

# 주제어 문장거리를 이용한 뉴스 편향성 분석 그래프 학습

조찬우<sup>o</sup> 조찬형  
부산대학교

kategosari@gmail.com, agamiboy@naver.com

## Graph Learning System for Analyzing Bias among News Using Keyword Distance Model

Cho Chanwoo<sup>o</sup> Cho Chanhung  
Pusan National University

### 요 약

문서에서 저자의 의도와 주제, 그 안에 포함된 감성을 분석하는 것은 자연어 연구의 핵심적인 주제이다. 이와 유사하게 특정 글에 포함된 정치적 문화적 편향을 분석하는 것 역시 매우 의미 있는 연구주제이다. 우리는 최근 발생한 한 사건에 대하여 여러 신문사와 해당 신문사에서 생산한 기사를 중심으로 해당 글의 정치적 편향을 정량화 하는 방법을 제시한다. 그 방법은 선택된 주제어들의 문장 공간에서의 거리를 중심으로 그래프를 생성하고, 생성된 그래프의 기계학습을 통하여 편향과 특징을 분석하였다. 그리고 그 그래프들의 시간적 변화를 추적하여 특정 신문사에서 특정 사건에 대한 입장이 시간적으로 어떻게 변화하였는지를 동적으로 보여주는 그래프 애니메이션 시스템을 개발하였다. 실험을 위하여 최근 이슈에 대하여 12개의 신문사에서 약 2000여 개의 기사를 수집하였다. 그 결과, 약 82%의 정확도로 일반적으로 알려진 정치적 편향을 예측할 수 있었다. 또한, 학습 데이터에 쓰이지 않은 신문기사를 활용하여도 같은 정도의 정확도를 보임을 알 수 있었다. 우리는 이를 통하여 신문기사에서의 정치적 편향은 작성자나 신문사의 특성이 아니라 주제어들의 문장 공간에서의 거리 관계로 특성화할 수 있음을 보였다.

할 수 있다.

### 1. 연구동기

본 연구에서는 몇 특정 주제어를 기준으로 그 주제어들이 특정 기사에서 나타난 “거리”의 멀고 가까움의 정도를 이용하여 그래프(graph)를 생성하는 방법을 이용하여 각 주제어를 정점(vertex)로, 주제어 간의 상관 정도를 간선(edge)의 가중치로 표시하는 그래프를 생성한다. 이 방법은 이미 이전 연구에서 시도된 바가 있는 방법이다. 우리는 이렇게 생성된 주제어 상관 그래프(Keywords Relation Graph)를 기본 데이터로 활용하여 그래프 학습(graph learning)을 통하여 각 주제어 그래프의 진보, 보수 정도를 레이블로 학습하는 모형을 구성한다.

우리는 주제어 그래프의 학습 모형으로 특정 기사의 정치적 지향을 판단하는 방법을 새롭게 제시하고자 한다. 이 실험을 위하여 최근 정치적으로 크게 보도된 “양평 고속도로 변경안”을 주제로 선정하였다. 이 실험을 위하여 위 주제와 관련하여 수집된 1000여 개 이상의 진보, 보수 기사를 대상으로 검증한 결과 82%이상의 정확도(accuracy)가 가능함을 보인다. 또한 학습 과정에 전혀 사용되지 않은 신문사의 기사를 대상으로 실험한 결과 보수 진보의 평가 측도가 일반적인 평가와 82%이상 일치함을 보일 수 있었다. 이는 신문 기사의 진보, 보수 편향은 기사에 사용된 주요 단어의 문장 거리상으로의 군집의 정도에 유의미하게 반영됨을 말해 주는 것으로 해석할 수 있다. 즉 진보나 보수가 즐겨 사용하는 주제어들의 출현 빈도와 패턴에 일정한 형식으로 나타남을 확인해준다고 할 수 있다. 이를 확장하면 정치적인 사안 뿐만 아니라 다양한 문화 현상 등에 나타난 편향성도 충분히 구분 가능함을 보인다고

### 2. 관련 연구

현재 문서의 편향성을 이용해서 문서의 특성을 판단하는 연구로는 단순히 사용된 단어 자체의 뜻이나 용도에 집중한 연구[4], 데이터 증강을 이용해서 하는 연구[1], 합성신경망과 순환신경망을 이용해서 문장의 정치적 편향성을 판별 해주는 연구[3]가 있었지만, 기사 자체의 편향성에 대해서는 정확도가 비교적 낮게 나왔었다. 우리는 이를 보완하기 위해서 중요한 단어들 간의 거리를 이용한[6] 연구를 참고하여 단어-거리 모형을 그래프 학습[5]를 이용해서 만들 고자한다.

