선진국방연구
Journal of Advances in Military Studies https://doi.org/10.37944/jams.v3i1.64
2020, Vol. 3, No. 1, 51-74.

빅데이터를 통해 알아본 '방위비 분담금'에 관한 인식: 트위터를 중심으로

김성곤*

≪국문초록≫

본 연구는 방위비 분담금에 대한 대중의 인식을 확인하고 이를 통해 한·미 동맹의 발전방안과 방위비 분담금 협상에서 고려해야 할 사항 등을 분석하는 데 목적이 있다. 이를 위해 소셜 미디어인 트위터를 선정하고 '방위비 분담금'을 키워드로 구름단어, 소셜 네트워크, LDA 기법을 적용하여 분석하였다. 그 결과 빈도수에 따른 주요 핵심어는 '사드', '주한', '인상', '부담', '협상' 등이었으며, 단어관계와 연결 중심성을 고려한 소셜 네트워크 분석결과 주요 핵심어는 '사드', '반대', '비용', '배치', '동맹' 등으로 나타났다. 또한, LDA 기법을 적용한 결과, 현재 '방위비 분담금'과 관련된 협상과정에서 이슈화 되고 있는 '사드 기지 공사비', '주한미군 철수', '주한미군 內 한국인 근로자의 무급휴직', '정부의 협상재개를 위한 노력의 내용' 등이 나타났다.

이를 통해 대중들이 '방위비 분담금'을 인식하는 주요 단어를 확인할 수 있었으며 더 나아가 정부는 성공적인 협상결과를 도출하기 위해 국민들이 인식하는 주요 단어와 주제를 고려하여 전략적인 협상정책을 추진하여야 하겠다.

주제어 : 방위비 분담금, 트위터, 구름 단어, 소셜 네트워크, 잠재 디리클레 할당(Latent Dirichlet Allocation)

* 육군사관학교 심리경영학과(광운대학교 방위사업학과 박사과정), okskkon@mnd.go.kr

Ⅰ. 서론

최근의 불확실한 안보환경에 맞춰 우방국과 안보협력을 강화하고 상황을 이끌기 위해 국방력 강화에 관심을 가지고 있다. 특히, 4차 산업혁명과 병력감축 등의 외부 상황은 '몸집은 줄어도 전투력은 더 강한 군'을 구현하고 무인전투체계 보급(1,148억원)과 드론・AI・빅데이터 등의 지능정보기술의 국방운영 접목(573억원)으로 진행 중에 있다. 이러한 국방예산과 더불어 한미 방위비 분담금도 한미 동맹과 국방분야에 중요한 이슈이다. 방위비 분담금은 미군의 한국 주둔비용 중 일부 또는 전부를 한국 정부가 분담하도록 규정한 한・미 양국 간 협정으로 미군이 한국에서 고용하는 근로자의 인건비(비중 약 40%), 군사건설 및 연합방위 증강사업(40%), 군수지원비(20%) 등의 명목으로 지원된다.(박기학, 2017) 한국은 1991년부터 '한미상호방위조약에 따른 주둔군지위협정(SOFA)' 제5조의 예외적 조치로서 미국과 체결한 '방위비분담금특별협정(SMA)'에 근거하여 주한미군의 방위비 일부를 분담해 왔다. 그동안 미국은 안보환경 제공과 한국의 경제 수준을 고려하여 지속적으로 분담비의 증액을 요청했고 협상이 타결되었다. 2017년도 이후, 미국 우선주의(America First)를 내건 트럼프 정부가 자국의 이익을 위해 방위비분담금 인상을 지속적으로 요구하고 있으며 이번 2020년도 협상도 쉽지 않을 것이라 예상된다.

현재까지 방위비 분담금에 대한 기존연구는 '협상', '인상요인', '외국 사례' 등 체결과정과 결과에 따른 쟁점분석 위주로 진행되었다. 한국의 방위비분담과 일본과의 비교를 통해 방위비분담 협상에 관한 "총액협상방식"과 "항목협상방식"의 장・단점을 분석하고(박휘락, 2015), 방위비 분담금 지급현황과 각종 요인들을 실증자료로 구축하여 방위비 분담금과 관계된 요인들을 토대로 상관관계를 분석하였다.(정원열・김종수, 2016) 또한, 각종 소요를 고려하여 분담금 규모를 결정하고 방위비 운용 측면에서 일본의 주도적인 의사결정을 분석하였으며(백재옥, 2017), 한국의 방위비분담의 현황과 쟁점을 분석하고 개선사항을 도출하였다(신종호, 2009). 트럼프 정부 출범 이후, 방위비 관련 연구는 트럼프 행정부의 동맹정책을 기반으로 협상을 위한 구체적인 분석(박원곤, 2019)과 트럼프 정부 출범에 따른 한미 군사관계의 주요 쟁점이 무엇이고 쟁점별 한국의 대응방향을 제시하였다(김열수, 김경규, 2017).

기존 연구는 방위비 분담금 협상에 관한 방법론과 방위비 분담금에 영향을 끼치는 요인, 쟁점별 대응 방향 등 협상 당사자와 예산과 비용을 고려하여 주로 분석하였으며 대중이 지니고 있는 방위비 분담금에 대한 인식으로 인한 한미 관계 분석과 협상과정에서 국내여론의 추이를 반영하는데 한계점을 지니고 있었다. 따라서 본 연구는 방위비 분담금에 대한 대중의 인식을 확인하고 이를 통해 국내여론이 반영된 협상전략 수립을 발전시키는 데 목적이 있다. 이를 위해, 소셜 빅데이터 분석 기법인 단어구름, 소셜 네트워크, LDA를 활용하여 트위터의 '트윗'을 통해 대중의 인식을 확인하고 분석하였다. 방위비 분담금에 대한 대중의 인식을 확인하여 방위비 분담금 협상에서 전략적 의사소통(SC) 계획에 반영될 필요성을 제시하였으며 국내 여론의 추이를 확인한 결과, 양국의 우호

적인 국민적 공감대 형성이 필요하다는 것도 확인하였다.

본 연구의 구성은 다음과 같다. 제1장은 서론은 방위비 분담금에 대한 선행연구와 연구 배경, 연구 목적을 제시하였으며 제2장은 연구를 위한 방위비 분담금, 소셜 빅데이터 분석의 개념을 고찰하였다. 제 3장은 '방위비 분담금' 키워드를 워드클라우드, 소셜 네트워크, LDA 기법을 적용하는 연구 방법을 설명하였으며 제4장은 세부 분석결과를 제시하였다. 마지막으로 제5장에서 연구의 의의, 시사점 및 한계점, 향후 연구방향 등을 설정하였다.

Ⅱ. 이론적 배경

2.1 방위비 분담금

방위비 분담금은 주한미군 주둔군지위협정(SOFA) 제5조(시설과 구역)에 대한 특별협정에 근거하여 지원하는 협의의 분담금을 의미하며 특별협정은 2-5년 주기로 체결되며 한・미간 협상에 의하여 분담액이 결정된다. 분담금은 주한미군의 인건비를 제외한 비 인적 주둔비용 중 일부를 분담하지만 흔히 주한미군에 대한 한국의 방위분담, 또는 지원이라고 하면 특별협정에 의한 방위비분담금을 포함하여 직・간접지원을 모두 일컫는 광의의 분담(cost-sharing)을 의미한다. 매년 주한미군에 지원하는 방위비 분담금 현황을 알 수 있는 기초자료로서 의미가 크다고 할 수 있으며 세부 내용은 아래 표 1과 같다.

