

23년간(1996~2018) 국내학술연구의 ‘고령’ 키워드로 살펴본 빅데이터 분석*

오민정(단독저자)
조선대학교 경영학부 시간강사
(lemona13@daum.net)

본 연구는 23년간의 국내 DB에 게재된 국내저널을 중심으로 토픽모델링과 네트워크 분석을 활용해 기존의 고령연구에 대한 관심주제의 주요한 어구들을 발견해내고 이에 대한 빅데이터 분석을 통해 고령연구에 대한 경영학적 시사점을 제시 하는데 연구 목적이 있다. 논문 제목에 ‘고령’의 단어를 포함하고 있는 사회과학연구 1291편의 논문을 추출하였으며, 토픽모델링 방법을 사용하여 연구 동향을 파악하였다. 토픽 모델링은 텍스트 본문에 있는 숨겨져 있는 의미를 파악하여 각 개별연구의 범위를 줄이면서 전체적인 연구의 초점을 다양한 관점에서 파악할 수 있는 기법으로 인식되고 있다. 이러한 토픽 모델 분석을 통해 5개의 토픽을 확인할 수 있었고 각 토픽별 세부 관심주제를 통해 고령 연구동향과 연구흐름을 명료하게 확인할 수 있었다.

그동안 고령연구가 주로 미시적인 접근에 국한되었다면 본 연구는 보다 거시적인 관점에서 고령연구를 살펴보고 있으며, 논문의 숨겨진 구조를 파악할 수 있는 LDA(Latent Dirichlet Allocation) 접근 방법을 활용하고 있으며, 기존의 연구에서 명확히 확인할 수 없었던 핵심단어에 초점을 둘 수 있었다. 이를 통해 우리는 초록들이 어떤 비율로 어떤 주제를 다루는지 알 수 있었으며 주제의 비율을 통해 주제의 인기도와 흐름 역시 예측할 수 있었다. 또한 LDA분석의 가장 중요한 목적은 ‘연구’ 또는 ‘분석’과 같이 흔하지만 비유리적인 명사구를 제거하면서도 중요한 명사구는 유지하는 방법이며 추출된 명사구의 TF-IDF(term frequency-inverse document frequency)를 계산하여 특정 초록에서 가장 많이 사용된 단어를 보여 준 후 빈도의 비율을 제시해주는 것이다. 그 결과 5개의 토픽별로 총 25개의 주요 명사구를 작성할 수 있었다. 이러한 결과를 통해 현재까지 고령연구의 흐름과 앞으로의 고령연구의 방향을 제시하면서 향후 고령에 관한 경영학적 연구에 시사점을 제시하고자 한다.

주제어: 고령, 빅데이터 분석, 네트워크 분석, 토픽 모델링, LDA, 텍스트 마이닝

1. 서론

UN은 국가의 전체인구에서 65세 이상 인구 비율이 7%이상~14%미만이면 고령화 사회(aging society), 14%이상이면 고령사회(aged society), 20%이상이면 초고령 사회(post-aged society)로 구분하고 있다. 그런데 대한민국의 인구는 2000년에 7.2%로 고령화 사회에 진입하였고 2018년 현재 우리나라의

65세 인구고령화 비율은 14.3%로 이미 고령사회로 진입했으며 2030년 24.3%로 초고령 사회로의 진입을 예상하고 있다(한계레신문, 2018년 8월 27일자). 이러한 변화는 대부분의 선진국이 고령화사회에서 초고령사회로 진입하는데 80년이 소요된데 반해 우리나라의 경우 겨우 30년 만에 발생한 것으로 파악되고 있다(통계청, 2017). 국가의 인구 분포는 경제지표의 중요한 척도가 되고 있으며, 초고령 사회로 급격히 변화하는 과정에서 고령에 대한 연구는

그 동안 어떤 관점을 중시해왔으며 이와 관련된 연구의 흐름을 파악하여 향후 고령 및 고령 소비자 등 고령 산업에 필요한 연구 방향성을 제시하는데 주된 초점을 두고 있다.

고령 관련 연구에서 고령자는 사회적 약자로 인식되고, 경영학적인 관점보다 노년학이나 사회복지 관점에서 주로 다루어져 왔다. 즉 노년학이나 사회학적인 연구들은 고령자를 경력 단절(career plateau)된 비경제활동자로서 사회적 지원을 필요로 하는 복지수혜자로 주로 인식하고 있다(정태연, 2008). 하지만 경영학적 관점에서의 고령자는 새로운 소비창출의 주체이며, 기업에서는 새로운 노동력의 원천으로 논의된다. 일본의 단카이 세대와 같이 한국의 베이비부머들은 자기부양 능력이 있는 소비여력을 갖추고 있으며 외식, 오락, 문화 등을 즐기고자 하는 액티브 시니어(active senior)로 간주된다(이나현, 하지수, 2015). 이에 경영학 연구들은 이러한 액티브 시니어에 대한 재조명을 통해 새로운 고령친화산업 등의 활성화 방안들에 대해 관심을 기울이고 있다.

그동안 진행된 고령 연구를 살펴보면 고령소비자, 실버소비자, 시니어 소비자에 대한 연구가 1990년대부터 꾸준히 진행되고 있으며 고령연구의 주제 역시 다양한 관점에서 연구되고 있음을 확인할 수 있었다(〈표 1〉 참조). 하지만 기존의 고령연구는 주로 미시적 접근에 중점 되었고, 고령연구 동향이나 흐름의 거시적 관점에 대한 연구는 충분히 이루어지지 못하고 있다. 다만 2008년 한국노년학회지의 논문들은 그동안의 고령연구에 대한 흐름을 상세히 기록해놓고 있었다. 한국노년학회는 1980년~2008년까지 한국노년학회지에 실린 논문을 대상으로 각 분야별 노인연구에 대한 이론과 접근시각, 연구방법론, 연구 주제 등을 정리해서 발표했으며 이를 통해 한국노년학의 변화과정을 제시하고 있다. 구체적으

로 살펴보면, 먼저 박충선, 손화희, 전해정(2008)의 가족분야에서 노년학의 접근은 부양부담, 노인 학대, 노인 스트레스, 사회적 지원 등으로 스트레스 모델과 사회교환이론을 언급하고 있다. 하지만 연구범위가 매우 미시적이라는 한계점을 갖는다고 밝히고 있다. 두 번째, 김미혜(2008)의 복지 분야에서 노년학의 접근은 제도와 정책, 노인문제, 노인의 사회참여, 사회 서비스 실천, 노인과 부양의 연구주제가 제시되었으며, 특히 2000년 중반에 와서는 제도와 정책 중 노인고용과 취업, 인력활용 문제가 이슈화되고 있음을 거론하였다. 셋째, 원영희, 모선희(2008)의 사회학 분야에서 노년학의 접근은 사회문제, 사회일탈, 사회변동 및 사회제도 그리고 노인권익, 지역사회활동, 거주 형태 등 다양한 연구주제가 등장하고 있다. 하지만 특별히 인기 있는 주제에 대한 집중성은 보이지 않는다고 밝히고 있다. 넷째, 정태연(2008)의 심리학 연구 분야에서 노년학의 접근은 정신건강, 삶의 만족 및 여가 부모부양 및 가족관계, 노년기 인지적 변화가 그 주를 이루고 있다고 제시하고 있다.

