

품질지표기반 정치 후원금 지원을 위한 국회의원 추천시스템 연구

정현우* · 윤형준** · 이시은*** · 박솔희*** · 손소영****†

* 연세대학교 경영학과 학사과정

** 연세대학교 천문우주학과 학사과정

*** 연세대학교 산업공학과 학사과정

**** 연세대학교 산업공학과 교수

Quality Indicator Based Recommendation System of the National Assembly Members for Political Sponsors

Jung, Hyun Woo* · Yoon, Hyung Jun** · Lee, See Eun*** · Park, Sol Hee*** · Sohn, So Young****†

* School of Business, Yonsei University

** Department of Astronomy, Yonsei University

*** Department of Industrial Engineering, Yonsei University

**** Professor of department of Industrial Engineering, Yonsei University

ABSTRACT

Purpose: During 2015–2019, the average amount of political donation to the national assembly members in Korea was 1,000 won per person. Despite its benefits such as receiving tax credits, the donation system has not been actively practiced. This paper aims to promote political donations by suggesting a recommendation system of national assembly members by analysing the bills they proposed.

Methods: In this paper, we propose a recommendation system based on two aspects: how similar the newly proposed or ammended bills are to the sponsors' interest (similarity index) and how much effort national assembly members put into those bills (intensity index). More than 25,000 bills were used to measure the recommendation quality index consisted with both the similarity and the intensity indices. Word2vec was used to calculate the similarity index of the bills proposed by the national assembly member to the sponsor's interest. The intensity index is calculated by diving the number of newly proposed or entirely revised bills with the number of senators who took part in those bills. Subsequently, we multiply the similarity index by the intensity index to obtain the recommendation quality index that can assist sponsors to identify potential assembly members for their donation.

● Received 17 November 2020, 1st revised 3 December 2020, 2nd revised 26 January 2021, accepted 22 February 2021

† Corresponding Author(sohns2_@hanmail.net)

© 2021, Korean Society for Quality Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-Commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Results: We apply the proposed recommendation system to personas for illustration. The recommendation system showed an average f1 score about 0.69. The analysis results provide insights in recommendation for donation.

Conclusion: In this study, the recommendation system was proposed to promote a political donation for national assembly members by creating the recommendation quality index based on the similarity and the intensity indices. We expect that the system presented in this paper will lower user barriers to political information, thereby boosting political sponsorship and increasing political participation.

Key Words: Political Funding, Recommendation System, Word2vec, Legislation Intensity

1. 서론

정치후원금이란 “후원회에 기부하는 금전이나 유가증권 그 밖의 물건(「정치자금법」 제3조)”을 말하며, “정치자금을 필요로 하는 자가 직접 정치자금을 받을 경우, 제공자와 제공받는 자 간에 정치자금을 매개로 각종 비리가 발생할 우려가 있으므로 후원회라는 별도의 단체를 통하여 정치자금을 조달할 수 있도록 하기 위한 제도”이다(정치후원금센터, n.d.). 후원인이 후원할 수 있는 금액은 연간 2천만 원을 초과할 수 없으며, 하나의 후원회에 500만 원 이상 후원할 수 없다. 또한, 후원 금액에 따라 10만 원 이하는 전액세액공제, 10만 원을 초과할 경우 15~25%의 세액을 공제해준다.

하지만 정치후원금을 통해 정치활동에 간접적으로 참여할 수 있고, 세액공제를 받을 수 있는 등 긍정적 효과에도 불구하고 중앙선거관리위원회의 보도자료에 따르면 2015~2019년의 대한민국 국회의원 후원금액은 350~550억 정도로 유권자당 약 1,000원 정도밖에 되지 않는 수준이다. 따라서 본 논문에서는 후원자들의 후원 동기와 후원 참여율이 저조한 이유를 분석하여 정치후원금 제도를 활성화하는 방안을 제시하고자 한다.

Welch(1974; 1980)는 기부자 유형을 ‘이익추구형 기부자’(quid pro quo contributor), ‘이념형 기부자’(ideological contributor)로 분류하였다. 그중 이념형 기부자의 경우 자신과 유사한 이념적, 정책적 입장을 가진 의원에게 후원하거나 그 후보를 당선 시키기 위해 후원한다고 하였다. 이와 유사한 연구인 Park(2014)는 기부자 유형을 선택형 기부자와 설득형 기부자로 분류하였고, 소액 후원금은 선택형 기부자가 다수를 이루며, 선택형 기부자는 자신의 선호와 유사한 성향의 정치인을 골라 후보가 정치적으로 성장할 수 있도록 지원하는 기부자라고 설명하고 있다. 즉, 소액 후원자의 다수를 이루는 선택형(이념형)기부자는 자신의 선호와 유사한 정치인에게 후원을 하는 경향이 있으므로, 후원하기 전에 자신의 선호와 유사한 정치인을 찾는 과정이 필요하다. 하지만 20대 국회 기준 국회의원들은 개인당 임기 내 평균 900개, 많게는 4,200개의 법안을 발의한다. 따라서 개인이 국회의원들의 법안을 모두 읽어보고 자신의 선호와 유사한 정치인을 찾기는 어렵다. 즉, 후원자가 후원할 국회의원을 찾는 과정에서 문제점이 발생한다.