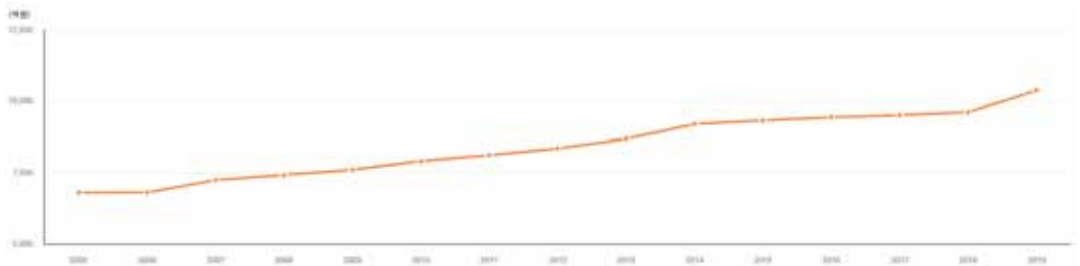
<표 1> '방위비 분담금' 현황

단위 : 억원

연 도	합의액	원화비율(%)	연 도	합의액	원화비율(%)	연 도	합의액	원화비율(%)
1991	1,073	0	2001	4,882	62	2011	8,125	100
1992	1,305	0	2002	6,135	88	2012	8,361	100
1993	1,694	0	2003	6,690	88	2013	8,695	100
1994	2,080	0	2004	7,465	88	2014	9,200	100
1995	2,400	0	2005	6,804	100	2015	9,320	100
1996	2,475	0	2006	6,804	100	2016	9,441	100
1997	2,904	0	2007	7,255	100	2017	9,507	100
1998	4,211	57	2008	7,415	100	2018	9,602	100
1999	4,408	58	2009	7,600	100	2019	10,389	100
2000	4,685	60	2010	7,904	100			

* 출처 : 방위비분담특별협정

위 표 1에서 합의액이 1991년 1천 억여원에서 점차 증가하여 2019년에는 1조가 넘어간 표를 알 수 있으며 원화비율도 점차 증가함을 표를 통해 알 수 있다. 특히, '05～'06년도 방위비분담 특별협정에서 그 간의 점증적 증가방식을 지양하고 2년간 인상률을 동결하기로 함으로써 동 기간동안 분담금액이 동결되었으며 '05년부터 방위비 분담금액을 전액 원화로 지급하였다. 정부는 방위비 분담 집행에 있어 투명성과 책임성 강화를 위해 5개 분야에서 포괄적 제도개선을 도출하였다. 합의액을 기준으로 변동 추이를 살펴보면 아래 그림 1과 같으며 1991년 이래 방위비 분담금액이 매년 점진적인 증가추세를 보여 왔다.



* 출처 : 방위비분담특별협정

<그림 1> 방위비 분담금 규모 추이

위의 그림 1을 구체적으로 살펴보면 1991년도 이후 방위비 분담금액이 증가한 원인은 80년대 이후 미국의 무역적자 누적과 이로 인한 국방비 삭감으로 동맹국의 재정지원 확대를 내용으로 하는 방위비분담 정책을 추진하였다. SOFA 제5조에 의거 우리나라가 주한미군에 시설과 구역을 제공하여야 하나, 구역(토지)만 제공하고 시설은 제공하지 못하여 방위비분담협정 체결을 통해 그간 거의 부담하지 않거나 일부 항목별로 지원하던 비용을 분담하기 시작한 것이므로 분담금액은 매년 증가하는 양상을 보여왔다. 방위비 분담금 규모 외에 과거 방위비 분담금 특별 협정 간 주요 내용과 관련된 쟁점들을 살펴볼 필요가 있다. 아래 표 2는 1차에서 5차간의 특별 협정의 주요 내용이다.

<표 2> '방위비 분담금' 특별 협정(1-5차) 주요내용

차 수	적용연도	주요 쟁점	합의 내용
1	1991-1993	-	· 한국이 주한미군 노동자 임금 일부 부담, 필요하면 기타 경비 부담 · 한국이 분담액 결정하고 통보
2	1994-1995	-	· 주한미군의 원화 경비 지출의 3분의 1 수준까지 증액 결정
3	1996-1998	-	· 전년도 기준 매해 10%씩 증액 결정 · 1998년 분담금은 IMF 구제 금융 고려하여 8500만 달러 감액
4	1999-2001	-	· 현물제공의 경우 세금 면제 가격을 적용(면제)

차 수	적용연도	주요 쟁점	합의 내용
5	2002-2004	-	· 한국의 총액 결정 조항 삭제 · 2002년 총액을 6132억원으로 확정 · 방위비분담 항목이 4가지로 정리됨

* 출처: 「트럼프 시대, 방위비 분담금 바로알기」

위 표 2를 구체적으로 살펴보면, 1차 협정간 '기타 경비 부담'은 인건비 외의 다른 어떤 비용이든 부담할 수 있다는 뜻으로 한국에게 불리할 수 있는 항목이다. 2차 협정내용 중 증액기준이 되어야 할 총 주둔 비용을 주한미군이 공지하지 않은 문제점이 발생되었으며, 3차 협정은 전반적으로 방위비 분담금의 증액 비율이 높고 1998년 당시 환율을 적용하여 원화로 계산하면 1997년보다 40.5% 인상되었다. 4차 협정은 한미 SOFA 개정과 연계하지 못한 협상전략을 적용했다는 의견이 많으며 5차는 과도한 인상률에 대한 문제와 항목별 금액 배분 결정 방식이 아닌 총액 결정 방식에 대해 문제가 제기되었으며 예산 통과 뒤 특별협정이 비준되어 국회 예산 심의권 침해 문제가 발생되었다.

<표 3> '방위비 분담금' 특별 협정(6-9차) 주요내용

차 수	적용연도	주요 쟁점	합의 내용
6	2005-2006	· 미국이 C4I 현대화 비용, 주택 임대료, 시설 유지비 등을 추가 요구	· 2005년, 2006년 모두 6804억 원으로 총액 확정 · 미국 요구 사항 중 시설 유지비 수용
7	2007-2008	· 전략적 유연성으로 인한 주한미군 역할 변경 문제 제기 · 미군 기지 이전비 전용의 불법성 제기	· 2007년 총액 7255억 원, 2008년 총액 7415억 원 확정, 전년 대비 6.6% 인상
8	2009-2013	· 이자 발생 문제 공론화 · 분담금 전용 "양해"문제 공론화	· 군사 건설비를 2011년까지는 전면 현물 지원하고, 전체 금액의 12%만 설계 및 감리 비용으로 현금 지급
9	2014-2018	· 미집행액의 연례적 발생 문제 공론화 · 이자 귀속처 및 탈세 문제 제기 · 한국 정부가 총액 삭감과 군사건설비 전용 방지를 협상 목표로 제시	· 2014년 총액 9200억 원, 2015년 9320억 원, 2016년 9441억 원, 2017년 9501억 원으로 합의 · 포괄적 제도 개선 합의

(출처: 「트럼프 시대, 방위비 분담금 바로알기」)

표 3을 살펴보면 6차 협정에서 주한미군 감축(1만 2500명)과 이라크 파병 비용을 감안하지 못한 합의를 했으며 분담금 총액은 줄었으나, 주한 미군 1인당 지원 비용은 2.7% 증액 되었다. 7차 협정은 방위비 분담금으로 기지 이전 비용을 충당하는 것은 불합리하다는 의견이 제기되었으며 8차 협정은 정부의 전용 "양해"는 공식 외교 문서가 아닌 구두 합의에 의한 것으로 법적 근거 없다는 문제가 야기되었다. 9차는 정부의 협상 목표가 달성되지 못했으며 현물 지원 88%비율 규정을 삭제하고 이자환수에 대한 내용이 없었다.

위의 내용으로 방위비 분담금에 대한 다양한 쟁점과 문제점을 확인 할 수 있었으며 앞서 다룬

다수의 선행 연구들은 협상 대상자를 대상으로 하고 있기 때문에 대국민 인식 및 여론의 영향정도를 인식하지 못한 한계점이 있어 본 연구에서는 소셜 빅데이터 분석을 통해 대중의 인식을 확인하려고 한다.

2.2 소셜 빅데이터 분석

빅 데이터(big data)란 디지털 환경에서 생성되는 데이터로 규모가 크고, 생성 주기가 짧으며 형태도 다양한 대규모 데이터를 말한다. 빅데이터는 시대가 지날수록 데이터의 규모가 커지며 종류도 다양해져 SNS를 통해 생각과 의견까지 축적되고 있다.(정용찬, 2013) 이러한 빅데이터를 분석하는 소셜 빅데이터 분석은 대량의 정형 또는 비정형의 데이터 집합조차 포함한 (James Manyika & Michael Chui, 2011) 데이터로부터 가치를 추출하고 결과를 분석하는 기술이다. (John Gantz & David Reinsel, 2011) 다양하고 대규모 데이터에 대해 수집하고 분석하여 표현할 수 있으며 이를 통해 사회현상 및 여론을 정확하게 예측하여 새로운 기술을 출현시키기도 한다. 빅 데이터 분석은 데이터를 사회 전 분야에서 유의미하고 중요한 정보로 제공하다는 것을 우리에게 제시하고 있다.