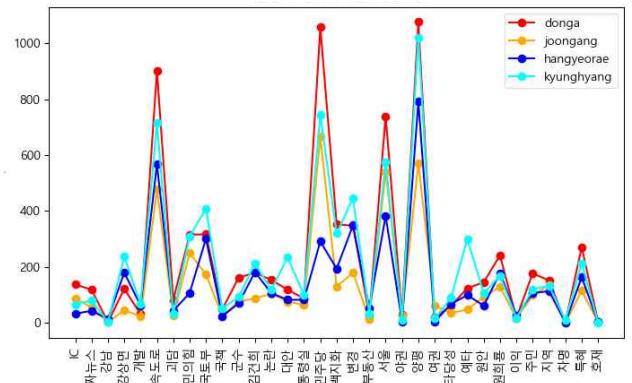


그림1. 신문사 별 단어의 출현 빈도

그림1은 각 언론사 별로 관련 기사들의 총 단어 출현 빈도를 분석한 결과이다. 이때 대부분의 단어의 빈도가 진보,

보수 가릴 것 없이 비슷하게 출현한다. 단순하게 단어 출현 빈도를 이용해서는 기사의 정치적 편향성을 특정 짓기에는 힘들 수 있다. 이를 해결하기 위해서 단어의 빈도 수가 아닌 문장 공간 안에서의 단어와 단어 간의 거리를 관계로 한 새로운 모델을 이용해야 한다.

또한 이를 설명하기 위해서 문서를 가시화 하고자 한다 문서 가시화 방법은 매우 다양한데 이들을 정리한 연구결과가 Web으로 제공되고 있다. 본 연구와 같은 방향으로 문맥의 내용을 고려한 텍스트 가시화에 관한 연구가 제시되었다[7]. 그러나 이 연구에서는 주제어를 특정하지 않았기 때문에 그래프 구조라기 보다, 클라우드 형식으로 제시되어 세부적인 내용을 관찰하기에는 어려움이 있다.

### 3. 신문기사의 동적 변화와 특성 분석

우리는 Keyword Relation Graph의 애니메이션을 통하여 각 기사의 주제가 어떻게 변화하는지를 가시화하려고 한다. 이를 위해서 우리는 주목해야 하는 주제어를 자동으로 선택하고 이를 가시화의 주요 대상으로 선정하여 그 변화에 집중이 가능한 새로운 애니메이션 시스템을 보인다. 앞서 우리는 두 주제어의 관련 정도  $weight(i, j | G)$ 를 계산하는 방법을 설명하였다. 그런데 각 주제어의 중요도는 그 주제어와 연관된 이웃 주제어들과의 weight의 합으로 표현할 수 있다. 즉 만일  $k_1$ 와 edge로 연결된 이웃 주제어가  $\{k_2, k_4, k_9\}$ 와 같이 3개라고 할 때 그래프  $G$ 에서  $k_1$ 의 중요도, 즉  $Kwt(i | G)$ 는 아래와 같이 정의한다. 아래 식에서  $N(i | G)$ 는 그래프  $G$ 에서  $i$ 의 모든 neighbor vertex 집합을 의미한다. 즉 이 값은 해당 vertex의 인접 모든 edge의 weight의 합이다.

$$Kwt(i | G) = \sum_{u \in N(i | G)} weight(u, i | G)$$

vertex  $i$ 의  $Kwt(i | G)$ 는 시점에 따라서 항상 변화한다. 이 값을 시간 기준으로 볼 때 특정 시간 동안 가장 많이 증가한 vertex(keyword)는 새롭게 나타난 주제어라고 볼 수 있고, 가장 많이 감소한 주제어는 더 이상 기사에서 언급되지 않는, 즉 중요도가 떨어진 주제어라고 볼 수 있다. 만일 시점  $t_1$ 에서 주제어의 weight를 위 식을 확장하여  $Kwt(i | G, t_1)$ 로 표현할 때 특정 시간 동안 변화한 가중치 변화도  $Change(i | t_1, t_2)$ 는 다음과 같이 계산된다.

$$Change(i | t_1, t_2) = (Kwt(i | G, t_2) - Kwt(i | G, t_1)) / |t_1 - t_2|$$

이 기준으로  $t_e$  시점을 기준으로  $t_s$ 까지의 기간 중에서 변화값이 최고인  $k_i$ 를 선택한다. 축구에서 골을 넣은 선수를 특정 한 뒤 그 시점 이전부터 그 선수의 움직임을 관찰할 수 있도록 표시하는 것과 같은 방식이다.

