콘텐츠 플랫폼의 비정형적 분류 개선을 위한 장르 구조화 연구

임채진[1], 한정수[2], 이현섭[2*]
[1]백석대학교 일반대학원 소프트웨어융합 전공 박사과정, [2]백석대학교 컴퓨터공학부 교수

A Study on Genre Structuring to Improve Unstandardized Classification in Content Platforms

Chae-Jin LIM[1], Jung-Soo HAN[2], Hyun-Seob LEE[2*]
[1]PhD candidate, Software Convergence, Baekseok University
[2]Professor, Division of Computer Engineering, Baekseok University

요 약 콘텐츠 플랫폼은 자사의 콘텐츠 분류에 작품명과 작가명과 함께 장르와 태그를 중요한 메타 데이터로 활용한다. 그 목적은 콘텐츠의 기본 속성을 명확하게 표시하고 비슷해 보이는 콘텐츠끼리 묶어 이용자들의 접근성을 높이는 데에 있다. 하지만 이 두 정보는 서비스 업체 또는 유통 업체가 임의로 할당하는 경우가 많아 정형성과 일관성에 문제점을 드러낸다. 본 연구는 이러한 문제를 개선하는 방법을 찾기 위해, 한국의 대표적 웹툰 플랫폼인 네이버 웹툰의 2005~2024년 연재작 정보를 기반으로 요소 간 통계적 연관성을 나타내는 지표인 PMI(Pointwise Mutual Information, 점별 상호 정보량) 수치로 산출한 장르 간 유사도를 통해 비정형적 장르 구조를 계층적으로 클러스터링했다. 특히 직관적인 연계성 지표지만 실제 증거의 절대적인 빈도를 반영하지 않는 단점을 지닌 PMI를 자체 보정하여 혼동도가 높은 장르들을 적정한 수준으로 클러스터링했다. 실험 결과, 재정의된 클러스터→장르 구조에서 줄거리 임베딩을 통한 장르 예측 성능은 91.4%로 나타나 재정의하지 않았을 때보다 1.49배 향상된 것으로 나타났다. 본 연구는 비정형 데이터의 분류를 위한 지표의 개량을 제시했다는 점, 그리고 웹툰에 국한하지 않은 범용적 콘텐츠 분류 체계로의 확장 및 추천 시스템의 정비에도 기여한다.

주제어 : 분류 구조 재정의, 콘텐츠 클러스터링, PMI, 창작·기획 어시스턴트, 메타데이터

Abstract The content platform utilizes genres and tags as important metadata along with the name of the work and the name of the writer in its content classification. Its purpose is to clearly indicate the basic properties of the content and to combine similar-looking content to increase the accessibility of users. However, these two pieces of information are often randomly assigned by service companies or distributors, revealing problems in structure and consistency. In order to find a way to improve this problem, this study hierarchically clustered the unstandardized genre structure by using the similarity between genres calculated through the PMI (Pointwise Mutual Information), an indicator of statistical correlation between elements, based on the serial information of Naver Webtoon, a representative webtoon platform in Korea from 2005 to 2024. As a result of the experiment, the genre prediction performance through plot embedding in the redefined cluster → genre structure was 91.4%, which was 1.49 times better than when it was not redefined. This study contributes to the improvement of indicators for the classification of unstructured data, the expansion of the general content classification system not limited to webtoons, and the maintenance of the recommendation system.

Key Words : redefine classification, content clustering, PMI, creative/planning assistant, metadata

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-지역지능화혁신인재양성사업의 지원을 받아 수행된 연구임(IITP-2025-RS-2024-00436765)
*교신저자 : 이현섭(hyunseob@bu.ac.kr)
접수일 2025년 07월 04일 수정일 2025년 07월 24일 심사완료일 2025년 08월 05일

1. 서론

콘텐츠 플랫폼은 장르와 태그를 작품 분류를 위한 중요한 메타 데이터로 활용한다. 하지만 실제 운용되는 콘텐츠 분류 체계는 서비스 플랫폼 또는 유통사의 자의적 기준에 따라 설정되는 경우가 많아, 콘텐츠의 특성을 체계적으로 반영하지 못하는 한계가 있다.