이처럼 고령화에 대한 기존 연구는 주로 노년학에서 접근해오고 있지만 미시적 접근에 그치고 있다는 한계점을 지적한다. 따라서 본 연구의 목적은 첫째, 기존의 선행연구의 한계점을 극복하기 위해 '고령'을 키워드로, 1996년부터 2018년까지 사회과학을 중심으로 DB에 등록된 빅데이터를 대상으로 거시적인 접근을 시도하였다. 둘째, 국내 연구 자료를 통해 토픽 모델링과 텍스트 네트워크 분석을 활용하여 주제를 분류하고 각 토픽별 핵심 키워드를 도출하여 그 영향력이나 중요성을 확인하고자 한다. 이러한 연구결과는 지금까지 고령 연구의 흐름을 파악하고 더 나아가 향후 연구자들에게 연구의 흐름을 제시함과 동시에 앞으로 경영학적인 관점에서의 연구방향을 제안할 수 있는 자료로 활용될 수 있을 것으로 기대해본다.

II. 이론적 배경

2.1 고령연구동향

고령연구에 대한 동향이나 흐름을 한눈에 파악할 수 있는 연구는 찾아보기 어렵지만 오민정, 정진철, 이미현(2012)은 국내 실버소비자를 중심으로 고령 소비자에 대한 연구의 흐름을 정리하고 있다. 구체적으로 1990년대 연구 주제는 소비문제, 소비자 사회화, 불평행동, 구매행동, 정보탐색, 마케팅전략, 그리고 인지연령으로 그 주제를 요약할 수 있으며, 2000년대부터 저축행동, 여가 스포츠, 소비자 문제, 광고, 실버타운과 실버비즈니스전략으로 실버소비자에 대한 접근이 개인적 차원에서 점차 사회, 문화적으로 바뀌고 있음을 언급하고 있다(〈표 1〉 참조). 하지만 이 연구 역시 2011년까지의 데이터만을 제시하고 있고, 연구자가 지극히 주관적으로 주제를 분류하는 한계점을 가지고 있다. 또한 2011년 이후 달라진 고령연구의 흐름을 한눈에 파악하기란 쉽지

않은 일이다. 따라서 본 연구에서는 DB화된 데이터를 통해 고령연구를 살펴보았다. 기존의 빅데이터를 활용한 네트워크 분석의 경우 분석대상은 키워드 연구가 주를 이루었으며, 최근에 와서야 박준석, 김창식, 광기영(2016)의 연구에서 호텔연구동향 분석을 위해 초록을 대상으로 연구가 진행되고 있음을 알 수 있다. 그러나 이 논문을 제외하고는 초록에 대한 연구는 거의 수행되지 않고 있는 실정이다(김규환, 장보성, 이현정, 2009). 하지만 논문의 키워드나 제목의 경우 단순히 논문의 제한적인 정보만을 제공하는 한계를 가지고 있기 때문에 논문에 대해 많은 정보를 제공하는 초록을 대상으로 한 연구가 필요하다고 판단하였다. 따라서 본 연구에서는 최근까지 고령연구가 어떻게 진행되고 있는지 초록을 중심으로 토픽별 주제를 분류해보았다.

2.2 빅데이터 분석

빅데이터 분석은 크게 내용분석(content based analysis)과 구조분석(structure based analysis)

〈표 1〉 1991년~2011년간 고령연구(고령소비자) 주제 요약

연도별	연구주제
1991~1999	소비자문제, 소비자 사회화, 불평행동, 구매행동, 정보탐색, 마케팅전략, 인지연령, 노인 소비자 특성 및 실태, 실버마케팅전략
2000~2004	저축행동, 여가, 스포츠, 소비자문제, 광고, 실버타운, 실버산업광고, 연령과 호칭, 실버비즈니스전략, 경제적 영향력
2005~2008	구매행동, 소비자교육, 인지연령, 컬러마케팅, 실버타운, 시니어그룹성향, 만족도, 온라인 소비자정보, 모바일 폰, 교육프로그램, 광고디자인전략, 라이프 스타일, 스포츠, 소비행태, 소매점선택, 친환경제품 교육프로그램, 실버화장품, 관광지, 의료비, 노후 삶, 구매문제, 노인복지시설, 연금자산, 요실금, 고령친화산업, 광고, 디자인, 노인소비자제품
2009~2011	광고디자인전략, 라이프 스타일, 스포츠, 소비행태, 소매점선택, 친환경제품 교육프로그램, 실버화장품, 관광지, 의료비, 노후 삶, 구매문제, 노인복지시설, 연금자산, 요실금, 고령친화산업, 광고, 디자인, 노인소비자제품

출처: 오민정, 정진철, 이미현(2012) 논문 재구성

으로 구분해볼 수 있으며 내용분석에 토픽모델링과 텍스트 마이닝, 구조분석에 네트워크 분석이 해당된다(〈그림 1〉 참조). 먼저 (소셜) 네트워크 분석이란 개인 및 집단들 간의 관계를 노드(node)와 링크(link)로 모델링하여 그 위상구조, 확산, 진화과정을 계량적으로 분석하는 방법론을 의미한다. 본 연구가 사용하는 텍스트 네트워크 분석은 텍스트를 하나의 데이터로 인식하여 모델링하는 것이며 연결된 구조적 특성을 파악하기 위해서는 밀도(density), 중심성(centrality), 집중성(centralization) 등의 지표가 사용된다(Hansen, Shneiderman, and Smith, 2011). 반면 토픽모델링은 비정형 데이터를 정형화된 데이터로 변형하는데 텍스트 마이닝 기법이 활용되며 텍스트 내의 단어를 분석하여 주제를 도출하는 방법론이다(Blei, 2012). 특히 텍스트 마이닝은 텍스트 형태의 데이터에 마이닝 기법을 적용한 것으로 대규모 비정형 형태의 텍스트 안에 숨겨져 있는 의미 있는 지식을 찾아내는데 탁월한 기법이다. 따라서 본 연구에서는 초록들을 대상으로 드러나지 않는 숨은 의미를 찾아내기 위해 텍스트 마

이닝을 활용한 토픽모델링을 사용하였으며, 특히 네트워크 분석과 비교했을 때 토픽모델링은 노드만 가지고도 시각화가 가능하다는 이점을 가지고 있어 적절한 분석방법이라고 판단하였다.

토픽 모델링에서 단어의 중요도 측정은 문서 내 또는 문서 간 등장 빈도, 총 등장 빈도 등 다양한 방법으로 측정할 수 있는데 그 중에서 본 연구에서 활용하고 있는 TF-IDF(Term Frequency-Inverse Document Frequency)는 어떤 문서에서의 특정 단어의 중요도를 측정할 수 있어 매우 널리 활용되는 방법 중 하나이다. 구체적으로 특정 문서에서의 단어 빈도가 높을수록 그리고 전체 문서 중 그 단어가 등장한 문서가 적을수록 TF-IDF값이 높게 나타난다. 즉 TF-IDF값이 높은 단어는 해당 문서 내에서 핵심적인 메시지를 담고 있을 확률이 높다고 할 수 있다(Wang, Bendle, and Cotte, 2015).

본 연구에서는 그동안 연구되어왔던 23년간의 고령 연구의 초록에서 연구 주제를 확인하기 위해 확률론적 주제 모델링 절차인 LDA접근법을 사용하였다. 이는 모든 초록 데이터를 사용해서 논문의 숨겨



출처: Sebei, Hadj Taieb, and Ben Aouicha(2018)

〈그림 1〉 빅데이터 분석

진 구조를 밝혀내는데 적합한 방식으로 Generative model(output정보를 가지고 input정보를 역추적 할 수 있는 모델)이며, 문서의 주제 분류를 위해 고안된 알고리즘이며 디리클레 분포(Dirichlet distribution)를 사용한다.

