

구조적 토픽모델을 이용한 품질경영분야 연구동향 분석*

박 만 희(제1저자)

부산가톨릭대학교 경영학과 (교수)

Analysis of Research Trends in Quality Management Field using Structural Topic Model

Park, Man Hee(First Author)

Department of Business Administration, Catholic University of Pusan (Professor)

Abstract

This study analyzed the research patterns and trends of domestic quality management-related research conducted from 2013 to 2022 using the structural topic model. Based on the collected paper data, we applied structural topic model and regression analysis to analyze research patterns and research trends, and applied a modularity algorithm to analyze the community of the topic network. The results of this study are summarized as follows. First, according to research results applying the structural topic model, research trends in the quality management field were found to consist of 20 topics. Second, the process improvement topic and customer satisfaction topic were analyzed as cold topics with a decreasing research proportion, while the data analysis topic and prediction model topic were analyzed as hot topics with an increasing research proportion. Third, according to the results of community analysis of the topic network, the 20 topics that make up the research trend were found to form 7

* 이 논문은 2020년 대한민국 교육부와 한국연구재단의 인문사회분야 중견연구자지원사업의 지원을 받아 수행된 연구임(NRF-2020S1A5A2A01044104)

communities. This study is expected to provide researchers with help in predicting trends by identifying research trends and exploring research directions in the field of quality management.

Keywords : Structural Topic Model, Text Mining, Quality Management, Regression Analysis, Research Trends

접수일(2024년 01월 27일), 수정일(2024년 02월 05일), 게재확정일(2024년 02월 09일)

I. 서론

품질경영은 기업이나 조직이 제품이나 서비스의 품질을 향상시키기 위해서 수행하는 모든 경영 활동을 의미하며, 고객 만족도를 제고하고 기업 경쟁력을 강화하여 지속가능한 성장을 달성하기 위한 필수적인 요인이면서 핵심 경쟁우위 요소라고 할 수 있다. 기업이나 조직은 품질경영을 통해 고객의 요구사항을 충족시키고, 시장에서의 경쟁력과 브랜드 가치를 강화할 수 있다. 품질경영을 통해 불량률을 감소시키고, 생산 효율을 향상시킴으로써 비용을 절감시킬 수 있고, 품질에 대한 조직의 인식수준을 높이고 품질향상 활동을 통해 조직의 효율성을 향상시킬 수 있다. 품질경영을 성공적으로 수행하기 위해서는 기업의 모든 구성원들이 품질에 대한 중요성을 인식하고 다양한 개선활동과 노력을 통해 지속적인 개선을 유지할 필요가 있다. 기업이 지속가능한 성장과 발전을 달성하기 위해서는 품질경영에 대한 지속적인 투자와 노력이 필요하다고 판단된다.

전문가 식견이나 지식을 바탕으로 연구동향이나 추세를 예측하는 것은 미시적인 관점이나 특정 기술의 동향과 추세를 분석하고 예측하는 경우에는 가치가 있지만 분석대상 자료가 방대한 경우 분석에 많은 시간과 비용이 소요되므로 전문가 지식과 판단에 의존하기에는 한계점이 존재한다. 분석대상 텍스트나 문헌의 양이 방대하거나 빅데이터 수준일 경우 적용 가능한 분석기법으로 텍스트 마이닝을 이용할 수 있다. 텍스트 마이닝(Text mining)은 문서, 소셜 미디어 게시물 등 다양한 형태의 텍스트 데이터 분석을 통해 의미 있는 패턴이나 정보를 추출하는 기법이다. 텍스트 마이닝 기법 중 하나인 토픽 모델(Topic model)은 텍스트 데이터에 나타나는 단어들의 동시 사용 패턴을 기반으로 분석 대상 텍스트를 대표하는 주제나 이슈를 추출하는 기법이다. 토픽 모델은 비즈니스 인사이트 분석, 소셜 미디어 분석, 자연어 처리 분야에서 폭넓게 활용되고 있다. 자연어 처리 분야에서는 문서 요약, 토픽 분석, 질의응답, 감성

분석 등 다양한 자연어 처리 작업에 활용되고 있다. 본 연구에서는 품질경영분야 연구 동향과 추세 분석을 위하여 최근 10년 동안 품질경영학회지에 게재된 논문의 서지 정보를 바탕으로 구조적 토픽 모델을 이용하여 품질경영분야의 연구 동향과 추세를 분석하고 시사점을 제시하였다.

II. 선행 연구 검토

품질경영 관련 문헌이나 논문 자료를 바탕으로 품질경영분야의 연구동향을 분석하고 연구추세를 예측한 연구들 중 본 연구와 관련성이 높은 연구들의 연구방법론을 분류하면 다음과 같다. 문헌분석을 바탕으로 전문가 식견을 이용한 동향분석, 텍스트 마이닝 기반 키워드 빈도분석을 통한 동향분석, 텍스트 마이닝 기반 동시단어분석과 군집분석을 통한 동향분석, 토픽모델을 이용한 동향분석 등 크게 4가지 방법론으로 정리할 수 있다. 4가지 방법론을 이용하여 품질경영 관련 연구동향을 분석한 연구들 중 비교적 최근에 발표된 연구들의 연구 내용을 정리하면 다음과 같다.

김연성(2016)은 1989년부터 2014년까지 품질경영학회지에 게재된 논문들 중 서비스품질 분야에 해당하는 71편의 논문을 대상으로 키워드 분석과 내용 검토를 거쳐 서비스 프로세스 매트릭스 분석에 의한 적용 대상 업종의 변화와 특성을 파악하고 관련 연구 동향을 분석하였다(김연성, 2016). 분석결과를 바탕으로 인더스트리4.0 시대를 맞아 보다 다양한 업종이 등장하고 새로운 비즈니스 모델이 실험될 것으로 예측하였고, 새로운 업종에서도 서비스품질이 주요 이슈가 된다면 적용 방향을 찾을 필요가 있음을 제시하였다(김연성, 2016). 기존의 서비스 수익체인에서 제시하는 서비스품질과 고객만족, 고객충성의 관계를 넘어서 비즈니스 성과까지 포괄하는 다단계의 다차원 분석이 필요하고, 이를 위해서는 명확한 비즈니스 모델에 대한 이해, 빅데이터와 사물인터넷(IoT) 등을 활용하여 정확한 데이터를 확보하고 다양한 관점의 심도 있는 분석이 진행될 것으로 예측하였다.

정규석 등(2016)은 TQM 관련 122편의 논문을 ISO 시스템, 품질 보증, 품질경영 이론, 품질경영 실증연구, 품질상, 품질경영 등 6개 범주로 구분하고 연구내용을 분석하였다. 품질경영의 종합적 혁신도구로서의 특성을 감안할 때 다른 여타 혁신기법과의 관계나 통합적 접근방법에 대한 연구는 품질경영의 실무적 적용성을 높이고 품질경영 범위의 외연적 확대에 기여할 것으로 예측하였다. 또한 국가품질상 이외에도 품질경영 부문상이나 기타 품질관련 여러 상들에 실증적 성과연구도 품질경영의 다양한 실천에 도움을 제공할 것으로 분석하였다.