본 연구는 이와 같은 문제를 해결하고자, 실제 서비스되고 있는 작품의 장르와 태그 간 통계적 연관성을 기반으로 콘텐츠 분류 구조를 재정비하는 방법론을 제안한다. 먼저 네이버 웹툰[1] 작품들에 할당된 장르와 태그 사이에 동시 출현 빈도(Co-occurrence)[2] 기반 연관성 지표인 PMI를 적용하여 장르-태그 벡터 행렬을 구성한다. 둘째, 장르-태그 벡터 간 코사인 유사도를 산출하고, 셋째, 와드 방법(Ward's Method)을 통한 계층적 클러스터링(Hierarchical Clustering) 분석을 통해 유의미한 장르 클러스터의 수를 도출한다. 최종 단계에서는 이상의 검토를 거쳐 장르 구조를 재정의하고, 장르와 태그를 예측하는 학습기를 설계한다. 이 연구는 비정형 콘텐츠 분류 체계를 데이터 기반으로 구조화하고 재정의할 경우 예측 및 검증 결과에 어떠한 향상이 일어나는지를 실증함으로써 창작 보조 및 추천 시스템 정비로의 확장 가능성을 보인다.

본 연구의 구조는 다음과 같다. 2장에서는 PMI와 문장 단위 임베딩 모델, 계층적 클러스터링 등 본 연구에 선행하는 이론적 배경을 소개한다. 3장에서는 실험 환경과 데이터셋에 대한 소개부터 클러스터링 등의 분석 절차와 설계 및 구현 과정을 소개하고, 4장에서는 분류 재정의에 대한 실험적 검증과 그 결과를 정리한다. 5장에서는 결론과 이후 연구의 확장 가능성을 제시한다.

2. 관련 연구 및 이론적 배경

본 연구에서는 네이버 웹툰 내 장르와 태그 간 배치를 파악하기 위해 PMI를 활용한다. Damani(2013)에 따르면 PMI는 단어 간 의미적 연관성을 측정하는 통계적 지표로, 단순 동시 출현 빈도보다는 두 단어가 기댓값 대비 얼마나 함께 자주 등장하는지를 정량화한다[3].

벡터 간의 상호 유사성을 확인하는 데에는 코사인 유사도가 쓰인다. 코사인 유사도는 두 벡터의 유사도를 측정하는 데에 가장 기본적으로 쓰이는 방식이다. Upadhyay 외 4인(2022)의 연구는 코사인 유사도를 두 벡터 간 내적을 통해 유사도를 측정하는 방법이라고 설명한다[4]. 이후 어떤 장르가 명확하게 구분되지 못하는가를 판단하기 위한 클러스터링을 시도하는데, 본 연구에서는 와드 방법(Ward's Method)에 따른 클러스터링을 활용한다. Kaufman 외 1인(1990)은 와드 방법에 대해 계층적 클러스터링의 각 단계에서 발생하는 클러스터 내 오차제곱합(ESS, Error Sum of Squares)의 증가량을 최소로 만드는 알고리즘으로 설명한다[5]. 이는 최근 Wani(2025)의 연구[16]에서도 동일한 정의로 제시된다. 오차제곱합의 증가량을 측정해 클러스터링하는 과정을 계층화하면 계층적 클러스터링이 된다.

이러한 분류 구조 재정의가 기계 학습의 정확도에 끼치는 영향을 확인하고 실용적 활용 방안을 찾기 위해 학습 모델을 구축하는 데에는 임베딩과 학습 모델을 활용한다. 본 연구는 임수린 외 1인(2025)의 연구[6]를 참조해 다양한 SBERT 모델로 성능을 시험해 웹툰 연재작 정보의 특색에 맞는 임베딩 모델을 찾는다.