III. 연구방법

3.1 분석대상

이 분석에 사용된 데이터는 먼저 국내 고령연구의 동향을 파악하기 위해 KISS의 학술DB 논문 검색 사이트를 활용하였으며, 초록에 '고령'이 들어간 연구를 검색하였으며, 연구범위를 사회과학으로 설정하여 1996년부터 2018년 상반기까지 23년간 발표된 논문 1291편을 대상으로 하였다.

3.2 분석방법

본 연구의 목적은 1291편의 논문 초록을 대상으로 텍스트 마이닝, 네트워크 분석을 통해 기간별 고령연구의 연구동향을 파악하고 그 경영학적인 시사점을 제시하는데 있으며, 이를 위해 토픽 모델링, 클라우드 분석, 연결중심성 분석 등을 시행하였다.

구체적인 분석방법으로는 논문 초록을 중요 명사 구별로 텍스트 마이닝을 하고, 토픽 모델링을 도입함으로써 고령 연구가 지금까지 어떻게 진행되었으며, 키워드와 주제는 무엇인지를 확인하고자 하였다. 구체적 분석 방법으로 넷마이너(NetMiner)를 활용하였으며 LDA접근을 사용해서 보다 구체화된

키워드 분석을 할 수 있었다. 이 접근법은 어떤 논문이 유사한 주제를 다루는지를 밝혀냄으로서 논문간의 관계를 파악할 수 있으며, 초록들이 어떤 비율로 어떤 주제를 다루는지 알 수 있으며 주제의 비율을 이해하면 주제의 주된 포커스와 그 연구 동향을 파악할 수 있다(Wang et al., 2015).

이 분석의 또 하나의 중요한 특징은 '연구' '고령'이라는 일반적이지만 비유익한 명사구를 제거하면서도 중요한 명사구는 유지할 수 있다는 것이다. 단어의 중요도 측정은 특정 문서 내에서의 특정 단어의 중요도를 측정할 수 있는 TF-IDF를 활용하였다. TF-IDF값이 높은 단어는 해당 문서 내에서 핵심적인 메시지를 담고 있을 확률이 높다.

구체적인 분석과정으로는 먼저 형태소 분석을 시행하였다. 형태소 분석이란 한글형태소 분석기를 통해 명사형 단어들을 키워드로 추출하고 사전을 활용하여 유의어, 지정어, 제외어를 입력하여 해당되는 단어들을 분석에서 제외시키는 방법이다. 예를 들어 위에서 언급했던 '연구', '고령'이라는 비유익한 명사구를 제거하기 위해서는 제외어에 '연구', '고령'이라는 단어를 입력해주면 이 단어를 제외한 명사구들이 도출되는 것이다. 다음으로 토픽수를 결정하기 위해 반복해서 토픽모델링을 시행하였다. 적합한 토픽수를 설정하는 방법으로 이수상(2016)의 연구에서는 2가지 방법을 제안하고 있는데 그중에서 연구자가 확인했을 때 범주화가 잘 되었다고 판단되는 토픽수를 결정하는 방법을 선택하였다. 그리고 핵심키워드 추출을 위해 의미 파악이 어려운 1글자 단어를 제외하였고, 문서에 흔하게 등장하는 단어를 다시 한번 제외시켰다. 예를 들어 워드 길이(Word length)는 2글자 이상인 단어만 추출하라는 값을 입력하였으며, 문서별 TF-IDF 값이 1이상인 단어만 추출하였다. 또한 한 문서에 모든 토픽이 비슷하게 할당할 수

있도록 α 값¹⁾으로 10을 입력하였으며, 한 토픽에 모든 단어가 비슷하게 할당되도록 β 값은 0.1을 사용하였다.

토픽분석은 토픽별 주요 단어를 참고하여 연구자가 직접 토픽명을 지정해야 하며 토픽이름을 지정할 때는 각 토픽에서 할당확률(영향력)이 높고 특징적인 단어를 중심으로 라벨링을 하였다(〈표 2〉, 〈표 3〉 참조).

IV. 분석결과

4.1 선호 주제

고령 연구가 시간이 지남에 따라 어떻게 바뀌었는지는 고령 연구가들에게 큰 관심거리이다. 분석을 위해 고령을 제목으로 한 초록을 대상으로 총 1291개의 논문이 분석에 사용되었다. 토픽 모델링의 가장 중요한 점은 개별 연구의 범위를 좁히는 동시에 고령 연구의 초점을 넓히고 다양화하는 것이라고 볼 수 있다. 적합한 토픽수를 설정하는 방법으로는 첫

번째, 연구자가 확인했을 때 범주화가 잘 되었다고 판단하는 것을 결정하는 방법이 있으며 두 번째, 토픽 확률 분포의 퍼플렉시티(perplexity)값을 사용하는 방법이 있다(이수상, 2016). 본 연구에서는 우선적으로 전자의 방법을 고려하여 총 4회(토픽수를 각각 10개, 7개, 5개, 3개로 지정한 횟수)의 토픽 모델링을 시행하여 모델링을 돌렸다. 그 결과 먼저 10개로 설정한 토픽 모델링에서는 토픽을 구성하는 단어의 할당값이 매우 낮은 단어들이(0.00)토픽에 도출되었다. 따라서 토픽수를 줄여서 반복진행을 한 결과 5개로 분류한 토픽수가 한 토픽에 유사한 단어들이 할당됨을 알 수 있었다. 따라서 도출된 핵심키워드와 할당 확률값을 기준으로 5개의 토픽을 적절한 토픽수로 결정하였으며 각각의 토픽에 라벨링을 하였다. 토픽 주제를 살펴보면 첫 번째 고령자 고용 및 일자리, 둘째 일본의 고령자와 저출산 세 번째 은퇴고령자들의 경제문제, 네 번째 고령자 복지 및 만족도, 마지막으로 고령자 서비스 인식과 교육으로 요약해 볼 수 있다. 또한 토픽모델링의 빈도는 각 단어가 등장한 등장논문의 수를 의미한다(〈표 2〉, 〈표 3〉 참고).

다음으로 23년 동안 고령연구에서 주로 다루고 있

〈표 2〉 논문 토픽 모델링 결과

토픽	핵심키워드	문서수
토픽1_ 고령자 고용 및 일자리	노동, 고용, 시장, 근로, 일자리	208
토픽2_ 일본의 고령자와 저출산	산업, 저출산, 성장, 일본, 인력	278
토픽3_ 은퇴 고령자들의 경제문제	소득, 재정, 보험, 자산, 은퇴	264
토픽4_ 고령자 복지와 만족도	건강, 참여, 만족도, 우울, 장애	297
토픽5_ 고령자 서비스 인식과 교육	서비스, 인식, 가족, 교육, 의료	244

1) α 는 50/토픽수, β 는 0.1을 사용한다. β 값이 커지면 모든 단어의 영향력이 같아지기 때문(Griffiths and Steyvers, 2004; 이수상, 2016)

〈표 3〉 토픽 키워드별 빈도수

토픽1		토픽2		토픽3		토픽4		토픽5	
키워드	빈도수	키워드	빈도수	키워드	빈도수	키워드	빈도수	키워드	빈도수
노동	83	산업	73	소득	135	건강	142	서비스	112
고용	77	저출산	68	재정	86	참여	95	인식	79
시장	61	성장	64	보험	68	만족도	75	가족	76
근로	56	일본	44	자산	62	우울	70	교육	67
일자리	39	인력	39	은퇴	41	장애	31	의료	55

는 연구주제를 한눈에 확인할 수 있도록 워드 클라우드로 제시하였으며(〈그림 2〉 참조), 토픽-주요 단어 시각화를 제시하였다(〈그림 3〉 참조). 〈그림 3〉에서 노드의 크기는 동일한 반면 선의 굵기가 다른 이유는 토픽모델링에서 사용하고 있는 TF-IDF가 문서별 단어의 가중치를 의미하고 단어 간 링크는 영향력을 의미하기 때문이다. 따라서 선의 굵기는 단어의 중요성 지표이며 토픽별 단어의 할당 확률값에 비례함을 알 수 있다.