아래는 4개의 주요 키워드가 특정 2개의 신문사에서 나타난  $Kwt(i | G)$  값의 변화를 표시한 그래프이다. 그림에

서 주제어 “강상면”은 갑자기 중요도가 증가하고 있음을 보여주고 지역은 처음에 보여준 중요도가 시간이 갈수록 줄어들고 있음을 보여준다.

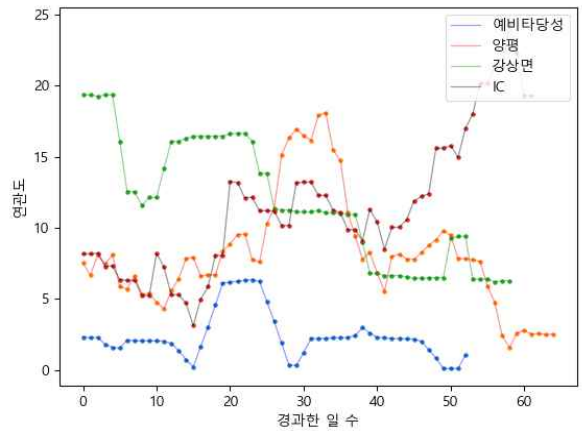


그림2. 조선일보의 4개의 주제어 변화 그래프

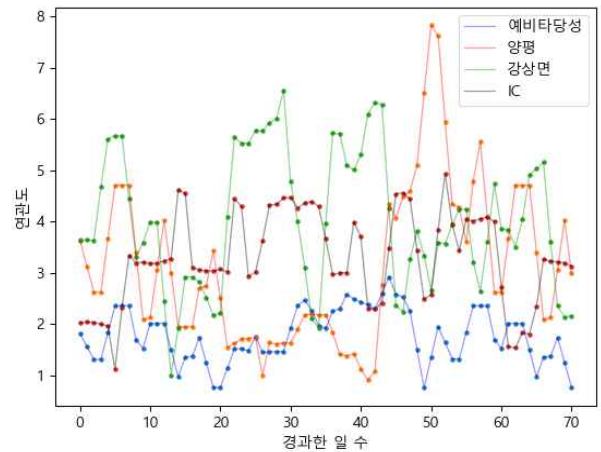


그림3. 경향신문의 4개의 주제어 변화 그래프

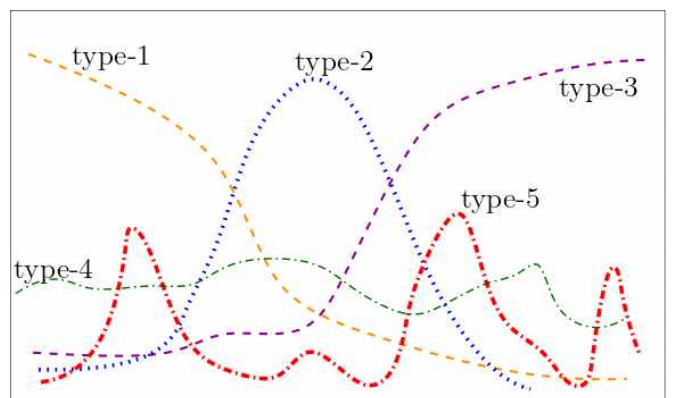


그림4. 문장 가시화의 예시

이전 문장의 가시화는 단순히 문장에 나타난 단어의 빈도(frequency)만을 이용하여 이를 단어 구름(word cloud)이라는 단순 모형으로만 제시하였다. 그러나 특정 사안에 대한 개체(개인, 신문사, 정파)의 입장은 시간에 지남에 따라서

변화 함을 알 때, 이 동적인 과정(dynamic process)을 가시화하는 것은 중요한 문제이다. 즉 어떤 주제가 점점 더 중요하게 대두되는지 또는 어떤 주제어는 주요 이슈에서 사라지는지를 파악하는 것이 더 중요한 문제이다. 또한 앞서 존재하는 문장 가시화 방식은 단어 간의 관계를 보여주지 못한다. 단어의 중요도는 알 수 있지만 단어들의 위치는 무작위임을 알 수 있다. 본 연구에서는 이런 주제의 동적 변화를 가시적으로 보일 수 있는 Gehpi 기반의 가시화 애니메이션 시스템을 개발하고 소개한다.



