치와 등급별 점수를 다르게 하여 ESG Score를 계산하게 된다면, 상품의 ESG Score들이 바뀌게 되어 그와 연관된 추천 상품들이 점수에 따라 정렬할 때 바뀌게 되므로 점수 설정에 따라 시스템이 민감하게 반응하는 것을 알 수 있었다.



(그림 9) 크롬 확장 프로그램으로 구현된 결과

(Figure 9) The results implemented as a Chrome extension

추천에는 그림 9와 같이 두 가지 추천 방식을 두어 사용자가 맞춤형 추천을 받을 수 있도록 하였다. 첫 번째 추천은 상품과 연결된 모든 특징을 고려한 추천으로 해당 상품과 연결된 모든 특징을 고려해 전체적으로 가장 유사한 상품을 추천한다. 두 번째 추천은 상품과 연결된 특징 중에서 사용자가 선택한 특징을 중점적으로 고려한 추천으로 사용자가 원하는 특징을 하나 선택하면 그 특징에 맞춘 새로운 Contextual Embedding을 구해서 해당 특징을 중심으로 한 추천을 제공한다.

구현 결과, 쇼핑 데이터 세트에서 가져온 특정 상품으로 모델을 넣었을 때 친환경적인 대체재를 우선하여 추천해주는 것을 확인하였다.

5. 결론 및 향후 연구과제

본 논문에서는 N 포털 기업의 쇼핑 데이터 세트를 활용하여 ESG 지표와 MultiSAGE 모델을 적용하고 ESG를 적용한 추천 결과와 성능을 연구하였다. 제한한 모델은 환경 보호와 사회적 가치에 대한 고객들의 기대에 부응하며, 비즈니스의 장기적인 성공을 위한 기반을 마련할 수 있을 것으로 기대할 수 있다.

한편, 일반적인 친환경적인 특징을 가진 제품보다 ESG 기업 등급이 좋은 제품의 친환경 점수가 높은 것을 보았는데 이를 보완하기 위하여 제품의 친환경적인 특징들을 좀 더 세분화하거나 한국ESG기준원에서 조사한 ESG 기업 등급 외에 다른 기업들을 평가할 수 있는 요소를 활용한다면 더 좋은 모델 성능이 보여질 것으로 기대할 수 있다.

이번 연구에서는 상품 데이터 중 일부 이커머스의 일부 카테고리 데이터만 사용했으며, 기업별 ESG 등급 데이터도 상장사에 한해서만 수집할 수 있었다. 때문에 향후 연구에서는 사용자의 편의에 맞게 28개의 쇼핑 카테고리 확대하고 향후 다양한 이커머스로 확대하여 ESG 추천 기능을 제공할 예정이며, 보다 다양한 제조사들의 상품을 추천해 주기 위해 기업의 ESG 등급을 예측할 수 있는 알고리즘을 연구할 예정이다.

참고문헌(Reference)

- [1] Jho Hunkoog and Lee Bongwoo, "Analysis of News Articles about Climate Change Based on Text Mining for the Last Two Decades," Journal of Energy and Climate Change Education, Vol.11 No.2, 153-163, 2021. <http://doi.org/10.22368/ksecce.2021.11.2.153>
- [2] Hyun kyung Yoo, Sung ho Lee and Jung min Nam, "The Effect of Consumer s Demand for ESG Management on ESG Management Support," Entrepreneurship & ESG Research, Vol.1 No.2, 1-24, 2021. <https://m.riss.kr/link?id=A108039867>
- [3] S. Pasch and D. Ehnes, "NLP for Responsible Finance: Fine-Tuning Transformer-Based Models for ESG," 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, pp. 3532-3536, 2022. <http://doi.org/10.1109/BigData55660.2022.10020755>
- [4] D. Han and Y. Zhao, "Research and Countermeasures on ESG Information Disclosure of Listed Enterprises in China Under the Background of Internet Technology," 2021 2nd International Conference on Computer Science and Management Technology (ICCSMT), Shanghai, China, pp. 550-553, 2021. <http://doi.org/10.1109/ICCSMT54525.2021.00108>

