KorPatBERT와 CenterLoss를 활용한 KNN 기반 IPC 자동분류 연구

곽재원 · 백종민

강남대학교 ICT 공학부 소프트웨어전공
강남대학교 ICT 공학부 데이터사이언스전공

Abstract

국제특허분류(IPC)는 특허 문서의 체계적인 분류와 검색에 필수적이지만, 수작업으로 인한 시간 소모와 일관성 저하 문제가 있다. 기존의 기계학습과 자연어 처리(NLP) 기반 자동 분류 기법은 특허 텍스트의 복잡한 문맥을 충분히 반영하지 못하고, 클래스 불균형으로 빈도가 낮은 분야에서 성능이 저하되는 한계를 보인다.

본 연구에서는 한국어 특허 문서로 사전 학습된 KorPatBERT 모델을 사용하여 특징 벡터를 추출하고, CenterLoss 손실 함수를 적용한 KNN 알고리즘 기반 IPC 자동분류 방법을 제안한다. 강남대학교 특허코드 분류 경진대회 데이터셋에서 G06F, G06Q, G16H의 서브클래스 레벨 분류를 수행한 결과, 91.03%의 정확도를 달성하였으며, 이는 동일한 하이퍼파라미터로 KorPatBERT만을 적용한 경우보다 3.82% 향상된 성능이다.

제안된 방법은 클래스 불균형 문제를 효과적으로 완화하여 빈도가 낮은 분야에서도 우수한 분류 성능을 보여주었으며, IPC 서브클래스 수준의 분류 시 NLP 기법에서 발생하는 모델 과적합을 방지하는 데 기여함을 확인할 수 있다.

1 서론

국제특허분류(IPC) 코드는 특허 문서를 체계적으로 분류하고 검색하는 데 필수적인 도구로서, 전 세계적으로 통용되는 표준화된 분류 체계이다. IPC 코드는 특허 발명의 기술적 특성을 다계층 구조로 체계화하여, 기술 동향 파악, 선행기술 조사, 권리 범위 해석 등에 활용된다. 국내에서는 특허청이 특허, 실용신안 및 PCT 국제출원의 기술 내용을 면밀히 분석하여 적절한 IPC 코드와 협력특허분류(CPC) 코드를 부여하고 있다. 그러나 이러한 분류 작업은 수작업으로 이루어져 많은 시간과 비용이 소요되며, 전문가의 주관에 따라 분류 결과의 일관성이 저하될 수 있다는 한계가 있다.

이를 해결하기 위해 방대한 특허 문서의 자동 분류를 지원하는 다양한 데이터 마이닝 및 기계학습 알고리즘이 적용되고 있다. 초기에는 키워드 빈도나 TF-IDF 등을 기반으로 특허 문서를 벡터화하고 분류하는 통계적 방법이나 머신러닝 기

법이 시도되었다. 하지만 이러한 방법들은 특허 텍스트의 복잡한 문맥적 의미를 충분히 반영하지 못하며, 긴 문장과 복잡한 표현, 전문 용어로 이루어진 특허 문서 특유의 특성을 처리하기에 한계가 있다. 또한 새로운 기술 분야나 용어에 유연하게 대응하기 어렵다는 문제가 존재한다.

이러한 한계를 극복하기 위해 자연어 처리(NLP) 분야의 발전으로 등장한 딥러닝 기반의 언어 모델이 주목받고 있으며, 그중에서도 BERT(Bidirectional Encoder Representations from Transformers) 모델은 사전 학습된 거대한 언어 모델로서 양방향에서 문맥을 이해하여 단어의 의미를 파악할 수 있다. 그러나 NLP 기반 방법 역시 클래스 불균형으로 인해 빈도가 낮은 분야, 즉 'Cold Fields'에서 높은 분류 성능을 달성하기 어렵고, 데이터 편향으로 인해 모델이 특정 클래스에 과적합될 위험성이 존재한다. 이는 모델의 일반화 능력을 저하시켜 새로운 데이터나 희귀한 클래스에 대한 분류 성능을 떨어뜨린다. 따라서 클래스 불균형 문제를 완화하고 모델의 일반화 성능을 향상시키는 새로운 접근이 요구되고 있다.

2 배경 이론 및 선행 연구

2.1 선행 연구 조사

2.1.1 KNN 기반 IPC 자동 분류

이재안 등은 특허 문서 검색 결과를 활용한 KNN(K-Nearest Neighbors) 기반의 IPC 자동 분류 시스템을 제안하였다 [1]. 이 연구에서는 방대한 학습 데이터와 복잡한 분류 체계로 인해 발생하는 자질(feature) 추출과 모델 생성의 어려움을 해결하기 위해, 분류 대상 특허 문헌을 쿼리로 사용하여 검색 엔진을 통해 상위 K개의 유사한 특허 문헌을 검색하였다. 이후, 검색된 문헌들과 분류 대상 문헌 간의 유사도를 계산하고, 이들의 IPC 코드를 기반으로 대상 문헌의 분류를 결정하는 방법을 적용하였다.