(a)



(b)

그림5. wordcloud.kr에서 만드는 가시화된 문장

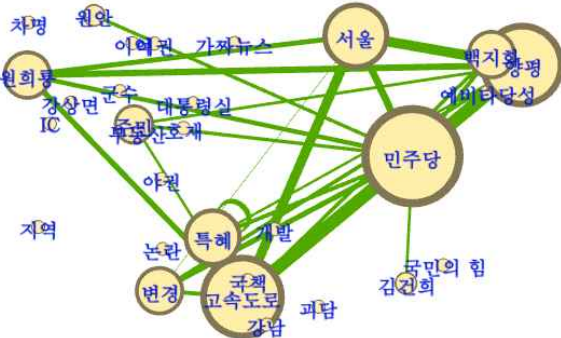


그림6. 관계를 고려하지 않고 만들어진 그래프

다음 사진(그림5, 그림6)은 일반적인 기사의 가시화를 나타낸 것이다. 단어의 빈도가 크기에 반영이 됨을 알 수 있지만 노드들의 배치에는 아무런 관련이 없다.

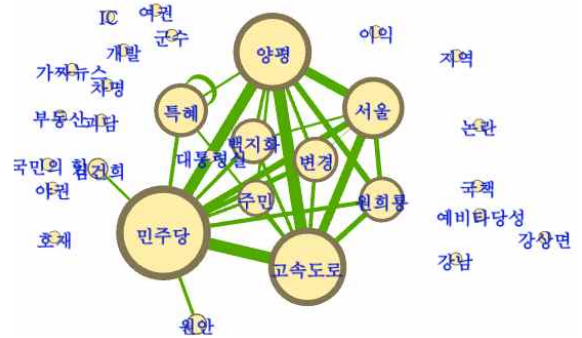
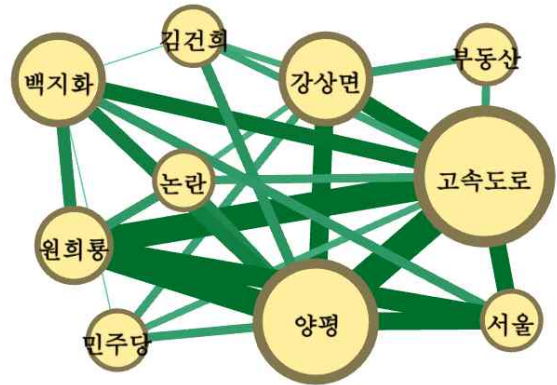
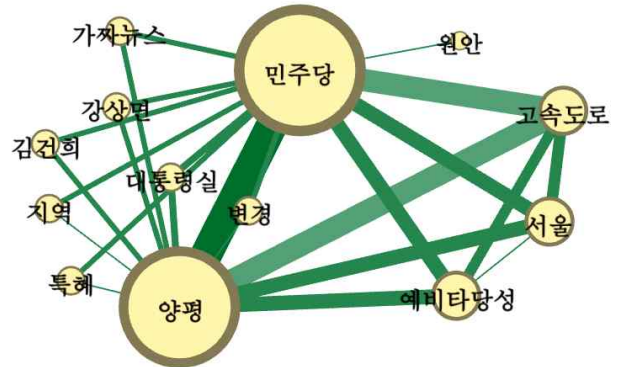


그림7. 동일한 텍스트(그림5의 뉴스)를 본 논문의 시스템으로 구성한 결과

노드 간의 중력을(Edge들의 합) 시뮬레이션하여 노드끼리 가까워지도록 한다. 연결된 노드 간의 스프링력을 시뮬레이션하여 연결된 노드가 서로 가까워지도록 했다. 그림 7은 설명한 방법을 사용하여 네트워크의 구조를 보존하면서 시각적으로 보기 좋은 그래프 레이아웃을 생성한 것이다. 나온 결과는 다음과 같다. 또한 실시간으로 값이 변함에 따라 중력이 변함을 알 수 있다.



(a) 조선일보 7월 11일 기사



(b) 조선일보 7월 25일 기사

그림8. 가시화 애니메이션에 사용된 그래프

그림 (a) 와 (b)는 조선일보 가시화 애니메이션의 한 장면이다(Snapshot). (a)는 초기의 기사, (b)는 그로부터 14일



경과한 후 나온 기사이다. 위의 그림들은 33개의 키워드 중에 상위 Edge를 가진 키워드들만 나타내었다. 시간이 경과함에 따라 조선일보에서는 다양한 키워드들이 등장하였고 복잡해짐을 알 수 있다. “양평” 키워드의 경우 꾸준히 등장했음을 알 수 있다. “원희룡”, “논란” 키워드의 경우 점차 사라졌는데 이 과정을 gephi를 이용하여 애니메이션으로 만들었고 특정 기사의 뉴스 생성 시점에서 주제어들의 가중치를 계산하고 빈 시간대의 가중치는 선형으로 보간하였다 (linear interpolation). 기존의 기사 시각화의 경우 키워드들의 변화를 알 수 없었지만 본 연구에서는 특정 시점에서 키워드들의 중요도 변화와 상호관계를 보여준다.