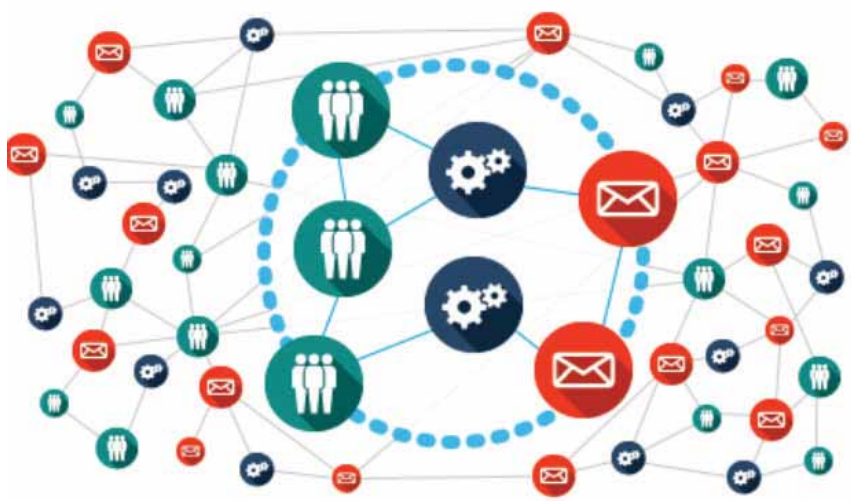
특히, 다양한 소셜 네트워크 서비스가 일반화된 현 시대에 방대한 정보가 텍스트로 생성되고 있으며 이에 대한 관심이 높아지고 있다. 비정형 텍스트 데이터(뉴스기사, SNS, 논문, 책 등)를 데이터 전처리 (단어 추출, 단어 품사 / 중요도 판별, 단어의 문서 내 위치 코딩, 사전(유사어/복합어) 처리, 단어 등장빈도 측정, 단어 간 인접관계 추출 등) 과정을 거쳐 시각적 분석(단어 구름), 소셜 네트워크 분석(개념구조), 텍스트 마이닝(토픽 추출) 등으로 텍스트 분석이 가능하다.

위에 제시된 주요 3가지 분석기법을 세부적으로 살펴보면 첫째로 시각적 분석인 단어 구름(word cloud)이다. 대상자의 인터뷰, 댓글이나 트윗을 수집하여 단어단위로 분해한 후, 단어별 빈도수를 통해 단어크기를 정하여 그림 2와 같이 시각적으로 알기 쉽게 제시하는 기법이다.

* 출처 : http://www.netminer.com/product/features.do

<그림 2> 단어 구름(Word Cloud)

둘째로 소셜 네트워크 분석이다. 텍스트에 내재된 언어와 지식은 단어와 그들간의 관계 네트워크로 모델링 될 수 있다는 점과 텍스트 네트워크 내에서 개념들의 위치와 연결 패턴을 통해 텍스트의 의미 또한 중요한 주제에 대해 이해할 수 있다는 점을 기본가정으로 기존 소셜 네트워크 분석을 확장하여 텍스트도 네트워크 분석을 가능케 하였다. 텍스트 네트워크 분석을 위해 그림 3과 같이 네트워크 시각화 방법을 적용한 단어 관계망(Kamada & Kawai, 1989) 기법을 적용하여 연결망의 구조적 특성과 단어들의 연결성을 분석하여 단어의 응집성 및 중앙성을 기준(김용학, 2016)으로 연결망에 반영해 네트워크를 가시화하여 텍스트 사이의 연결구조와 응집구조, 중심구조를 분석할 수 있다.



* 출처 : http://www.cyram.com/product/product_00_01.jsp

<그림 3> 소셜 네트워크 분석

이러한 연결구조와 응집구조, 중심구조를 분석하기 위한 여러 중심성 지표 중 연결중심성, 근접중심성, 매개중심성이 주로 많이 사용되고 있다. 연결 중심성은 각 단어가 네트워크 안에서 얼마나 중심에 위치해 있는지를 나타내는 지표로 키워드가 연결하고 있는 다른 연결점의 수로서 연결 수에 비례하여 나타나는 중심정도를 말하며, 한 점에 직접적으로 연결되어 있는 점들의 합으로 얻어진 값이 클수록 네트워크 안에서 중심에 있다는 것을 나타낸다. 즉, 단어 네트워크 내 특정 단어의 연결중심성이 높다는 것은 '다양한 단어와 자주 근접하여 등장하는 단어' 라고 해석하면 된다.

매개 중심성은 네트워크 안에서 한 키워드가 다른 키워드들 사이에 위치하는 정도를 측정하는 것으로, 중개역할을 얼마나 잘 수행하는지를 나타내는 지표이다. 여기서는 중개 역할이 '중심'으로 간주 되며 이 값이 클수록 주요 키워드들 사이에 위치하며 네트워크 안에서 정보 흐름에 영향력을

크게 가질 수 있다.

근접 중심성은 키워드 사이의 거리를 나타내는데, 키워드 간의 가까운 정도를 알아보는 지표이다. 즉, 네트워크 내의 다른 키워드들과 얼마나 근접하게 연결되어 있는가를 알아보는 것으로 해당 키워드가 얼마나 네트워크 중앙에 위치해 있는가를 측정하는 것으로 이를 통해 네트워크 안에서 가장 중심이 되며 일반적인 영향력을 가진 키워드가 무엇인지 파악할 수 있다. 근접중심성의 경우 단어 네트워크에는 다소 적용하기 어려운 부분이 있어 자주 활용하진 않는다.

이러한 텍스트 네트워크 분석의 이점은 시각적 요약, 핵심 개념과 하위주제 발견, 새로운 아이디어 발견, 비선형적 읽기 방식, 정량적이고 정성적인 접근이 가능하다는 것에 있다. 이러한 분석은 뉴스 토픽 분석, 미디어 여론 분석, 브랜드 이미지 분석, 기술 트렌드 분석, 연구 동향 분석, 문서 의미구조 분석 등 널리 이용되고 있다.

셋째로 텍스트 마이닝 분석이다. 마이닝 기법은 근래에 많이 활용되는 기법으로 데이터에서 유의미한 의미를 도출하여 정보를 생성하는 방법이며 아래 표 4와 같이 분류할 수 있다.

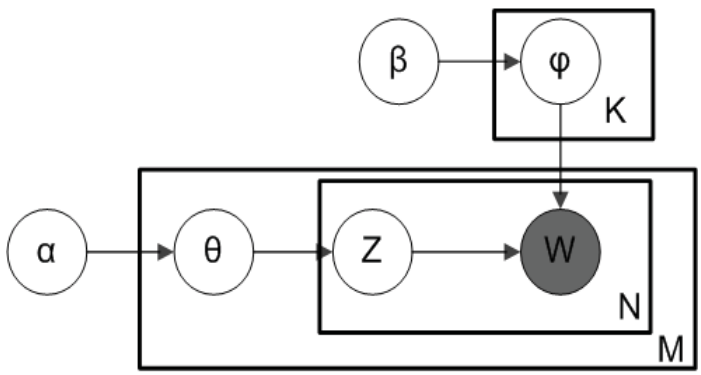
<표 4> 마이닝 기법의 분류

구 분	내 용
데이터 마이닝	· 대용량 데이터에 숨겨진 지식이나 예측하지 못한 경향, 새로운 규칙 등 유용한 정보를 발견하는 기법 · 정보의 연관성(순차 패턴, 유사성 등)을 파악함으로써 가치있는 정보를 도출하여 의사결정에 적용
텍스트 마이닝	· 자연어 처리(Natural Language Process) 기술에 기반하여 비정형 텍스트 데이터에서 관계를 도출하여 유의미한 정보를 찾아내는 기법
웹 마이닝	· 인터넷 상의 정보를 수집하여 데이터 마이닝 기법을 활용
소셜 마이닝	· 소셜미디어의 글, 사용자 분석으로 소비자 흐름이나 패천 분석 · 마케팅 분 아니라 사회 트렌드와 그 변화 추세를 읽음
현실 마이닝	· 사람들의 행동패턴 예측을 위해 관련 정보를 기기(휴대폰, GPS 등)를 통해 획득하여 분석

* 출처: 「소셜 빅데이터 분석을 통해 알아본 대중의 '미드'에 관한 인식 및 자기주도적 영어학습 행태: 유투브와 트위터를 중심으로」

위 표 4의 내용 중 텍스트 마이닝을 통해 비정형 텍스트 데이터에서 유의미한 정보를 찾아 낼 수 있으며 텍스트 마이닝의 기법 중 하나로써 주제를 추출하는 토픽 모델링은 실제 관찰 가능한 문서 내 단어를 이용하여 토픽의 단어분포와 문서의 토픽분포를 추정하는 과정이다. 여기서 텍스트는 여러 개의 '토픽'으로 구성되며, '토픽' 은 복수개의 단어들과 연관되어 있다. 문서에서 '토픽'은 사용된 단어들로 표현되며 사람이 문서를 생성하는 과정을 '토픽'이라는 개념을 도입하여 확률적으로 모델링 한 개념으로 '토픽'은 관찰된 데이터가 아니라 숨겨진 변수로 표현된다. 토픽 모델링의 가정사항으로 문서내의 단어들은 무작위적으로 선택되는 것이 아니라는 것이다. 토픽의 단어 분

포와 문서의 토픽 분포의 결합으로 문서 내 단어들이 생성된다고 가정된다. 이러한 토픽 모델링은 LDA(Latent Dirichlet Allocation) 기법으로 많이 활용되고 있으며 아래 그림 4과 같다.



