

데이터마이닝을 활용한 기업 R&D역량 특성에 관한 탐색 연구*

김상국

한국과학기술정보연구원
(sgkim@kisti.re.kr)

임정선

한국과학기술정보연구원
(jsunnylim@kisti.re.kr)

박 완

과학기술연합대학원대학교
(dare2win@kisti.re.kr)

글로벌 경영환경 변화로 기술개발과 시장니즈의 불확실성이 커지고 기업 간 상호 경쟁이 심화되면서 개별 기업들의 연구개발 활동에 대한 관심과 요구가 증가하고 있다. 이러한 환경변화에 대응하기 위하여 연구개발 기업들은 설비투자에 더욱 신중을 가하면서 연구개발의 질적인 경쟁력을 제고시키기 위한 수단 중 하나로 연구개발 투자를 강화하고 있다. 결과적으로 설비나 연구개발 투자 요소는 연구개발 기업들의 입장에서는 미래 불확실성을 떠안아야하는 부담이 될 수 밖에 없다. 단지 연구개발 역량을 제고시키기 위한 수단으로 연구개발 투자를 증가시키는 경영 전략은 기업성과측면에서 불확실성이 높은 것이 사실이다. 본 연구에서는 데이터마이닝 기법을 활용하여 기업들의 연구개발 역량에 영향을 주는 특성들을 기술경영능력, 연구개발능력, 그리고 기업분류 속성 관점에서 탐색하고 이러한 개별 요인들이 연구개발 역량의 수준에 따라 나타나는 특성들을 탐색하였다. 이를 위해서 국내 연구개발 기업 전체를 대상으로 증거데이터에 근거해 군집분석과 실험결과를 제시하였다. 상기의 3개 관점마다 세부 평가지표를 각각 7개, 2개, 4개로 구성하여 해당 영역에서의 개별적인 수준을 정량적으로 측정하고자 하였다. 기술경영능력과 연구개발능력의 경우 현행 기술력 평가기관들이 주도적으로 활용하고 있는 소항목 평가지표를 참조하였으며, 이때 정량적으로 자료 확보가능인지 여부를 고려하여 최종적인 세부 평가지표를 새롭게 구성하였다. 기업분류 속성의 경우에는 가장 기본적인 기업 분류 프로파일 정보를 고려하여 구성하였다. 특히 연구개발 역량수준의 동질성 파악을 위해서 기술경영능력과 연구개발능력의 세부평가지표를 활용하여 개별기업별 종합점수를 부여하였으며, 이때 역량수준을 5개의 등급으로 분류하여 군집분석 결과와 비교하였다. 분석된 군집과 역량수준 등급과의 비교평가에 따른 의미를 부여하기 위해서 군집별로 연구개발 역량수준이 높은 경향과 낮은 경향이 존재하는 군집들을 탐색하였다. 이후 해당 군집에서 세부 평가지표에 따른 특징들을 분석하였다. 이와 같은 연구수행 방법을 통해 연구개발 역량수준이 높은 군집이 2개, 낮은 군집이 1개로 분석되었으며, 나머지 2개의 군집들은 역량수준이 거의 높은 발생 빈도로 유사하게 나타났다. 결과적으로 본 연구에서는 역량수준이 높은 2개 군집과 낮은 1개의 군집들을 대상으로 세부 평가지표에 따른 개별적 특징들을 분석하였다. 본 연구의 결과가 제시하고 있는 시사점은 기술변화 속도와 시장수요의 변화에 효과적으로 대응할 수 있는 전문 경영자의 교체주기가 짧을수록 연구개발 역량 제고에 기여할 가능성이 높다는 점이다. 개인기업의 경우에 법인기업으로의 전환을 통해 연구개발 인력들의 기업에 대한 소속감을 제고시킴으로써 연구개발 역량의 투입강도를 높일 필요가 있으며, 조직적 측면에서도 팀단위의 조직구성을 통해 책임과 권한의 정확성을 제공할 필요가 있다는 점이다. 기술상용화 실적진수나 기술인증진수는 역량제고에 기여하는 경우와 그렇지 않은 경우 모두 발생되고 있어, 경영자 입장에서 연구개발 역량제고를 위한 중요 인자로 검토하는데 한계가 있는 것으로 확인되었다. 마지막으로 실용신안출원의 경험 여부는 연구개발 역량에 중요한 영향을 미치는 요인으로 파악되어, 연구개발 역량 제고를 위해서는 실용신안출원 장려를 위한 동기부여를 제공할 필요성을 확인하였다. 이처럼 본 연구결과는 개별 기업들의 연구개발 역량 제고를 위한 기업 경영전략의 중요한 시사점을 제공할 수 있을 것으로 기대된다.