실험 결과, 서브 클래스 수준에서 F-measure 40.21%로 검색 결과의 순위 정보와 문서 간 유사도 계산 방식을 고려함으로써 분류 성능이 향상되었다. 그러나 이 방법은 검색 결과의 순위와 유사도 계산 방식에 성능이 크게 의존하며, 세부 분류 수준에서 정확도가 낮아지는 한계가 있었다. 또한, 분류 체계의 복잡성과 클래스 불균형 문제로 인해 일부 분야에서 모델의 성능이 저하되는 문제점이 발견되었다.

2.1.2 NLP 기반 IPC 자동 분류

김기용 등은 Word2Vec과 2계층 양방향 장단기 메모리 신경망(BLSTM)을 활용한 IPC 자동 분류 시스템을 제안하였다 [2]. 이 연구에서는 특허 문서의 복잡한 분류 체계와 방대한 데이터로 인해 발생하는 학습 시간 증가와 모델 성능 저하 문제를 해결하기 위해, 특허 문서의 요약 필드에 집중하여 Word2Vec을 통한 단어 임베딩과 2계층 BLSTM 신경망을 적용하였다.

실험 결과, 기존의 CNN 및 단일 계층 LSTM을 이용한 방법에 비해 분류 정확도가 각각 약 7%와 5% 향상되었으며, 전통적인 분류 모델인 나이브 베이즈, 로지스틱 회귀, 서포트 벡터 머신보다도 우수한 성능을 보였다. 그러나 입력 시퀀스의 길이 증가와 신경망의 계층 수 확대에 따라 학습 시간과 메모리 사용량이 크게 증가하는 한계가 있었다. 또한, 특정 IPC 섹션 간 데이터 불균형으로 인해 일부 섹션에서 상대적으로 낮은 분류 정확도를 보이는 문제점이 발견되었다.

2.1.3 BERT 기반 IPC 자동 분류

Jieh-Sheng Lee와 Jieh Hsiang은 사전 학습된 BERT 모델의 미세 조정을 통한 IPC 자동 분류 시스템을 제안하였다 [3]. 이 연구에서는 약 305만 건의 대규모 특허 데이터셋을 활용하여 BERT-Base 모델을 특허 분류 작업에 맞게 미세 조정하였으며, 이를 IPC 분류에 적용하였다. 특히, 특허 문서의 다양한 부분 중 '청구항(claims)'만을 사

용하여 분류를 수행하였는데, 이는 특허 청구항만으로도 효과적인 분류가 가능함을 입증하고자 한 것이다. 실험 결과, CPC 서브클래스 수준(총 656개 레이블)에서 F1-score가 Top 1 기준으로 66.83%를 달성하였으며, 정확도는 84.26%로 기존의 CNN과 워드 임베딩을 활용한 DeepPatent 보다 우수한 성능을 보였다. 또한, 테스트 데이터의 크기를 세 배로 증가시켜도 성능이 안정적으로 유지됨을 확인하였다. 그러나 테스트 데이터의 연도가 훈련 데이터보다 멀어질수록 F1-score, 정확도 및 재현율이 약간씩 감소하는 경향을 보였으며, 이는 모델의 시간에 따른 일반화 성능 개선을 위한 추가 연구가 필요함을 시사한다.

본 연구는 사전 학습된 언어 모델과 미세 조정을 통해 특허 분류에서의 새로운 가능성을 제시하였으며, 특히 특허 청구항의 중요성을 부각시켰다.

2.2 배경 지식

2.2.1 특허 데이터

특허 데이터는 기술 혁신과 산업 동향을 파악하는 데 핵심적인 역할을 하는 중요한 기술 문서로서, 기술 발전의 방향성을 이해하고 전략을 수립하는 데 필수적이다. 특허 데이터는 공개된 정보를 기반으로 하므로 누구나 접근하고 활용할 수 있으며, 이를 통해 기술 발전의 흐름, 경쟁사의 연구 동향, 시장의 기술 격차 등을 분석할 수 있다. 또한, 특허 데이터 분석을 통해 핵심 기술, 부상 기술, 공백 기술 등을 식별하여 효과적인 연구개발(R&D) 전략 수립에 기여할 수 있다. 이러한 이유로 연구자와 기업은 특허 데이터를 적극적으로 활용하여 제품 설계, 시장 진입 전략, 지식재산권 확보 등을 추진하고 있다.

특허 데이터에 접근하는 방법으로는 무료 및 유료 데이터베이스가 있으며, 무료로는 한국특허정보원이 운영하는 KIPRIS가 있고, 유료로는 WIPS, Thomson Innovation 등이 있다. 이러한 데이터베이스를 활용하면 효율적인 기술 검색과 분석이 가능하며, 최신의 기술 정보를 손쉽게 확보할 수 있다.

2.2.2 IPC 시스템

IPC(International Patent Classification) 시스템은 전 세계 특허 문헌을 통일된 기준으로 분류하기 위해 마련된 국제적인 특허 분류 체계이다. 이 시스템은 기술 분야를 체계적으로 분류하여 특허 검색과 분석의 효율성을 높이며, 국제적인 특허 정보 교환과 협력을 촉진한다.