#### 4. 실험 및 성능 평가

##### 4.1 실험 데이터 준비

먼저 실험을 위하여 특정 사건으로 검색된 신문 기사를 크롤링을 통하여 신문사 별로 수집하였다. 수집된 하나의 뉴스 기사에는 그것을 생산한 신문사(news company)와 게시 시점(temporal information)이 text 정보와 함께 기록된다. 하나의 news text를  $T_i$ 로 나타낸다. 그리고 이 뉴스를 생산한 신문사를  $paper(T_i)$ , 그 뉴스의 생산 시점을  $time(T_i)$ 으로 표시한다. 본 논문에서 이 시점은 관찰 시작 시점을 정수 0으로 하여, 시간(hourly)단위의 정수로 표시하였다. 따라서 30일간 기사를 수집한 경우라면  $time(T_i)$ 은  $[0, 24*30]$  범위의 정수가 된다.

본 연구에서 우리는 특정한 사건에 한정하여 분석을 하기 때문에 관련된 주제어(keywords)를 따로 결정하였다. 이번 실험은 문제의 “양평 도로 변경” 관련 사건이라 여러 정파적 주장과 관점을 기준으로 아래와 같이 33개의 주제어를 선택하였다. 앞으로  $key_i$ 는 하나의 주제어를 나타낸다. 이 주제어 집합을  $K=\{k_i\}$ 로 나타낸다.

'IC','가짜뉴스','강남','강상면','개발','고속도로','괴담','국민의힘','국토부','국책','군수','김건희','논란','대안','대통령실','민주당','백지화','변경','부동산','서울','야권','양평','여권','예비타당성','예타','원안','원희룡','이익','주민','지역','차명','특혜','호재'

표1. 선택한 주제어 33개

우리 주제어를 선택한 이유는 최근에 일어난 사건 중에서 짧은 기간 동안 여러 신문사에서 많은 기사가 나왔기 때문이다. 특히 그곳에 등장하는 키워드가 언론의 성향이나 상황에 따라서 동적으로 변하기 때문이다.

우리는  $T_i$ 와  $K=\{k_i\}$ 를 바탕으로 주제어 연관 그래프(Keyword Relation Graph, KRG)를 생성한다. 이제 그 방법을 설명한다.  $T_i$ 는  $r$ 개의 문장으로 구성되어 있다. 즉 기사는  $r$ 개의 문장, 즉  $s_1 \oplus s_2 \oplus s_3 \dots s_i \oplus s_{i+1} \dots \oplus s_r$ 으로 표현된다. 여기에서  $\oplus$ 는 문장(statement,  $s_j$ )의 접합(concatenation)을 의미한다. 그런데 한글은 조사의 변화가 매우 다양한 언어이므로 특정 주제어의 포함을 보다

쉽게 계산하기 위하여 우리는 전체  $s_i$ 는 원래 문장에서 공백과 특수 기호를 삭제한 하나의 문자열(character string)로 변환하였다.

본 논문의 주된 아이디어는  $k_i, k_j$  두 개의 주제어의 연관된 정도(the degree of working relation)는 특정 뉴스 기사  $T_i$ 의 내부 문장에서의 “거리”를 기준으로 표현될 수 있다는 것이다. 즉 두 주제어가 서로 강하게 연관될수록 두 주제어의 문장 거리(statement distance)는 짧아진다. 이러한 거리-공간 모형은 이미 발표된 연구 논문에서 착안하였다[조환규]. 즉 두 주제어가 강하게 연관된 경우 둘은 하나의 문장에서 동시에 나타난다. 즉 공기(cooccur)되는 경향이 두드러진다.

서로 다른 주제어  $k_i$ 와  $k_j$ 의 연관 강도 가중치 즉, 특정 뉴스 text  $T_x$ 에서의 가중치  $weight(i, j | x)$ 는 다음과 같이 정의한다. 만일 특정  $k_i$ 를 포함하고 있는 문장의 index를  $u$ 라고 하고  $k_j$ 를 포함하고 있는 한 문장의 index를  $v$ 라고 하고 이 경우  $u \leq v$ 라고 한정한다. 즉 문장  $s_v$ 는  $s_u$  다음에 나타난 한 문장이라고 가정하자.