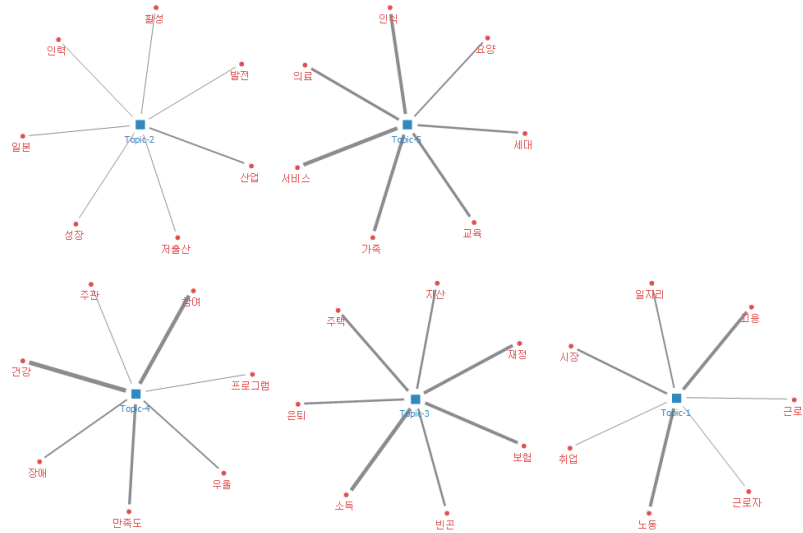
4.2 시간경과에 따른 인기도 변화

다음은 시간 경과에 따른 논문 주제의 변화를 한눈에 알아보기 약 5년 주기를 기준으로 기간을 분류

하여 토픽모델링을 수행하였다. 2000년 이전 논문의 경우 DB로 등록된 논문이 17편에 불과하여 논문 주기를 크게 4주기로 나누었다. 구체적으로 1996년~2004년, 2005년~2009년, 2010년~2014년, 2015년~2018년으로 나누었으며, 1996년~2004년까지의 연구논문은 109편, 2005년~2009년은 175편, 2010년~2014년은 상대적으로 많은 608편, 마지막 2015년~2018년은 399편의 연구논문이 분석대상으로 분류되었다. 토픽모델링을 통해 시기별 중요하게 언급되고 있는 핵심키워드를 도출하기 위해 각 시기별 논문 초록을 대상으로 토픽 모델링을 분석하였으며 그 결과 시기별 가장 많이 등장한 핵심키워드를 확인할 수 있었다(〈표 4〉, 〈표 5〉, 〈그림 4〉, 〈그림 5〉, 〈그림 6〉, 〈그림 7〉 참조).



〈그림 2〉 워드 클라우드 분석 결과



Legend
 ○ : topic, △ : main word, — : word influence on topic

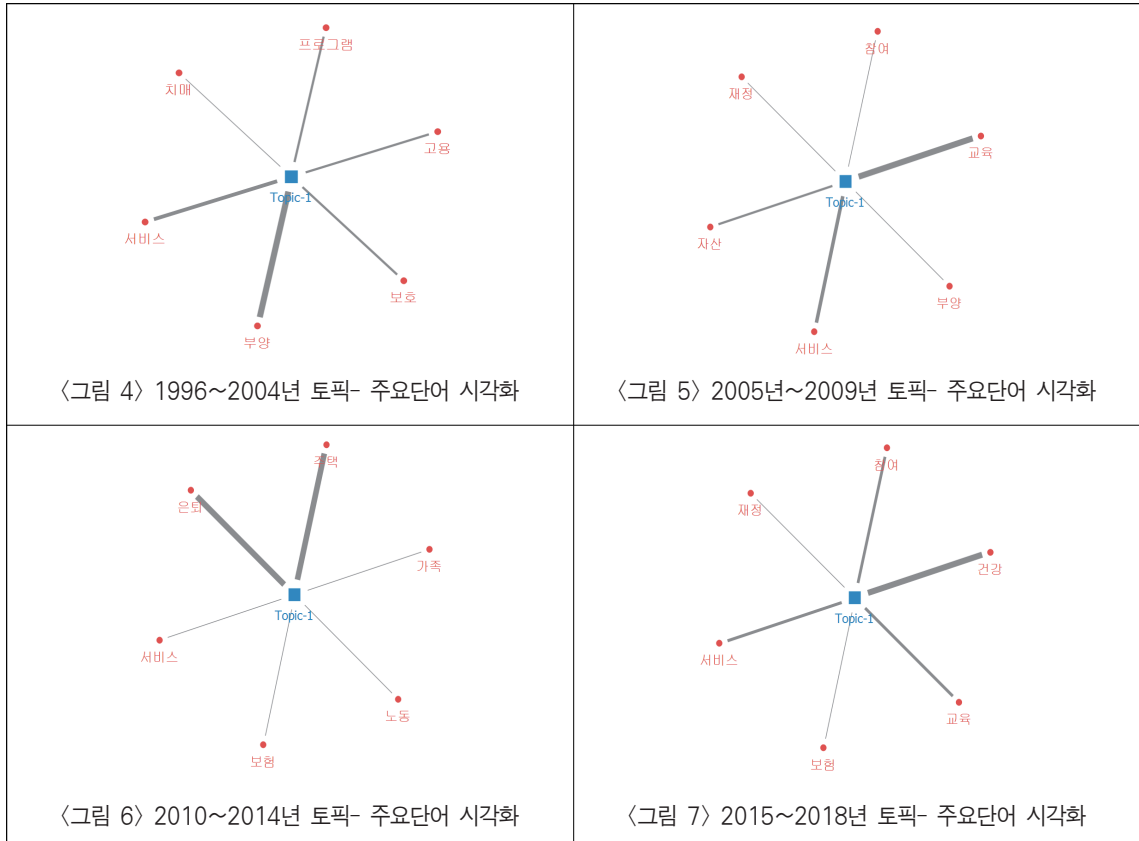
〈그림 3〉 토픽- 주요단어 시각화

〈표 4〉 시기별 핵심키워드

년도	핵심키워드	등장문서수
1996년 ~ 2004년	부양, 서비스, 고용, 보호, 프로그램, 치매	109
2005년 ~ 2009년	교육, 서비스, 자산, 참여, 재정, 부양	175
2010년 ~ 2014년	은퇴, 주택, 서비스, 가족, 노동, 보험	608
2015년 ~ 2018년	건강, 소득, 참여, 재정, 보험, 노동	399

〈표 5〉 시기별 핵심키워드 빈도분석 결과

기간	단어	빈도	단어	빈도
1996년~2004년	부양	100	보호	63
	서비스	75	프로그램	59
	고용	64	치매	52
2005년~2009년	교육	153	참여	83
	서비스	101	재정	81
	자산	90	부양	76
2010년~2014년	은퇴	304	가족	266
	주택	271	노동	255
	서비스	270	보험	251
2015년~2018년	건강	260	재정	205
	소득	259	보험	182
	참여	242	노동	179