박현아 등(2023)은 온라인 서비스 품질에 관한 높은 관심과 연구에도 불구하고 온

라인 서비스 품질에 대한 정의나 측정 등이 체계화되지 않았음을 인식하고, 온라인 서비스 품질 연구에 대한 종합적 고찰을 수행하였다(박현아, 2023). 2001년부터 2021년까지 마케팅 분야의 주요 학회지에 게재된 온라인 서비스 품질 관련 연구들을 대상으로 국내 온라인 서비스 품질의 연구 경향을 종합적으로 고찰하였으며, 49편의 연구들을 바탕으로 온라인 서비스 품질의 정의 및 측정 방법에 대해 살펴보고 온라인 서비스 품질과 관련된 다양한 변수들과의 인과관계를 분석하였다.

이상복(2019)은 1977년부터 2018년까지 40년 동안 품질경영학회지에 게재된 논문 초록을 대상으로 텍스트 마이닝 기법을 적용하여 단어 빈도수 기반 워드 클라우드(Word cloud) 기법을 바탕으로 연구동향을 분석하였다. 지난 40년간의 연구 동향은 시대 변화에 따라 중요 단어가 변화하고 있으나, 품질경영 전체 이론에서 보면 큰 변화가 없는 것으로 분석되었으며, 100회 이상 사용된 22개 단어는 품질 범주에서 많이 사용되는 익숙한 단어들이나 process, system, model, service 단어는 600번 이상 사용 되었으며, management, data, methods 단어는 400번 이상 사용된 것으로 분석되었다.

박만희(2016)는 2002년부터 2014년까지 품질경영학회지에 게재된 학술논문의 서지정보를 바탕으로 품질경영분야 지식구조의 변화행태와 연구주제의 변화추세를 분석하였으며, 분석을 위해 동시단어분석과 군집분석, 전략맵 등의 기법을 이용하여 단어들의 네트워크 구조와 지식구조를 가시화하였다(박만희, 2016). 시간의 흐름에 따른 전략맵 분석결과에 따르면 품질경영분야의 전통적인 연구주제인 신뢰성, 6시그마, 관리도 관련 연구주제는 비중이 감소하고 연구의 중요성도 쇠퇴하고 있는 것으로 분석되었으며, 제품설계와 고객만족 관련 연구주제는 전체 분석기간에 걸쳐 중요한 연구 주제로 자리잡고 있는 것으로 분석되었다. 경영혁신 관련 연구주제는 새롭게 부상하고 있으며, 프로세스 모델 관련 연구주제는 시스템 성과 관련 연구 주제로 연구범위가 확대된 것으로 분석되었다(박만희, 2016).

Wawak et al.(2020)은 품질경영 관련 8개 저널로부터 4,833편의 논문을 수집한 후 논문 전체 내용을 대상으로 텍스트 마이닝을 수행하고 군집 분석을 통해 연구동향을 분석하였다. 추세분석을 통해 장기적으로 지속되고 있는 연구주제가 17개, 추세가 하락하고 있는 연구주제가 4개, 새롭게 부상하고 있는 연구주제가 11개, 등락을 반복하는 연구주제가 13개라는 것을 밝혔다.

김민준(2023)은 품질 4.0 관련 연구의 서로 다른 관점 및 여러 연구 간의 통합적 이해의 필요성에 따라, 품질 4.0 관련 논문 1,234개를 텍스트 마이닝 및 기계학습을 활용해 분석하고, 그 결과를 해석함으로써 품질 4.0 관련 연구들의 동향을 파악하였다(김민준, 2023). 품질 4.0 관련 연구로부터 27개 연구토픽을 도출하고, 이를 5개

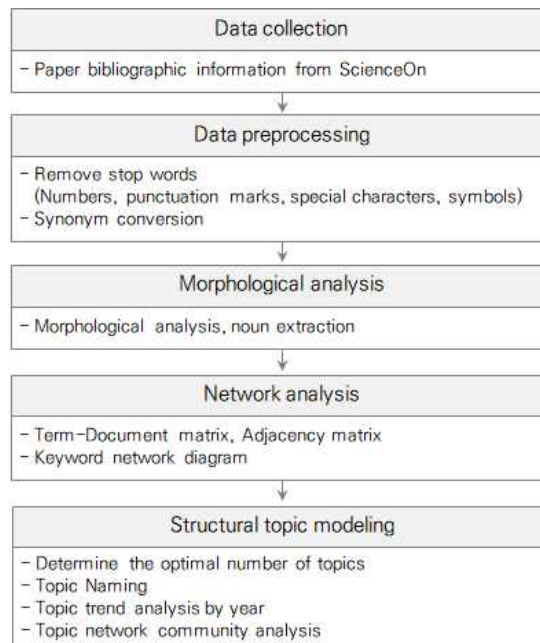
레벨로 분류(연결 인프라 기술-지능화, 자동화 및 최적화 기술-방법론 및 전략-달성 목표-적용 산업)하여 멀티레벨 프레임워크를 제안하였으며, 27개 연구토픽간의 관계 분석을 통해 연관성을 파악하고 관계를 해석하였다(김민준, 2023).

이상에서 살펴본 바와 같이 전문가 식견이나 키워드 분석을 이용하여 품질경영분야의 연구동향을 제시한 연구들은 진행되어 왔으나, 분석 데이터의 메타정보를 바탕으로 토픽들의 변화 형태와 추이를 제시할 수 있는 구조적 토픽모델을 적용한 연구는 진행되지 못하였다. 따라서 본 연구에서는 품질경영 관련 저널에 게재된 최근 10년 동안의 논문 서지정보를 바탕으로 구조적 토픽 모델을 적용하여 연구 동향을 분석하고, 연구 비중이 증가하는 추세에 있는 hot topic과 연구 비중이 감소하는 추세에 있는 cold topic을 제시하였다.

III. 연구방법론

본 연구에서는 구조적 토픽모델을 이용하여 품질경영분야의 연구 동향과 추세를 분석하기 위하여 <그림 1>과 같은 연구 방법론을 적용하였다.

<그림 1> 연구 모형



분석 대상 자료인 논문의 서지정보는 KISTI에서 운영 중인 과학기술 지식 인프라 (ScienceOn) 사이트의 상세검색 기능을 이용하여 2013년부터 2022년까지 품질경영 학회지에 게재된 논문의 논문명, 저자, 발행 연도, 키워드, 영문초록을 수집하였다. 상세검색 기능의 검색항목 중 저널명은 품질경영학회지, 발행연도는 2013과 2022로 설정하고 검색을 수행한 후 535건의 논문서지정보를 수집하였다. 자료 전처리 단계에서는 논문서지 정보들 중 영문초록을 텍스트 마이닝 대상 항목으로 선정하고 분석결과에 무의한 정보를 제공하는 특수문자, 공백문자, 구두점, 기호 등의 불용어를 삭제하였고, 품질경영 분야에서 활용도가 높은 동의어 목록을 생성한 후 29개의 동의어를 변환하였다.