또한, 후원자가 국회의원이 발의한 법안의 입법 형태 비율을 확인하기 어렵다는 문제점도 존재한다. Park (2013)에 따르면, 의원발의 법률안의 입법형태는 일부개정법률안이 85.8%, 제정법률안이 10.9%, 전부개정안이나 폐지법안은 2% 정도밖에 미치지 못하는 이유를 법안 발의 건수를 늘리기 위해 현행 법률의 몇 조문만을 고치거나 다른 제출 법률안을 베끼기 때문이라고 하였으며, 의원입법의 질적 제고의 필요성을 제기하였다. 즉, 입법의 질이 낮아지는 상황에서, 후원자의 선호와 얼마나 유사한 법안을 발의하는지 뿐만 아니라, 발의한 법안의 개정 수준 또한 국회의원 선택의 중요한 요소일 것이다.

따라서 본 연구에서는 법안 데이터와 참여 국회의원 데이터를 사용해 개인이 모든 법안을 살펴보지 않고도 ‘국회의원 후원 추천 품질지수’를 알 수 있는 시스템을 제시하고자 한다. ‘국회의원 후원 추천 품질지수’의 기준으로는 후원자의 선호와 유사한 법안을 얼마나 발의하였는지, 발의한 법안의 개정 수준이 어느 정도인지를 고려하고자 한다. 결과적으로, 국회의원 추천 입법품질을 후원자에게 제시함으로써 소액정치후원금을 활성화시키고자 한다.

본 논문의 구성은 다음과 같다. 2장에는 이론적 배경과 선행연구에 대해 설명하고, 3장에서 제안하는 프레임워크와 데이터에 대해 설명하였다. 4장에는 국회의원 추천 예시를 통해 제안된 시스템의 효능을 보였고, 마지막으로 5장에서는 본 연구의 결론 및 기대효과와 한계점에 대해 서술하였다.

2. 이론적 배경 및 선행연구

추천시스템에는 크게 History Data-based Recommendation(HDBR), Content-Based Recommendation(CBR), Collaborative Filtering(CF)이 있다(Shu et al., 2016). HDBR은 사용자의 과거 데이터를 활용하여 추천하며 다른 추가 정보를 사용하지 않는다. CBR은 사용자와 상품에 대한 세부정보를 바탕으로 추천을 해주는 추천 기법이다. 상품의 세부정보에는 텍스트, 그림, 오디오, 동영상 정보 등을 활용할 수 있다. CF는 사용자들이 매긴 별점 등에 기반하여 유사한 사용자들간에 상품들을 추천해주는 방식이다(Polatidis et al., 2016). 이 중에서 텍스트 등 비정형 데이터를 활용할 수 있는 CBR이 빅데이터 시대 추천시스템으로 활발하게 사용되고 있다.

구글이 2013년 발표한 word2vec은 텍스트와 같은 비정형 데이터를 다차원의 벡터로 변환시켜주어 semantic 정보들을 학습하기에 적합한 모델(Mikolov et al., 2013; Musto et al., 2015)로 CBR에 적용되고 있다. Kim (2016), 과 Lee et al.(2017)는 word2vec 모델을 문서간 유사도 분석, 감성분석 등에 사용하였다. Ma et al.(2017)은 강좌 설명(course description)을 word2vec으로 벡터화시켜 학생들에게 강좌를 추천해주는 시스템을 제시하였다. 강의의 실라버스를 벡터로 만든 뒤, 학생이 관심있는 강의와 다른 강의의 의미적 유사성(semantic similarity)을 비교하여 학생들에게 강좌를 추천해주는 시스템을 제시하였다. Ozsoy(2016)은 상품의 구매 순서를 문장으로 간주하고 word2vec으로 상품 벡터를 구하여 구매자가 기존에 구매했던 상품 과 높은 코사인 유사도를 가지는 상품들을 추천해주는 시스템을 제시하였다. Musto et al.(2015)는 word2vec을 활용한 CBR을 제시하였다. Word2vec으로 상품 특성 단어 벡터를 만든 뒤 상품 특성에 있는 모든 단어 벡터들을 합하여 상품 벡터를 만들고, 사용자가 좋아하는 상품들의 벡터를 전부 더하는 방식으로 사용자 벡터를 만들어 사용하였다. 이렇게 만든 사용자 벡터들을 활용하여 상품과의 유사도를 측정해 Top-N 상품을 추천해주는 CBR시스템을 제시하였다.

정치후원금 제도의 활성화에 관한 다양한 연구들도 진행되었다. Ji et al.(2018)은 정치 후원금에 대한 무관심의 원인으로 ‘정치자금 비리 사건 반복으로 인한 부정적 이미지와 효능감 저하’, ‘적극적인 제도 수용 노력 부족’, ‘새로운 지불방식에 대한 수용 노력 부족’ 등을 제시하고, 다양한 기술을 수용하는 제도 개선을 통한 해결책을 제시하였다. Kim(2011)은 소액기부의 저해요인으로 과도한 규제를 제시하고, 정치자금 제도의 규제 완화의 필요성을 제시하였다. 또한, 제도 개혁에 앞서 정치자금의 수요 및 공급자들의 의도와 동기를 파악하는 것이 선행되어야 한다고 주장하였다.