이 경우의 두 주제어  $k_i$ 와  $k_j$ 에 추가되는 weight는 두 문장의 거리를 exponent로 하는  $a^{|v-u|}$ 로 정의된다. 여기에서  $a$ 는 tuning parameter로서 얼마나 멀리 떨어져 있는 문장까지 영향을 주는 것 인지를 결정하는 값이다.  $a=1$ 이면 문장거리에 상관없이 추가되는 weight는 +1이 된다. 만일  $a=1/2$ 로 설명하면 같은 문장에 두 주제어가 동시에 나타난 경우에는 +1, 1문장 거리로 떨어진 경우 0.5, 2문장 거리로 떨어져 있으면 0.25가 된다. 즉 한 문장 떨어질 때 마다 그 연관강도는 1/2씩 연속 감소한다. 따라서 만일 10문장에 떨어져 있는 경우라면 그 연관강도는  $1/1024 < 0.001$ 이된다. 만일  $a=0.1$ 로 한다면 3개 문장거리로만 떨어져 있어도 추가되는 연관강도는 0.001로 매우 미미하게 된다. 실제 구현과정에서는 8문장 보다 먼 거리의 단어들은 계산을 하지로 않기로 했다.

$a$ 가 1로 가까워질수록 먼 거리의 주제어까지 영향력을 인정하게되고,  $a \approx 0$ 에 되면 공기된 경우, 즉 같은 문장에 동시에 나타난 경우가 아니면 무시하게 되는 경향을 보인다. 따라서 두 주제어  $k_i, k_j$ 가 특정 문수  $T_x$ 에서 가지는 연관 가중치(weight)는 다음과 같이 정의된다. 아래 식에서  $distset(i, j)$ 은 두 주제어의 모든 문장 거리의 집합을 의미한다.

$$weight(i, j | x) = \sum_{k_i, k_j \in K_x} a^{distset(i, j)}$$

예를 들어 다음의 뉴스 기사에서 선택된 서로 다른 문장거리 집합은 다음과 같다. 기사는 문장 별로 각 칸에 배치하였다.

No	<i>i</i> 번째 문장의 내용
$s_1$	2008년 민간 제안으로 시작된 서울-양평고속도로 사업의 종점은 13년간 추진→무산'을 거듭하는 가운데서도 줄곧 양평군 양서면'이었다.
$s_2$	상습 정체 구간으로 악명이 높은 6번국도(남양주-팔당대교-양평 구간) 교통 정체 해소를 위해 인근인 양서면에 고속도로 분기점(JCT)을 만들어 교통량을 분산시킬 필요성이 컸기 때문이다.
$s_3$	경제성이 부족해 사업이 번번이 퇴짜를 맞았을 때도 여야는 한목소리로 '6번 국도 정체 해소'를 위해 사업이 이어져야 한다고 주장했다.
$s_4$	양평고속도로 사업은 지난 2008년 한신공영 등 6개 민간 사업자들이 경기도에 사업을 제안하면서 처음으로 공론화됐다.
$s_5$	당시 업체들은 서울 송파구 오금동을 출발해 양평군 양서면 도곡 나들목(IC)까지 이어지는 노선을 고안했고, 이 노선이 향후 경기 하남과 양평을 잇는 고속도로 노선의 뼈대가 됐다.
$s_6$	여당과 <조선일보> 등이 더불어민주당 소속인 정동균 전 양평군수가 양서면 종점 일대 토지를 보유하고 있다며 의혹을 제기하고 나섰지만, 해당 종점이 거론된 것은 국민의힘 소속 지자체장 시절이었다.

표2. 기사의 문장 6개와 사용된 keyword keywords  $k_i$ 와  $k_j$ 의 distset( $i, j$ )

1	2	3	4
양평	고속도로	서울	IC

표3. 번호에 따른 keyword

$i, j$	distset( $i, j$ )	$i, j$	distset( $i, j$ )
1, 2	1,2,3	1, 4	0,1,4
1, 3	0,1,3	2, 4	0,1,3
2, 3	0,1,3	3, 4	0

표4. keyword들 간의 문장 사이의 거리 집합

$i, j$	$\alpha = 0.5$	$\alpha = 0.7$	$i, j$	$\alpha = 0.5$	$\alpha = 0.7$
1, 2	3.875	4.933	1, 4	2.0625	2.6401
1, 3	1.625	2.043	2, 4	1.625	2.043
2, 3	1.625	2.043	3, 4	1	1

표5. keyword간의 weight

문장 공간 속의 keyword vector는 위의 표와 같이 만들어진다. alpha 값이 커질수록 문장 사이의 거리가 더 가까운 단어에 대해서 가중치가 커진다.

