텍스트 네트워크분석을 활용한 국방분야 연구논문 지식구조 분석

이용규[O], 윤성웅[*], 이상훈[*]
[O*]국방대학교 컴퓨터공학과
e-mail: yongq7@gmail.com[O], ysw1209@gmail.com[*], hoony@kndu.ac.kr[*]

Knowledge Structure Analysis on Defense Research Using Text Network Analysis

Yong-Kyu Lee[O], Soung-woong Yoon[*], Sang-Hoon Lee[*]
[O*]Dept. of Computer Science, Korea National Defense University

● 요 약 ●

본 연구에서는 텍스트 네트워크분석을 활용하여 국방분야 연구의 핵심 주제어와 연구주제를 분석하고 이를 통해 전체 지식구조를 파악하고자 하였다. 이를 위해 2010년부터 2017년까지의 국방대학교 학위과정 논문을 대상으로 국방분야 연구현황을 진단하고 지식구조를 구성하였다. 8년간 누적된 논문 710건의 초록을 분석하여 총 6,883개의 단어를 추출한 후, 단어의 논문 등장 빈도수와 단어간 링크수를 파레토 법칙에 따라 상위 20%의 기준으로 총 270개의 단어로 추출하였고, 컴포넌트 분석을 통해 최종 170개의 핵심 주제어를 도출하였다. 이 핵심 주제어를 통해 중심성 분석과 응집구조를 분석하여, 국방분야에 대한 총 6개의 지식구조 그룹을 도출하였다.

키워드: 텍스트 네트워크분석(Text Network Analysis), 지식구조(Knowledge Structure)

I. Introduction

국방분야와 관련한 수많은 연구가 그동안 진행이 되어왔으나, 국방분야 연구현황 분석과 전체적인 지식구조에 대한 분석은 미진한 부분이 있다. 지식구조 분석은 국방분야의 전제적인 연구의 현황을 파악할 수 있게 하여, 이를 통해 향후 연구의 방향을 제시해 줄 수 있는 역할을 할 수 있다. 따라서 본 연구에서는 2010년부터 2017년까지의 국방대학교 학위과정 논문의 초록을 대상으로 국방분야의 지식구조를 분석하는 것을 목적으로 하였다. 이를 위해 초록 데이터에서 단어를 추출하고 전/후처리를 거쳐 핵심 주제어를 도출하였고, 이를 통해 중심성 분석, 응집구조 분석을 시행하여 국방분야의 중심 연구개념과 연구영역 현황을 파악하고자 한다.

II. Related works

기존의 문헌분석 방법은 많은 시간과 노력이 필요하고, 최근의 방대해진 데이터 양에서는 현실적인 한계가 더욱 두드러진다. 텍스트 네트워크분석은 광범위한 양의 텍스트 데이터를 효율적으로 분석하는데 효과적이다. 이러한 텍스트 네트워크분석의 장점은 텍스트에서 실제적이고 잠재적인 의미를 파악할 수 있다는 점이다. [1] 또한 언어의 구조를 시각적으로 구현함에 따라 단어 간의 관계를 쉽게 파악할 수 있으며, 네트워크분석을 통해 단순히 단어가 등장한 빈도수만 파악하는 것이 아닌 다른 단어와의 동시등장빈도 분석을 통해 단어 간의 관계를 파악할 수 있다. [2]

III. The Proposed Scheme

본 연구에서는 국방분야 연구 자료로 국방대학교 학위과정 논문의 초록을 분석 대상 데이터로 설정하고, 이러한 비정형 텍스트에서 의미형태소 단어를 추출한 후, 단어 간 동시등장행렬을 작성하여 네트워크분석을 진행하였다.

3.1 대상 데이터 선정

논문은 국방대학교 학위과정의 논문으로 선정하였다. 이는 국방대학위과정이 각 군의 소요연구 진행을 위해 군별 위탁교육생을 선발하여 연구가 진행되었다는 점을 반영한 것이고, 논문의 발행연도 기간은 2010년~2017년으로 제한하였다. 이는 국방대도서관의 논문 서지데이터 중 초록 데이터가 2010년부터 전산화 입력된 점을 고려한 것이다. 수집된 논문은 총 710편으로 해당 논문의 초록을 분석에 사용하였다.