시기별 토픽 주요단어의 핵심 토픽이 어떻게 변화되었는지를 다시 한번 살펴보면 초창기 2004년까지의 연구에서는 고령자 부양제도, 고령자 서비스, 고용, 치매 문제에 대한 연구가 주를 이루었으며, 2009년까지 연구에서 눈에 띄는 키워드는 부양, 서비스에서 더 나아가 스스로 자립할 수 있는 고령자 교육, 참여, 서비스에 대한 관심이 높아지고 있음을 알 수 있었다. 그리고 2010년~2014년에는 본격적인 베이비부머들의 은퇴시기와 맞물려 고령자 은퇴, 주택문제, 가족, 노동, 보험과 같은 경제적·실무적인 연구가 이루어지고 있음을 알 수 있었다. 마지막으로 2018년 최근에 와서는 건강, 소득, 참여, 재정,

보험, 노동, 서비스, 교육의 키워드를 통해 구체적인 일자리 마련과 고령자들의 건강과 웰빙을 강조하는 연구가 많이 이루어지고 있음을 확인할 수 있었다.

4.3 네트워크 분석

다음으로 연구대상 논문 초록의 네트워크 분석을 실시하였다. 이를 위해 중심구조 분석 중 하나인 연결 중심성(Degree Centrality)분석을 실시하였다. 중심구조분석이란 어떤 노드가 가장 중요한 노드인지를 찾고, 네트워크 구조가 얼마나 소수의 중요한 노드에게 집중되어 있는지 집중화 정도를 파악

할 수 있는 있는 분석방법이다. 즉 중요한 역할을 담당하는 사람(노드)는 누구이며? 영향력이 얼마나 소수에게 집중되어 있는 구조인지? 를 파악하고자 하는 연구에 활용된다. 중심구조분석은 연결 중심성(Degree Centrality), 근접 중심성(Closeness Centrality), 매개 중심성(Betweenness Centrality), 아이겐벡터 중심성(Eigenvector Centrality) 4가지로 구분되는데, 그 중 텍스트 네트워크 분석에서는 연결 중심성(Degree Centrality)과 매개 중심성(Betweenness Centrality) 방법이 주로 활용되고 있다. 먼저 연결 중심성이란 직접 연결된 이웃 노드가 많을수록 연결 중심성이 높아지며 직접적인 영향력의 크기를 측정할 수 있는 장점이 있다. 따라서 연결 중심성 값이 높다는 것은 다양한 단어와 자주 근접하여 등장했다는 의미를 포함한다. 따라서 본 연구에서 단어(키워드)의 영향력을 파악하기 위해서는 연결 중심성을 확인하는 것이 적합하다. 반면 매개 중심성은 다른 노드간의 최단 경로에 많이 등장할수록 중심성이 높다고 해석된다. 매개 중심성이 높은 노드는 정보흐름에 대한 통제력을 가지며 이 노드가 제거될 경우 네트워크 전체 연결과 흐름에 큰 영향을 미친다. 따라서 노드간의 상호작용이나 통제력의 역할을 확인하는데 적합한 분석방법이다. 근접 중심성은 다른 노드들과의 거리가 짧을수록 노드의 중심성이 높다고 해석된다. 근접 중심성이 높은 노드는 가장 빨리 다른 노드에 영향을 주거나 받을 수 있다. 따라서 정보전달이나 운송, 교통, 질병 전파와 같은 확산과 전파의 연구에 주로 활용될 수 있다. 마지막으로 아이겐벡터 중심성이란 중심성이 높은 노드와 많은 연결을 갖을수록 중심성이 높아진다. 네트워크 전반에 미칠 수 있는 파급효과 또는 영향력이 가장 큰 노드라고 할 수 있다. 또한 연결 중심성과 성격이 유사하나 연결중심성이 지엽적 영향력을 고려한다면

아이겐벡터 중심성은 전체적인 영향력을 검증한다는 점에서 글로벌한 영향력을 검증하는데 더 적합한 분석 방법이다.

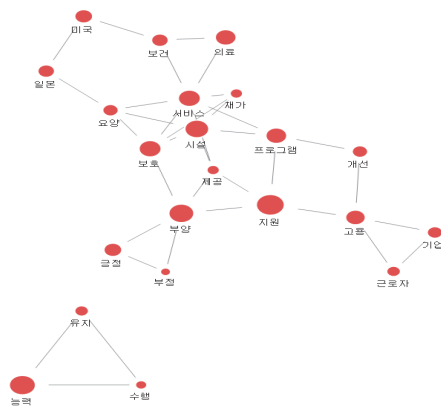
따라서 본 연구에서는 연결중심성 분석을 통한 네트워크 분석을 실시하기 위해 <표 4>와 동일하게 시기를 구분하여 시기별 네트워크 중심성 분석을 실시하였다. 구체적인 연구동향 파악을 위해서 링크수를 줄이고 시각화를 높이기 위해 필요한 사전 작업을 진행하였다. 구체적으로 의미를 파악하기 힘든 1글자 단어를 제외하기 위해 워드 길이 값으로 2글자 이상인 단어만 추출하였고, 동시 등장횟수가 높은 단어 간 링크만 추출하기 위해 가중치 값으로 2이상인 링크를 산출하도록 하였다. 그 결과 <그림 8>, <그림 9>, <그림 10>, <그림 11>과 같은 각각의 시각화 그림을 확인할 수 있었다. 이때 노드는 중심성 값을 의미하며, 중심성 값이 클수록 원의 크기가 크게 나타나며, 동시등장횟수로 링크수를 확인할 수 있다. 구체적인 연구흐름을 이해하기 위해 <표 6>과 같이 시기별 핵심키워드와 연결 중심성 분석값을 정리하였다.

4.3.1 1996년~2004년

1996년부터 2004년까지 분석대상 논문의 수는 109편이었으며, 이 연구논문의 초록을 대상으로 한 네트워크 중심성 분석 결과 <그림 8>과 같이 다른 시기에 비해 연구주체의 범위를 한눈에 확인할 수 있었다. 구체적으로 살펴보면 고령자 지원정책(0.24), 고령자 부양제도(0.20), 고령자 서비스(0.18), 고령자 보호(0.18), 고령자 프로그램(0.17), 고령자 고용제도(0.16)가 높은 순위를 차지하는 것으로 확인되었다.