기존의 선행연구들에서는 정치후원금 제도의 활성화를 위해 제도적, 기술적 관점에서의 해결책을 제시하였다. 하지만 후원자들이 후원을 하는 의도, 동기 및 후원을 할 때 겪는 문제점등을 고려하지 않은 해결책을 제시하였고, 이는 후원금 제도를 근본적으로 활성화할 방안이 될 수 없다는 한계가 있다. 따라서 본 연구에서는 후원자들의 동기에 따라 후원하고자 하는 국회의원을 추천 할 수 있는 시스템을 제안하여 정치후원금 제도를 활성화하는 방안을 제시

하고자 한다. 이를 위해 후원동기 외에 후원자들의 추가적인 정보가 없는 상태에서 국회의원이 발의한 법안 내용을 word2vec으로 분석하여 후원 추천에 활용 할 수 있는 content-based recommendation 모델을 사용하였다.

3. 제안 방법

3.1 데이터 구성 및 전처리

본 연구에서는 ‘열려라 국회’ 사이트 크롤링을 통해 20대 국회 국회의원 295명(2020.4 기준)이 발의한 모든 법안과 국회의원의 인구통계학적 데이터(이름, 소속정당, 소속위원회)를 사용하였다. 법안은 21,447개로 일부개정법률안 93.12%, 전부개정법률안 0.41%, 제정법률안 4.92%, 폐기법률안 1.55%로 구성되어있다.

Table 1. The number of co-proposers by bills types

	Under 10 proposers	10~19 proposers	20~30 proposers	30~40 proposers	more than 40 proposers
Revised bills	2 (0.01%)	19,036 (95.31%)	649 (3.25%)	161 (0.81%)	125 (0.63%)
Newly proposed bills	3 (0.28%)	866 (82.01%)	86 (8.14%)	38 (3.60%)	63 (5.97%)
All amendments bills	0 (0.00%)	68 (78.16%)	6 (6.90%)	5 (5.75%)	8 (9.20%)
Abolished bills	4 (1.21%)	298 (90.03%)	16 (4.83%)	2 (0.60%)	11 (3.32%)

입법형태에 따라 법안들의 공동발의 의원 수분포는 Table 1과 같다. 주로 10-19명 정도의 의원이 공동발의 하는 것으로 나타났으며, 제정법률안, 전부개정안과 같이 법안을 새로 만들거나 대부분을 수정해야 하는 경우 다른 법안들에 비해 공동으로 발의하는 의원 수가 많은 것을 확인할 수 있다.

법안 토큰화에는 사용자 사전을 추가할 수 있는 Mecab Tokenizer를 사용하였다. Mecab은 한국어 문장을 토큰화하기 위해 널리 사용되는 형태소 분석기 중 하나다(Kim, 2020). 사용자 사전은 단어를 원하는 품사로 지정해 주는 것으로, 사용자 사전 추가를 통해 토큰화 과정에서 단어 본연의 뜻을 잃어버리는 것을 방지할 수 있다. 예를 들면, 사용자 사전을 추가하지 않고 단어를 토큰화할 경우 ‘지방자치단체’의 경우에는 ‘지방’, ‘자치’, ‘단체’로 토큰화 되고, ‘헌법재판소’는 ‘헌법’, ‘재판소’로 토큰화가 된다. 이 경우 ‘지방자치단체’와 ‘헌법재판소’ 단어의 본래의 의미를 잃어버리게 된다. 따라서 사용자 사전에 ‘지방자치단체’, ‘헌법재판소’를 고유명사로 추가하여 토큰화를 방지함으로써 본래 단어의 의미를 잃어버리지 않도록 하였다. 사용자 사전에는 약 2,100개 단어를 추가하였다. 또한, ‘을’, ‘를’, ‘이’, ‘가’, ‘하였’, ‘졌으며’, ‘한다’ 등 조사와 같은 의미전달에 불필요한 단어들을 제거하였다.

3.2 분석 방법

Figure 1에 정리된 국회의원 추천 과정은 다음과 같다. 우선, 법안 데이터를 word2vec 모델로 벡터화 시키고 각 국회의원이 발의한 법안을 기준으로 국회의원 벡터를 만들었다. 유사도 측정을 위해 후원자가 문장으로 입력한 관심사로 후원자 벡터를 만들고 후원자 벡터와 국회의원 벡터들 간의 코사인 유사도를 계산한다. 이때, 코사인 유사도가 국회의원과 후원자의 ‘유사도 지수’가 된다. ‘입법 참여 지수’는 각 국회의원의 발의 법안 개정 수준을 나타내는 값으로, 입법의 질이 높다고 판단되는 제정법률안, 전부개정안의 합을 각 법안의 공동 발의 의원수로 나누어 계산하였다. 이후, 0과 1사이의 값을 가지도록 minmax scaling 하였다. 이때, 개정 수준은 제정 법률안, 전부개정안 비율을 기준으로 한다. 최종적으로, ‘입법 참여 지수’와 ‘유사도 지수’를 곱해 ‘국회의원 후원 추천 품질지수’를 계산해 후원자에게 품질 지수 상위 5명을 추천해준다.