학습기로는 랜덤포레스트(RandomForest)를 활용했는데, 이 방식은 ① 학습용 데이터에서 복원 추출에 의해 부트스트랩 방식으로 무작위 생성한 여러 개의 표본을 두고 ② 각 단계마다 역시 무작위로 예측 변수들을 선택한 후 표본에 대해 분류(혹은 예측) 트리를 적합한 후 ③ 각 트리로부터 얻은 예측과 분류 결과를 각기 평균화와 투표를 통해 결정함으로써 예측을 향상시킨다[7]. 이상에서 랜덤포레스트는 입력 대상인 줄거리의 분량이 일정하지 않고 태그의 수 또한 제각기여서 몹시 불균형한 데이터인 웹툰 연재 정보를 자동으로 분류하는 데에 적절한 것으로 평가할 수 있다. 태그의 경우 클러스터 또는 클러스터 내 세부 장르와 달리 작품 당 여럿 배정돼 있기 때문에 이를 해결하기 위해 OVR (One-Vs-Rest) 방식을 도입하였다. Ashwinkumar 외 2인(2024)의 연구에 따르면 OVR은 다중 라벨 분류에서 널리 쓰이는 접근 방식이며, 데이터셋 내 각 라벨마다 이진 분류기를 별도로 구축하는 방식이다[8].

3. 방법론

3.1 데이터셋 소개

본 연구에서는 한국 웹툰 사이트인 네이버 웹툰의 2005~2024년 연재작 정보 2,179편을 이용했다. 해당 데이터는 2025년 7월 현재 서비스 계약 종료로 제외된 작품을 제외한 총 7,218건의 정보 가운데 연도별 중복분

을 제외한 것으로, 전처리와 해석 과정을 거쳐 저장되었다. 장르는 개그, 일상, 감성, 드라마, 무협/사극, 액션, 판타지, 스포츠, 로맨스, 스릴러 등 모두 10개이며 태그는 총 250개다. 데이터셋 가운데 실제로 활용한 변수는 작품명과 작가명을 제외한 ① 줄거리 ② 장르 ③ 태그이며 그 중 장르는 작품 당 1개, 태그는 장르에 구애받지 않고 다중 설정되어 있다.

3.2 분석 절차 및 구현 방법

본 연구의 흐름은 다음과 같다. 먼저 네이버 웹툰의 연재작 데이터를 수집해 적재한다. 그중 장르와 태그, 줄거리 정보만을 추출한 후, 이를 불용어 처리 등을 거쳐 정제한다. 그다음 장르와 태그의 연관성을 판별하는 알고리즘을 거쳐 장르-태그 연관성 벡터를 만들고 유사도를 측정해 클러스터링을 시도한다. 이 클러스터링 자료를 통해 네이버 웹툰의 장르 구분을 확인한 후, 재정의한 클러스터-장르 구조의 성능을 시험한다.

먼저 장르와 태그 간 정량적 연관성을 파악하기 위해, 각 장르별로 등장한 태그들의 동시 출현 빈도를 측정한 후, 통계적 연관 지표인 PMI를 산출한다. [Eq. 1]은 PMI의 기본식이다[3].

$$\mathrm{PMI}(g,t) = \log\left(\frac{P(g,t)}{P(g)\times P(t)}\right)$$

[Eq. 1] PMI Formula

[Eq. 1]에서 P는 확률(probability), g는 특정 장르, t는 특정 태그다. $P(g)$는 전제 작품 중 특정 장르인 작품의 비율, $P(t)$는 전체 작품 중 특정 태그가 있는 작품의 비율이다. $P(g, t)$는 특정 장르를 가진 작품 가운데에서 특정 태그가 함께 붙어 있는, 다시 말해 전체 작품 중에 이러한 조합이 등장할 확률이다. 이 확률에서 전체 작품 수가 있는 분모를 제거한 것이 동시 출현 빈도로, 곧 특정 장르에 특정 태그가 붙은 빈도수다.

그러나 Damani(2013)에 따르면 PMI에는 확률만 다루고 실제 증거의 절대적인 빈도를 무시하는 단점이 있다[3]. 또한 희귀한 단어쌍의 점수를 과도하게 가치 있게 평가하며, 스케일이 넓다는 점도 지적된다. 즉 PMI 값의 보완은 빈도를 적용하고 스케일을 조정하여 실제적인 변별력을 반영하는 방향으로 맞춰져야 한다. 이에 따라 본 연구는 PMI에 빈도값의 로그값을 곱하는 형태의 보정을 가했다. 이러한 보정의 형태는 Manning 외 2인(2009)의 연구 등에서 TF-IDF에 가했던 TF 대체 로그 스케일 가중치를 응용한 것이다[9].