<그림 4> LDA 그래픽 모델(Blei 외 2인, 2003)

그림 4을 세부적으로 살펴보면, α와 β가 주어지면 α를 통해 Θ를 구하고, β를 통해 φ를 구하는 과정에서 분석되는 문서를 모델링하는 과정이 시스템이 얼마나 원본을 잘 재현해 내는지를 구하는 과정이다. 모델링 결과와 원본의 차이를 최소화하기 위한 Θ와 φ를 추론하는데 베이즈 추론을 사용한다. 즉, K 값이 커질수록 결과가 특정적이고 작을수록 광범위해진다. 그림 1에 제시된 의미를 구체적으로 살펴보면 K는 토픽의 수를 의미하며, α와β는 하이퍼 파라미터를 의미한다. N은 문서의 단어 수를 M은 문서의 수이다. 또한 φ(i)는 문서(i)의 토픽분포(문서-토픽 비중 / 생성확률)를 의미하며 w(i,j)는 I번째 문서의 j번째 단어를 의미하고 z(i, j)는 w(i,j)에 대한 토픽 할당을 의미한다. φ(k)는 토픽(k)의 단어 분포(단어-토픽 비중/생성확률)를 의미한다. 이 중 Θ와 φ는 Dirichlet 분포이고 z와 w는 multinominal 분포이다. LDA는 문서를 작성하는 과정에 관한 Generative model이며, 문서의 주제 분류를 위해 고안된 알고리즘이다.

각 단어를 결정할 때, Θ에 대한 다항 분포로 토픽을 할당한 후 그 토픽으로부터 단어를 추출하여 작성자가 문서를 작성할 때, 문서의 토픽 비율은 최대한 한 토픽에 치우치게 되며 토픽에 사용하는 단어들도 특정 단어에 치우치는 경향을 갖는다고 가정한 상태에서 문서의 토픽 비율에 따라 토픽을 고르고, 선택된 토픽에서의 단어 비율에 따라 단어를 고르는 과정을 통해 문서를 생성하는 것이다.

Ⅲ. 연구방법

3.1 자료수집 대상 및 기간

본 연구는 ‘방위비 분담금’에 대한 대중의 인식을 분석하기 위하여 소셜 네트워크 서비스(SNS) 중 트위터를 대상으로 연구를 진행하였다. 트위터는 단문의 글을 통해 소통할 수 있는 온라인 플랫폼으로 텍스트 데이터가 빠른 속도로 축적되고 있다. 또한 사용자의 의견 및 감정, 사실 또는 정보가 공유됨으로써 사회현상을 이해할 수 있다.

트위터의 오픈API(Application Programming Interface) 정책에 따라 ‘방위비 분담금’을 키워드로 SNS Data Collector 사용하여 2020년 2월 19일에 검색을 실시하였으며 트윗 3,000개를 수집하였다. 위 수집된 내용을 Netminer 4.4.1.C의 ‘Filter and Dictionart Settings’ 기능에서 ‘방위비 분담금’, ‘미국’, ‘한국’, ‘미군’의 4개를 단어를 제외어로 등록하여 비정형 텍스트 데이터인 트윗 3,000개에서 연관된 단어를 추출하였다. 트위터의 텍스트 자료의 종류는 ‘트윗’이며, 수집기간은 기준일의 최근 7일이며, 수집량은 3,000 ‘트윗’으로 한정하였다.

3.2 분석도구

본 연구에서는 Netminer 4.4.1.C 버전을 사용하여 분석하였으며 확장된 기능의 모듈형식으로 제공되는 SNS Data Collector에서 트위터의 댓글이나 영상의 제목, 사용자 아이디 등을 자료수집 하였다. Netminer 4.4.1.C은 SNS(인스타그램, 트위터, 페이스북, 유투브) 데이터 수집이 가능하고 오픈 API 정책를 활용하여 프로그래밍을 통해 수집할 수 있다는 특징을 가지고 있다. 본 연구에서는 SNS Data Collector를 사용하여 트위터 API 정책에 따라 ‘방위비 분담금’ 라는 키워드로 수집된 트윗을 기초로 Netminer 4.4.1.C를 통해 시각적 분석, 소셜 네트워크 분석, LDA 기법을 적용하였다.

3.3 연구방법 및 절차

트위터의 ‘트윗’(비정형 텍스트 데이터)에 대해 ‘방위비 분담금’이라는 특정 단어를 SNS Data Collector를 통해 검색한다. 수집된 단어 중 등장빈도와 종류를 고려한 단어의 크기가 결정되진 단어 구름으로 시각적 분석을 실시한다. 동시에 방위비 분담금과 관련된 단어 간 연결구조, 응집구조, 중심구조를 네트워크 모형으로 설계하기 위해 소셜 네트워크 분석을 실시한다. 소셜 네트워크 분석은 트윗 내용 중 어떠한 단어들이 밀접하게 응집 또는 서로 연결되어 있으며, 네트워크 모형의 중앙에 특정 단어가 위치하는지를 확인할 수 있다.

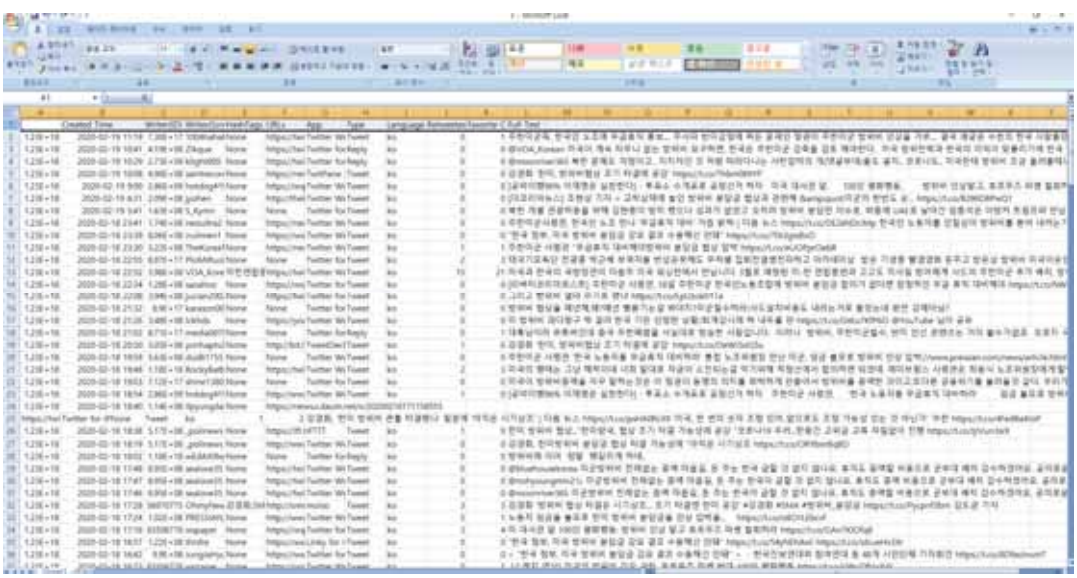
마지막으로 단어들의 품사와 중요도를 판별하고 단어의 문서 내 위치를 코딩하여 주요 토픽을

추출하는 텍스트 마이닝 분석을 실시하였다. 트윗 내용의 특징을 분석하기 위해 텍스트 마이닝 기법 중 하나인 LDA 모델링을 기반하여 트윗 3,000개의 트윗을 분석하여 주제를 분류하고 주제별 구성단어와 트윗갯수를 확인함으로써 주요 내용을 설명하는 방법을 적용하였다.

Ⅳ. 분석결과

4.1 텍스트 데이터 수집

본 연구는 분석도구인 SNS Data Collector를 활용하여 트위터에서 제공하는 오픈 API 정책에 따라 검색 기준일인 2020년 2월 19일로 최대 7일간의 트윗에 대해 텍스트 데이터를 수집하였다. 이를 통해 3,000개의 트윗을 140자의 단문 문서로 수집하였으며 아래 그림 5와 같다.