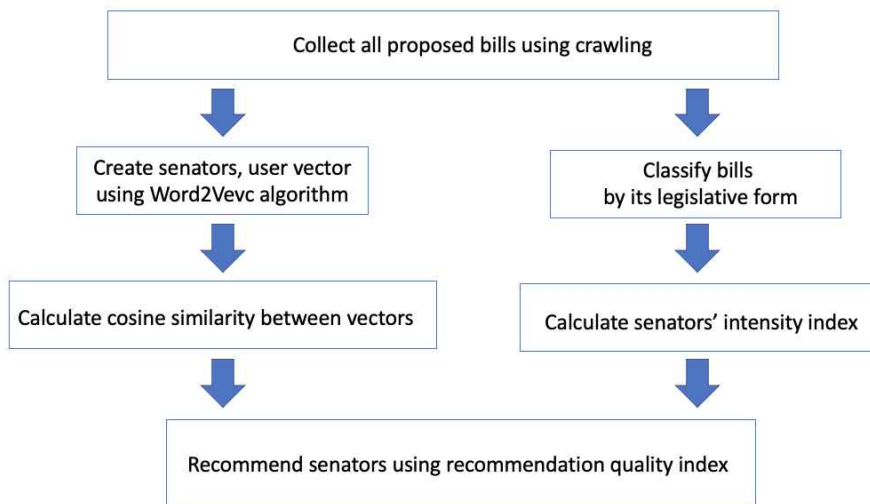
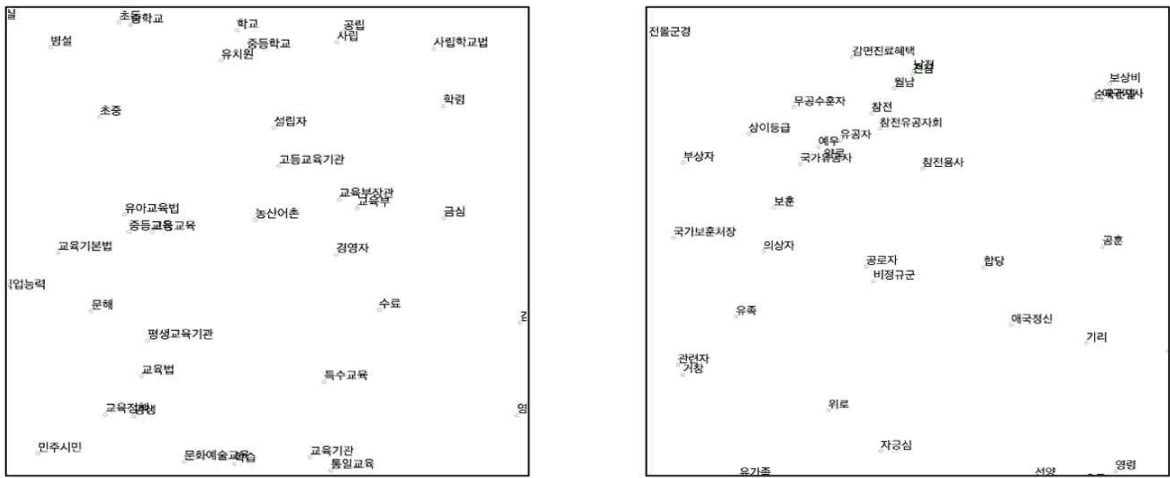


Figure 1. Model Framework

3.3 분석 과정 및 결과

전처리된 단어데이터는 word2vec을 이용해 벡터로 변환하였다. 총 100차원으로 벡터를 생성하였고, window size는 3으로 설정하였다. 또한, 최소한 10번 이상 등장한 단어에 한해서만 학습을 진행하였고 skip-gram을 사용하여 9,000번 학습을 시켰다. 최종적으로 약 10,000여 개 단어를 벡터화 하였으며, 벡터화된 100차원의 단어 벡터를 t-sne를 이용해 차원축소하여 2차원 상에 나타낸 결과는 Figure 2와 같다. (a)는 교육, 학교와 관련된 단어들이 모여 분포하고 있고, (b)는 보훈과 관련된 단어들이 모여 분포하고 있다. 이를 통해 서로 유사한 의미를 가지는 단어들끼리 유사한 벡터값을 가지도록 잘 학습이 되었음을 알 수 있다.



(a) Education, school

(b) Veteran

Figure 2. Words distribution on 2D space

앞서 구한 단어 벡터를 활용하여 법안 벡터를 생성하였다. 법안 벡터는 법안에 들어가는 모든 단어의 벡터를 더하여 계산하였다. 동일한 방식으로 국회의원이 임기 동안 발의한 모든 법안의 벡터를 합해서 국회의원 벡터를 만들었다. 만들어진 국회의원의 벡터를 t-sne로 2차원으로 차원을 축소하여 시각화한 결과는 Figure 3과 같다.