$$\mathrm{weighted\_PMI}(g,t) = \mathrm{PMI}(g,t)\times\log(1+\mathrm{count}(g,t))$$

[Eq. 2] weighted_PMI formula

[Eq. 2]은 이상과 같은 가중치를 PMI에 적용하여 보정된 PMI인 weighted_PMI를 얻는 식이다. count(g, t)은 [Eq. 2]의 count(g, t)와 동일한 동시 출현 빈도다. 즉 [Eq. 2]는 PMI의 단점인 빈도 무시를 부드럽게 보정하는 역할을 한다. 이렇게 산출한 weighted_PMI로 장르와 태그의 연관 행렬을 구성한 후 장르 간 코사인 유사도를 측정하고, 이를 통해 계층 구조를 파악해 장르 구조를 재정의한다. 마지막으로, 클러스터→장르로 재정의된 구조가 재정의되지 않은 구조에 비해 얼마나 우수한지를 학습기로 예측/검증한다.

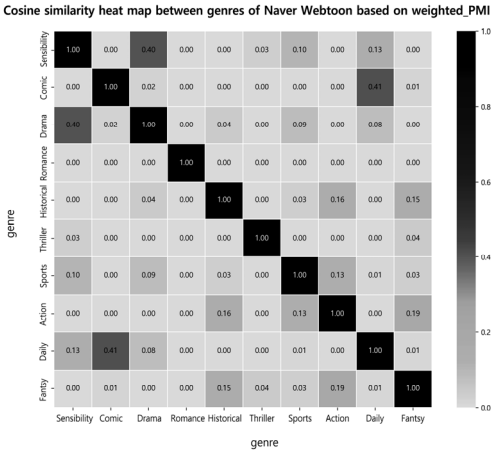
본 연구에서 설계한 학습기는 클러스터를 기준으로 재분류된 장르에 따라 줄거리(시놉시스)의 임베딩을 200개의 결정 트리를 사용하는 랜덤포레스트로 학습한다. 이때 클러스터로 묶인 장르의 경우는 별도로 학습을 진행하여 혼동도를 줄인다. 이때 줄거리에 기반한 태그 학습의 경우, 전체 데이터셋 2,179개 중 태그가 미설정된 작품 1,191건(54.56%)은 학습 대상에서 제외한다. 정제를 거친 태그 242개 중 등장 빈도가 5회 이하인 태그 53개(21.9%) 또한 제외한다. 네이버 웹툰은 장르별 작품 수, 작품당 태그 할당 수 면에서 심한 불균형을 보이기 때문에 오버 샘플링으로 데이터를 보정한다. 또 태그가 다중일 때 지도 학습의 효율이 상승하므로, 작품당 할당된 태그 수에 따라 행을 복사하여 학습시킨다. 이때 OVR을 적용해 태그별로 줄거리를 학습한다.

4. 실험

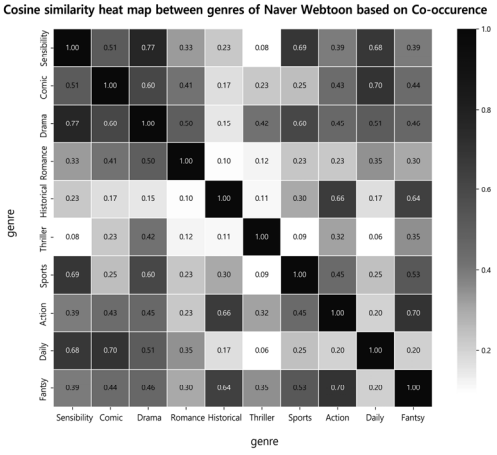
4.1 실험 환경

본 연구의 분석은 Python 3.10 환경에서 수행되었으며, 주요 라이브러리로는 scikit-learn 1.6.1, gensim 4.3.3, pandas 2.2.2, matplotlib 3.9.0, sentence_transformer 5.0.0 등이 사용되었다. 데이터 전처리 및 형태소 분석에는 Okt(Open Korean Text) 형태소 분석기가 활용되었다. 실험은 Windows 10 운영체제를 사용하는 Intel i7-8700 CPU(3.20GHz), 32GB RAM의 사양을 갖춘 PC에서 수행되었다.