IPC는 총 8개의 섹션(section)으로 구성되며, 각 섹션은 다시 클래스(class), 서브클래스(subclass), 메인 그룹(main group), 서브 그룹(subgroup)으로 세분화되는 계층적 구조를 갖추고 있다. 각 특허 문서는 해당 기술 분야에 맞는 IPC 코드를 부여받아 분류되며, 이를 통해 연구자와 기업은 특정 기술 분야의 특허를 손쉽게 검색하고 기술 동향 및 발전 추이를 효과적으로 파악할 수 있다. 또한, IPC 시스템은 세계 각국의 특허 분류 체계와 연계되어 국제 특허 협력과 정보 교환에 중요한 역할을 한다. 최근에는 IPC와 함께 Cooperative Patent Classification(CPC)과 같은 더욱 세분화된 분류 체계가 도입되어 특허 정보의 접근성과 분석 능력을 향상시키고 있다. 이러한 발전은 기술 혁신의 가속화와 산업 경쟁력 강화에 기여하고 있다.

2.2.3 Center Loss

Center Loss는 딥러닝 기반의 얼굴 인식 및 특징 학습에서 동일 클래스 내의 특징 벡터를 하나의 중심으로 모으는 것을 목표로 하는 손실 함수로, Wen 등에 의해 제안되었다 [4]. 이 방법은 특징 공간에서 클래스 내 분산(intra-class variance)을 최소화하여 분류 성능을 향상시키는 데 효과적이며, 이는 KNN과 유사한 원리를 가진다.

Center Loss는 각 클래스 y에 대한 중심 벡터 c_y를 학습하며, 각 샘플의 특징 벡터 x_i와 해당

클래스의 중심 c_{y_i} 사이의 유클리드 거리를 최소화한다. 이를 수식으로 표현하면 다음과 같다.

$$L_c = \frac{1}{2}\sum_{i=1}^{m} \|x_i - c_{y_i}\|_2^2 \tag{1}$$

여기서 m은 미니배치의 크기, x_i는 i번째 샘플의 특징 벡터, c_{y_i}는 해당 클래스의 중심 벡터이다. Center Loss는 일반적으로 Softmax 손실 함수 L_s와 결합하여 전체 손실 함수 L을 구성한다.

$$L = L_s + \lambda L_c \tag{2}$$

여기서 λ는 Softmax 손실과 Center Loss 간의 균형을 조절하는 하이퍼파라미터이다.

Center Loss의 장점은 다음과 같다.

동일 클래스 내 분산 최소화: 특징 벡터들이 해당 클래스 중심 주변에 밀집되어 분류기의 성능이 향상된다.

클래스 간 분별력 향상: 클래스 간 거리가 상대적으로 증가하여 모델의 일반화 능력이 향상된다.

그러나 단점도 존재한다.

추가적인 메모리 및 계산 비용: 각 클래스 중심을 저장하고 업데이트해야 하므로 추가적인 자원이 필요하다.

하이퍼파라미터 설정의 어려움: λ 값을 적절하게 설정하지 않으면 학습 과정에서 손실 함수 간의 균형이 깨질 수 있다.

Center Loss는 얼굴 인식, 객체 분류, 사람 재식별 등 다양한 분야에서 적용되어 왔으며, 특히 클래스 분포가 복잡한 데이터에서 효과적이다. 본 연구에서는 이미지 데이터에서 벡터화하여 계산한다는 Center Loss의 아이디어를 착안하여, 한국어 특허 문서로 사전 학습된 KorPatBERT 모델에서 특징 벡터를 추출하고 Center Loss를 적용하여 클래스 불균형 문제를 해결함으로써 IPC 자동 분류의 성능을 향상시키고자 한다.

2.2.4 KorPatBERT

KorPatBERT는 한국특허정보원이 개발한 특허 분야의 한국어 자연어 처리 사전 학습 언어 모델로서, 특허 분야에서의 한국어 자연어 처리 문제를 해결하고 특허 산업 분야의 지능정보화 인프라를 구축하기 위해 개발되었다 [5]. 이 모델은 Google의 BERT(base) 아키텍처를 기반으로 하며, 국내 특허 문헌 406만 건에서 추출한 약 4억 6천만 개의 문장(약 120GB)을 학습하여 구축되었다.

KorPatBERT의 가장 큰 특징은 특허 문헌에 특화된 토크나이저와 사전을 사용하여 특허 문헌의 언어적 특성을 효과적으로 반영했다는 점이다. 약 1,000만 개의 주요 명사 및 복합 명사를 추출하여 한국어 형태소 분석기 MeCab-ko의 사용자 사전에 추가하고, SentencePiece를 활용한 MSP(MeCab-ko SentencePiece Patent) 토크나이저를 사용하였다. 이를 통해 특허 텍스트의 복잡한 용어와 구조를 보다 정확하게 분석할 수 있게 되었다.

모델의 성능 평가 결과, KorPatBERT는 특허 문헌의 분류 태스크에서 기존 모델들보다 우수한 성능을 보였다. 예를 들어, CPC 서브클래스 분류 태스크에서 KorPatBERT-base 모델은 정확도 76.32%, KorPatBERT-large 모델은 77.06%의 정확도를 달성하였다. 이는 Google BERT의 72.33%나 KorBERT의 73.29%에 비해 높은 수치로, 특허 분야에 특화된 학습이 성능 향상에 크게 기여했음을 보여준다. 또한, CPC 메인그룹 분류 태스크에서도 KorPatBERT-large 모델은 Top-1 정확도 62.89%를 달성하여 높은 수준의 분류 능력을 입증하였다.