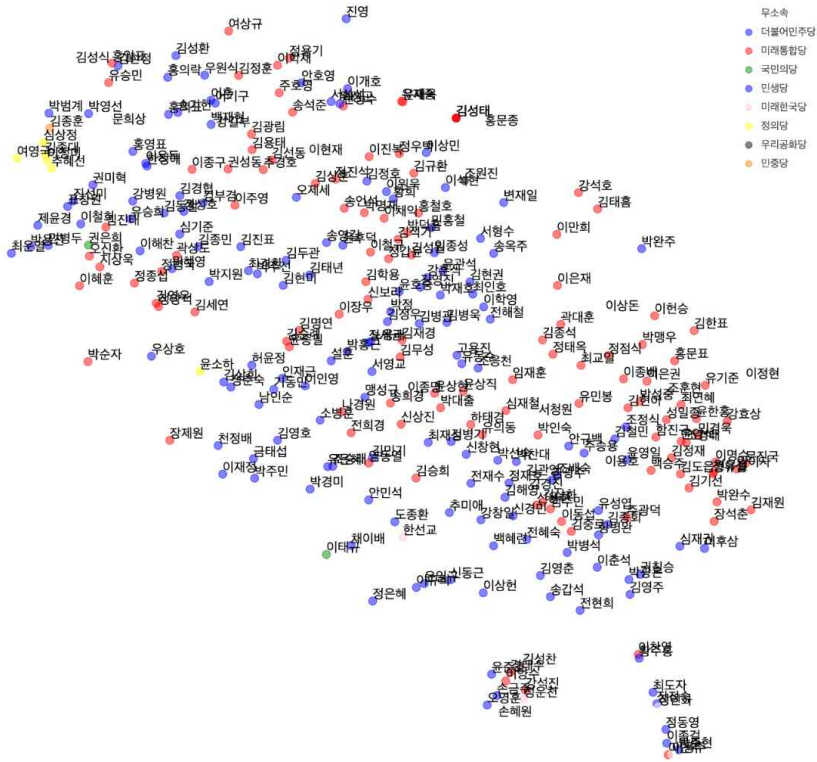


Figure 3. Distribution of members of congress on 2D (Each color stands for political party)

Figure 3은 Park & Jang(2017)에서 제시한 공동 발의 법안을 기준으로 네트워크 분석을 한 결과와는 다른 양상을 보였다. 네트워크 분석은 특정 이해와 연관되어 네트워크로 묶을 수 있는 개인, 집단, 사회의 상호 관계를 파악하기 위하여 활용되는 기법이다(Choi et al., 2019). 공동 발의를 하는 의원들 간의 네트워크는 같은 정당 의원끼리 가까운 곳에 위치하였지만, 법안의 내용을 바탕으로 시각화한 Figure 3은 정당을 제외하고 서로 다른 정당의 의원들이 섞여서 분포하고 있는 것을 확인할 수 있다. 이는 국회의원의 관심분야는 소속된 당파 관련이 적음을 의미한다.

추가적으로 가까운 공간에 위치하는 국회의원들의 발의 법안 특징을 확인해보기 위해서 국회의원 벡터를 2차원에 표현한 Figure 3을 clustering 한 뒤, 각 cluster 별로 word cloud를 만들어서 군집분석을 비교해보았다. 군집분석은 데이터 중에서 유사성을 가지는 집단으로 군집하여 각 군집의 유형별 특성을 파악하여 데이터 구조를 이해하는 분석기법이다(Koh et al., 2019). cluster의 개수는 Dendrogram의 y축(closeness of either individual data points or clusters)을 25 수준으로 유지하면서 가장 높은 silhouette score를 보인 13개를 선택하였다. cluster의 개수를 13개로 설정한 뒤 k-means clustering을 시행하여 시각화한 결과는 Figure 4와 같다.



Figure 4. Distribution of members of congress on 2D
(Each color stands for cluster number)



Figure 5. Word Cloud of Cluster 1, 4, 6

Fig 5는 왼쪽부터 1, 4, 6번 cluster에 속한 국회의원들끼리 같이 발의한 법안의 word cloud를 보여준다. Word cloud는 텍스트로부터 지식을 발견하고 추출하는 텍스트마이닝 기법 중 가장 단순한 기법이다(Ree, 2019). Word cloud를 통해 각 Word cloud의 키워드들을 살펴보면 cluster 1은 기간제 근로자, 직장 성희롱, 근로자 권익이 주 키워드로 나타났다. cluster 1의 입법 참여 지수는 0.443로 전체 평균(0.49)보다는 낮은 수준이었다. 이 cluster에는 유승민 의원, 박영선 의원, 오신환 의원, 표창원 의원과 정의당 의원 전원이 속해있다. cluster 4은 여성 경제, 여성인사처우 개선, 유리 천장 현상 해소 등을 주 키워드로 법안을 발의하는 cluster이다. cluster 4의 입법 참여 지수는 0.448이었다. 이 cluster에는 이언주 의원, 최도자 의원, 정동영 의원, 이찬열 의원 등이 속해있었다. cluster 6는 교육, 학교, 문화예술교육 등을 주요 키워드로 가진다. 이 cluster의 입법 참여 지수는 0.54로 전체 평균보다는 높은 수준이다. 이 cluster에는 나경원 의원, 유은혜 의원, 이태규 의원, 채이배 의원이 속해있다.

이 외에도 다른 cluster들의 word cloud를 살펴보면 공공기관, 경제, 산업과 관련된 법안을 많이 발의한 cluster 5, 청문회에서 후보자의 결격사유를 지적하는 법안을 많이 발의한 cluster 12, 의료 시스템과 관련된 법안을 많이 발의한 cluster 13 등 각 cluster별로 발의하는 법안의 주제가 상이했다. 이와 같이 cluster별로 발의하는 법안의 개정 수준, 법안의 성격이 확연하게 구분되는 것은 본 연구가 제안하는 모델이 후원자에게 유의미한 국회의원 추천을 해줄 수 있다는 사실을 보여준다.