4.2 동시 출현 빈도와 PMI, weighted_PMI의 비교

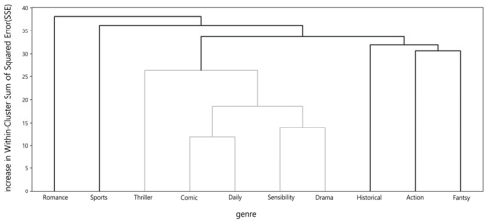


[Fig. 1] Cosine Similarity Heatmap (weighted_PMI)

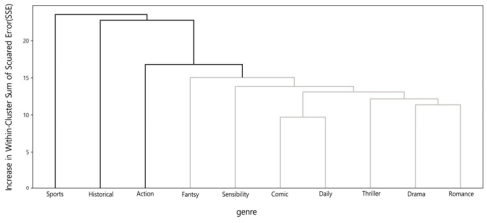


[Fig. 2] Cosine Similarity Heatmap (Co-occurrence)

이 두 도표는 네이버 웹툰의 장르 간 코사인 유사도를 나타낸 히트맵이다. [Fig. 1]는 weighted PMI 기반이고 [Fig. 2]는 PMI의 요소기도 한 동시 출현 빈도 기반이다. [Fig. 1]는 비교적 양 축의 장르별 구분을 명확히 하면서 사용 태그에 따라 경계가 흐린 장르들만 보이는 데 비해 [Fig. 2]은 장르 간의 구분이 거의 되지 않는 양상을 보인다. 즉 PMI는 비정형적으로 정리된 데이터를 분석하는 데에 단순 빈도 기반보다 우수하다.



[Fig. 3] Clustering Dendrogram (weighted_PMI)



[Fig. 4] Clustering Dendrogram (PMI)

[Fig. 3]와 [Fig. 4]의 덴드로그램은 각각 weighted_PMI와 PMI를 기반으로 와드 방법을 통해 계층적 클러스터링을 진행한 결과다. 각 덴드로그램의 X축은 장르, Y축은 클러스터 병합 시의 오차제곱합 증가량을 뜻한다. [Fig. 3]에서는 Y축이 40에 가까운 데 비해 [Fig. 4]의 Y축은 20에 가깝다. 또한 [Fig. 3]에서는 10개 장르 중 4개가 묶이는 데 비해 [Fig. 4]에서는 10개 장르 중 7개가 묶이는 것으로 나타난다. 이는 weighted_PMI가 PMI보다 태그 분포에 따른 장르의 특색을 잘 구분함을 나타낸다.

〈Table 1〉 Genre Clustering Result

Cluster No	Genre
1	Sensibility, Comic, Drama, Daily
2	Thriller
3	Action
4	Fantasy
5	Historical
6	Sports
7	Romance

〈Table 1〉은 이상의 결과를 반영해 10개 장르를 7개 클러스터로 재정의한 결과다. 유사도가 높아 타 장르에 비해 서로 강하게 연결되는 장르 네 개(감성, 개그, 드라마, 일상)를 상위 분류인 'Cluster 1'로 묶었다. 즉 Cluter 1은 네 개의 세부 장르를 보유하게 된다.

4.3 구조 재정의 전후 임베딩 모델별 성능 비교

weighted_PMI를 적용한 구조 재정의의 성능을 확인하기 위해 원래의 장르 구분 그대로를 학습하여 예측/검증의 정확도를 비교했다. 임수린 외 1인의 연구(2025)는 2024년 11월 온라인 커피 유통 전문 기업 W사에서 제공받은 커피 리뷰 데이터를 정제해 클러스터링을 진행하는데[6], 비정형적 한국어 리뷰 문장을 임베딩하기 위해 SBERT(Sentence-BERT) 중 한국어에 맞는 임베딩 모델로 KR-SBERT[10]와 BGE-m3-Ko[11]를 주요 모델로 선정하고, 다국어 표현 포괄 가능성과 모델 범용성을 위해 구글의 다국어 임베딩인 distilUSE[12]와 openAI[13]의 임베딩 모델인 text-embedding-3-small[14]까지 포함해 성능을 시험했다. 본 연구에서는 비정형적인 데이터를 임베딩하려는 선행 연구의 문제 인식을 이어받아 같은 SBERT 모델들을 통해 클러스터링 전후의 분류 성능을 비교함으로써 줄거리 정보를 바탕으로 특정 장르 및 주요 태그를 파악하는 창작·기획자용 지원 도구 개발에 쓰일 임베딩 모델을 찾았다. 단 text-embedding-3-small은 유료 모델이어서 intfloat/multilingual-e5-small[15]로 대체했다.