<그림 5> 트위터 데이터 수집 화면

수집한 3,000개의 트윗을 다시 Netminer 4.4.1.C의 'Filter and Dictionart Settings' 기능을 통해 언어는 Korean, 동시출현 단위는 Sentence, Document, 추출 품사는 명사로 설정한 후 '방위비 분담금', '미국', '한국', '미군'의 4개를 단어를 제외어로 등록하여 비정형 텍스트 데이터인 트윗 3,000개에서 연관된 단어를 추출하였다.

실제 3,000개의 트윗 중 '미군', '미국', '한국', '방위비 분담금' 의 4가지 명사를 제외어로 설정한

이유는 직접적으로 언급된 동일 단어인 ‘방위비 분담금’을 제외하였고 협상 당사국인 ‘미국’, ‘미군’, ‘한국’을 제외하였다. 그러나 ‘주한’이라는 단어는 이번 연구의 목적상 Ⅱ. 이론적 배경 부분에서 2.1 방위비 분담금의 정의를 고려하여 대중들이 ‘방위비 분담금’의 대상 및 주체의 인식 정도를 확인하고자 제외하지 않았다. ‘방위비 분담금’의 협상 당사국이 아닌 ‘중국’도 제외단어로 선정하지 않고 분석하였다.

4.2 시각적 분석결과

본 연구에서는 시각적 분석을 위해 단어구름을 사용하였으며 ‘방위비 분담금’라는 단어가 들어가 있는 트윗을 통해 트위터 사용자들의 다양한 생각을 분석하였다. Netminer 4.4.1.C을 통해 분석한 결과 단어구름은 그림 6과 같다. 3,000개의 트윗 중 1,058개 단어를 확인할 수 있다.

<그림 6> ‘방위비 분담금’에 대한 단어 구름

그림 6에서 단어의 크기가 가장 크고 눈에 잘 띄는 단어는 ‘사드’, ‘인상’, ‘부담’, ‘주한’, ‘공사비’ 등 이다. 단어 구름의 단어 크기는 빈도수로 반영되어 시각화 해주는 특징을 지니고 있지만 수치를 표현해주지는 못한다. 단어구름에 제시된 주요 단어들의 빈도수를 아래 표 5를 통해 확인할 수 있다.

<표 5> '방위비 분담금'에 대한 빈도수 기준 상위 20개 단어

순 위	단어	빈도수	순 위	단어	빈도수
1	사드	64	11	국민	25
2	주한	37	12	충당	24
3	인상	34	13	반대	24
4	부담	33	14	민중	24
5	협상	31	15	협의	23
6	다음	31	16	돈	20
7	뉴스	31	17	트럼프	19
8	공사비	29	18	철수	19
9	기지	28	19	비용	19
10	정부	26	20	동맹	19

표 5와 같이 '방위비 분담금'에 대한 트위터 트윗 내용의 주요 20개 단어의 빈도수가 큰 단어는 '사드', '주한', '인상', '부담', '협상' 이 빈도수가 높게 나타났다. 그 이유는 현재 방위비 분담금 협상이 진행 중에 언론 및 뉴스 등에 협상 주요 쟁점사항이 대중에게 전파되기 때문이다. 사드 기지 건설, 주한미군 철수, 방위비 분담금 인상, 인상률 부담, 교착된 협상진행 등이 현재 방위비 분담금 협상의 주요 이슈이다. 이러한 결과를 토대로 볼 때 '방위비 분담금'에 대해 현재 대중의 인식은 정부가 '방위비 분담금' 협상에 난항을 겪고 있으며 대중들은 이러한 인식이 트위터에 잘 반영되어 있다고 볼 수 있다.

4.3 소셜 네트워크 분석결과

방위비 분담금과 관련된 단어 간 연결구조, 응집구조, 중심구조를 네트워크 모형으로 설계하기 위해 소셜 네트워크 분석을 실시하였다. 소셜 네트워크 분석은 트윗 내용 중 어떠한 단어들이 밀접하게 응집되고 연결되고 어떠한 단어가 중앙에 위치하는지를 확인할 수 있게 네트워크 모형으로 제시하고 나타난 단어들를 분석하고 추정하여 방위비 분담금에 대한 유의미한 의미를 설명할 수 있다. 여러 중심성 지표 중 연결중심성, 근접중심성, 매개중심성이 주로 많이 사용되고 있으며 본 연구도 위 3가지 지표를 적용하여 분석을 실시하였다. Netminer 4.4.1.C을 통해 응집성과 중앙성이 높은 단어들을 우선 분석 하였으며, 소셜 네트워크 분석간 중심구조를 설명할 수 있는 '방위비 분담금'라는 단어를 검색하여 얻은 트윗 3,000개에 사용된 단어들의 연결 중심성 지수는 표 5와 같다.

<표 6> '방위비 분담금'에 대한 트윗에 포함된 단어들의 연결 중심성

순 위	단어	연결 중심성	순 위	단어	연결 중심성
1	사드	0.041667	11	협상	0.013889
2	반대	0.034722	12	학생	0.013889
3	후보	0.027778	13	표	0.013889
4	출마	0.027778	14	평화	0.013889
5	비용	0.027778	15	파병	0.013889
6	배치	0.027778	16	파기	0.013889
7	선언	0.020833	17	통일	0.013889
8	동맹	0.020833	18	추진력	0.013889
9	회계	0.013889	19	증액	0.013889
10	확장	0.013889	20	중국	0.013889

표 6의 분석결과, 단어 구름 분석 결과와 유사하게 상위 20개 단어에 '사드', '비용', '배치', '협상' 관련 단어가 다수 포함되어 있는 것을 알 수 있다. 특히, '사드' 관련 단어들이 높은 연결 중심성 지수를 갖고 있는 것을 통해 볼 때 '방위비 분담금'에 대한 주요 쟁점으로 사용되고 있다는 것을 확인할 수 있었다. 또한 '반대', '후보', '출마', '학생', '표' 관련 단어들이 나타난 이유는 검색일을 기준으로 7일간 '방위비 분담금' 관련 이슈와 관련된 단어로, 미 대사 관저에서 월담시위를 한 김유진 민중당 비례후보의 방위비 분담금 인상 반대 관련 내용으로 분석할 수 있다.

두 번째로 매개 중심성을 이용하여 '방위비 분담금'에 있어서 주요 역할을 하는 키워드를 확인하였다. 매개 중심성은 특정 키워드가 네트워크 내 다른 키워드 쌍의 최단경로상에 위치하는 정도를 측정하여 계산된다. 이를 통해 네트워크 내의 정보교환 또는 자원흐름에 대한 통제 능력을 분석할 수 있으며 매개 중심성 지수는 표 7과 같다.

<표 7> '방위비 분담금'에 대한 트윗에 포함된 단어들의 매개 중심성

순 위	단어	매개 중심성	순 위	단어	매개 중심성
1	사드	0.084241	11	합리	0.03541
2	동맹	0.059618	12	이치	0.03498
3	국민	0.055306	13	구속	0.033197
4	부담	0.048651	14	수개월	0.032862
5	외교	0.045484	15	인상	0.030832
6	요구	0.044066	16	비용	0.030725
7	협상	0.043981	17	공정	0.030647
8	반대	0.041558	18	기만	0.030618
9	중국	0.041342	19	선언	0.030577
10	국가	0.039985	20	문제	0.030474

즉, 매개 중심성이 높은 키워드는 '방위비 분담금'에서 여러 키워드들과 함께 중요한 키워드로 이용되고 매개 역할을 하고 있다. 높은 매개 중심성 지표값을 가지는 키워드는 '사드'로 연결 중심성 지표와 매개 중심성 지표 모두 높게 나타나는데, 이는 해당 키워드가 많은 트윗에서 언급됨과 동시에 다른 키워드와의 연계를 위한 역할에서도 크게 기여하고 있다고 이해할 수 있다. 한편, 대체적으로 연결 중심성이 높은 키워드가 매개 중심성 지표에서도 높은 값을 가지는 경향을 보이고 있다. 결론적으로 매개중심성이 높은 단어는 '방위비 분담금' 이란 타이틀에 하위 주제들을 이어주는 역할을 하고 있다.

세 번째로 근접 중심성 분석이다. 근접 중심성은 키워드 사이의 거리를 나타내는데, 키워드 간의 가까운 정도를 알아보는 지표이다. 즉, 키워드와 키워드가 연결되기 위해 지나가는 키워드 수의 합을 말한다. 즉 네트워크 내의 다른 키워드들과 얼마나 근접하게 연결되어 있는가를 알아보는 것으로 해당 키워드가 얼마나 네트워크 중앙에 위치해 있는가를 측정하는 것으로 표 8과 같다.