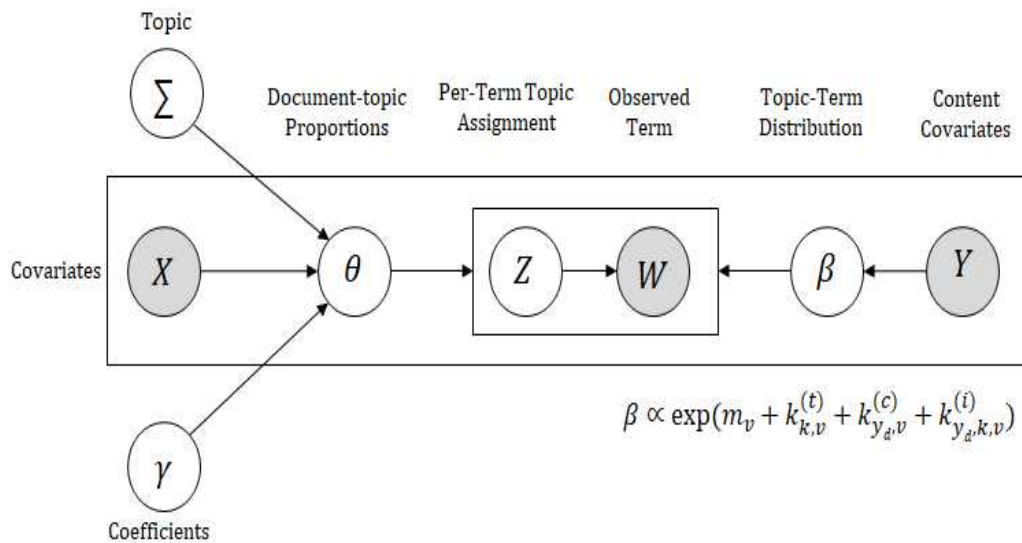
품질경영 분야의 연구동향을 분석하므로 품질과 경영이라는 단어를 제거하였고, 영문초록에 자주 등장하는 purpose, methods, results, conclusion 등을 포함하여 24개 단어를 제거하였다. 형태소 분석단계에서는 텍스트 마이닝 대상 항목인 영문초록에 대하여 최소의 어미단어인 형태소로 분해하고, 개별 형태소의 품사를 분석하는 형태소 분석을 수행하였다. R에서 이용 가능한 형태소 분석기는 다양하게 존재하지만 본 연구에서는 속도 측면에서 좋은 성능을 나타내는 RccpMeCab 패키지를 이용하여 형태소 분석을 수행하였으며, 형태소 분석결과로부터 품사가 명사인 단어만 추출하여 분석에 이용하였다. 키워드 네트워크 분석 단계에서는 형태소 분석 결과로부터 단어-문서행렬을 생성하고 네트워크 분석에 필요한 인접행렬(Adjacency matrix)을 도출한다. 인접행렬을 이용하여 네트워크 다이어그램을 생성하는데 행렬의 행과 열은 노드가 되고, 개별 원소의 값은 노드간의 관계의 강도를 나타낸다. 단어 네트워크 다이어그램을 생성하여 시각화하면 논문에서 등장한 단어들의 전체적인 분포와 형태를 파악할 수 있다.

대량의 텍스트 데이터로부터 의미 있는 토픽들을 찾아내는데 이용되는 토픽 모델링 방법 중 가장 대표적인 기법은 LDA(Latent Dirichlet Allocation)라고 할 수 있으며, LDA는 문서의 단어 분포를 분석하여 개별 문서가 여러 토픽들의 혼합으로 구성되어 있다고 가정하고 문서 집합에서 숨겨진 토픽을 찾아내는 확률적 생성 모델이다(Blei et al., 2003). LDA는 문서 집합 내의 문서들이 시간적으로 어떻게 변하는지를 고려할 수 없으며, 개별 문서의 단어 선택이 서로 독립이라고 가정하고 있으므로 단어들간의 의존성을 고려할 수 없다는 단점을 가지고 있다. STM(Structural Topic Model)은 LDA에 비해 텍스트 데이터에서 개별 토픽 간의 상호 연관성을 파악할 수 있고 문서들 간의 상호작용을 바탕으로 보다 정확한 토픽 모델을 제공할 뿐만 아니라, 시간의 흐름에 따른 토픽들의 변화 형태도 파악할 수 있다는 장점을 가지고 있다. STM은 분석 대상 데이터에 함께 제공된 메타데이터를 이용하여 공변량(Covariate) 토픽 모

델 분석을 수행할 수 있다. 메타데이터는 문서에 기록된 정보로서 기사나 보고서 등을 나타내는 문서 유형, 문서의 주제 분야를 나타내는 문서 분류, 문서가 작성된 시간적 정보를 나타내는 작성시점 등을 의미한다. 메타데이터를 독립변수, 토픽의 출현확률을 종속변수로 설정하고 회귀분석을 수행하면 공변량에 따른 변화 형태를 분석할 수 있다.

STM의 구조는 <그림 2>와 같고 단어 크기가 V개인 개별 문서(d)로부터 K개 주제를 생성하는 프로세스를 정리하면 다음과 같다(Roberts et al., 2016).

<그림 2> 구조적 토픽 모델



(단계 1) 문서로부터 주제 추출

개별 주제는 문서 공변인(X_d) 벡터를 기반으로 로지스틱 정규 일반 회귀모형으로부터 추출한다.

$$\theta_d \sim \text{LogisticNormal}_{k-1}(\Gamma' x_d, \Sigma)$$

여기서 x_d 는 $1 \times p$ 벡터, Γ 는 $p \times (K-1)$ 계수 벡터, Σ 은 $(k-1) \times (k-1)$ 공분산 행렬을 나타낸다.

(단계 2) 주제에 대한 단어 분포 생성

문서수준의 공변량(content covariate) y_d 가 주어졌을 때, 단어 분포(m), 주제 편차($k_{k,v}^{(t)}$), 공변량 집단편차($k_{y_d,v}^{(c)}$), 두 편차간의 교호작용($k_{y_d,k,v}^{(i)}$)를 이용하여 개별 주제

(k)를 나타내는 단어분포를 생성한다.

$$\beta_{d,k,v} = \frac{\exp(m_v + k_{k,v}^{(t)} + k_{y_d,v}^{(c)} + k_{y_d,k,v}^{(i)})}{\sum_v \exp(m_v + k_{k,v}^{(t)} + k_{y_d,v}^{(c)} + k_{y_d,k,v}^{(i)})}$$

(단계 3) 주제 할당 및 단어 추출

개별 문서에서 특정 주제들이 발현할 확률분포를 바탕으로 주제를 할당한다.

$$z_{d,n} \sim \text{Multinomial}_K(\theta_d)$$

주제가 주어지면 주제에 해당하는 단어(w_d)들을 추출하여 할당한다.