3.2 단어 추출 및 정제

논문별 서지데이터(청구기호, 제목, 저자, 발행년도, 초록)를 엑셀파일로 구성한 후 SNA 분석 프로그램인 Netminer(4.4.0.b) [3]를 이용하여 논문별 초록에서 단어를 추출하였다. 추출된 단어는 총 6,883개로 이 중에서 분석 데이터로 부적절한 의미없는 1글자 단어, 일반적인 개념의 단어 등을 제외하여 6,597개로 1차 정제하였다. 네트워크분석에서 모든 단어를 대상으로 의미를 도출하고 지식지도를 구현하는 것은 제한되기 때문에 단어의 적절한 등장빈도 및 링크수를 정제 기준으로 적용한다. 따라서 본 연구에서는 단어의 등장분포에 따른 기준으로, 상위 20%의 데이터가 80%의 영향력을 가진다는 파레토 법칙을 적용하여, 등장빈도 8회 이상(비중 19.5%), 단어간 링크 상위 20%로 정제하여 총 270개의 단어를 추출하였다. 이후 270개 단어의 컴포넌트 분석을 통해 <표 1>과 같이 최종 정제된 단어로 174개가 추출되었다.

Table 1. Frequencies of keyword

Rank	Keyword	n	Rank	Keyword	n
1	미국	288	11	능력	190
2	체계	283	12	군사	186
3	수행	275	13	전략	183
4	관계	266	14	위협	179
5	국가	234	15	운용	179
6	정책	224	16	경제	177
7	환경	220	17	지원	174
8	개발	216	…	…	…
9	인식	197	…	…	…
10	효율	196	174	이지스	8

3.3 분석 및 시각화

네트워크분석을 위해 먼저 단어에 대한 등장빈도를 도출하고, 다음으로 단어 간 동시등장행렬을 작성하여 한 편의 논문에 동시에 등장한 단어를 분석한다. 이는 단어가 함께 등장한 논문의 빈도수가 높다면 단어들 간의 연관성이 높다는 것을 의미한다. 다시 말해 논문-단어의 2모드 네트워크를 작성한 후, 이를 단어-단어의 1모드 네트워크로 변환하는 것으로, 이 단어 동시등장행렬이 텍스트 네트워크가 된다. [4] 이를 통해 중심성 분석, 응집그룹 분석을 진행하고 국방분야의 지식지도를 시각화하였다.

IV. Results

4.1 중심성 분석

중심성 분석은 네트워크에서 단어가 중심에 위치하는 상대적 정도를 나타내며, 중심성이 높은 단어일수록 핵심적인 단어로 볼 수 있다. [5] 본 연구에서는 다른 단어들과 가장 많이 연구되어진 정도를 나타내는 연결중심성(degree centrality), 하부 연구영역을 연결해주는 중개자 역할을 얼마나 수행하는지를 측정하는 매개중심성(betweenness centrality)을 분석하였다.

Table 2. Centrality of keyword

Rank	Degree Centrality		Betweenness Centrality	
	Keyword	Ratio(%)	Keyword	Ratio(%)
1	매개	2.82	위협	9.15
2	체계	2.82	미국	7.58
3	국가	2.26	국가	5.51
4	미국	2.26	안보	5.13
5	탄도	1.98	북한	4.24
6	중국	1.69	체계	4.09
7	전략	1.69	김정은	3.98
8	김정일	1.41	관계	3.86
9	요격	1.41	조직	3.54
10	세습	1.41	구성원	3.33
11	위협	1.41	매개	3.29
12	후계	1.41	김정일	2.84
13	김정은	1.41	중국	2.76
…	…		…	
…	…		…	
174	역할	0.28	간부	0.00

최종 추출된 단어 174개에 대한 중심성 분석을 수행한 결과, <표 2>와 같이 연결중심성, 매개중심성 순위를 도출하였다. 먼저 다른 단어들과 활발히 연구되고 있는 연결중심성(degree centrality)이 높은 단어로는 매개(2.82%), 체계(2.82%), 국가(2.26%), 미국(2.26%), 탄도(1.67%), 중국(1.98%), 전략(1.69%) 순으로 확인하였고, 다른 연구분야와의 교량역을 하는 매개중심성이 높은 단어로는 위협(9.15%), 미국(7.58%), 국가(5.51%), 안보(5.13%), 북한(4.24%), 체계(4.09%) 순으로 나타났다.