〈표 6〉 시기별 네트워크 중심성 분석(연결 중심성)

1996년~2004년		2005년~2009년		2010년~2014년		2015년~2018년	
핵심 키워드	연결 중심성	핵심 키워드	연결 중심성	핵심 키워드	연결 중심성	핵심 키워드	연결 중심성
지원	0.24	교육	0.34	지역	0.61	소득	0.59
능력	0.21	재정	0.26	서비스	0.53	건강	0.51
부양	0.20	서비스	0.26	고용	0.50	노동	0.42
서비스	0.18	참여	0.22	은퇴	0.47	재정	0.41
보호	0.18	취업	0.21	노동	0.47	참여	0.35
프로그램	0.17	부양	0.19	근로	0.46	고용	0.35
고용	0.16	자산	0.16	주택	0.43	보험	0.35



〈그림 8〉 1996년~2004년까지 네트워크 중심성 분석 결과(연결 중심성)

4.3.2 2005년~2009년

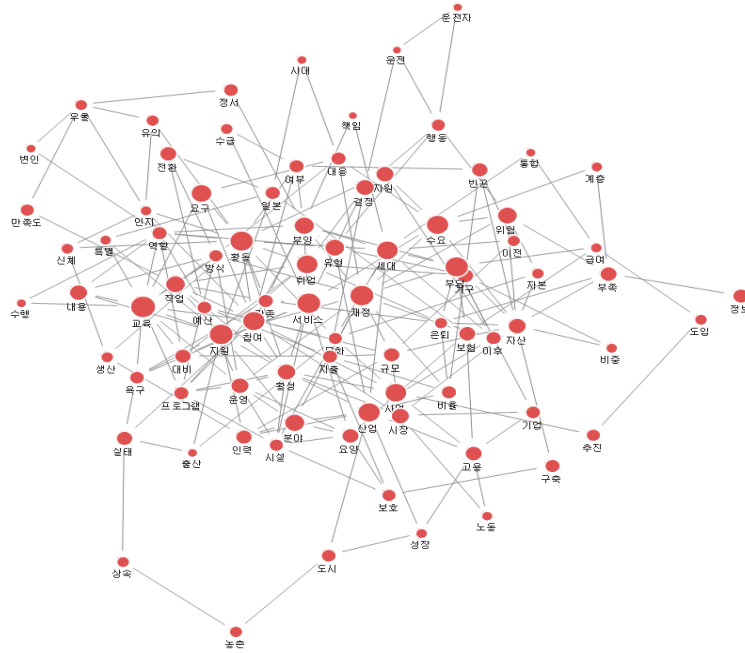
2005년부터 2009년까지 분석대상 논문의 수는 175편이었으며, 이 연구논문의 초록을 대상으로 한 네트워크 중심성 분석 결과 〈그림 9〉와 같이 다양한 연구주제가 연구되고 있음을 확인할 수 있었다. 구체적으로 살펴보면 고령자를 위한 교육제도(0.34), 재정 정책(0.26), 고령자 서비스(0.26), 고령자 참여(0.22), 고령자 취업(0.21), 고령자 자산(0.16)의

순위를 확인할 수 있었으며, 1996년부터 2004년까지의 연구에서 중요하게 거론되었던 고령자 서비스와 정부의 부양제도는 2009년까지도 중요한 연구 주제로 다루어지고 있음을 확인할 수 있었다.

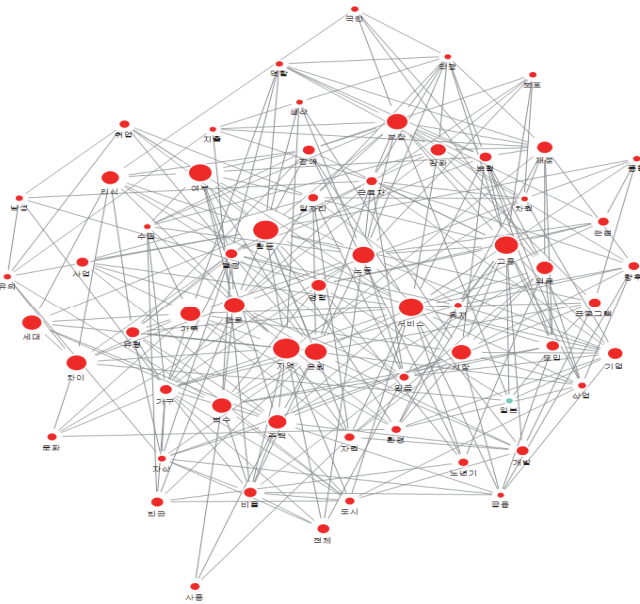
4.3.3 2010년~2014년

2010년부터 2014년까지 분석대상 논문의 수는 상대적으로 많은 608편이었다. 시기적으로 5년 간 고령연구가 폭발적으로 증가하고 있음을 확인할 수 있었으며 이를 입증하듯 중심성 분석 결과 노드와 링크수가 다른 시기에 비해 엄청나게 증가하고 있어 복잡한 네트워크를 띄고 있음을 확인할 수 있었다.

이 시기의 연구논문의 초록을 대상으로 한 네트워크 중심성 분석 결과 〈그림 10〉과 같이 다양한 연구주제가 연구되고 있음을 확인할 수 있었다. 구체적인 연구 흐름을 보면 지역특색을 반영한 지역별 고령자 문제(0.61), 고령자 고용(0.50), 고령자 은퇴(0.47), 고령자 노동정책(0.47), 고령자 근로제도(0.46), 고령자 주택마련(0.43), 고령자 의료(0.41) 등이 높은 순위를 차지하는 것으로 나타났다. 이러한 연구흐름을 볼 때 이 시기부터 본격적인 고령자들의



〈그림 9〉 2005년~2009년까지 네트워크 중심성 분석 결과(연결 중심성)



〈그림 10〉 2010~2014년까지 네트워크 중심성 분석 결과(연결 중심성)

경제적 처우 문제, 은퇴, 고용에 대한 실무적 접근이 이루어지고 있음을 짐작해 볼 수 있었다.

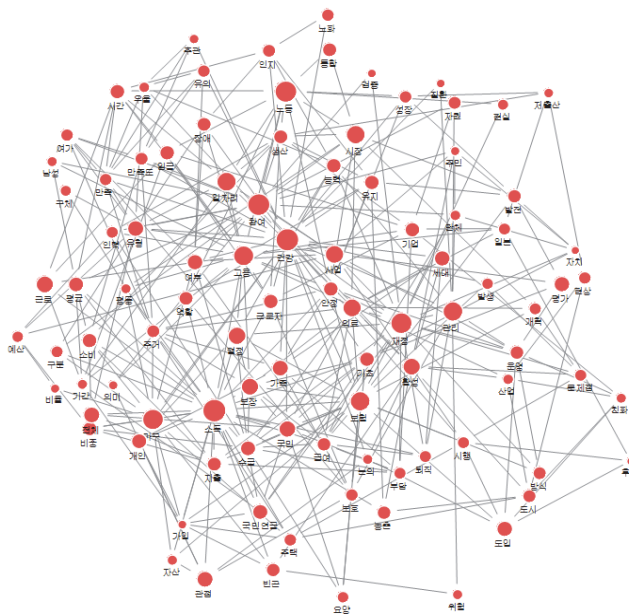
4.3.4 2015년~2018년

2015년부터 2018년 상반기까지 분석 대상 논문 수는 399편이었으며 이 시기의 연구논문의 초록을 대상으로 네트워크 중심성 분석 결과 <그림 11>과 같이 실무적인 제도와 정책 연구가 이루어지고 있음을 확인할 수 있었다. 구체적인 연구 흐름을 보면 고령자를 위한 소득(0.59), 고령자들의 건강(0.51), 고령자 노동(0.42), 고령자 재정(0.41), 고령자 참여(0.35), 고령자 고용(0.35), 고령자 보험(0.35)의 순으로 나타났다. 즉 최근에 와서는 고령자들을 위한 서비스 정책에서 벗어나 실질적인 소득과 재정 마련에 관심이 증가하고 있으며, 의학기술의 발달과

더불어 고령자들의 건강과 관련된 웰빙을 강조하고 있었다. 이를 종합해보면 구체적인 일자리 마련, 건강과 의료 문제를 해결하기 위한 보험 혜택을 마련하여 고령자들을 보호하고 지원하는 범위에서 벗어나 스스로 일자리를 통한 소득창출을 통해 도래하는 고령사회에 대한 대비의 필요성을 제시하고 있는 것으로 확인되었다.