추가적으로 서로 비슷하지만 반대의 의미를 가지는 사용자 관심사 투입에 따른 추천의 적합성 판단을 위해 서로 의미가 반대인 사용자 투입을 넣어 결과값을 확인하였다. 사용자 관심사 투입은 ‘저는 정부가 법인세 등을 감면해주고, 기업 투자를 확대할 수 있게 혜택 등을 주어 기업하기 좋은 환경을 만들어야 한다고 생각합니다. 투자를 할 때 세액공제를 확대하거나 혹은 법인세지원을 해주는 법안을 지지합니다.’와 ‘저는 정부가 기업 세금을 깎아주는 것에 반대합니다. 대신 기업 규제를 더 강력하게 해야 한다고 생각합니다. 사회적으로 소득불평등이 점점 심해지는 상황에서 양극화 문제를 해결하기 위해서는 기업에 대한 규제와 규율을 강화하여 부당하게 부를 축적하지 못하게 해야

한다고 생각합니다.’ 두 가지로, 각각 기업의 규제 완화, 강화를 원하는 반대의 의미를 가지는 사용자 투입으로 구성하였다. 두 사용자 투입과 코사인 유사도가 높은 법안들을 살펴본 결과는 Table 2와 같다.

Table 2. Similar Bills by Different User Input

	name of bill	cosine similarity		name of bill	cosine similarity
Enterprise friendly users	Restriction of Special Taxation Act	90.23%	Users who are not enterprise friendly	Trade Union and Labor Relations Adjustment Act	85.28%
	Restriction of Special Taxation Act	89.97%		Capital Markets and Financial Investment Business Act	83.95%
	Restriction of Special Local Taxation Act	89.96%		Corporate tax Act	83.44%

각 법안의 세부 내용을 살펴본 결과 <조세특례제한법>, <지방세특례제한법> 등은 세액공제 기한을 연장하거나, 기업의 투자를 촉진하기 위한 법안들인 반면 <자본시장과 금융투자업에 관한 법률>, <법인세법> 등은 법인세 인상, 기업의 사회적 책임을 강조하는 법안들이었다. 이처럼 사용자의 투입이 서로 반대되면 유사도가 높은 법안의 성격 또한 달라지는 것을 확인할 수 있다.

결과적으로, 법안의 내용 기반 국회의원 분석을 통해 국회의원의 정당과는 관련 없이 오로지 해당 국회의원의 입법 참여 지수와 입법주제에 따라 국회의원의 특징을 파악할 수 있게 된다. 이를 통해 사용자는 개인의 관심사와 가장 잘 매칭되는 국회의원을 발의하는 법안의 주제와 입법 참여 지수에 따라 추천 받을 수 있다.

4. 국회의원 추천 예시

구현된 시스템 평가를 위해 두 가지 페르소나 시나리오를 설정하여 실험하였다. 첫 번째 페르소나는 ‘장애아이를 둔 부모’로 관심사 사용자 투입으로 ‘저는 장애인 아이를 키우고 있는 부모입니다. 아이가 더 좋은 복지서비스를 받아 한 사회의 구성원으로서 살아갔으면 합니다. 특수교육과 장애인복지시설 등에 관심이 있습니다.’를 입력하였다. 추천 결과 이찬열, 김철민, 신창현, 황주홍, 박정 의원이 추천되었다. 해당 의원들이 발의했던 법안을 살펴보면 ‘장애아동 복지지원법’, ‘발달장애인 권리보장 및 지원에 관한 법률’, ‘아이돌봄 지원법’ 등이 있다. 각 국회의원 별 유사도 지수와 입법 참여 지수, 이를 종합적으로 고려한 국회의원 후원을 위한 추천 품질지수는 Table 3과 같다.

Table 3. Example of Recommended Result 1

Name of senators	Similarity Index	Intensity Index	Recommendation Quality Index
Lee Chan-yeol	93.54	99.66	93.22
Kim Chul-min	91.89	97.96	90.02
Sin Chang-hyun	91.29	98.30	89.74
Hwang Ju-hong	98.33	99.32	88.73
Park Jung	87.83	98.98	86.94

두 번째 페르소나는 ‘대학 졸업 후 창업을 하고자 하는 청년’으로 사용자 투입으로 ‘저는 대학을 졸업한 청년입니다. 대학 졸업하고 소셜 벤처를 창업하고자 합니다. 소셜 벤처와 청년창업을 지원해주셨으면 합니다.’를 입력하였다. 추천 결과 정인화, 김상희, 이찬열, 유성엽, 노웅래 의원이 추천되었다. 해당 의원들이 발의했던 법안을 살펴보면 ‘청년창업 활성화 및 청년창업기업 지원에 관한 특별법안’, ‘청년창업기업 육성 및 지원에 관한 특별법안’ 등이 있다. 각 국회의원의 유사도 지수와 입법 참여 지수, 이를 종합적으로 고려한 국회의원 후원을 위한 추천 품질지수는 Table 4와 같다.

Table 4. Example of Recommended Result 2

Name of senators	Similarity Index	Intensity Index	Recommendation Quality Index
Jeong In-Hwa	83.24	94.91	79.18
Kim Sang-hee	75.74	96.61	73.17
Lee Chan-yeol	73.18	99.66	72.93
Yoo Sung-yeop	75.20	96.27	72.39
No Ung-rae	73.58	97.62	71.83

추천모델의 성능을 평가해보기 위해 입법모니터링 서비스 ‘캣벨’에서 제공하는 법안 주제 중 10가지를 선택해 모델을 검증해보았다. 법안 주제로는 직장 내 성희롱 예방, 신재생 에너지 확대, 사회적 참사 진상규명, 어려운 한자어 쉽게 바꾸기, 군인 인권보호, 폐기물 관리, 범죄인 인도법, 콘텐츠 산업 진흥 등 10가지를 선정하였다. 성능 평가를 위해 사용자가 입력한 문장과 관련된 법안을 발의한 의원 중 발의 건수가 median 이상인 N명을 실제 참 값으로 설정하고, 국회의원 후원 추천 품질 지수 상위 N명을 예측 참 값으로 설정하였다. 본 연구의 데이터는 참 값과 거짓 값이 불균형 하므로 성능 평가 지표로는 f1 스코어를 사용하였다. 결과적으로 10개의 주제에 대해 평균 0.69의 F1 스코어를 얻을 수 있었다.