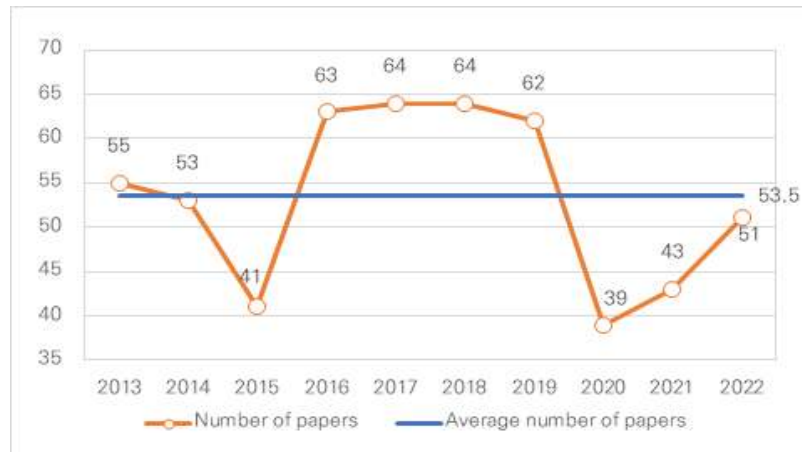
$$w_{d,n} \sim \text{Multinomial}(\beta, z_{d,n})$$

토픽 모델에서 의미 일관성(Semantic coherence)은 개별 토픽 내 단어들이 얼마나 의미적으로 일관성이 있는지를 나타내고, 배타성(Exclusivity)은 서로 다른 토픽 간의 구분이 얼마나 명확한지를 나타낸다. 본 연구에서는 의미 일관성과 배타성 지표를 이용하여 최적 토픽 수를 결정하는데, 두 지표는 서로 상쇄관계(trade-off)에 있으므로 두 지표간의 균형을 고려하여 최적의 토픽수를 결정한다. 논문 게재연도를 독립변수, 토픽의 발현확률을 종속변수로 설정하고 회귀분석을 수행하면 연구비중이 증가하는 경향을 보이는 hot topic과 연구비중이 감소하는 경향을 보이는 cold topic을 도출할 수 있다. 구조적 토픽모델로부터 도출된 특정 문서에 대한 주제 분포를 나타내는 θ 값을 바탕으로 유사도계수를 이용하여 네트워크를 생성하고 modularity를 기반으로 커뮤니티(Community)를 생성하고 가시화한다.

IV. 실증분석 결과

본 연구에서는 품질경영분야의 연구동향과 추세분석을 위해 KISTI에서 운영 중인 과학기술 지식인프라(ScienceOn) 사이트로부터 2013년부터 2022년까지 품질경영학회지에 게재된 535편의 논문 서지정보를 수집하여 분석에 이용하였다. <그림 3>의 연도별 논문 게재 수 추이를 살펴보면 연평균 53.5편의 논문이 게재되고 있으며, 1년 동안 가장 많이 게재된 논문 수는 64편이고 가장 작게 게재된 논문 수는 39편으로 분석되었다.

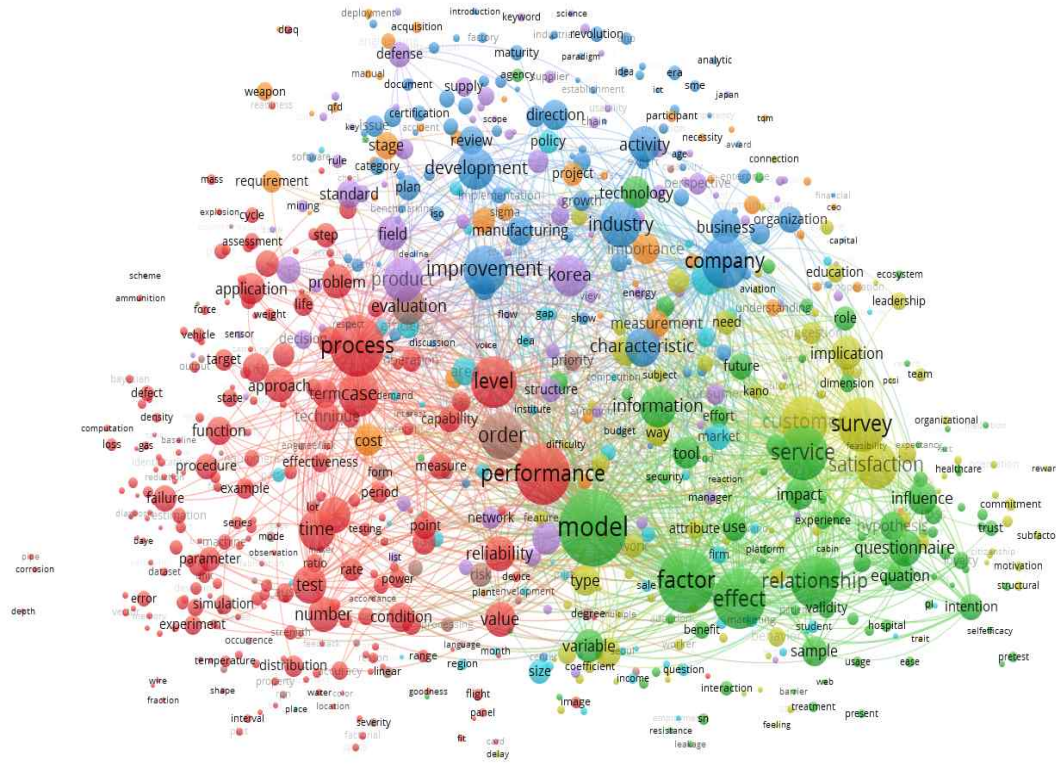
<그림 3> 연도별 논문현황



수집된 논문의 영문초록 정보를 바탕으로 형태소 분석을 수행한 후 단어문서행렬을 생성하고 인접행렬을 생성하여 네트워크 다이어그램으로 표현한다. 단어문서행렬을 인접행렬로 변환하여 네트워크 다이어그램으로 표현하면 노드와 연결이 너무 많아 구분이 어렵다는 문제를 해결하기 위하여 단어문서행렬의 희소성을 0.994로 설정하고, 빈도수가 낮은 일부 단어를 제거하여 차원이 751인 인접행렬을 생성하고 VOSviewer SW를 이용하여 단어 네트워크로 가시화하면 <그림 4>와 같다. 단어 네트워크 다이어그램을 통해 535개 논문에 등장한 단어들의 분포 형태와 연결성을 확인할 수 있다.

VOSviewer는 노드들 간의 연관 강도(Association strength)를 기준으로 계층적 군집분석 알고리즘을 적용하여 군집분석을 수행하는데 분석결과에 따르면 8개 클러스터를 형성하는 것으로 분석되었다. 군집별로 빈도수가 높은 상위 단어를 기준으로 군집명을 부여하면 <표 1>과 같다. 8개 군집은 공정관리(Process management), 서비스 모델(Service model), 개선활동(Improvement activity), 고객만족(Customer satisfaction), 제품표준(Product standard), 품질전략(Quality strategy), 중요성 측정(Importance measurement), 위험관리(Risk management) 이고, 군집 1이 가장 많은 키워드를 포함하고 있으며 군집 8이 가장 작은 수의 키워드를 포함하고 있는 것으로 나타났다. VOSviewer와 같이 동시단어출현행렬을 기준으로 군집을 형성하는 SW는 군집결정에 이용되는 지표의 설정 값에 따라 군집수가 상이하게 형성된다는 문제점과 함께 연도에 따른 연구추세를 파악할 수 없다는 문제점을 가지고 있다.