V. 결론 및 경영학 연구의 시사점

본 연구에서는 1996년부터 2018년까지 23년간의 고령연구에 대한 국내 연구 동향분석을 위해 사회과학분야 등재지 논문의 초록을 대상으로 텍스트 마이닝, 토픽모델링, 네트워크 분석방법을 활용하였



<그림 11> 2018년까지 네트워크 중심성 분석 결과(연결 중심성)

다. 분석 결과 5개의 토픽으로 고령연구가 진행되고 있음을 확인할 수 있었다. 구체적인 토픽 주제를 살펴보면 첫째, 고령자 고용 및 일자리라는 토픽으로 노동, 고용, 시장, 일자리, 근로에 대한 핵심키워드가 등장하였으며, 둘째 일본의 고령자와 저출산이라는 토픽으로 성장, 저출산, 산업, 일본, 인력의 핵심키워드가 등장하였으며, 세 번째 은퇴고령자들의 경제적 측면이라는 토픽에는 소득, 재정, 보험, 자산, 은퇴의 핵심키워드가 도출되었고, 네 번째 고령자 복지와 만족도의 토픽에는 건강, 참여, 만족도, 우울, 장애의 핵심키워드가 등장하였으며, 마지막으로 고령자 인식과 교육토픽에서는 서비스, 인식, 가족, 교육, 의료의 핵심키워드가 등장하고 있었다.

다음으로 시간경과에 따른 고령연구의 인기도 변화를 추적하기 위해 약 5년을 주기로 4개의 구간으로 나누어 연구를 진행하였다. 구체적으로 1)1996년~2004년 2)2005년~2009년 3)2010년~2014년 4)2015년~2018년 네 기간으로 분류하여 분석결과를 중요한 키워드 중심으로 살펴본 결과 1996년부터 2004년까지는 '부양', '서비스', '시설', '고용', '프로그램', '치매'가 매우 눈에 띄는 명사구였으며, 2005년~2009년에는 '교육', '서비스', '자산', '참여', '재정', '부양'이 주류를 이루고 있으며, 2010년~2014년에는 '은퇴', '주택', '서비스', '가족', '노동', '보험'의 키워드로 진행되고 있었다. 마지막으로 2015년부터 2018년까지는 '건강', '소득', '참여', '재정', '보험', '노동'의 키워드로 연구가 진행되었다. 기간별 연구주제의 흐름을 보면 고령자들을 위한 서비스가 지속적으로 관심 있는 주제로 등장하고 있었으며 2010년 이후부터 고령자들의 경제적 처우 문제, 은퇴문제, 일자리 문제에 대한 경제적·사회적 접근이 두드러짐을 알 수 있었다.

다음으로 본 연구의 이론적 시사점을 제시하면 다

음과 같다.

첫째, 본 연구는 그동안 연구동향 분석이 최신 네트워크 분석과 LDA를 활용한 토픽 모델링을 통해 보다 해석이 용이한 시각적인 표현으로 고령연구동향의 이해도를 높였다는데 의의가 있다.

둘째, 기존 고령연구에 대한 미시적 접근(개인적인 행동변수 등)에서 벗어나 거시적인 고령연구의 출발점을 제시하고 있다는 점이다. 이를 통해 그동안의 진행되어온 주제의 중복에서 벗어나 지금까지 다루어지지 않은 연구주제를 찾아낼 수 있다는데 의의가 있다.

다음은 본 연구의 실무적 시사점이다.

첫째, 현재 고령사회로 접어든 시점에서 더 이상 고령화의 문제점을 진단하는 연구에만 국한하지 말고 고령화의 문제를 기회로 바꾸는 인식의 제고가 필요하다는 사실을 확인하였다. 지금까지 고령화에 따른 문제점을 진단하였다면 이제는 그 해결책을 제시할 수 있는 솔루션을 제시하는 연구가 필요하다. 이를 위해 핵심키워드에서도 제시된 바와 같이 고령자들의 은퇴문제와 일자리 문제와 같은 고용측면의 연구에서 더 나아가 노동인력의 감소가 불가피한 상황에서 시니어 인재를 활용하거나 여성인재를 적극적으로 활용하여 기업의 경쟁력을 높일 수 있는 방안을 제시할 수 있는 연구가 필요하다고 생각한다. 예를 들면 최근 김다솔, 백지연(2018)의 준고령자와 고령자의 경력개발에 대한 연구 등을 들 수 있다. 다시 말해 고령자를 사회복지의 대상이나 은퇴의 관점에서 고려하기보다는 액티브 시니어의 개념을 적용하여 서드 에이지(third age)로서 새로운 경력계획(career planning)이나 조직 내 시니어 경력개발 프로그램(career development program) 등을 제시하는 연구가 필요하다. 김문정, 김홍기(2018)의 연구가 60세 이상의 고령자들에 대한 생산성 평가

에 미치는 요인 등을 파악하고 있는 것도 이러한 맥락이다. 고령인력의 기업의 재취업이나 새로운 노동력으로 재진입하는 이슈는 사회학적으로 가계부채의 증대에 따른 소득 증대 효과를 가져 올 수 있고 저출산 등으로 인한 노동력의 감소 문제 등을 보완할 수 있는 대체인력이 될 수 있다. 더 나아가 이러한 고령인력 등에 대한 재교육이 어떻게 인력개발 차원에서 추진되어야 할지 등은 새로운 연구 주제가 될 것이다(오민홍, 2017). 또한 박지성(2018) 연구는 고령인력의 신지식 창출의 한계로 인해 조직혁신에 비선형의 역U형이 존재하지만 연령 다양성은 이들 관계를 조절한다고 보고하고 있다. 결국 고령인력의 인력활용 문제에 대한 새로운 논점을 제시할 필요가 있다고 볼 수 있는데, 고령인력과 밀레니엄 세대 간에는 사고와 문화의 격차가 존재하지만 이를 효과적으로 활용한 인력 다양성 제고 방안을 검토할 필요가 있다. 조직이론의 양면조직(ambidextrous organization)이 탐색과 활용을 필요로 하고 기계적 조직과 유기적 조직의 효과적인 혼합이 제시되는 점과 유사한 맥락이다(이흥, 김찬모, 2004).

둘째, 경영 및 사회학적인 관점에서 고령인력 활용의 문제에 대한 인식이 필요하다. 고령인력에 대한 정년이 존재하고 있는 현실에서 은퇴한 고령인력이 사회의 비영리조직에 적극적인 참여 방안에 대한 연구 제안이 요청된다. 최근 SK는 사회적 기업 및 사회적 가치의 중요성을 강조하고 있다. 공유가치창출(creating shared value)나 사회적 경제 이슈 등이 점차 강조되는 상황에서 은퇴한 전문고령인력들이 사회의 비영리 혹은 NPO 조직 등에서 그들의 전문기능을 활용할 수 있음으로서 사회적 생태계는 보다 건장한 모습으로 전환되고 기업과의 전방위적인 연대관계가 형성되는 조직 생태계가 형성될 수 있다.