<표 8> '방위비 분담금'에 대한 트윗에 포함된 단어들의 근접 중심성

순 위	단어	근접중심성	순 위	단어	근접중심성
1	사드	0.224289	11	협의	0.203075
2	비용	0.219746	12	반대	0.202477
3	인상	0.218118	13	요구	0.202477
4	국민	0.217199	14	강요	0.20208
5	증액	0.214486	15	압박	0.200702
6	추가	0.214263	16	북한	0.200507
7	부담	0.210114	17	트럼프	0.200507
8	협상	0.20926	18	예년	0.19973
9	관저	0.209048	19	미사일	0.199537
10	대사	0.208836	20	합리	0.199537

근접중심성은 네트워크 내의 주제어들 사이의 최단거리를 기준으로 하는 중심성으로 연결중심성과는 달리 직접적인 연결 관계를 고려한 지표이다. 따라서 주제어의 근접중심성이 높다는 것은 연결망 내의 다른 주제어들과의 최단거리가 짧아 '방위비 분담금'에 가장 빠르게 영향을 미칠 수 있는 위치에 있다는 것을 의미한다. 위 표 8과 같이 '사드'가 근접중심성이 가장 높게 나타났으며 '비용', '인상', '국민' 등이 네트워크의 중심에 위치하고 있어, '방위비 분담금' 협상간 주요 쟁점사항인 국방력과 예산 등이 '방위비 분담금'을 설명하는데 중심적 역할을 함을 알 수 있다. 이러한 분석결과는 단어 단위의 분석으로 연결된 전반적인 사고를 추정하기에는 한계가 있다. 따라서 네트워크망으로 이러한 단어들의 연결을 분석할 필요가 있다.

본 연구의 네트워크망 구성은 Kamada와 Kawai(1989)가 제안한 네트워크 구성 방법을 적용하여

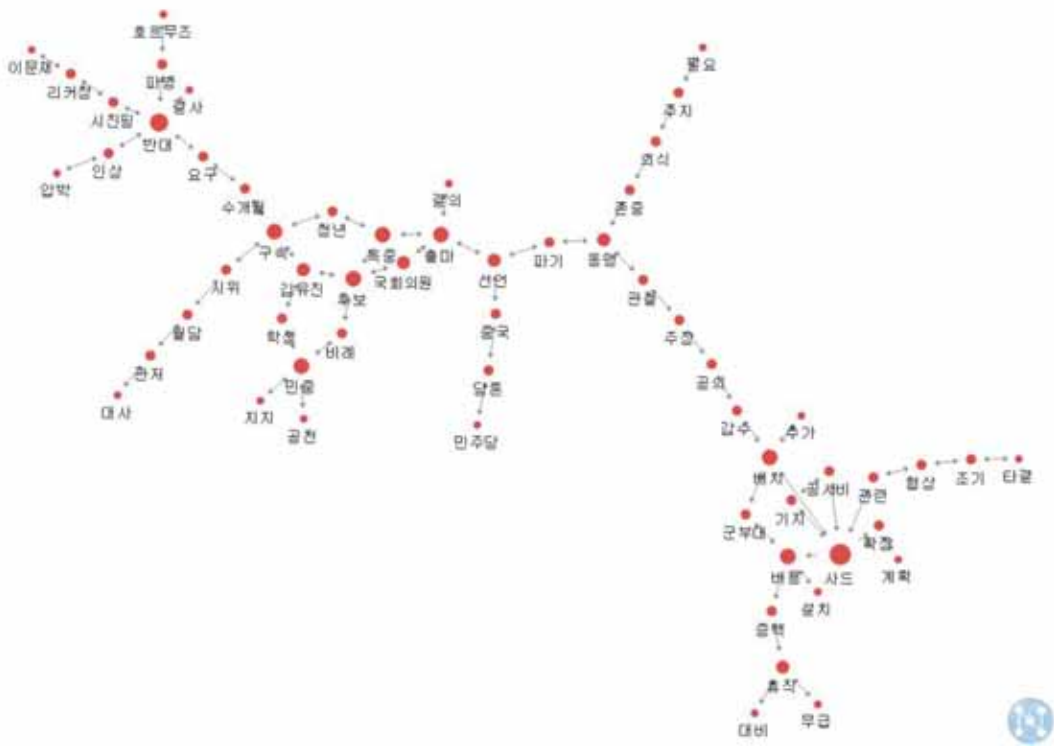
단어들 간의 네트워크를 구성하였으며 '방위비 분담금' 키워드로 트윗 3,000개의 단어들을 네트워크망으로 분석한 결과는 그림 7과 같다.

<그림 7> '방위비 분담금'에 대한 단어관계망 (2회 이상 함께 나온 단어)

그림 7과 같이, 응집성과 중앙성을 나타내는 연결 중심성은 원 모양으로 나타내며 원 지름이 큰 단어들의 연결을 확인해 보면 트위터에서 '방위비 분담금'을 분석할 수 있다. 원지름의 크기는 연결 중심성의 지수와 비례하며 트윗의 전반적인 네트워크는 '사드', '부담', '비용', '트럼프', '배치', '기지' 들이 서로 연결되어 있음을 알 수 있다. 또한, '인상', '문제', '비용', '고집', '시진핑', '반대' 들이 서로 연결되어 있음을 알 수 있으며 '국민', '북한', '안전', '미사일', '연계', '투표' 들이 서로 연결되어 있고 '협상', '예산', '타결', '이해', '조기', '통상' 들이 서로 연결되어 있음을 알 수 있다.

그림 7에서 제시된 단어 중 '신무' 등의 단어는 SNS Data Collector에서 텍스트를 수집시 140자 단어까지 수집되어 이후 문장은 수집되지 않은 단어로 트위터 원본에서 '신무기'로 확인되었다. 또한 '때', '긴말', '청' 등의 단어는 추출과정에서 트윗 작성자가 문장작성 시 문장구조가 완벽하지 못한 트윗에서 수집된 내용으로 확인이 되었다. 즉, '방위비 분담금'의 네트워크 구조는 크게 '사드', '주변국', '북한', '협상' 등의 관련 글로 주로 구성이 되어있음을 알 수 있다. 그림 7은 2회 이상 함께 나온 단어를 기준으로 네트워크 모형을 분석하였지만 좀 더 구체적이고 관련이 적은 단어를

제외하기 위해 그림 8은 3번 이상 같이 나온 단어들로 단어관계망 분석하였다.



<그림 8> '방위비 분담금'에 대한 단어관계망 분석 결과 (3회 이상 함께 나온 단어)

그림 8에 제시된 네트워크 모형의 결과는 '사드', '공사비', '확장', '기지', '비용'들이 서로 연결되어 있음을 알 수 있다. 또한, '반대', '시진핑', '파병', '인상', '요구' 들이 서로 연결되어 있음을 알 수 있으며 '후보', '김유진', '국회의원', '옥중', '비례'들이 서로 연결되어 있음을 알 수 있다. 즉, '방위비 분담금'의 네트워크 구조는 크게 '사드', '반대', '후보' 등의 관련 주제로 구성되어 있음을 알 수 있다.

4.4 텍스트 마이닝 분석결과

본 연구에서는 텍스트 마이닝 분석을 위해 토픽 모델링인 LDA 기법을 적용하여 '방위비 분담금'과 관련된 단어를 포함하는 트윗을 3,000개를 대상으로 분석을 실시하였다. LDA 기법을 적용하기 위해 최적의 토식수를 산정하는 방법 중 여러 개의 토픽수를 부여하여 얻어지는 결과를 가지고 주제 범주화가 가장 잘 되었다고 분석자가 판단하는 토픽수를 선택하는 방법으로 선정하였다(이수

상, 2016). 이 경우 연구자가 스스로 판단해야 하는 부담이 있을 수 있으나 토픽수를 30개, 25개, 20개, 15개, 10개, 5개 순으로 분류한 후 위에서 분석한 시각적 분석과 소셜 네트워크 분석을 통해 확인된 단어를 고려하여 최종 토픽수를 5가지로 선정하였다. 이후, LDA 기법에 토픽의 수를 5개로 설정하여 연구를 진행하였다. 그 결과는 그림 9와 같이 5개의 주제로 분류되었다.