KorPatBERT는 특허 문헌의 자동 분류, 유사 특허 검색, 질의응답 시스템 등 다양한 AI 기반 지식재산(IP) 정보 서비스에 활용될 수 있다. 이를

통해 특허 분야의 자연어 처리 정확도를 향상시켜 특허 정보의 분석과 활용을 더욱 효율적으로 지원하며, IP 정보 서비스 산업의 디지털 전환과 혁신에 기여할 수 있다.

따라서 본 연구에서는 KorPatBERT를 활용하여 한국어 특허 문헌의 특징 벡터를 추출하고, Center Loss를 적용하여 KNN 기반의 IPC 자동 분류 성능을 향상시키고자 한다. 이를 통해 특허 문헌의 클래스 내 분산을 최소화하고, 클래스 간 분별력을 높여 IPC 자동 분류의 정확도를 개선하는 것이 목표이다.

3 제안 방법

본 연구에서는 한국어 특허 문서로 사전 학습된 KorPatBERT모델을 활용하여 특징 벡터를 추출하고, CenterLoss 손실 함수를 적용한 KNN 기계학습 알고리즘 기반의 IPC 자동분류 방법을 제안한다.

3.1 데이터 소개 및 분석

본 연구에서 사용된 데이터셋은 2024 강남대학교 데이터분석 경진대회에서 제공된 것으로, 각 특허 문서의 다양한 정보를 포함하고 있다. 주요 변수는 다음과 같다.

발명의 명칭: 발명의 전체적인 주제와 분야를 나타내며, 핵심 내용을 함축적으로 표현한다. **요약**: 발명의 핵심 내용과 목적을 요약하며, 주요 특징과 효과를 간단히 기술한다.

대표청구항: 발명의 법적 보호 범위를 정의하고, 기술적인 상세 사항을 포함한다.

메인IPC2: 발명의 주요 기술 분야를 나타내는 IPC 코드의 첫 두 자리로, 본 연구에서의 종속 변수이다.

본 연구의 목표는 주어진 특허 출원 데이터를 기반으로 “메인IPC2” 코드를 정확하게 분류하는 것이다. “메인IPC2”는 국제 특허 분류(IPC) 코드 중 G06F(전기 디지털 데이터 처리), G06Q(데이터 처리 시스템 또는 방법), G16H(의료/헬스케어 정보학)의 세 가지로 구성되며, 이를 세 개의 카테고리로 분류하는 문제이다.

데이터 분석 과정에서 다음과 같은 문제점이 확인되었다.

첫째, 특허 문서의 텍스트 데이터는 복잡한 언어 구조와 도메인 특화된 전문 용어를 다수 포함하고 있다. 이러한 특성은 전통적인 기계학습 알고리즘으로는 효과적으로 특징을 추출하고 분류하기 어렵게 만든다. 특히, 특허 문서는 긴 문장 구조, 복합적인 어휘, 그리고 문맥에 따라 의미가 달라지는 용어들을 포함하고 있어 단순한 Bag-of-Words나 TF-IDF와 같은 기법으로는 텍스트의 의미를 제대로 파악하기 어려울 수 있다. 이는 모델이 정확한 분류를 수행하는 데 장애가 된다.

둘째, 데이터 클래스의 불균형 문제가 심각하였다. 각 IPC 코드에 해당하는 특허 문서의 수는 G06F가 398개, G06Q가 1,791개, G16H가 215개로 크게 차이가 난다.

소수 클래스는 새로운 기술이 발전할 수 있는 기회의 분야이므로, 이들의 정확한 분류는 신기술 탐지 및 특허 전략 수립에 매우 중요하다. 따라서 전체적인 정확도가 높더라도 소수 클래스를 놓치게 되면 특허 분류 시스템의 가치가 감소하므로, 모델링 과정에서 반드시 극복해야 할 요소이다.

셋째, 클래스 내 변동이 크고 클래스 간 차이가 작다는 문제가 있었다. 구체적으로 살펴보면:

클래스 내 변동(Within-Class Variance)이 크다. 동일한 IPC 클래스 내에서도 다양한 주제와 표현의 특허 문서가 포함되어 해당 클래스의 대표적인 특징 벡터를 정확하게 추출하기가 어렵

다. 예를 들어, G06Q 클래스는 금융, 물류, 의료 등 다양한 분야의 비즈니스 방법을 포함하고 있어 각 샘플의 텍스트 표현이 매우 상이할 수 있다.

클래스 간 차이(Between-Class Difference) 가 작다. 서로 다른 IPC 클래스 간에도 유사한 용어나 표현이 사용되어 모델이 이를 구분하기 어려운 상황이 발생한다. 예를 들어, G06F와 G06Q 클래스 모두 "데이터 처리", "시스템", "장치" 등의 유사한 단어를 포함할 수 있어 모델이 두 클래스를 정확히 분류하는 데 어려움을 겪을 수 있다.

이러한 문제를 극복하기 위해, 본 연구에서는 KorPatBERT 모델을 활용하여 특허 문서의 문맥적 의미를 효과적으로 파악하고, CenterLoss 손실 함수를 적용하여 클래스 내 변동을 줄이며 클래스 간 분별력을 높이고자 한다.

3.2 모델 생성

데이터 분석 과정에서 파악된 주요 문제점은 다음 세 가지로 요약할 수 있다.