언론사	기사	시작일	종료일	성향
중앙일보	155	07.05	08.01	보수
동아일보	210	07.05	08.01	보수
조선일보	230	07.06	08.01	보수
한겨레신문	120	06.30	08.04	진보
경향신문	160	07.03	08.04	진보
노컷뉴스	210	07.05	08.04	진보
오마이뉴스	101	07.07	08.04	진보
프레시안	112	07.13	08.04	진보
시사포커스	99	07.12	08.07	보수
데일리안	127	07.04	08.07	보수
파이낸셜뉴스	112	07.08	08.07	보수
한국경제	110	07.09	08.11	보수

표6. 언론사 별 데이터 통계

연구를 위해서 수집한 기사는 약 2000개로 python의 selenium, beautifulsoup을 이용해서 “양평 고속도로” 키워드를 포함하는 기사들을 7월 초 부터 8월 초 까지 수집하였다. 언론사의 정치적 성향이 비교적 잘 알려져있는 경우를 수집하였고, 이 중에서 가장 규모가 큰 조선일보, 중앙일보, 동아일보, 노컷뉴스, 한겨레신문, 경향신문의 기사들은 모델의 데이터를 학습 시키기 위해서 사용하였고, 그 외의 언론사들의 성향을 예측하였다.

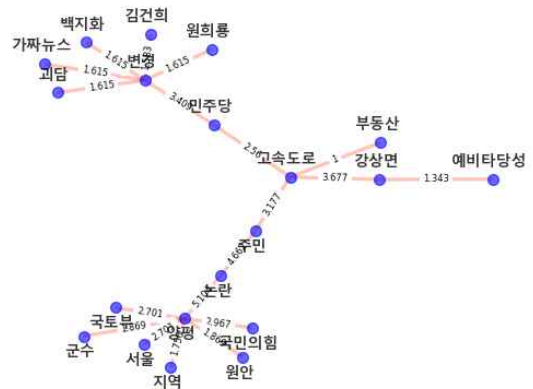


그림9. 중앙일보 07/13 주제어 연관그래프

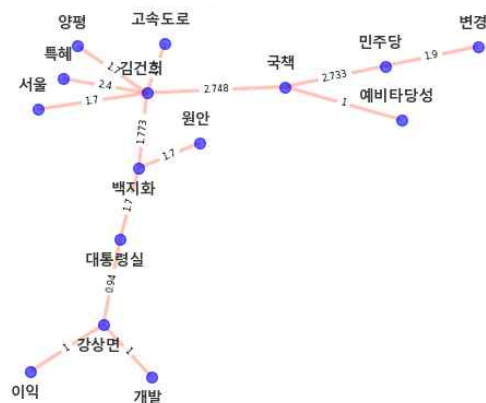


그림10. 경향신문 07/22 주제어 연관그래프

위의 그래프는 중앙일보와 경향신문의 keyword vector

graph를 maximum spanning tree로 구현한 것이다. 각 언론사의 성향에 따라서 중심이 되는 node들과 이웃 node들이 다르게 나타남을 알 수 있다.