<그림 1>과 <그림 2>는 174개 단어들의 연결중심성과 매개중심성을 네트워크로 시각화한 것으로, 각 중심성에 대한 판단이 쉽도록 연결중심성은 1.41% 이상, 매개중심성은 1.82% 이상이 되는 단어들을 각각 청색과 녹색으로 별도 표시하였다.

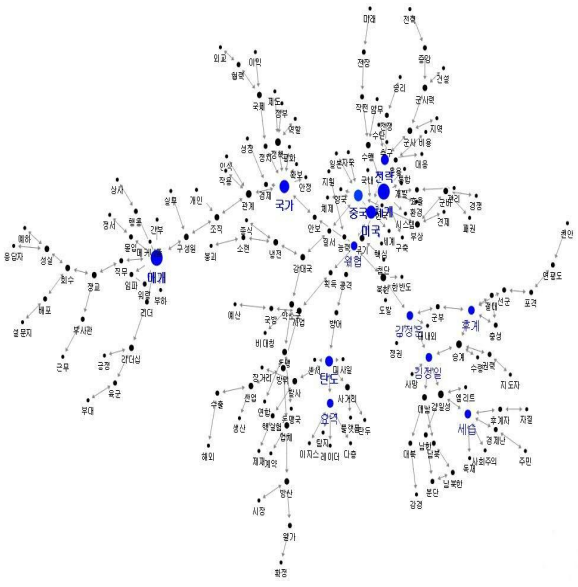

Fig. 1. Network of Degree Centrality

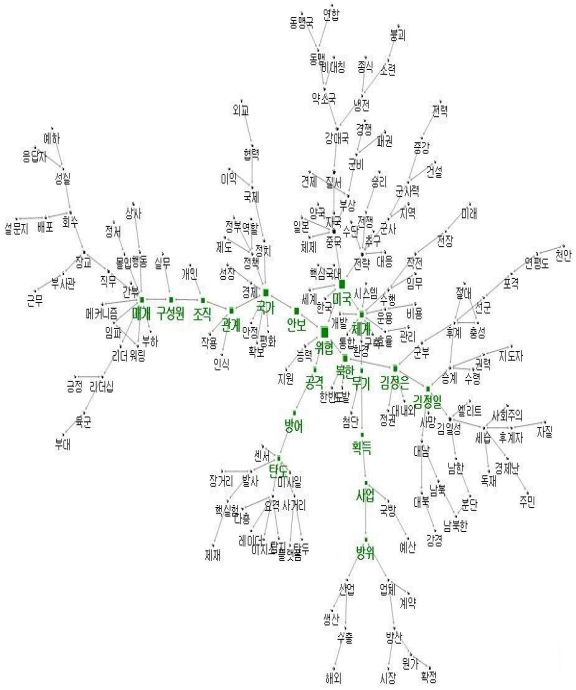

Fig. 2. Network of Betweenness Centrality

<그림 1>에서와 같이 연결중심성은 네트워크의 중심부와 분화되는 지점의 중심부에서 연결중심성이 높은 단어들이 위치하고 있는데, 이를테면 '미국' 주변에는'체계', '전략', '중국', '위협' 등의 연구가 상당히 활발하게 이루어지고 있는 네트워크의 중심부에 해당되겠고, '탄도' 주변은 '요격', '탐지', '사거리', '발사' 등의 북한의 탄도탄 위협에 대한 연구영역을 이루고 있는 것을 알 수 있다. <그림 2>는 중심부와 함께 하부 연구분야로 분화되는 지점 간에 매개중심성이 높은 단어들이 나타나고 있는데, '북한'→'김정은'→'김정일'로 이어지는 북한 권력에 대한 연구영역의 분화와 '체계'→'무기'→'획득'→'사업'→'방위'로 이어지는 방위사업 관련 연구로의 분화 등이 매개중심성을 통해 구분되어지고 있다.