이처럼 본 연구는 궁극적으로 ‘국회의원 후원 추천 품질’을 알려줄 수 있는 시스템을 제안하고자 했다. 국회의원 후원 추천품질은 크게 후원자의 관심사와 얼마나 유사한 법안을 발의하는지, 발의한 법안들의 입법 참여 지수를 종합적으로 고려하였다. 후원자의 관심사가 뚜렷하지 않은 사람들도 간단한 단어를 키워드를 입력함을 통해서 자신과 관심사가 유사한 국회의원을 추천받을 수 있고, 후원자는 추천받은 국회의원의 법안이 후원자의 관심사와 얼마나 유사한지, 발의하는 법안에 참여도는 어느 정도인지 가늠해 후원할 때 필요한 정보를 제공받을 수 있다.

Lee(2019)는 정치참여에 영향을 미치는 요인으로 정치관심, 정치정보, 정치효능감을 뽑았다. 본 연구는 국회의원의 입법 내용을 손쉽게 제공해줌으로써 정치 정보에 대한 장벽을 낮춰 대중들의 정치 관심도를 높이고, 정보의 접근성을 높일 수 있을 것이다. 이렇게 정치 참여도가 높아지면 자연스럽게 정치후원에 대한 관심이 높아지고, 대중들은 본인의 관심 분야에서 묵묵히 힘써주는 국회의원들의 의정 활동에 후원이 늘 것이다. 후원을 받은 의원들이 자신의 전문분야에서 더 적극적으로 활동한다면 법안의 혜택은 곧 후원한 국민에게 돌아오니 정치 후원을 통해 선순환 구조가 형성되고, 전반적인 국회의원의 입법품질 개선에 힘을 실어줄 것이다.

5. 결론 및 시사점

본 연구는 소액 후원자의 다수를 이루는 선택형(이념형) 기부자가 자신의 선호와 유사한 정치인을 찾는 과정에서

생기는 어려움을 해결하여 정치후원금 제도의 저조한 참여율을 개선하고자 하였다. 기존 연구들이 제도적, 기술적 해결책을 제시하는데 그쳤던 반면, 본 연구에서는 국회의원들이 발의한 법안과 후원자의 선호분야 유사도와 국회의원의 법안발의 참여도를 고려한 품질지표를 바탕으로 국회의원을 추천해줌으로써 후원자가 후원을 할 때 겪는 근본적인 문제를 해결하고, word2vec을 활용한 추천시스템을 입법과 정치후원이라는 새로운 도메인에 적용하여 그 가능성을 확인했다는 점에서 의의가 있다.

기존의 법안 빅데이터는 공동 발의, 키워드 검색 등의 한정된 방법으로만 이용되어왔다. Park & Jang (2017)에서 제시한 공동 발의 법안을 기준으로 한 네트워크 분석의 결과에서는 정당 간의 분포만 확인할 수 있었던 반면, 본 연구에서 제시한 국회의원들의 벡터 분포에서는 의원들의 발의법안 주제별 분포를 확인할 수 있었다. 유권자에게 이와 같은 법안 중심의 정보를 제공한다면 기존의 정당 정치, 색깔 정치가 아닌 법안이 중심이 되는 정치로의 변화를 이끌어 낼 수 있을 것이다. 즉, 법안 데이터의 텍스트 분석 연구가 지속적으로 이루어진다면 지역주의에서 벗어난 정치 선진화를 이룰 수 있을 것이라 기대한다.

본 연구는 소액 후원금 활성화를 위해 후원자가 직관적으로 이해할 수 있는 국회의원 후원 추천 품질지표를 제시하였다. 품질지표에서는 후원자의 관심사와 얼마나 유사한지, 국회의원이 얼마나 열심히 법안발의에 참여를 하는지를 복합적으로 고려하였다. 또한, 품질 지표를 기준으로 후원자에게 국회의원 추천을 해줌으로써 후원자가 후원 대상을 찾는 과정에서 발생하는 어려움을 해결해 소액 정치후원금 활성화에 기여할 수 있을 것이라고 기대한다.

본 연구는 위와 같은 기대 효과가 있지만 몇몇 한계 또한 존재한다. 첫 번째 한계는 입법 참여 지수 도출 과정에서의 한계이다. 본 연구에서는 제정법률안, 전부개정안의 개수와 평균 공동발의 의원수를 이용해 입법 참여 지수를 계산하였다. 이는 각 법안에서 의원별 기여도를 반영하지 못했다는 한계가 있다.

두 번째 한계는 입법 데이터만 반영하여 평소 의원들의 성향을 완벽하게 반영하지 못한다는 점이 있다. 향후 의원들의 회의록 및 SNS 데이터와 같이 의원들의 성향을 파악할 수 있는 데이터를 추가적으로 학습 시켜 이와 같은 한계를 보완해 나갈 수 있을 것이다.