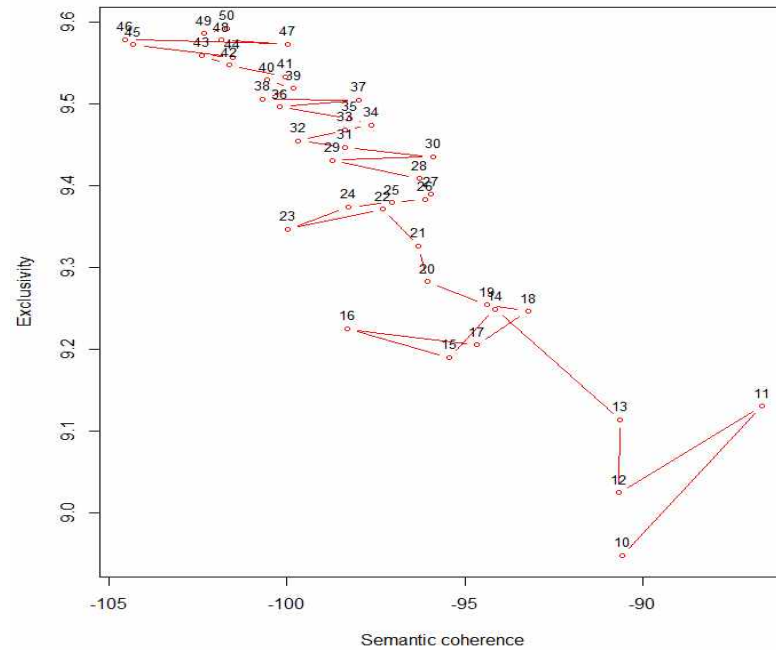
<그림 4> 단어 네트워크 다이어그램



<표 1> 단어 네트워크의 클러스터링 결과

Cluster	Top Keyword	Number of keyword	Cluster Name
Cluster 1	Performance, Process, Level, Case, Time	233	Process management
Cluster 2	Model, Factor, Effect, Service, Relationship	98	Service model
Cluster 3	Improvement, Company, Industry, Characteristic, Development	97	Improvement activity
Cluster 4	Customer, Satisfaction, Survey, Environment, Implication	94	Customer satisfaction
Cluster 5	Product, Korea, Field, Standard, Methodology	88	Product standard
Cluster 6	Strategy, Area, Market, Efficiency, Policy	60	Quality strategy
Cluster 7	Importance, Measurement, Stage, Item, Requirement	48	Importance measurement
Cluster 8	Order, Evaluation, Risk, Action, Form	33	Risk Management

<그림 5> 의미 일관성과 배타성



최적 토픽 수를 결정하기 위해서 R의 stm 패키지에서 제공하는 searchK 함수를 이용하였으며 토픽의 수를 10부터 50까지 변경하면서 개별 토픽 수를 위한 모델을 학습하고, 모델의 품질을 평가하기 위한 지표인 의미 일관성(Semantic coherence), 배타성(Exclusivity), 가능도 추정치(Held-out likelihood), 잔차(Residuals), 하한(Lower bound)을 계산하였다. 의미 일관성은 토픽에서 출현 가능성이 가장 높은 단어가 자주 발생할 때 최대가 되고 배타성은 출현 가능성이 높은 상위 단어가 특정 주제에서 고유할 때 값이 커진다(Griffiths and Steyvers, 2004). 본 연구에서는 <그림 5>에서 보는 바와 같이 의미 일관성 지표와 배타성 지표 측정값의 절충점인 20을 최적의 토픽수로 결정하였다.

최적 군집 수를 바탕으로 20개의 토픽을 형성하고 특정 단어가 개별 토픽에 속할 확률(Gamma)과 확률 값이 큰 상위 10개 단어를 도출하면 <그림 6>과 같다. 토픽명은 토픽에 속하는 빈도수가 높은 단어 조합으로 명확하게 부여되는 경우도 있지만 단어 조합으로 토픽명을 부여하기 모호한 경우에는 해당 토픽에 속하는 연구논문의 제목, 영문초록, 키워드 등을 고려하고 전문가 협의를 통하여 토픽명을 부여하였다.

<그림 6> 토픽별 단어 확률 분포



<표 2>의 토픽별 연구 내용을 살펴보면 다음과 같다. 위험관리(Risk management) 토픽은 제품이나 서비스 전주기상에서 발생할 수 있는 잠재적인 위험을 식별하고 평가하여 위험을 최소화하기 위한 관련 연구들로 구성되어 있다. FMEA기반 위험평가, 제품 위험성 평가모델, 위험평가 방안 및 대응방안, 감사기준 정립 등의 연구들로 구성되어 있다. 인과 모델(Casual model) 토픽은 다양한 품질속성이 고객, 결과, 경영성과 등에 미치는 영향요인과 영향의 정도를 분석하고 성과를 개선하기 위한 연구들로 구성되어 있으며 인과관계 모형과 구조 방정식 모형을 이용한 연구가 주를 이루고 있다. 공정 기술(Process technology) 토픽은 실험 방안개선과 모니터링 및 진단, 공정 설계수준 결정 방법, 공정 최적화를 위한 관련 연구들로 구성되어 있다. 개선 활동(Improvement activity) 토픽은 QFD, 6시그마, CMMI, BSC 등의 기법이나 개선활동을 통해 품질성과를 개선하고 관리하기 위한 관련 연구들로 구성되어 있다. 실험 설계(Experiment design) 토픽은 실험설계와 실험계획에 따른 신뢰성, 내구성, 품질개선, 영향분석 등을 통해 반응변수 최적화 관련 연구들로 구성되어 있다.

<표 2> 토픽 단어 및 토픽 명칭

Topic	Gamma	Terms	Topic name
1	0.0625	product, risk, order, development, failure, assessment, process, evaluation, factor, procedure	Risk management
2	0.0593	survey, level, model, industry, questionnaire, factor, policy, effect, service, user	Causal model
3	0.0451	production, process, problem, condition, simulation, design, product, ratio, cost, level	Process technology
4	0.0415	problem, improvement, activity, project, performance, development, company, business, goal, model	Improvement activity
5	0.0533	test, cause, life, design, material, condition, case, effect, experiment, temperature	Experiment design
6	0.1180	relationship, factor, effect, model, survey, equation, service, satisfaction, questionnaire, customer	Factor analysis
7	0.0598	industry, service, strategy, area, company, case, content, application, direction, category	Research review
8	0.0526	process, development, improvement, information, cost, change, order, weapon, component, engineering	Process improvement
9	0.0572	customer, satisfaction, kano, model, attribute, survey, service, index, factor, importance	Customer satisfaction
10	0.0444	test, process, time, case, sample, condition, value, distribution, size, point	Process optimization
11	0.0336	model, distribution, reliability, life, failure, parameter, year, rate, probability, process	Reliability
12	0.0270	competitiveness, characteristic, aspect, korea, index, level, role, definition, guideline, support	Quality indicators
13	0.0290	network, service, difference, approach, mining, word, classification, keyword, standard, trend	Data analysis
14	0.0701	effect, survey, company, performance, relationship, service, factor, satisfaction, leadership, implication	Performance improvement
15	0.0472	r, performance, time, difference, test, number, order, type, model, efficiency	Performance analysis
16	0.0437	process, factor, value, performance, capability, model, condition, index, case, simulation	Process capability
17	0.0585	model, time, algorithm, performance, control, design, example, variable, process, function	Forecasting model
18	0.0590	evaluation, case, certification, improvement, process, defense, development, safety, korea, performance	Quality certification
19	0.0257	level, model, number, question, survey, factor, interaction, attribute, effect, design	Quality improvement
20	0.0126	methodology, approach, parameter, product, function, control, bayesian, model, estimation, application	Parameter optimization