1. 특허 도메인 용어의 다수 포함: 특허 문서는 전문 용어와 복잡한 표현을 많이 포함하고 있어 일반적인 텍스트 처리보다 어려움이 있다.

2. 클래스 간 데이터 불균형 심화: 특정 IPC 코드에 데이터가 편중되어 있어 모델 학습 시 편향이 발생할 수 있다.

3. 클래스 내 변동이 크고 클래스 간 차이가 작음: 동일한 클래스 내에서도 다양성이 크며, 서로 다른 클래스 간에는 유사성이 높아 분류가 어렵다.

첫 번째 문제를 극복하기 위해 본 연구에서는 BERT 기반의 한국어 특허 데이터셋으로 사전 학습된 자연어 처리 모델인 KorPatBERT를 활용하여 특허 문서의 특징 벡터를 추출하였다.

이때 중요한 고려 사항은 어떤 속성 값(열)을 사용할 것인가였다. 주요 후보로는 "발명의 명칭", "요약", "대표청구항"이 있었다. 본 연구에서는 다음과 같은 이유로 "발명의 명칭"과 "요약"을 결합하여 사용하였다.

핵심 정보의 충분한 반영: "발명의 명칭"과 "요약"은 특허의 주제와 핵심 내용을 간결하면서도 명확하게 담고 있어 IPC 코드의 세부 분류에 필요한 정보를 제공한다. 특허 문서는 발명의 주요 목적과 특징을 중심으로 작성되므로 해당 부분의 정보가 분류에 유용하다.

입력 길이 제한 문제 해결: BERT 모델의 최대 입력 토큰 수는 512이다. 그러나 "대표청구항"은 이 제한을 초과하는 경우가 많아 전체 문장을 모델에 입력할 수 없다. 이는 중요한 정보를 반영하지 못해 모델 성능 저하로 이어질 수 있다. 반면, "발명의 명칭"과 "요약"을 결합한 텍스트는 대부분 이 제한 내에 포함되어 입력 길이 문제를 효과적으로 해결할 수 있다.

효율적인 특징 추출: "발명의 명칭"과 "요약"은 발명의 핵심을 담고 있으므로 불필요한 세부 정보나 중복 정보를 배제하고 효율적으로 특징 벡터를 추출할 수 있다. 이는 모델의 학습 속도를 향상시키고 과적합을 방지하는 데 도움이 된다.

두 번째와 세 번째 문제를 해결하기 위해 Center Loss 손실 함수를 적용하여 단순한 Cross-Entropy 손실 함수 기반의 분류가 아닌 KNN 기반의 분류를 수행하였다. 이를 통해 동일한 클래스 내의 특징 벡터가 중심으로 모이도록 유도하여 클래스의 대표성을 강화하고, 서로 다른 클래스 간의 특징 벡터 거리를 증가시켜 모델이 클래스 간 차이를 더 명확히 할 수 있었다.

모델링의 전체 과정은 다음과 같다.

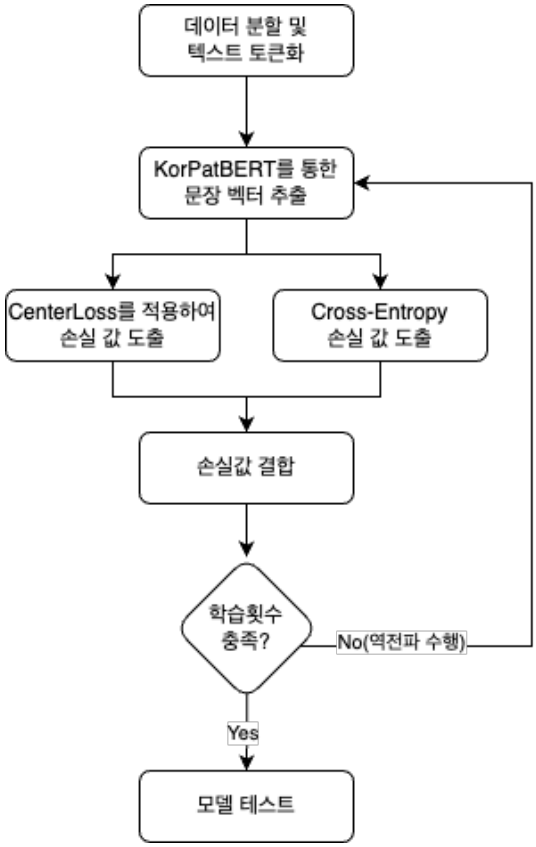

Figure 1: KorPatBERT와 Center Loss를 활용한 IPC 자동분류 모델 아키텍처

본 연구에서는 KorPatBERT와 Center Loss를 활용하여 한국어 특허 문서의 분류 성능을 향상시키는 방법을 제안하였다. 먼저, 데이터 전처리 단계에서 한국어 특허 문서에 특화된 KorPatTokenizer를 적용하여 텍스트 데이터를 토큰화하였으며, 토큰화된 단어는 사전에 정의된 어휘(vocabulary)를 기반으로 인덱스로 변환하였다. 입력 시퀀스는 최대 시퀀스 길이(MAX_SEQ_LEN)를 정의하고, 이를 초과하는 시퀀스는 잘라내고(truncation), 길이가 부족한 경우 패딩(padding)을 적용하여 구성하였다. 이에 따라 입력 IDs는 다음과 같이 표현된다:

$$\begin{aligned}\text{Input IDs} = {} & [\text{CLS}] + \text{Tokenized Sequence} \\ & + \ [\text{SEP}] + \text{Padding}\end{aligned} \tag{3}$$