4.2 응집성 분석

네트워크의 응집성 분석은 먼저 컴포넌트 분석으로 응집 컴포넌트를 추출하고 이에 대한 클러스터 분석을 통해 하부 연구영역을 도출해 낼 수 있다. 클러스터링을 위해 NetMiner에서 응집성(cohesion) 기능을 사용하여 가장 큰 하위 컴포넌트를 추출하고 커뮤니티 분석을 수행하였다. 응집성 분석에서는 커뮤니티 분석결과 지수를 사용하고 그룹 내에서 단어의 연결이 그룹 간의 연결보다 많도록 하여 자주 함께 연구되는 주제어들을 세부연구영역으로 구분지어 준다. 응집성 분석결과 아래 <그림 3>과 같이 국방분야 연구영역은 총 6개의 그룹으로 구분되었다.

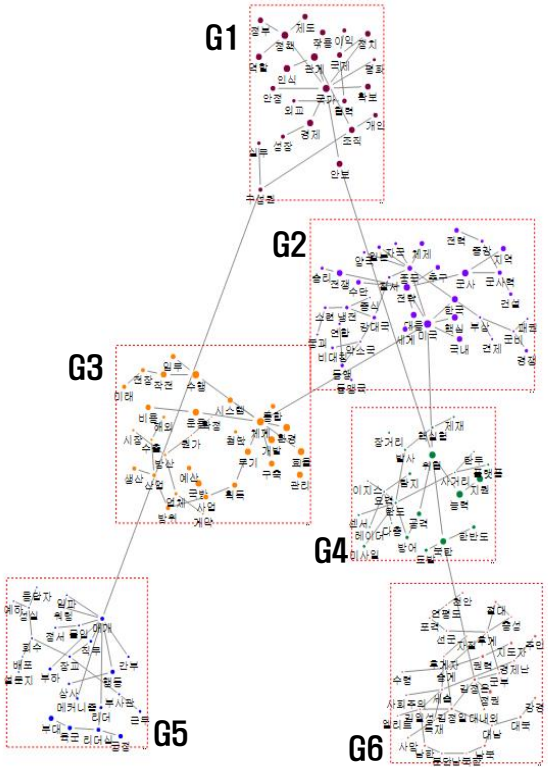

Fig. 3. Network of Subtopic Groups

본 연구의 국방분야 연구논문의 지식구조는 총 6개 그룹으로 구분되어 졌으며, 각 그룹의 주제어는 <표 3>과 같다.

Table 3. Group of keyword

Group	Keyword
G1 (국가안보 및 국방정책 연구)	개인, 경제, 관계, 구성원, 국가, 국제, 성장, 실무, 안보, 안정, 역할, 외교, 이익, 인식, 작용, 정부, 정책, 정치, 제도, 조직, 평화, 협력, 확보
G2 (국제정세 및 대외관계 연구)	강대국, 건설, 견제, 경쟁, 국내, 군비, 군사, 군사력, 냉전, 대응, 동맹, 동맹국, 미국, 부상, 붕괴, 비대칭, 세계, 소련, 수단, 승리, 약소국, 양국, 연합, 일본, 자국, 전략, 전력, 전쟁, 종식, 중국, 증강, 지역, 질서, 체제, 추구, 패권, 한국, 핵심
G3 (무기체계 및 방위산업 연구)	개발, 계약, 관리, 구축, 국방, 무기, 미래, 방산, 방위, 비용, 사업, 산업, 생산, 수출, 수행, 시스템, 시장, 업체, 예산, 운용, 원가, 임무, 작전, 전장, 첨단, 체계, 통합, 해외, 확정, 환경, 획득, 효율
G4 (북핵·탄도탄 위협 대비 연구)	공격, 능력, 다층, 도발, 레이더, 미사일, 발사, 방어, 북한, 사거리, 센서, 요격, 위협, 이지스, 장거리, 제재, 지원, 탄도, 탄두, 탐지, 플랫폼, 한반도, 핵실험
G5 (군 리더십 및 조직 연구)	간부, 근무, 긍정, 리더, 리더십, 매개, 메커니즘, 몰입, 배포, 부대, 부사관, 부하, 상사, 설문지, 성실, 예하, 워링, 육군, 응답자, 임파, 장교, 정서, 직무, 행동, 회수
G6 (남북관계 및 북한 정권 연구)	강경, 경제난, 군부, 권력, 김일성, 김정은, 김정일, 남북, 남북한, 남한, 대남, 대내외, 대북, 독재, 분단, 사망, 사회주의, 선군, 세습, 수령, 승계, 엘리트, 연평도, 자질, 절대, 정권, 주민, 지도자, 천안, 충성, 포격, 후계, 후계자