#### 4.2 실험 결과 및 분석

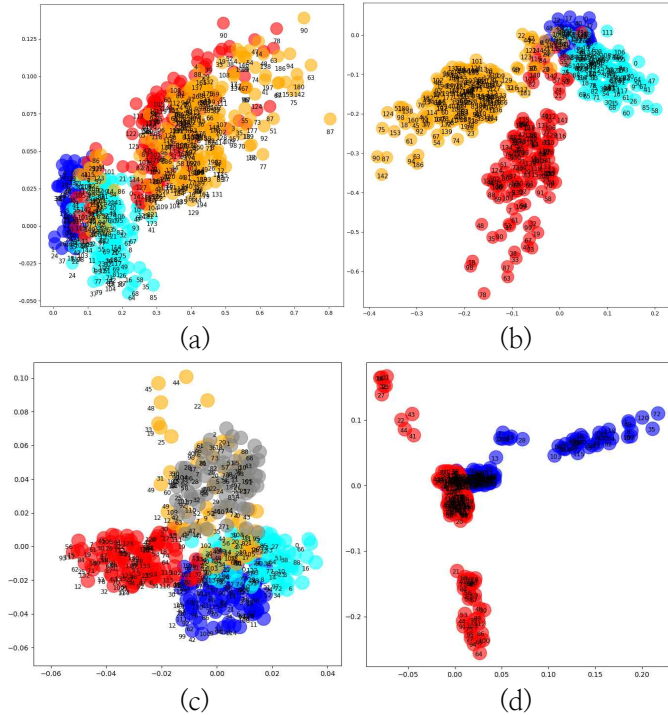


그림 11. (a)-{P,H,J,D}, (b)-{H,K,C,D}  
(c)-{K,H,J,D, Da}, (d)-{진보, 보수}

우리는 위에서 수집한 데이터를 karateclub의 graph2vec 모델에 학습을 시킨 후, 언론사들의 그래프들을 분류하는 작업을 하였다. Epoch 값은 40, dimension은 91로 하였다. 그림4의 (a)는 프레시안(P, 파랑), 한겨레신문(H, 하늘), 중앙일보(J, 주황), 동아일보(D, 빨강), (b)는 한겨레신문(H, 파란), 경향신문(K, 하늘), 조선일보(C, 주황), 동아일보(D, 빨강)을 학습 시킨 결과로, 진보, 보수 계열의 언론사들이 눈에 띄게 분류가 된 모습이다.

그림4의 (c)는 (b)에서 사용된 모델에서 dimension을 120으로 바꾼 후, 데일리안(Da, 회색)을 학습 시켰을 때 분포를 나타낸 것으로, 같은 보수 언론사 계열끼리 치우치는 경향을 보여준다. (d)에서는 (c)에서 사용된 신문사를 학습만 시키고, 나머지 7개의 언론사를 새롭게 학습시킨 다음, 그 언론사들만 분포시켰다. 파란색이 진보, 빨간색이 보수 언론사들이다.

이 실험을 토대로 stellargraph의 Deep Graph CNN을 이용해서 보수, 진보 계열의 언론사를 각각 3개씩 학습시킨 후, 나머지 언론사의 정치적 성향을 예측하는 모델을 만든 결과 기사의 정치적 편향성을 약 82%의 확률로 예측하는데 성공하였다.

#### 5. 결론

본 연구를 통해서 밝혀낸 바는 다음과 같다.

- 주제어들의 연관 관계를 문장거리로 계산해서 나타낼 수 있는 모델을 새롭게 제시하였다.

- 문장공간에서의 주제어들의 연관 가중치를 결정하는  $\alpha$  값은 0.7에 가까울수록 편향성 측정의 정확도가 가장 높음을 알 수 있었다.
- 주제어들의 문장거리 그래프를 이용한 결과 약 82%의 확률로 기사의 정치적 편향성을 예측 할 수 있었다.

본 연구 이후 추후 연구 주제는 다음과 같다.

- 이번 그래프 학습에서 기사의 시간적 순서는 고려하지 않았다. 따라서 일반 그래프가 아닌 시간 그래프(temporal graph)기반의 기계학습 방법에 대하여 연구할 예정이다.
- 앞서 보인 것과 같이 기사의 특성을 결정하는 것은 문장 내에서의 주제어들의 절대적 위치가 아니라, 문장 공간에서의 상대적 거리이다. 본 방법은 언어 구조가 다른 외국어에 대해서도 충분히 의미가 있을 것이기 때문에 다른 언어권의 뉴스 대해서도 적용을 할 예정이다.
- 이러한 방법을 이용하여 웹을 통해서 정치적 편향을 분석하는 서비스를 만들어서 공개할 예정이다.

#### 참고문헌

[1]이재훈, 김미숙. "문서 편향성 및 감정분석을 활용한 가짜뉴스 탐지" 한국HCI학회, pp.593-597, 2023  
[\[2\]https://textvis.lnu.se/](https://textvis.lnu.se/)  
 [3]오승빈, 김현민, 김승재. "순환 신경망과 합성곱 신경망을 이용한 뉴스 기사 편향도 분석" 한국정보통신학회논문지 pp.999-1005, 2020  
 [4]오경환외 6인.거짓표현 분석을 통한 기사 사실 여부 판단. 한국정보기술학회, pp.439-442, 2020  
 [5]WU, Chuhan, et al. User-as-Graph: User Modeling with Heterogeneous Graph Pooling for News Recommendation. *IJCAI*. pp.1624-1630, 2021  
 [6]박경미외 2인, "소셜 등장인물의 텍스트 거리를 이용한 사회 구성망 분석," *한국콘텐츠학회논문지*, vol. 13, no. 4, pp.427-439, 2013.  
 [7]Cui, Weiwei, et al. "Context preserving dynamic word cloud visualization." 2010 IEEE Pacific Visualization Symposium (PacificVis).pp.112-122, 2010.