Report : LDA

• **Information**

PROCESS			
	Menu	Mining >> Text >> Topic >> LDA	
	Options	- # of Topics : 5 - Random Seed : 1	
	Time	February 19, 2020 12:44:39 PM KST~ February 19, 2020 12:44:40 PM KST 654 ms	
DATA			
	Project	Untitled Project	
	Workfile	[Content] '방위비'(2020-02-19 12:35:01)	
	Input	**2-mode Network**	Words-Tweets Network

• **Output Summary**

TOPIC INFO

	1st Keyword	2nd Keyword	3rd Keyword	4th Keyword	5th Keyword
Topic-1	인상	철수	트럼프	평화	북한
Topic-2	협상	주한	휴직	강경화	한국인
Topic-3	국민	정부	수준	미사일	김현종
Topic-4	사드	뉴스	다음	부담	공사비
Topic-5	민중	반대	대사	김유진	관저

DOCUMENT CLASSIFICATION STATISTICS

	# of documents
Topic-1	51
Topic-2	51
Topic-3	43
Topic-4	58
Topic-5	25

<그림 9> '방위비 분담금'과 연관성이 높은 3,000개 트윗에 대한 LDA 결과

LDA 모델링 결과 총 5개의 주제로 분류되었으며 이 5개의 주제의 총 트윗수는 228개의 트윗으로 분류되었다. 첫 번째 주제는 58개의 트윗이 속한 4번 주제로 '사드', '뉴스', '다음', '부담', '공사비'이다. 2번째 주제는 51개의 트윗이 속한 1, 2번 주제로 1번은 '인상', '철수', '트럼프', '평화', '북한'이며 2번은 '협상', '주한', '휴직', '강경화', '한국인' 이다. 3번째 주제는 43개의 트윗이 속한 3번이며 '국민', '정부', '수준', '미사일', '김현종' 이었다. 마지막으로 4번째 주제는 '민중', '반대', '대사', '김유진', '관저'로 나왔다. LDA 기법으로 주제별 단어를 확인할 수 있지만 이러한 정보가 대중이 인식하는 방위비 분담금을 정확하게 표현하는지 여부를 확인하는데 제한이 있다. LDA의 주

요 특징 중 제시된 단어 외에도 숨겨진 의미의 단어가 주제별로 선정될 수 있어 방위비 분담금과 관련된 뉴스와 보도문 등을 확인하여 토픽 수에 따라 분류된 단어를 활용하였으며 토픽주제는 토픽별 주요 단어를 통해 연구자가 직접 지정하였다. 토픽주제를 지정할 때는 토픽 내 순위가 높고 해당 토픽에만 나타난 특징적인 단어를 참고하였으며 그 내용은 표 9와 같다.

<표 9> LDA를 재구성한 토픽별 주제

순 번	주 제	소속 트윗(수)
1	사드 기지 공사비 부담	58
2	주한미군 철수 및 북한과의 평화협정을 고려한 트럼프 대통령	51
3	주한미군이 고용한 한국인 근로자의 무급휴직 가능성	51
4	한·미 미사일 지침 개정안을 조율	43
5	방위비 분담금 인상 반대	25

표 9과 같이 총 5개의 주제 중 5번 주제를 제외한 나머지 주제가 20년 방위비 분담금 협상간 주요 쟁점과 관련된 내용인 것을 확인할 수 있었으며 방위비 분담금이라는 키워드로 3,000개의 트윗를 추출하여 이를 5개의 주제로 구분하면 현재 '방위비 분담금'의 주요 쟁점사항과 관련된 내용으로 구성되어 있음을 확인할 수 있었다. 트윗의 내용으로 '방위비 분담금'의 대중의 인식을 재현할 때 '사회현상과 현재 이슈와 얼마나 일치하는가?' 에 대한 과정을 반복적으로 LDA 기법을 통해 수행함으로써 제일 근접한 위의 5가지 주제가 도출되었으며 실제로 검색일을 기준으로 7일간 언론보도가 많이 되었으며 대중이 인식과 트위터의 트윗 내용이 대부분 일치하다는 것을 유추할 수 있었다. 이는 트위터의 특징 중 개인의 의견을 트윗을 통해 전파하고 표출할 때 오랜시간을 거쳐 주제를 선정하고 장문으로 자신의 의견을 표출하는 것이 아니라 바로 보고 바로 표출하는 트윗의 기능도 반영된 것으로 볼 수 있다.

Ⅴ. 결론 및 논의

본 연구에서는 소셜 미디어의 빅데이터에 나타난 대중들의 '방위비 분담금'에 대한 인식을 알아보고자 하였으며 이를 위해 '방위비 분담금'에 대한 텍스트 데이터를 수집하고 정제과정을 거친 후 시각적 분석, 소셜 네트워크 분석, 텍스트 마이닝 분석을 실시하였으며 결과는 다음과 같다.

첫째, '방위비 분담금'에 대한 사회적 인식조사를 목적으로 트위터에서 수집된 데이터 중 빈도수 기준 상위 20개 단어를 산출한 결과 '사드', '주한', '인상', '부담', '협상'이 선정되었다. 이를 통해 대중이 방위비 분담금에 대한 협상대상 및 협상 주요 쟁점 등을 올바르게 인식하고 있음을 확인할

수 있었다.

둘째, 소셜 네트워크를 통해 연결 중심성, 매개 중심성, 근접 중심성을 분석한 결과 연결 중심성 지수가 높은 단어로 '사드', '주변국', '북한', '협상', '반대', '후보' 가 선정되었으며, 매개 중심성 지수가 높은 단어는 '사드', '동맹', '국민', '부담', '외교' 로 선정되었다. 또한 근접 중심성 지수가 높은 단어는 '사드', '비용', '인상', '국민', '증액'으로 선정되었다. 이는 '방위비 분담금'을 설명하는데 위의 단어들이 중심적인 역할을 하고 있으며 대중들이 '방위비 분담금'에 대한 인식을 나타내는 주요 단어들이라고 할 수 있겠다.

마지막으로 트윗이 어떤 내용으로 구성되어 있으며 주로 어떤 내용을 포함하는 트윗인지를 확인한 결과, 트윗 사용자의 방위비 분담금에 대한 인식이 20년 '방위비 분담금' 협상간 주요 쟁점과 관련된 내용인 것으로 확인할 수 있다. 주요 토픽은 '사드 기지 공사비 부담', '주한미군 철수 및 북한과의 평화협정을 고려한 트럼프 대통령', '주한미군이 고용한 한국인 근로자의 무급휴직 가능성', '한·미 미사일 지침 개정안을 조율', '방위비 분담금 인상 반대'로 토픽을 구분하였다.

본 연구에서는 정량적인 분석을 위해 소셜 미디어의 빅데이터 분석 중 시각적 분석, 소셜 네트워크 분석, 토픽 모델링을 활용하였으나. 트위터 사용자가 국민을 대표한다고 볼 수 없기 때문에 연구의 한계점을 가진다. 또한 트위터의 오픈 API 정책으로 약 7일간의 짧은 기간 동안 특정 소셜 미디어에 나타난 사용자 트윗을 분석한 결과가 소셜 미디어를 사용하지 않는 인원들의 의견이나 소셜 미디어에 나타나지 않은 기타인식과 표현은 배제되어 있다. 또한, 협상 당사국의 방위비 분담금에 대한 인식과 관련 연구가 미비하다는 한계를 지니고 있다. 그러나 텍스트로 구성된 자료를 정량적으로 계량하여 시각적으로 분석하였으며 3,000개의 트윗에 드러난 트위터 사용자의 일상적인 의사표현들은 기타 다른 방법으로는 얻기 힘든 중요한 정보를 식별하였다는 것이 의의가 있겠다. 이러한 내용이 방위비 분담금 협상에 고려되어 적용이 가능할 것이라 판단된다. 추가로 소셜 미디어 사용자들의 일상적인 표현과 반응을 시각적 분석과 소셜 네트워크 분석, 토픽 모델링으로 구분하여 서로 비교함으로써 의미를 도출하는 방법론을 제시하였다.

소셜 빅데이터 분석을 활용해 도출한 본 연구의 결과를 통해 '방위비 분담금'에 대한 대중들의 인식을 확인할 수 있었다. 한미 관계는 중요하고 '방위비 분담금' 역시 중요하다. 과거를 돌이켜보았을 때 지속적으로 이어질 과제라고 할 수 있다. '방위비 분담금'과 관련하여 국민의 공감대 형성과 더불어 소셜 빅데이터 분석을 활용하여 협상 후 대중의 인식을 확인하거나 협상국의 트위터를 분석하는 등 좀 더 실질적이고 현실성 있는 후속 연구가 필요함을 시사한다. 현재 방위비 분담금에 관심이 높아지고 있는 시점에서 정부는 전략적 의사소통(SC)을 효과적으로 활용하여 국내 여론과 양국의 국민적 공감대 형성이 필요할 것이다.