다음으로, 사전 학습된 KorPatBERT모델을 활용하여 특허 문서의 문맥적 의미를 추출하였다. KorPatBERT는 Transformer 기반의 언어 모델로, 한국어 특허 문서 코퍼스를 활용하여 사전 학습되었다. BERT의 출력에서 첫 번째 토큰인 [CLS] 토큰의 임베딩 벡터를 추출하여 분류기의 입력으로 사용하였으며, 이는 전체 문장의 대표적인 특징을 담고 있다. 모델의 일반화 능력을 향상시키기 위해 드롭아웃 레이어를 적용하여 일부 뉴런의 출력을 무작위로 0으로 설정하였다. 출력층은 소프트맥스 활성화 함수를 사용하여 각 클래스에 대한 확률 분포를 예측하도록 설계되었으며, 이는 다음과 같이 수식화된다:

$$\mathbf{o} = \text{softmax}(\mathbf{W}\mathbf{h}_{\text{CLS}} + \mathbf{b}) \tag{4}$$

여기서 $\mathbf{h}_{\text{CLS}}$는 [CLS] 토큰의 임베딩 벡터, $\mathbf{W}$는 가중치 행렬, $\mathbf{b}$는 바이어스 벡터이다.

분류 성능을 더욱 향상시키기 위해 Center Loss를 적용하였다. Center Loss는 각 클래스의 중심(center)을 학습하여 동일한 클래스의 샘플들이 특징 공간에서 서로 가깝게 위치하도록 유도한다. 이를 통해 특징 분포의 내재적 구조를 반영하고 분류 경계를 더욱 명확하게 할 수 있다. 총 손실 함수는 소프트맥스 손실(Cross-Entropy Loss)과 Center Loss의 결합으로 정의되며 다음과 같다:

$$L_{\text{total}} = L_{\text{softmax}} + \lambda L_{\text{center}} \tag{5}$$

여기서 λ는 두 손실 함수 간의 균형을 조절하는 하이퍼파라미터이다. 역전파 과정을 통해 총 손실 함수 L_{total}에 대한 그래디언트를 계산하고, 모델의 가중치 및 Center Loss의 중심 벡터를 업데이트하였다. 중심 벡터는 각 미니배치에서 샘

플들의 특징 벡터를 기반으로 업데이트되며, 이는 다음과 같이 표현된다:

$$\mathbf{c}_{y_i} \leftarrow \mathbf{c}_{y_i} - \alpha(\mathbf{c}_{y_i} - \mathbf{f}_i) \tag{6}$$

여기서 y_i는 샘플 i의 클래스 레이블, $\mathbf{f}_i$는 해당 샘플의 특징 벡터, α는 작은 값으로 설정된 학습률이다.

학습이 완료된 모델에 대해 테스트 데이터셋을 사용하여 최종 성능을 평가하였다. 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-스코어 등의 성능 지표를 계산하였으며, 혼동 행렬을 작성하여 각 클래스별 예측 결과를 분석하고 오류 패턴을 파악하였다.

마지막으로, 결과 해석 및 시각화를 통해 모델의 개선 사항을 도출하였다. 성능 평가 결과를 기반으로 모델의 한계점과 향후 개선 방안을 제시하였으며, t-SNE와 PCA를 활용하여 분류 결과를 시각화함으로써 이상치를 탐지하고 군집의 분포를 해석하였다. 이를 통해 모델의 예측 성능과 일반화 능력을 향상시킬 수 있는 방향성을 확인하였다.

4 결과

4.1 모델 결과

본 연구에서 제안한 KorPatBERT와 Center Loss를 결합한 모델을 총 5회 학습한 결과, 평균 91.03%의 정확도를 달성하였다. 표 1은 각 클래스별 분류 성능을 나타낸 분류 리포트로, 정밀도(Precision), 재현율(Recall), F1-스코어(F1-score), 지원(Support)을 포함하고 있다.

Table 1: 분류 리포트

	Precision	Recall	F1-score	Support
G06F	0.8588	0.7374	0.7935	99
G06Q	0.9427	0.9554	0.9490	448
G16H	0.7460	0.8545	0.7966	55
Accuracy			0.9103	602
Macro Avg	0.8492	0.8491	0.8464	602
Weighted Avg	0.9110	0.9103	0.9095	602

4.2 모델 결과 비교

4.2.1 기계학습 알고리즘과의 비교

제안한 모델의 성능을 검증하기 위해, TF-IDF를 텍스트 임베딩으로 사용한 대표적인 기계학습 알고리즘인 KNN, 나이브 베이즈(Naive Bayesian), SVM과의 성능을 비교하였다.

Table 2: 기계학습 알고리즘과 본 모델의 성능 비교

	Accuracy	Precision	Recall
KNN	0.7824	0.7758	0.7824
나이브 베이즈	0.7442	0.5538	0.7442
SVM	0.8289	0.8214	0.8289
KorPatBERT + Center Loss	0.9103	0.9110	0.9103

그 결과, 제안한 모델은 기존 기계학습 알고리즘 대비 약 8% 이상의 성능 향상을 보였다.