〈Table 2〉 Comparison Of Learning Accuracy Before And After Clustering By Embedding Model

Embedding Model	Accuracy (10 Genres)	Accuracy (7 Clusters)
BGE-m3-Ko (dragonkue/BGE-m3-ko)	0.6008	0.9062 (Cluster 1~7) 0.9809 (Cluster 1)
KR-SBERT (snunlp/KR-SBERT-V40K-klueNLI-augSTS)	0.6129	0.9140 (Cluster 1~7) 0.97 (Cluster 1)
distilUSE (distiluse-base-multilingual-cased-v1)	0.5948	0.8886 (Cluster 1~7) 0.9754 (Cluster 1)
text-embedding-3-small (intfloat/multilingual-e5-small)	0.5726	0.8857 (Cluster 1~7) 0.9809 (Cluster 1)

〈Table 2〉는 10개 장르 체제에서 장르와 줄거리를 학습한 결과를 7개 클러스터 체제로 재정의한 후와 비교한 결과다. 200개의 결정트리와 20%의 검증 사이즈를 적용한 랜덤포레스트를 이용하였고, 오버 샘플링 등의 조건도 동일하게 맞춘 상태에서 재정의 여부를 달리 한 것이다. 예측 중 정답과 일치한 비율인 정확도 산출 결과, 가장 높은 성능을 보인 임베딩 모델은 KR-SBERT로 〈Table 2〉에서 정의된 Cluster 1~7 전체 대상 91.4%의 정확도를 보였다. 이는 구조 재정의 이전의 0.6129에 비해 1.49배 높은 수치다. 다른 임베딩 모델들도 대체로 1.49~1.54배의 성능 차이를 보였다. 구조 재정의 이후 Cluster 1로 묶인 네 개 장르 '감성' '개그' '드라마' '일상'에 별도의 성능 시험을 한 결과 97~98%의 정확도로 나타났다. 특히 Cluster 1의 경우 전체와는 별도로 원래의 장르명과 줄거리 문장 임베딩을 진행했음에도 매우 높은 정확도 수치를 보였다. 이 결과는 변경된 구조에 최적화한 학습 모델을 구축할 시 명확한 성능 향상이 나타남을 보여준 결과로, 웹툰 연재 정보와 같은 비정형 데이터의 분류에 기준 재정의가 왜 필요한지를 실증적으로 입증하였다.

5. 결론

본 연구는 네이버 웹툰이 작품마다 설정한 장르와 태그가 어떤 분포를 이루고 있는가를 바탕으로, 운영 측이 임의적이고도 비정형적으로 정보를 분류하고 있을 때 각 정보의 관계가 명확한 형태로 전달될 수 있는가에 대한 질문에서 시작했다. 장르와 태그는 가장 기본적인 콘텐츠 분류법으로 널리 쓰이고 있지만 모두에게 통용될 명확한 기준이 제시되기는 어렵다. 본 연구가 보이고자 한 바와 지표의 보정은 비정형적인 형태를 보이는 네이버 웹툰의 데이터를 적정한 수준으로 묶을 수 있음을 실험적으로 증명하였다. 이 결과가 네이버 웹툰은 물론 웹소설과 음악 등 메타 데이터 분류가 아직 매우 중요한 역할을 하는 분야의 분류 구조 재정의로 확장되길 희망한다.

본 연구의 한계는 데이터의 공백을 여건상 채우지 못한 것이다. 3장에서 전술하였듯 네이버 웹툰에는 태그가 아예 없는 작품 수가 54.56%를 차지한다. 따라서 LLM을 통한 데이터 보강, 타 분야/업체의 콘텐츠 데이터를 통한 확장은 향후 추구해야 할 과제다. 나아가 장르 판별에 근거한 줄거리 첨삭, 배포형 어시스턴트 도구 개발 등도 이후 연구에서 추구하고자 한다.