- [5] H. Seo, D. H. Jo and Z. Pan, "Big Data Analysis of 'ESG' News Using Topic Modeling," in Proc. of 2022 IEEE/ACIS 7th International Conference on Big Data, Cloud Computing, and Data Science (BCD), Danang, Vietnam, pp. 183-187, 2022.
<http://doi.org/10.1109/BCD54882.2022.9900604>.
- [6] A. Al-Jalahma, H. Al-Fadhel, M. Al-Muhanadi and N. Al-Zaimoor, "Environmental, Social, and Governance (ESG) disclosure and firm performance: Evidence from GCC Banking sector," in Proc. of 2020 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, pp. 54-58, 2020.
<http://doi.org/10.1109/DASA51403.2020.9317210>.
- [7] T. R. Teor, I. A. Ilyina and V. V. Kulibanova, "The Influence of ESG-concept on the Reputation of High-technology Enterprises," 2022 Communication Strategies in Digital Society Seminar (ComSDS), Saint Petersburg, Russian Federation, pp. 184-189, 2022.
<http://doi.org/10.1109/ComSDS55328.2022.9769074>.
- [8] Ahmed Hamdi Abdurhman, Jihwan Lee, Donghyun Kim, ByeongSeok Yu. "Deep Learning Architecture for Choice-based Recommendation System: A Case Study of Flight Search Engine." Journal of Korea Multimedia Society, 26(8), 1027-1041, 2023.
<http://doi.org/10.9717/knms.2023.26.8.1027>
- [9] T. -D. Nguyen, T. Le-Cong, T. H. Nguyen, X. -B. D. Le and Q. -T. Huynh, "Toward the Analysis of Graph Neural Networks," 2022 IEEE/ACM 44th International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER), Pittsburgh, PA, USA, pp. 116-120, 2022.
<http://doi.org/10.1145/3510455.3512780>.
- [10] Y. -C. Lin and V. Prasanna, "HyScale-GNN: A Scalable Hybrid GNN Training System on Single-Node Heterogeneous Architecture," 2023 IEEE International Parallel and Distributed Processing Symposium (IPDPS), St. Petersburg, FL, USA, pp. 557-567, 2023.
<http://doi.org/10.1109/IPDPS54959.2023.00062>.
- [11] A. Qin, C. Liu, Z. Shang and J. Tian, "Spectral-Spatial Graph Convolutional Networks for Semel-Supervised Hyperspectral Image Classification," 2018 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), Chengdu, China, pp. 89-94, 2018. <http://doi.org/10.1109/ICWAPR.2018.8521407>.
- [12] L. Ciabini, A. Lucchetti and A. Rozza, "High-Order Subgraph Convolution Networks with Auxiliary Self-Supervised Task for Recommendation," 2022 International Conference on Algorithms, Data Mining, and Information Technology (ADMIT), Xi'an, China, pp. 104-109, 2022.
<http://doi.org/10.1109/ADMIT57209.2022.00025>.
- [13] S. N. Motevallian and S. M. Hossein Hasheminejad, "Using Trust Statements and Ratings by GraphSAGE to Alleviate Cold Start in Recommender Systems," 2021 12th International Conference on Information and Knowledge Technology (IKT), Babol, Iran, Islamic Republic of, pp. 139-143, 2021.
<http://doi.org/10.1109/IKT54664.2021.9685137>.
- [14] T. Liu, A. Jiang, J. Zhou, M. Li and H. K. Kwan, "GraphSAGE-Based Dynamic Spatial - Temporal Graph Convolutional Network for Traffic Prediction," IEEE Transactions on Intelligent Transportation Systems, 2023.
<http://doi.org/10.1109/TITS.2023.3279929>.
- [15] O. D. Kose and Y. Shen, "Fairness-aware Graph Attention Networks," in Proc. of 2022 56th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, pp. 843-846, 2022.
<http://doi.org/10.1109/IEEECONF56349.2022.10052007>.
- [16] C. Yang et al., "MultiSage: Empowering GCN with contextualized multi-embeddings on web-scale multipartite networks," in Proc. of 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 2434 - 2443, 2020. <http://doi.org/10.1145/3394486.3403293>
- [17] K. Chen, X. Li, T. Gong and D. Qiu, "A Graph Neural Network with Type-Feature Attention for Node Classification on Heterogeneous Graphs," in Proc. of 2022 International Conference on Automation, Robotics and Computer Engineering (ICARCE), Wuhan, China, pp. 1-5, 2022.
<http://doi.org/10.1109/ICARCE55724.2022.10046551>.
- [18] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L. Hamilton, and Jure Leskovec, "Graph Convolutional Neural Networks for Web-Scale Recommender Systems," 24th ACM SIGKDD International Conference on Knowledge

Discovery & Data Mining (KDD '18). Association for Computing Machinery, New York, NY, USA, pp.974-983, 2018.
<http://doi.org/10.1145/3219819.3219890>

[19] E. Myasnikov, "Using UMAP for Dimensionality

Reduction of Hyperspectral Data," 2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon), Vladivostok, Russia, pp. 1-5, 2020.

<http://doi.org/10.1109/FarEastCon50210.2020.9271656>.

○ 저 자 소 개 ○



김 현 우(Hyeon-woo Kim)

2018년~현재 서울과학기술대학교 산업정보시스템공학과 학사과정

2023년 고려대학교 데이터청년캠퍼스(빅데이터 기반의 지능 정보 시스템 개발 과정, 360시간) 수료

관심분야 : AI & Intelligent System, DB, Big Data & Data Mining, 자연어처리, 기계학습, Information Retrieve & Filtering

E-mail : sam.hyeonwoo.kim@seoultech.ac.kr



김 용 준(Yong-jun Kim)

2018~현재 가천대학교 응용통계학과 학사과정

2023년 고려대학교 데이터청년캠퍼스(빅데이터 기반의 지능 정보 시스템 개발 과정, 360시간) 수료

관심분야 : 데이터베이스, 데이터 분석, Big Data, 데이터 마이닝, R, Python, AI, Intelligent System, 자연어 처리, 기계학습, 딥러닝, 데이터 시각화 전문가, 데이터 사이언티스트, 분석 엔지니어, 빅데이터 운영

E-mail : qkdrkzx@gachon.ac.kr



유 길 상(Gil-sang Yoo)

2017년~2021년 서울디지털대학교 컴퓨터소프트웨어학과 객원교수

2010년 중앙대학교 영상공학과(박사)

2010년~현재 (사)한국컴퓨터게임학회 이사

2011년~현재 고려대학교 정보대학 정보창의교육연구소 / 지능정보 SW아카데미 교수

2023년~현재 (사)한국미디어아트산업협회 수석부회장

관심분야 : 데이터사이언스, 데이터 시각화, 빅데이터 분석, HCI, Game, 3D영상 콘텐츠 개발/분석, 콘텐츠 보안, 머신러닝, 딥러닝, 컴퓨터교육

E-mail : ksyoo@korea.ac.kr