각 그룹의 주제는 해당 그룹 내의 논문과 주제어 등을 반영하여 설정하였다. 그룹 1은 '국가안보 및 국방정책 연구' 그룹으로 관계, 국가, 정책, 안보 등의 거시적 주제어와 개인, 역할, 협력, 구성원 등 이를 실행하는 구성원에 대한 연구가 이루어지고 있다. 그룹 2는 '국제정세 및 대외관계 연구' 그룹으로 미국, 군사, 전략, 한국, 대응, 등의 세계 정세 연구와 동맹, 연합, 견제 등의 대외관계적 주제어가 포함된 연구가 진행되고 있음을 나타내고 있다. 그룹 3은 '무기체계 및 방위산업 연구' 그룹으로 무기, 작전, 구축 등 무기체계 및 운용에 대한 연구와 체계, 개발, 방산, 수출 등 방위산업에 관한 연구가 진행됨을 알 수 있다. 그룹 4는 '북핵·탄도탄 위협 대비 연구' 그룹으로 핵실험, 미사일, 탄도, 위협 등의 위협 분석 연구와 함께 능력, 탐지, 요격, 이지스 등의 적 위협에 대한 대응능력에 관한 연구가 진행되고 있다. 그룹 5는 '군 리더십 및 조직 연구' 그룹으로 리더십, 부하, 간부, 직무 등 리더십과 조직에 관한 연구와 함께 이에 대한 연구의 방법으로 설문방식을 많이 사용하고 있음을 나타내고 있다. 그룹 6는 '남북관계 및 북한 정권 연구' 그룹으로 정권, 대북, 남한, 연평도, 포격 등의 남북관계와 권력, 김정은, 세습, 독재 등 북한 정권체제에 대한 연구가 진행되고 있음을 알 수 있다.

IV. Conclusions

본 연구는 이전 8년 간의 국방분야에 대한 연구를 텍스트 네트워크로 분석하여 연구의 중점 주제어와 지식구조를 파악하였다. 2010년~2017년 기간 총 710건의 국방대학교 학위과정 논문의 초록을 대상으로 174개의 주제어를 선정하여 국방분야 연구와 관련한 총 6개의 지식구조를 구분하였고, 각각의 지식구조 그룹의 핵심 주제어를 도출하였다. 따라서 본 연구의 결과는 국방분야를 연구하는 연구기관 및 연구자, 각 군의 연구정책담당 및 위탁교육생의 연구성과 담당부서에서 향후 연구와 연구정책에 참조가 되는 방향을 제시하고자 하였다. 다만 본 연구는 비록 분석 프로그램에 의한 계량적분석이 이루어졌음에도 불구하고 그 결과에 대한 해석이 필수적이었으며, 이에 따라 연구자의 주관적인 해석이 개입될 수 있음이 제한점으로 판단될 수 있다. 더불어 국방분야 지식구조의 향후 연구 방향으로 연도별, 시기별 등의 시계열적 분석을 통해 연구의 변화 추이를 확인하여 그 경향을 파악하는 연구가 필요하며, 본 연구에서는 국방대 학위과정 논문에 한정하여 연구가 이루어진 바, 분석 대상 데이터의 다변화가 필요하다고 사료된다.

REFERENCES

[1] Park. CS, Jung. JW, "Text network analysis: sharing meanings among stakeholders of the policy through socio-cognitive network analysis." Journal of Governmental Studies. Vol. 19, No. 2, pp. 75-110, 2013.

[2] Paranyushkin. D, "Text Network Analysis" *https://issuu.com/deemeetree/docs/text-network-analysis*

[3] Cyram. NetMiner 4.4. Seoul: Cyram Inc. 2018.

[4] Ding. Y, Chowdhury. C. G, & Foo. S, "Bibliometric Cartography of Information Retrieval by Using Co-word Analysis." Information Processing and Management. Vol. 37, No. 6, pp. 817-842, 2001.

[5] Kim. YH, "Social Network Analysis. 3rd edition." Pakyoungsa, pp. 8-85. 2011.