REFERENCES

[1] Naver Webtoon, https://comic.naver.com/

[2] C.Wartena, R.Brussee, M.Wibbels, "Using Tag Co-occurrence for Recommendation", 2009 Ninth International Conference on Intelligent Systems Design and Applications, 2009.

[3] O.P.Damani, "Improving Pointwise Mutual Information (PMI) by Incorporating Significant Co-occurrence",

Seventeenth Conference on Computational Natural Language Learning, pp.20-28, 2013.

[4] A.Upadhyay, A.Bhatnagar, N.Bhavsar, M.Singh, and P.Motlicek, "An Empirical Comparison of Semantic Similarity Methods for Analyzing down-streaming Automatic Minuting task", Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation, pp.572-581, 2022.

[5] L.Kaufman, P.J.Rousseeuw, "FINDING GROUPS IN DATA", A JOHN WILEY&SONS, pp.230-234, 1990.

[6] S.R.Im, H.J.Lim, "Framework for Embedding and Clustering Product Review Texts with Cluster Interpretation Using Large Language Models", Journal of Digital Contents Society Vol.26, No.6, pp.1579-1587, 2025.

[7] G.Shmueli, P.C.Bruce, P.Gedeck, N.R.Patel, "DATAMINING FOR BUSINESS ANALYTICS", John Wiley & Sons, Hanbit academy(Korean Translate) p270, 2019,

[8] V.Ashwinkumar, P.P.Arage, JeyaR, P.Sudhakaran, "One-vs-Rest vs. Voting Classifiers for MultiLabel Text Classification: An Empirical Study", E3S Web Conf., Volume 491, 01014, 2024.

[9] C.D.Manning, P.Raghavan, H.Schütze, "An Introduction to Information Retrieval", Cambridge University Press, p128, 2009.

[10] Snunlp KR-SBERT Model, https://github.com/snunlp/kr-sbert

[11] J.Chen, S.Xiao, P.Zhang, K.Luo, D.Lian, and Z.Liu, "M3-Embedding: Multi-Linguality, Multi-Functionality, Multi-Granularity Text Embeddings through Self-Knowledge Distillation," in Findings of the Association for Computational Linguistics: ACL 2024, Bangkok, Thailand, pp.2318-2335, 2024.

[12] N.Reimers, I.Gurevych, "Making Monolingual Sentence Embeddings Multilingual Using Knowledge Distillation," arXiv:2004.09813, 2020.

[13] OpenAI, https://openai.com/

[14] OpenAI API Platform, https://openai.com/ko-KR/api/

[15] Hugging Face intfloat/multilingual-e5-small, https://huggingface.co/intfloat/multilingual-e5-small

[16] A.A.Wani, "Comprehensive analysis of clustering algorithms: exploring limitations and innovative solutions", PeerJ Computer Science, vol.10, e2286, 2024.

임 채 진(Chae-Jin, LIM) [정회원]

- 2004년 8월 : 백석대학교 정보통신학부 컴퓨터학 전공 (공학사)
- 2014년 2월 : 성공회대학교 문화대학원 미디어.문화연구전공 (문학석사)
- 2025년 3월 ~ : 백석대학교 일반대학원 소프트웨어융합 전공 (박사과정생)

〈관심분야〉
콘텐츠-정보통신 융합, 인공지능

한 정 수(Han, Jung Soo) [정회원]

- 1992년 8월 : 경희대학교 컴퓨터공학부(공학석사)
- 2000년 8월 : 경희대학교 대학원 컴퓨터공학부(공학박사)
- 2001년 3월 ~ 현재 : 백석대학교 컴퓨터공학부 교수

〈관심분야〉
AI 교육, 자율주행, 데이터 분석, SW 모델링

이 현 섭(Hyun-Seob Lee) [종신회원]

- 2007년 2월 : 한양대학교 컴퓨터공학과 (공학 석사)
- 2013년 2월 : 한양대학교 컴퓨터공학과 (공학 박사)
- 2012년 3월 ~ 2021년 2월 : 삼성전자 책임연구원
- 2021년 3월 ~ 현재 : 백석대학교 컴퓨터공학부 조교수

〈관심분야〉
인공지능, 저장시스템, 임베디드 시스템