주제어 : 데이터마이닝, 군집분석, 연구개발 역량, 기술경영역량, 기업분류속성

논문접수일 : 2020년 11월 4일 논문수정일 : 2021년 1월 28일 게재확정일 : 2021년 2월 2일

원고유형 : 일반논문 교신저자 : 김상국

* 이 논문은 2020년도 한국과학기술정보연구원에서 수행한 주요사업의 연구임(K-20-IA-0302-R).

1. 서론

4차산업혁명과 관련 기술들의 발달과 고객의 다양한 요구사항이 함께 증가하면서 여러 분야에 걸쳐 데이터의 중요성 또한 증가하고 있다. 이러한 기술들 중에서 빅데이터 분석, 데이터마케팅, IoT(Internet of Things), 로봇기술, 인공지능 등이 많은 역할을 수행하고 있다. 생산현장에서도 IoT와 빅데이터 등의 기술을 활용하여 많은 양의 데이터를 유용한 정보로 변환하여 기업의 이익에 기여를 하고 있다.

전 세계가 하나로 연결되는 글로벌 시대가 확산되고 있는 가운데, 기업들은 혁신적인 기술개발의 필요성과 시장니즈에 대응하기 위한 연구개발을 위하여 핵심역량 개발에 집중하고 있다. 이러한 기업의 궁극적 목표는 이윤극대화에 있으며, 기업이 보유하고 있는 내부역량과 외부환경이 조화를 이루면서 추진될 때 원활하게 달성될 가능성이 높다. 일반적으로 기업역량은 기업이 활동하는 모든 분야에서 창출 가능하나, 혁신적인 기술개발을 핵심역량으로 여기는 기업들의 입장에서는 자체 보유중인 기술역량을 중심으로 연구개발(R&D)역량과 내부역량으로부터 만들어지고, 이러한 결과가 기업성으로 이어지는 것으로 볼 수 있다. 특히, 혁신과 관련하여 기업의 R&D 활동은 혁신을 창출하는 중요한 원천이며 혁신과 관련된 가장 중요한 기업 활동 중 하나이다(Souitaris, 2002). Burgelman et. al(2004)은 기술혁신 역량을 기업의 기술혁신을 촉진하고 지원하는 일련의 기업의 특성으로 정의하였다. 또한 많은 선행연구들(Christensen, 1995; Cooper, 1996; Bobe and Bobe, 1998; Burgelman et al., 2004; Yam et al., 2004)에서는 기술혁신 역량의 영역으로 외부로부터 지식을 발견·흡수·활용하

는 학습역량과 연구개발 전략, 프로젝트 실행, 프로젝트 포트폴리오 관리, 연구개발 지출 등을 통합하는 능력인 연구개발 역량, 자원배분 역량, 생산 역량, 마케팅 역량, 조직 역량, 전략계획 역량을 제시하였다.

특히 김문선(2012)은 기업의 기술혁신 역량이 기업 내에서 이루어지는 직접적인 기술 활동뿐만 아니라, 이와 관련된 주변자원 활용 등을 광범위하게 포함하고 있는 것으로 분석하였다. 이와 함께 기술혁신 역량이 기업성공에 미치는 영향에 관한 분석을 수행하면서 매출액 규모와 수익성 같은 재무적 성과와 목표달성도, 시장점유율, 고객만족도, 서비스 품질 향상 정도와 같은 비재무적 성과로 구분하여 기업의 성과를 측정하였다.