요인분석(Factor analysis) 토픽은 다양한 분야에서 품질에 영향을 미치는 속성들을 고려하고 실증분석을 통해 영향요인을 도출하고 조절효과 분석을 통해 프로세스를 개선하기 위한 연구들로 구성되어 있다. 연구 동향(Research trend) 토픽은 2016년 품질경영학회 50주년 특별 호에 게재된 통계적 품질관리, 통계적 기법, 서비스품질, 품질혁신 등 품질경영 주제영역별로 수행된 연구동향 분석 연구들로 구성되어 있다. 프로세스 개선(Process improvement) 토픽은 프로세스 개선을 위한 우선순위 선정, 다양한 표준화 사례, 품질 개선과 품질관리를 위한 프로세스 개발 및 개선 등과 관련된 연구들로 구성되어 있다. 고객 만족(Customer satisfaction) 토픽은 KANO 모형, AHP 기법, SERVQUAL, PCSI 지수 등을 이용하여 고객 만족도를 분석하고 측정하여 개선방안을 제시하는 연구들로 구성되어 있다. 공정 최적화(Process optimization) 토픽은 공정능력 분석, 성능 개선, 시험방법 개선, 경제적 실험설계, 품질향상 방안, 최적 실험계획 등 공정 최적화 관련 연구들로 구성되어 있다.

신뢰성(Reliability) 토픽은 제품이나 서비스의 품질과 신뢰성을 측정하고 개선하기 위해 신뢰성 지수, 신뢰도 평가, 보전정책, 수명 및 신뢰도 추정모형 개발 관련 연구들로 구성되어 있다. 품질 지표(Quality indicators) 토픽은 제품 또는 서비스의 품질을 측정하고 개선시키기 위해 특정 목표나 표준을 충족하는 정도를 측정하고 개선방향을 제시할 수 있는 품질지표 및 지수 개발 관련 연구들로 구성되어 있다. 데이터 분석(Data analysis) 토픽은 빅데이터 분석, 텍스트 마이닝, 머신러닝, 계량서지학, 네트워크 분석 방법 등을 이용하여 현황분석 및 품질개선 방안을 제시하는 연구들로 구성되어 있다. 성과 개선(Performance improvement) 토픽은 경영성과에 영향을 미치는 품질 속성과 요인들을 선정하고 실증분석을 통해 결과변수나 성과변수에 미치는 영향을 분석하는 연구들로 구성되어 있다. 성과 분석(Performance analysis) 토픽은 예측 모형, 품질평가, 효율성 분석, 통계적 차이분석 등의 기법을 적용하여 품질 성과 분석을 위한 방법과 측정기법 관련 연구들로 구성되어 있다.

공정 능력(Process capability) 토픽은 강건 설계, 공정능력지수, 측정능력지수, 측정시스템 평가방법 등 공정능력을 개선하기 위한 관련 연구들로 구성되어 있다. 예측 모형(Forecasting model) 토픽은 머신러닝이나 다양한 최적화 기법을 적용하여 시스템의 품질지표나 성과지표를 예측하기 위한 연구들로 구성되어 있다. 품질 인증(Quality certification) 토픽은 품질개선, 인증기술, 인증제도, 인증체계, 경쟁력 요인, 신뢰성 보증방안 등 제품이나 서비스의 품질을 개선하고 품질관리 체계 준수를 포함하는 품질인증 관련 연구들로 구성되어 있다. 품질 개선(Quality improvement) 토픽은 서비스품질 개선방안, 최적 실험방법, 품질평가지표, 최적 파라미터 설계, 핵심 품질속성 등 지속적으로 고객 만족도를 높이고 효율성을 개선하여 경쟁우위를 유지하기 위한 품질개선 관련 연구들로 구성되어 있다. 모수 최적화(Parameter optimization)

토픽은 베이지 추정, 품질손실함수, 신경망, 데이터 마이닝, 감성분석 등의 기법을 적용하여 다양한 시스템의 최적 모수를 결정하는 연구들로 구성되어 있다.

연구 토픽별 hot topic과 cold topic을 도출하기 위하여 연구가 수행된 연도를 독립변수, 토픽의 출현확률(Prevalence)을 종속변수로 설정하고 회귀분석을 수행하였다. 회귀분석 결과들 중 유의수준 5%와 10%에서 통계적으로 유의한 토픽의 계수 값과 유의확률을 정리하면 <표 3>과 같다.

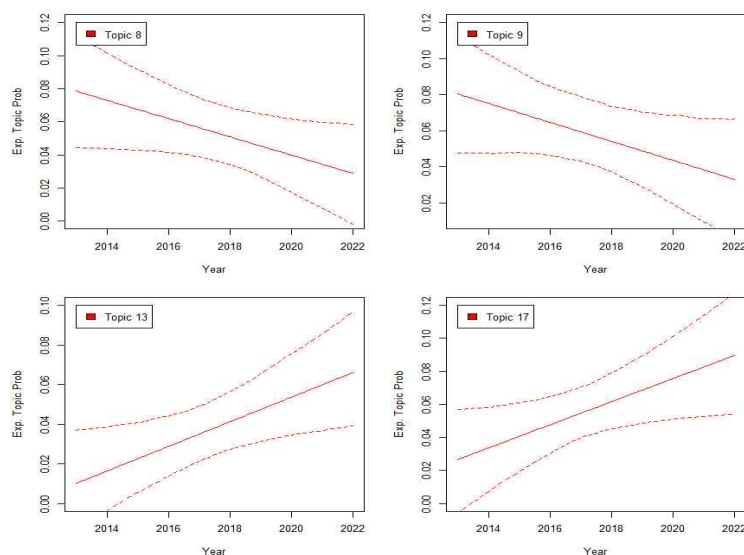
<표 3> 회귀분석 결과

Topic	Intercept		Slope	
	Coefficient	Pr(> t)	Coefficient	Pr(> t)
Topic 8	11.1957	0.0613*	-0.0055	0.0625*
Topic 9	10.6446	0.0906*	-0.0053	0.0924*
Topic 13	-12.3709	0.0207**	0.0062	0.0204**
Topic 17	-13.9144	0.042**	0.0069	0.0412**

* p<0.1, ** p<0.05

회귀분석 결과를 바탕으로 연도에 따른 cold topic과 hot topic의 변화 추세를 그림으로 나타내면 <그림 7>과 같다.