셋째, 마케팅측면에서는 고령자를 잠재고객으로 인

지하고 고령자 니즈에 선제적으로 대응하는 것이 필요할 것이며 고령화로 인한 소비마켓의 변화에 연동하여 금융상품, 레저 상품 등 고령친화산업에 대한 연구가 진행되어야 할 것이다. 그동안 마케팅 관측 활동의 대상으로 고령자는 중시되지 않았다. 하지만 최근 액티브 시니어 활동("꽃보다 할배"의 개념)이 부각되면서 마케팅의 새로운 목표시장 접근법 및 전략적 마케팅 전략에 대한 연구가 필요하다. 그 이유는 고령자들은 일반 젊은 소비자들과 다른 성향을 가진다. 예를 들면 젊은 소비자들은 사람들이 모이는 곳을 선호하지만 고령 소비자들은 그렇지 않은 경우가 많다. 예를 들어 일본의 경우 여행상품으로 조용한 휴양지만을 선정하여 고령소비자들을 위한 고급 크루즈 여행 상품을 출시하고 있으며 그곳이 인기를 모으고 있다. 따라서 고령 소비자를 새로운 소비 계층으로 인식하고, 이들에 맞춘 새로운 판촉 및 마케팅 활동 등이 강조될 필요성이 있다.

마지막으로 본 연구의 한계점은 다음과 같다.

첫째, 분석에 사용된 고령 데이터가 국내 DB만을 활용하였다는 점이다. 추후 연구에서는 국내·외 연구를 활용한 비교연구가 필요하다고 본다.

둘째, 방대한 빅데이터를 통해 결과를 도출하다 보니 구체적인 고령화의 형태적 내용을 재현하지 못했다는 한계점을 가지고 있다. 따라서 추후 연구에서는 거시적인 접근과 미시적 접근을 접목시켜 보다 구체적이고 현실적인 고령연구의 동향 분석이 필요하다고 본다. 이를 위해 추후 연구에서는 논문에만 국한하지 말고 실무적인 자료(보고서, 통계청 자료 등)를 분석대상에 포함시켜 이론과 실무의 갭을 줄이고 정책과 제도적 측면과 함께 현실적인 해결책을 제시해 줄 필요가 있다.

셋째, 본 연구는 사회과학의 연구 분야로 국한하여 연구가 진행되었다. 그러다 보니 각 저널의 성격

을 고려하지 못했다. 예를 들어 노년학 관점에서 살펴보고 있는 고령과, 경영학 관점에서 바라보는 고령에 대한 입장이 다를 수 있다. 따라서 추후 연구에서는 저널의 성격을 분류하여 저널의 성격에 따른 비교 연구도 필요할 것으로 본다.

참고문헌

- 김규환, 정보성, 이현정(2009), "우리나라 기록관리학 분야의 연구영역 분석," **한국문헌정보학회지**, 43(3), 417-439.
- 김다솔, 백지연(2018), "준고령자/고령자의 경력개발에 관한 탐색적 연구," **비서·사무경영연구**, 27(4), 87-113.
- 김문정, 김흥기(2018), "60+ 고령자의 생산성에 대한 기업의 평가와 영향 요인," **한국콘텐츠학회논문지**, 18(11), 571-580.
- 김미혜(2008), "한국노년학의 복지분야: 연구동향," **한국노년학**, 28(4), 733-752.
- 박지성(2018), "고령화와 조직 혁신: 비선형성과 연령 다양성의 조절효과," **한국산학기술학회논문지**, 19(6), 440-446.
- 박준석, 김창식, 광기영(2016), "텍스트마이닝과 소셜네트워크분석 기법을 활용한 호텔분야 연구동향 분석," **관광레저연구**, 28(9), 209-226.
- 박충선, 손화희, 전해정(2008), "한국노년학 가족분야 연구 30년," **한국노년학**, 28(4), 797-813.
- 오민홍(2017), "고령사회 대응 중고령자 인력 활용," **보건복지포럼**, 254, 50-66.
- 이나현, 하지수(2015), "한국 중년 남성의 젊음 추구와 유행스타일 변화," **한국의류학회지**, 39(5), 745-754.
- 이수상(2016), "독후감 텍스트의 토픽 모델링 적용에 관한 탐색적 연구," **한국도서관·정보학회지**, 47(4), 1-18.
- 이홍, 김찬모(2004), "양면조직(ambidextrous organization)의 작동 메커니즘: 사례연구를 통한 탐색," **인사조직연구**, 12, 167-197.
- 오민정, 정진철, 이미현(2012), "실버소비자 연구동향 분석과 향후 연구 방향성의 제언," **한국비즈니스 리뷰**, 5(3), 103-119.
- 원영희, 모선희(2008), "한국노년학의 사회학 연구동향 분석," **한국노년학**, 28(4), 753-772.
- 정태연(2008), "「한국 노년학」에서 다룬 노인의 심리적 특성," **한국노년학**, 28(4), 815-829.
- 통계청(2017), 장래인구추계: 2010-2060.
- Blei, D.(2012), "Probabilistic Topic Models," *Communications of the ACM*, 55(4), 77-84.
- Griffiths, T. L. and M. Steyvers(2004), "Finding Scientific Topics," *National Academy of Sciences*, 101, 5228-5235.
- Hansen, D., B. Shneiderman, and M. A. Smith(2011), *Analysing Social Media Networks with Node XL: Insights from a Connected World*, Morgan kaufman Burlington, MA.
- Sebei, H., Hadj Taieb, M. A., and M. Ben Aouicha(2018), "Review of Social Media Analytics Process and Big Data Pipeline," *Social Network Analysis and Mining*, 8(1).
- Wang, Xin, N. T. Bendle, F. Mai, and J. Cotte(2015), "The Journal of Consumer Research at 40: A Historical Analysis," *Journal of Consumer Research*, 42(1), 5-18.

A Topic Modeling Analysis from the 'Aging' Keyword of Domestic Academic Research

Min Jung Oh*

Abstract

The purpose of this study is to find the main phrases of interest topics in old research by using text mining centering on domestic journals published in the national DB for 23 years. 1327 papers were used in the analysis, and the journal was extracted from the social science research including the word 'aging, old' in the abstract, and the topic modeling method was used to show trends during that period. The most important aspect of topic modeling is to broaden and diversify the focus of aging research. As a result, five topics were identified, and the research trends and themes were confirmed through each topic.

In the past, research on aged research has been limited to a micro-approach, and this study examines aged research with a more macroscopic eye. The methodology is also based on the Latent Dirichlet Allocation (LDA) approach. Using analytical methods, we advanced the analysis one step further and focused on core word research that we have not been able to examine in the meantime.

Through this, we were able to know what percentage of the abstracts were covered by the abstracts, and the popularity and trends of the topics could also be predicted if we understood the proportion of themes. The most important goal of this analysis is to maintain important noun phrases while removing common but unfavorable noun phrases such as 'research' or 'analysis', and to calculate the term frequency-inverse document frequency (TF-IDF) of extracted noun phrases. It shows the percentage of frequency after showing the most used words in a specific abstract. As a result, a total of 25 major noun phrases could be created for each of the five topics. Through these results, we will present the trend of aged research and future direction of research.

* Part-time lecturer, Division of Business, Chosun University, First Author

Key words: aging, big data analysis, social network analysis, topic modeling, LDA, text mining

-
- 저자 오민정은 현재 조선대학교 경영학과 시간강사로 재직 중이다. 조선대학교에서 경영학 석사 및 박사를 취득하였다. 2012년과 2015년 한국소비문화학회에서 우수논문상을 수상한바 있으며, 박사학위 취득 후 2016년, 2017년, 2018년 연속 3회 한국연구재단의 연구지원사업에 선정되어 연구를 진행하고 있으며 이 연구 역시 2018년 한국연구재단에서 선정된 과제이다. 주요 연구분야는 고령, 기부, 사회적 배제, 도덕적 정체성 등이다.