참고문헌

강애띠, 강영옥 (2018). 트윗데이터를 활용한 스트레스 토픽의 지역별 특징 분석. 한국지도학회지, 18(2), 53-69. http://dx.doi.org/10.16879/jkca.2018.18.2.053

권은영, 이종길, 박선욱 (2019). 소셜 빅데이터 분석을 통해 알아본 대중의 '미드'에 과한 인식 및 자기주도적 영어학습 행태: 유투브와 트위터를 중심으로, 영상영어교육학회지, 20(3), 25-48. http://dx.doi.org/10.16875/stem.2019.20.3.25

권헌철 (2011). 주한미군의 가치 추정: 경제에 미치는 영향과 대체비용 추정. 국방연구, 54(2), 23-45. G704-001669.2011.54.2.006

김용우, 강행봉 (2016). 소셜 네트워크 서비스의 단어 빈도와 범죄 발생과의 관계 분석. 정보처리학회 논문지: 컴퓨터 및 통신시스템, 5(9), 229-236. http://dx.doi.org/10.3745/KTCCS.2016.5.9.229

김열수, 김경규 (2017). 트럼프 시대의 한미 군사쟁점과 대응방향. 신아세아, 24(1), 29-55. G704-000851.2017.24.1.006

김새록, 김주연 (2019). 텍스트마이닝을 통한 국내 전시회의 관람동기 분석 연구. 한국공간디자인학회, 14(2), 11-20. http://dx.doi.org/10.35216/kisd.2019.14.2.11

이승후, 김학선 (2018). 텍스트마이닝 기법을 활용한 유스호스텔의 활성화 방안에 관한 연구: 부산 유스호스텔 아르피나를 중심으로, 산업혁신연구, 34(2), 267-288. http://dx.doi.org/10.22793/indinn.2018.34.2.010

박원곤 (2019). 한미 방위비 분담 협상을 위한 소고. 한국국가전략, 4(2), 145-173.

박자현, 송민 (2013). 토픽모델링을 활용한 국내 문헌정보학 연구 동향분석. 정보관리학회지, 30(1), 7-32. http://dx.doi.org/10.3743/KOSIM.2013.30.1.007

박휘락 (2015). 한국과 일본의 대미 방위비분담 비교: 분담금 협상방식을 중심으로. 입법과 정책, 7(2), 5-32. G704-SER000003935.2015.7.2.018

백재옥 (2017). 일본의 방위비분담 분석 및 정책적 시사점. 국가전략, 23(3), 125-152. http://dx.doi.org/10.35390/sejong.23.3.201709.005

심준식, 김형중 (2017). LDA 토픽 모델링을 활용한 판례 검색 및 분류방법. 전자공학논문집, 54(9), 67-75.

신종호 (2009). 한국의 방위비분담의 현황과 쟁점 및 개선방안. 국가안보와 전략, 9(1), 35-66. G704-SER000010557.2009.9.1.002

이응균, 서민구 (2019). 워드클라우드를 이용한 국내・외 BIM 연구 동향 분석. 한국건축시공학회지, 19(1), 95-103. http://dx.doi.org/10.5345/JKIBC.2019.19.1.095

이수상 (2016). 독후감 텍스트의 토픽 모델링 적용에 관한 탐색적 연구. 한국도서관・정보학회지, 47(4), 1-18. http://dx.doi.org/10.16981/kliss.47.4.201612.1

정원열, 김종수 (2016). 주한미군 방위비 분담금에 대한 실증적 요인분석. 한국방위산업학회, 23(1), 75-89. G704-SER000001543.2016.23.1.003

채희찬, 이종욱, 최윤아, 박대희, 정용화 (2018). 트윗 텍스트 마이닝 기법을 이용한 구제역의 감성분석. 정보처리학회논문지. 소프트웨어 및 데이터 공학, 7(11), 419-426. http://dx.doi.org/10.3745/KTSDE.2018.7.11.419

김용학・김영진 (2016). 사회 연결망 분석. 서울: 박영사.

대한민국 국방부(2018). 국방백서.

대한민국 국방부(1994). 주한미군을 위한 한국정부의 방위비분담.

박기학 (2017). 트럼프 시대, 방위비분담금 바로 알기. 서울: 한울아카데미.

이시영, 한태준 (2000). 주한미군의 경제적 가치측정 및 평가. 한국전략문제연구소 연구보고서, 323, 25-36.

정용찬 (2013). 빅테이터. 서울: 커뮤니케이션북스.

한국 국방연구원(1998). 한미동맹의 경제적 역할 평가 및 정책 방향

함유근, 채승병 (2012). 빅데이터, 경영을 바꾸다. 서울: 삼성경제연구소.

Blei, D., Ng, A, Y.,& Jordan, M (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3, 993-1022.

Bonacich, P. (2007). Some unique properties of eigenvector centrality. Social Networks, 29(4), 466-484. https://doi.org/10.1016/j.socnet.2007.04.002

Cheong, S. H., & Si, Y. W. (2016). Accelerating the Kamada-Kawai algorithm for boundary detection in a mobile ad hoc network. ACM Transactions on Sensor Networks (TOSN), 13(1), 1-29. https://doi.org/10.1145/3005718

Defense Technical Information Center. (2004). 2004 Statistical compendium on Allied Contributions to the Common Defence.

Freeman, L, C. (2000). Visualizing social networks. Journal of Social Structure, 1(1), 4. https://www.cmu.edu/joss/content/articles/volume1/Freeman.html

International Institute for Strategic Studies (2016). Military Balance 2016. New York: Routledge.

James Manyika & Michael Chui. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, 1, 1.

John Gantz & David Reinsel (2011). Extracting Value from Chaos. IDC IVIEW, 1142(2011), 1-12. https://www.whizpr.be/upload/medialab/21/company/IDC_1142.pdf

Maizels, Aflfred and Machiko K. Nissanke. (1986). The Determinants of Military Expenditures in Developing Countries. World Development, 14(9), 1125-1140. https://doi.org/

10.1016/0305-750X(86)90115-4

Netminer.(2019a). Product overview (Version 4.4.1.C) [Computer software]. Retrieved from http://www.netminer.com/product/overview.do

Netminer.(2019b). SNS data collector (Version 2.5.0) [Computer software]. Retrieved from http://www.netminer.com/product/extentions.do

Scott, J. (2012). Social Network Analysis 3rd. SAGE, 41p-52p

Wang, J.-C.,Chiang, C.-H., & Lin, S. -W. (2010). Network structure of innovation: Can Borkerage or closure predict patent quality?. Scientometrics, 84(3), 735-748. https://doi.org/10.1007/s11192-010-0211-y

Xie, h., & Sun, W. (2013). Word cloud visualization with abstract. In Society for Computational Design and Engineering (Ed.), Proceedings of the 2013 International Conference on CED (pp. 224-231). Seoul : Society for Computational Design and Engineering.

원 고 접 수 일 2020년 02월 27일
원 고 수 정 일 2020년 04월 01일
게 재 확 정 일 2020년 04월 22일

Abstract

https://doi.org/10.37944/jams.v3i1.64

Recognition of ‘Defense Cost Sharing’ through Big Data: Focusing on Twitter

Sungkon Kim*

(Korea Military Academy)

To study, Twitter, the social media, was selected and analyzed by applying word cloud, social network and LDA techniques as keywords. The results showed that the main key words based on frequency were “THADD”, “resident in Korea,” “rise,” “burdened,” and “negotiation,” while the main key words considering word relationship and centrality were “THADD,” “opposite,” “cost,” “deployment” and “aligned alliance.” In addition, LDA technology has shown that ‘THADD base construction costs’, ‘the withdrawal of U.S. troops from South Korea’, ‘Recommendation of unpaid leave of Korean workers’ and ‘the government’s efforts to reopen negotiations’ have become an issue in the current negotiation process related to ‘defense cost sharing’.

Through this, it was possible to identify the key words that the public recognizes “defense-sharing expenses,” and furthermore, the government should pursue strategic negotiation policies in consideration of the key words and topics recognized by the people in order to achieve successful negotiation results.

Keywords : Defense Cost Sharing, Twitter, Word Cloud, Social network, Latent Dirichlet Allocation (LDA)

This work is licensed under a Creative Commons Attribution 4.0 International License.

* Korea Military Academy (Kwangwoon University, Defense Acquisition Program, Ph.D student)