<그림 7> Cold topic & Hot topic

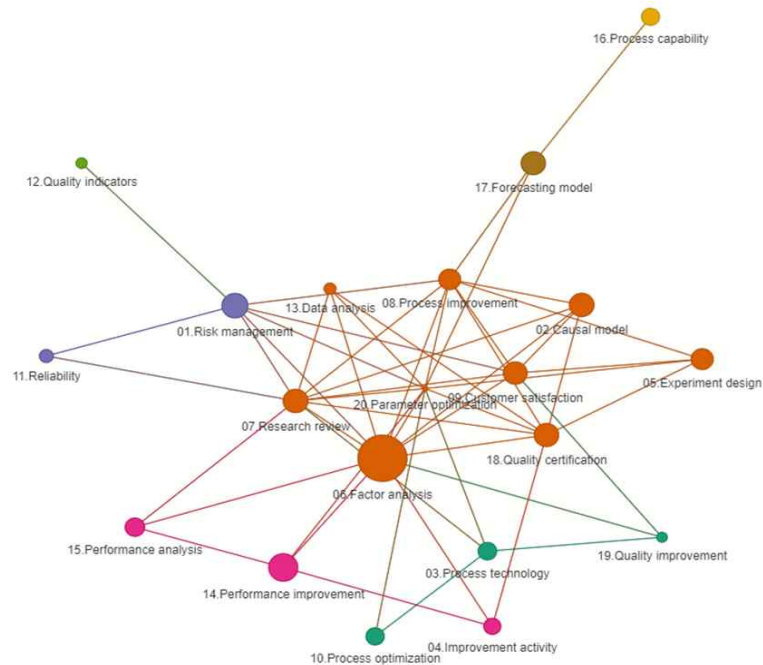


공정 개선(Topic 8) 토픽과 고객 만족(Topic 9) 토픽은 연구비중이 감소하는 추세에 있는 cold topic으로 나타났다. 이는 공정개선과 고객만족 주제는 품질경영 분야에서 지속적으로 진행되어 온 연구주제인 관계로 연구자들의 관심이 낮아지고 이로 인해 연구의 비중이 감소한 것으로 판단된다. 데이터 분석(Topic 13) 토픽과 예측 모형(Topic 17) 토픽은 연구비중이 증가하는 추세에 있는 hot topic으로 분석되었다. 이는 품질경영 분야에서도 빅데이터, 머신러닝 등의 기법을 적용하여 다양한 품질문제를 분석하고 최적의 해결방안을 제시하기 위한 연구들이 수행되고 있으며, 향후 불량 예측, 품질 이상 감지, 품질검사 자동화, 고객 반응 분석, 프로세스 최적화 등의 분야에 폭넓게 적용될 것으로 예측된다. 또한 현재까지는 강화 학습(Reinforcement learning), 딥러닝(Deep learning) 등의 인공지능 기법을 적용한 연구는 미진하지만 장기적으로 관련 기법을 적용한 연구가 활발하게 진행될 것으로 판단된다.

품질경영 분야의 연구동향 분석과 추세 분석결과를 바탕으로 향후 서비스 경영 분야에서 연구의 관심도와 비중이 증가할 것으로 예상되는 연구주제를 정리하면 다음과 같다. 첫째, 빅데이터와 딥러닝을 활용하여 서비스 특성이나 고객 피드백 데이터를 분석하고, 분석결과를 바탕으로 고객 만족도를 예측하는 연구가 활성화 될 것으로 예상된다. 둘째, 빅데이터 분석을 통해 개별 고객의 선호도, 행동 패턴 등을 이해하고, 딥러닝을 활용하여 개인화된 맞춤형 서비스를 제공하는 방법 연구도 활성화 될 것으로 예상된다. 셋째, 딥러닝을 활용한 자연어 처리 모델을 통해 고객의 문의를 자동으로 처리할 수 있는 시스템 구축 방안과 고객지원 서비스 품질 제고방안에 대한 연구도 활성화 될 것으로 예측된다.

토픽 모델 적용을 통해 도출된 문서가 특정 주제에 속할 확률을 나타내는 θ 값에 대하여 cosine 유사도 계수를 적용하고, 최대 발현확률에 대한 개별 토픽의 발현확률의 비율로 노드의 크기를 할당하여 토픽 네트워크를 구성하였다. 네트워크 다이어그램의 가독성을 높이기 위하여 일정한 크기의 동일한 가중치를 부여하여 노드의 크기를 조정하고 modularity 알고리즘을 적용하여 토픽 네트워크의 커뮤니티를 형성하면 <그림 8>과 같으며, 형성된 7개 커뮤니티를 구분하기 위하여 동일한 커뮤니티에 속하는 토픽들은 동일한 색상을 부여하였다. 7개 커뮤니티에 속하는 토픽을 정리하면 <표 4>와 같으며, 9개 토픽을 형성하는 것으로 분석되었다. 9개 토픽으로 구성된 규모가 큰 커뮤니티부터 1개 토픽으로 구성된 작은 커뮤니티를 형성하는 것으로 나타났다. 형성된 7개 커뮤니티를 정리하면 프로세스 및 제품 품질, 성과 개선, 공정 최적화, 신제품 개발 위험관리, 품질 지표, 공정 능력, 예측 모형으로 요약할 수 있다.

<그림 8> 토픽의 네트워크 커뮤니티



<표 4> 토픽 커뮤니티

Community	Topics	Number of topic	Community name
1	02.Causal model 05.Experiment design 06.Factor analysis 07.Research review 08.Process improvement 09.Customer satisfaction 13.Data analysis 18.Quality certification 20.Parameter optimization	9	Process & Product quality
2	04.Improvement activity 14.Performance improvement 15.Performance analysis	3	Performance improvement
3	03.Process technology 10.Process optimization 19.Quality improvement	3	Process optimization
4	01.Risk management 11.Reliability	2	Reliability & Risk management
5	12.Quality indicators	1	Quality indicators
6	16.Process capability	1	Process capability
7	17.Forecasting model	1	Forecasting model

V. 결 론

본 연구에서는 텍스트 마이닝 기법 중 하나인 구조적 토픽 모델을 이용하여 품질경영 분야의 연구 동향과 추세를 분석하였다. 본 연구의 결과는 연구자들에게 현재의 연구 동향을 이해하고 미래 연구를 위한 연구의 방향성을 설정하는데 도움을 제공할 수 있으며, 연구의 우선순위를 조정할 통해 효율적인 연구가 진행될 수 있도록 지원할 수 있다는 점에서 연구의의가 있다. 본 연구를 통해 도출된 연구결과를 요약하면 다음과 같다. 첫째, 구조적 토픽모델을 이용하여 2013년부터 2022년까지 최근 10년 동안 품질경영 분야의 연구동향을 분석한 결과에 따르면 품질경영 관련 연구는 위험관리, 인과 모델, 공정 기술, 개선 활동, 실험 설계, 요인분석, 연구 동향, 프로세스 개선, 고객 만족, 공정 최적화, 신뢰성, 품질 지표, 데이터 분석, 성과 개선, 성과 분석, 공정능력, 예측 모형, 품질 인증, 품질 개선, 모수 최적화 등 20개 토픽을 형성하는 것으로 분석되었다. 둘째, 연구가 수행된 연도와 토픽 출현확률간의 회귀분석 결과에 따르면 공정 개선 토픽과 고객 만족 토픽은 연구비중이 감소하는 추세에 있는 cold topic으로 나타났고, 데이터 분석 토픽과 예측 모형 토픽은 연구비중이 증가하는 추세에 있는 hot topic으로 분석되었다. 셋째, 토픽 네트워크에 대한 커뮤니티 분석결과에 따르면 토픽 네트워크는 프로세스 및 제품 품질, 성과 개선, 공정 최적화, 신뢰성과 위험관리, 품질 지표, 공정 능력, 예측 모형 등 7개 커뮤니티를 형성하는 것으로 나타났다.