4.2.2 KorPatBERT와 Center Loss 적용에 따른 비교

앞서 언급한 세 가지 주요 문제점을 해결하기 위해 본 논문에서는 한국어 특허 데이터로 사전 학습된 KorPatBERT와 Center Loss 기법을 적용하였다. 두 요소의 적용 유무에 따른 성능 차이를 비교하기 위해, KorPatBERT를 적용하지 않은 모델로는 한국어에 특화된 KoBERT를 사용하였다. 표 3은 각 모델의 정확도, 정밀도, 재현율을 비교한 결과이다.

Table 3: 각 모델에 따른 성능 비교

	Accuracy	Precision	Recall
KoBERT	0.8688	0.8718	0.8688
KoBERT + Center Loss	0.8787	0.8771	0.8787
KorPatBERT	0.8721	0.8709	0.8721
KorPatBERT + Center Loss	0.9103	0.9110	0.9103

비교 결과, KorPatBERT에 Center Loss를 적용한 모델이 가장 우수한 성능을 보였다. KoBERT에 비해 KoBERT + Center Loss의 성능이 향상되었고, KorPatBERT에 Center Loss를 적용한 경우에도 성능이 개선됨을 통해 Center Loss의 적용이 모델에 긍정적인 영향을 미침을 확인할 수 있었다. 또한, KoBERT보다 KorPatBERT의 성능이 더 우수한 것을 통해 특허 분야의 전문 용어를 사전 학습한 언어 모델의 효과를 확인하였다. 따라서 데이터의 세 가지 주요 문제점 해결을 위해 적용한 두 가지 방법 모두 효과적임을 알 수 있다.

4.2.3 Center Loss 적용에 따른 재현율 비교

데이터셋의 클래스 불균형 문제를 완화하기 위해 Center Loss를 적용하여 단순 Cross-Entropy 기반 분류가 아닌 유클리드 거리 기반의 KNN 알고리즘을 통해 분류를 수행하였다. 표 4는 Center Loss 적용 전후의 각 클래스별 재현율(Recall)을 비교한 것이다.

Table 4: Center Loss 적용에 따른 각 클래스별 재현율 비교

	G16H	G06F	G06Q
KorPatBERT	0.5818	0.7677	0.9308
KorPatBERT + Center Loss	0.8545	0.7374	0.9554

IPC 코드 분류의 경우, 실제로 해당 클래스인 것들 중에서 모델이 정확하게 예측한 비율인 재현율이 중요하다. 이는 실제 사례를 놓치지 않고 잘 분류하는 것이 핵심이기 때문이다. 테스트 데이터에서 G16H, G06F, G06Q 클래스의 샘플 수는 각각 55, 99, 448개로 클래스 간 데이터 불균형이 존재한다. 특히 데이터 양이 가장 적은 G16H 클래스의 경우, Center Loss를 적용함으로써 재현율이 0.5818에서 0.8545로 약 27% 향상되었다. 이를 통해 Center Loss가 클래스 불균형 문제를 완화하고 모델의 성능을 개선하는 데 효과적임을 확인할 수 있다.

또한, Center Loss를 적용함으로써 특징 공간에서 각 클래스의 분포를 시각화하여 분석할 수 있었다. Figure 2와 Figure 3은 t-SNE와 PCA를 사용하여 각 클래스의 특징 벡터를 2차원 공간에 시각화한 결과이다.

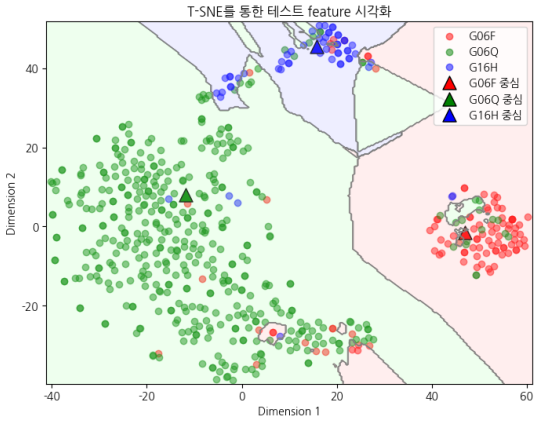

Figure 2: Center Loss 적용 후 특징 벡터의 T-SNE 시각화 결과

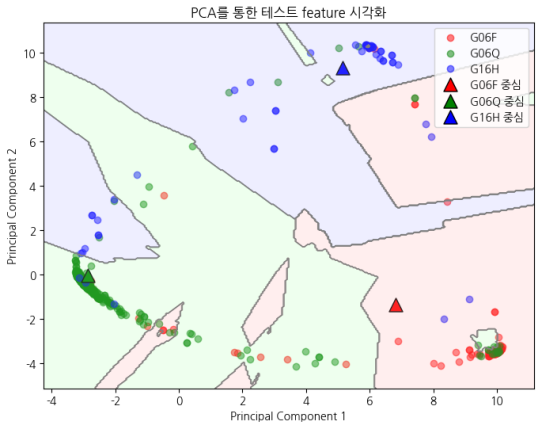

Figure 3: Center Loss 적용 후 특징 벡터의 PCA 시각화 결과

위의 시각화 결과에서 볼 수 있듯이, 샘플 수가 많은 G06Q 클래스(초록색)는 데이터의 다양성이 높아 이상치도 다양하게 분포하고 있음을 알 수 있다. 또한 다양한 분야에서의 데이터 분석 관련 내용을 가지는 G06Q는 다른 군집 중심 부근에도 샘플들이 존재함을 확인할 수 있다.