REFERENCES

- Caselles-Dupre, H., Lesaint, F., and Royo-Letelier, J. 2018. Word2vec applied to recommendation: Hyperparameters matter. *RecSys* 18:352-356.
- Central Election Commission, Pree Release(2018. 2. 27), 「Disclosure of fundraising details for the 2017 Central Party Sponsorship Committee and the National Assembly Sponsorship Committee」.
- Central Election Commission, Pree Release(2019. 2. 26), 「Disclosure of fundraising details for the 2018 Central Party Sponsorship Committee and the National Assembly Sponsorship Committee」.
- Central Election Commission, Pree Release(2020. 2. 27), 「Disclosure of fundraising details for the 2019 Central Party Sponsorship Committee and the National Assembly Sponsorship Committee」.
- Choi, J. J., Chung, S. S., and Kim, K. S. 2019. A Study on the Standards Evaluation Method Using Standards Networks. *The Journal of Korean Society of Quality Management* 47(2):315-325.
- Ji, J., Cho H., and Kim S. 2018. A Study of Policy Conditions for the Popularization of Digital Grassroots Donations and Micro Fundraising. *Journal of Humanities* 47(0):97-123.
- Kim, J. D. 2011. The Effect of Small Donation and Factors of Causing Its Negative Effects. *The Korean Journal of Area Studies* 29(1):134-158.

- Kim, W. 2020. Comparative Study of Tokenizer Based on Learning for Sentiment Analysis. *The Journal of Korean Society of Quality Management* 48(3):421–431.
- Kim, W., Kim, D., and Jang, H. 2016. Semantic Extension Search for Documents Using the Word2vec. *The Journal of the Korea Contents Association* 16(10), 687–692.
- Koh, K. J., Kim, N. Y., and Kim, Y. S. 2019. Improving the Yield of Semiconductor Manufacturing Processes using Clustering Analysis and Response Surface Method. *The Journal of Korean Society of Quality Management* 47(2):381–396.
- Lee, C., Yoo, K. H., Mun, B. M., and Bae, S. J. 2017. Informal Quality Data Analysis via Sentimental analysis and Word2vec Method. *The Journal of Korean Society Quality Management* 45(1):117–127.
- Lee, J. C. 2019. Comparison of Political Participation: Conventiojnal and Unconventional Political Participations in the Metropolitan Area of Korea Vol.12 No.3: 35–67.
- Ma, H., Wang, X., Hou J., and Lu Y. 2017. Course recommendation based on semantic similarity analysis. 2017 3rd IEEE International Conference on Control Science and Systems Engineering (ICCSSE) pp.638–641.
- Mikolov, T., Sutskever, I., Chen, L., Corrado, G. S., and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems* 26, pp.3331–3119.
- Musto, C., Semeraro, g., De Gemmis, M., and Lops, P. 2015. Word Embedding Techniques for Content-based Recommender Systems: An Empirical Evaluation. In *recsys posters*.
- Ozsoy, M. G. 2016. From Word Embeddings to Item Recommendation. *Arxiv*, abs/1601.01356.
- Park, C., and Jang, W. 2017. Cosponsorship networks in the 17th National Assembly of Republic of Korea. *The Korean journal of applied statistics* v.30 no.3, 2017, pp.403 – 415.
- Park, H. S. 2014. Motivation of Political Donations for Legislators: Supporters' Associations in the 18th National Assembly of Korea. *Korean Journal of Legislative Studies* 43(0):68–88.
- Park, S. H. 2013. Study of Legislation Evaluation in Legislation by the National Assembly. Master's thesis., Ewha Womans University.
- Polatidis, N., and Georgiadis, C. K. 2016. A multi-level collaborative filtering method that improves recommendations. *Expert Systems with Applications*, 48, 100–110.
- Ree, S. 2019. Analysis of Research Trends in Journal of Korean Society of Quality Management by Text mining Processing. *The Journal of Korean Society of Quality Management* 47(3):597–613.
- Shu, J., Shen, X., Liu, H., Yi, B., and Zhang, Z. 2018. A content-based recommendation algorithm for learning resources. *Multimedia Systems*, 24(2), 163–173.
- Welch, William P. 1974. "The Economics of Campaign Funds." *Public Choice*. Vol. 20, winter. 83–87.
- Welch, William P. 1980. "Allocation of Political Monies: Economic Interest Groups." *Public Choice*. Vol. 35, No. 1. 97–120.

저자소개

정현우 연세대학교에서 경영학을 전공하고 있다. 연구 관심분야는 계량 마케팅, 빅데이터 마이닝, 자연어처리이다.

윤형준 현재 연세대학교 산업공학과 석사과정이며, 연구분야는 텍스트 마이닝이다.

이시은 연세대학교 산업공학과에 재학 중이며, 연구 관심분야는 데이터 엔지니어링, 백엔드 개발, 스파크 앱 개발, 쿼리 옵티마이저 등이다.

박슬희 연세대학교에서 산업공학과에 재학 중이며, 연구 관심분야는 데이터 분석, 빅데이터 관리, 품질 경영, 시장 분석, 머신러닝 등이다.

손소영 연세대학교 산업공학과 교수로 재직하고 있으며 최근 관심분야는 공간 빅데이터 마이닝, 기술경영, 품질공학으로 상세 정보는 <http://isl.yonsei.ac.kr>에서 확인할 수 있다.