본 연구에서 제시한 구조적 토픽모델을 이용한 연구동향 분석 방법론의 이론적 기여도를 정리하면 다음과 같다. 첫째, 구조적 토픽모델과 통계적 기법을 적용하여 연구 토픽들을 도출하므로 문헌분석이나 키워드 빈도분석을 이용한 기존의 연구동향 분석 방법에 비해 분석결과에 대한 신뢰도가 높고 연구자의 편견을 줄일 수 있다는 장점이 있다. 둘째, 연구 비중이 증가하는 추세에 있는 연구주제와 연구비중이 감소하는 추세에 있는 연구주제를 제시하므로 연구자가 집중할 연구 분야를 선정하거나 새로운 연구주제를 발견하는데 도움을 제공할 수 있다. 셋째, 본 연구에서 고려한 연도 이외에도 문헌에 포함된 다양한 메타정보를 바탕으로 공변량 분석을 수행할 수 있으므로 기존 방법론에서 분석할 수 없었던 인과관계를 보다 세부적으로 분석할 수 있다. 또한 본 연구에서 제시한 구조적 토픽모델을 이용한 연구동향 분석 방법론의 실무적 기여도를 정리하면 다음과 같다. 첫째, 구조적 토픽모델을 적용하면 기사, 소셜 미디어, 문헌 등의 텍스트 데이터로부터 기업이 관심을 가지고 있는 특정 도메인에 대한 주요 이슈와 트렌드를 파악할 수 있으므로 기업의 전략적이고 효율적인 의사결정을 지원할 수 있을 것이다. 둘째, 소셜 미디어를 통해 생성되는 방대한 양의 텍스트 데이터로부터 고객 의견이나 감정, 이슈를 파악하고, 제품, 서비스, 브랜드, 이벤트 등에 대한 피

드백을 실시간으로 분석하여 활용할 수 있으므로 비용과 시간을 절감할 수 있을 것이다.

본 연구의 한계점과 추후 연구 과제를 정리하면 다음과 같다. 첫째, 본 연구에서는 품질경영분야 연구동향 분석을 위해 국내 저널을 분석대상으로 이용하였지만 제한된 범위의 자료를 이용하였다는 점에서 한계점이 존재한다. 품질경영분야 국내외 대표 저널을 대상으로 분석을 수행한다면 국내외 연구동향과 연구추세에 대한 비교분석이 가능하고 이를 통해 전략적 시사점을 제공할 수 있을 것으로 판단된다. 둘째, 토픽 모델을 적용하는 과정에서 가장 어려운 사항은 토픽명을 부여하는 과정이라고 판단된다. 일반적으로 연구자의 전문지식이나 전문가 협의를 통해 토픽명을 부여하지만 여전히 연구자들의 주관성이 개입될 수 있다는 제약점을 가지고 있다. 따라서 머신 러닝이나 딥러닝 기법 등을 적용한 토픽명 자동할당 방법에 대한 연구도 필요하다고 판단된다.

참고문헌

- 과학기술 지식인프라(ScienceOn), <https://scienceon.kisti.re.kr>.
- 김민준 (2023), 텍스트마이닝을 활용한 품질 4.0 연구동향 분석, *품질경영학회지*, 51(3), 403-418.
- 김성혁, 권상미 (2010), 서비스 품질의 연구 동향 : 1989-2008, *관광연구*, 25(1), 205-223.
- 김연성 (2016), 품질경영학회 50주년 특별호: 서비스품질 분야 연구 리뷰, *품질경영학회지*, 44(2), 265-276.
- 정규석, 신완선, 백동현, 주용준 (2016), 품질경영학회 50주년 특별호: 품질경영 분야 연구 리뷰, *품질경영학회지*, 44(1), 43-60.
- 박만희 (2016), 동시단어분석을 이용한 품질경영분야 지식구조 분석, *품질경영학회지*, 44(2), 389-408.
- 박현아, 이청림, 이유재 (2023), 국내 온라인 서비스 품질 연구에 관한 종합적 고찰: 2001-2021년의 연구를 중심으로, *마케팅 연구*, 38(8), 45-90.
- 이상복 (2019), 텍스트 마이닝 처리로 품질경영학회지 연구동향 분석, *품질경영학회지*, 47(3), 597-613.
- 홍석기 (2016), 서비스경영 연구 동향에 대한 계량서지학적 분석, *서비스경영학회지*, 17(3), 133-149.
- Blei, D. M., Ng, A. Y., Jordan, M. I. and Lafferty, J. (2003), Latent Dirichlet Allocation, *Journal of Machine Learning Research*, 3, 993-1022.
- Carnerud, D. (2018), 25 years of quality management research - outlines and trends, *International Journal of Quality & Reliability Management*, 35(1), 208-231.
- Griffiths, T. and Steyvers, M. (2004), Finding scientific topics, *Proceedings of the National Academy of Sciences*, 101(1), 5228-5235.
- Roberts, M. E., Stewart, B. M. and Airoldi, E. M. (2016), A model of text for experimentation in the social sciences, *Journal of the American Statistical Association*, 111, 988-1003.
- Roberts, M. E., Stewart, B. M. and Tingley, D. (2019), stm: R Package for Structural Topic Models, *Journal of Statistical Software*, 191(2), 1-40.
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B. and Rand, D. G. (2014), Structural topic models for

- open-ended survey responses, *American Journal of Political Science*, 58(4), 1064-1082.
- Siva, V., Gremyr, I., Bergquist, B., Garvare, R., Zobel, T. and Isaksson, R. (2016), The support of quality management to sustainable development: a literature review, *Journal of Cleaner Production*, 138, 148-157.
- Spasojevic, I., Havzi, S., Stefanovic, D., Ristic, S. and Marjanovic, U. (2021), Research Trends and Topics in IJIEEM from 2010 to 2020: A Statistical History, *International Journal of Industrial Engineering and Management*, 12(4), 228-242.
- VOSviewer, Visualizing scientific landscapes, <https://www.vosviewer.com>.
- Wawak, S., Rogala, P. and Dahlgaard-Park, S.M. (2020), Research trends in quality management in years 2000-2019, *International Journal of Quality and Service Sciences*, 12(4), 417-433.
- Zonnenshain, A. and Kenett, R. S. (2020), Quality 4.0-the challenging future of quality engineering, *Quality Engineering*, 32(4), 614-626.

* 저자소개 *

· 박 만 희(mhpark@cup.ac.kr)

성균관대학교 산업공학과에서 박사학위를 취득하였고, 대우정보시스템, KAIST 연구교수, KISTI 선임연구원으로 근무한 경력이 있으며 현재는 부산가톨릭대학교 경영학과 교수로 재직 중이다. 주요 관심분야는 빅데이터 분석, 딥러닝, 품질경영, 최적화 등이다.