4.2.4 어텐션 스코어를 통한 토큰 가중치 계산

모델의 예측 과정에서 어텐션 스코어를 활용하여 각 토큰이 분류에 미치는 영향을 분석하였다.

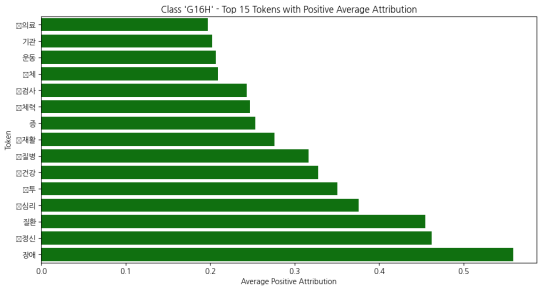

Figure 4: 각 클래스의 어텐션 스코어 시각화 결과

어텐션 스코어를 통해 각 클래스에 영향을 미치는 주요 토큰을 선정하였으며, 이를 바탕으로 표 5와 같이 각 IPC 클래스별로 영향력이 큰 키워드를 정리하였다.

Table 5: IPC 클래스별 영향력이 큰 키워드

IPC 코드	설명	키워드
G06F	전기 디지털 데이터 처리	장애, 보안, 공격, 표준, 저장, 진단, 결과, 표현
G06Q	데이터 처리 시스템 또는 방법	건설, 매장, 저렴, 행복, 테크, 관광, 자치, 학부모, 의약품, 직무, 에어컨, 의류, 행사, 경기
G16H	의료/헬스케어 정보학	장애, 정신, 질환, 심리, 건강, 질병, 재활, 체력, 검사, 운동, 기관, 의료

G16H 클래스(의료/헬스케어 정보학)의 경우 해당 분야에서 사용되는 용어들이 다수 등장함을 확인할 수 있다. 특히 "장애"의 경우 평균 어텐션 스코어가 0.5 이상으로, G16H 클래스를 분류하는 데 큰 영향을 미쳤다. 반면, 다양한 분야에서 데이터를 처리하는 시스템을 다루는 G06Q 클래스는 여러 분야의 키워드를 포함하고 있어 분류에 어려움이 있을 수 있다.

4.3 기대효과 및 발전 가능성

본 연구는 KNN 기반의 Center Loss 기법을 활용하여 IPC 자동분류 모델을 개발하였다. 주어진 경진대회 데이터셋은 단일 라벨 분류 문제였으나, 향후 임계값 이상의 거리를 기준으로 클래스 판별을 수행함으로써 다중 라벨 분류로의 확장이 가능하다 [6]. 또한, 본 연구는 2,404개의 데이터셋을 통해 학습이 진행되었지만, 더 큰 규모의 데이터셋과 다양한 클래스를 활용할 경우 KorPatBERT와 Center Loss의 적용 효과를 더욱 극대화할 수 있을 것으로 기대된다.

또한 본 연구를 통해 개발된 IPC 자동분류 모델은 특허 분류 작업의 효율성과 정확성을 향상

시킬 수 있다. 이는 특허청의 심사 과정에서 업무 부담을 경감시키고 처리 시간을 단축하여, 신속한 특허 출원과 심사를 가능하게 도와 줄 수 있다.

더 나아가, 자동화된 IPC 분류 시스템의 특허 정보 검색의 효율성을 높여, 연구자들과 기업들이 필요한 기술 정보를 빠르게 획득하고 연구 개발을 추진하는 데 기여할 수 있다.

결론적으로, 본 연구의 IPC 자동분류 모델은 특허 분야에서의 업무 효율성 증대와 혁신 촉진을 통해 사회 전반에 걸쳐 긍정적인 영향력을 발휘할 것으로 기대된다.

References

[1] 이재안, 서형국, 한규열. (2011). 특허 문서 검색 결과를 이용한 KNN 기반의 특허 분류 시스템. *한국정보과학회 학술발표논문집*, 서울.

[2] 김기용, 박찬정. (2019). Word2Vec과 2계층 양방향 장단기 기억 네트워크를 이용한 특허 문서의 자동 IPC 분류. *한국차세대컴퓨팅학회 논문지*, 15(2), 50-60.

[3] Lee, Jieh-Sheng & Hsiang, Jieh. (2020). Patent classification by fine-tuning BERT language model. *World Patent Information*, 61, 101965.

[4] Wen, Y., Zhang, K., Li, Z., Qiao, Y. (2016). A Discriminative Feature Learning Approach for Deep Face Recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) *Computer Vision – ECCV 2016*. ECCV 2016. Lecture Notes in Computer Science(), vol 9911. Springer, Cham.

[5] 한국특허정보원. (2021). KorPatBERT: 특허 분야 한국어 사전학습 언어모델. `https://github.com/kipi-ai/korpatbert`

[6] Bekamiri, H., Hain, D. S., & Jurowetzki, R. (2024). PatentSBERTa: A deep NLP based hybrid model for patent distance and classification using augmented SBERT. *Technological Forecasting and Social Change*, 206, 123536.