최근 급변하는 경영환경과 시장의 수요를 창출할 수 있는 혁신제품을 개발해야만 하는 수요자 중심의 시장 환경에서는 기업의 R&D역량이 더욱 중요한 것으로 인식되고 있으며, R&D를 통한 기술혁신은 기업의 경쟁력과 성장의 핵심요소로 여겨지고 있다. 기업의 핵심역량을 형성하는데 있어서 중요성이 더욱 확대되고 있는 R&D 투자는 기업의 성과창출 측면에서 극복해야 될 불확실성이 크다. 일반적으로 기업 R&D투자에 따른 기업성과의 선순환 구조는 기본적으로 연구개발, 제품개발, 마케팅 등의 성공을 전제로 하는 것이며, 이러한 사업화 단계별 불확실성을 극복하기 위한 가능성은 지극히 낮은 것이 현실이다.

본 논문에서는 기업들의 R&D역량을 제고시키기 위한 수단으로 R&D투자를 증가시켜야 하는 위험을 최소화하기 위한 차원에서, R&D역량 제고에 영향을 미치는 기술경영능력, R&D능력, 그리고 기업 속성에 관한 영향요인들을 데이터

마이닝 기법을 활용하여 탐색하고 궁극적으로 기업 R&D역량 제고를 위한 전략적 시사점을 제시하고자 한다. 특히 기술경영능력 및 연구개발 능력은 기업보유 기술력을 평가하는 평가지표 중 하나로, 현재 국내 기술평가기관들이 공통적으로 사용하고 있는 평가지표이다.

본 연구에서는 국내 개별기업들의 R&D역량 수준을 기술경영능력과 연구개발능력의 세부평가지표를 구성하고, 평가지표별 최근 5년 동안의 국내 시장에서 형성된 증거데이터를 적용하여 기술개발 기업들의 상대적 R&D역량 수준을 분석하였다. 이러한 R&D역량 수준을 개별기업별로 레이블링(labeling) 하고, 기술경영능력과 연구개발능력의 평가속성에 기업 속성을 추가하여 R&D역량 특성을 탐색하기 위한 속성으로 활용하였다. 해당 속성들을 대상으로 최근 10년 동안 개별 기업들의 증거데이터를 구성하고, R&D역량 특성을 탐색하기 위한 군집분석(clustering analysis)을 수행하였다. 군집분석은 표본들간의 유사도에 기반하여 동질성이 높은 표본들을 하나의 그룹으로 분류하는 탐색적 분석을 말한다(Wilks, 2011).

이외에도 기술경영능력 및 연구개발능력 중심의 증거데이터를 기반으로 하는 R&D역량 수준 분류와 기업 속성이 추가된 상태에서의 군집분석의 결과가 비슷한 결과가 도출되는지에 대해서도 비교분석을 시행하였다.

2. 선행연구 동향

2.1 R&D역량 분석방법

R&D역량이 기업성이나 기술혁신 성과, 그리

고 기술사업화 성과 등에 미치는 영향에 관한 선행연구는 다수 존재하고 있다.

(Yoon, 2018) 해당 논문에서는 국내 중소기업의 연구개발 및 기술혁신 역량과 기술혁신 지원제도 활용이 기업 성과에 미치는 영향을 확인하면서, 연구개발 및 기술혁신과 관련된 여러 역량 중에서 집중적으로 활용해야 할 역량과 정부가 보완해야 할 지원제도를 분석하였다. (Kim and Kim, 2017) 국내 양재지역에 소재하고 있는 ICT 기업들을 대상으로 R&D활동을 조사 분석하면서 R&D역량 강화를 위한 방안에 대해서 탐색연구를 수행하였다. (Song, 2013) 국내 수출 중소기업 내부역량이 기업 R&D성과에 미치는 영향에 관한 연구에서, 내부 R&D역량이 경제적/사회적/기술적 성과에 미치는 영향을 다중회귀분석을 활용하여 분석하였다.

(Jeong, 2015) 논문은 국내 정부출연연구원과 기술협력을 하고 있는 중소기업만을 대상으로, R&D역량 및 기술사업화역량이 기술혁신성과에 미치는 영향을 분석하였다. 이 연구에서는 R&D역량, 기술사업화역량, 기술혁신성과를 대상으로 21개 변수를 구성하고 가설검증을 위한 요인분석과 다중회귀분석을 수행하였다. (Choi, 2012) 논문에서는 정보기술 관련 창업기업들을 대상으로 연구개발 역량과 경영성과와의 관계에 관한 연구를 통해 재무적/비재무적 성과와의 상관성을 찾는 연구를 구조방정식 모형을 활용하여 수행하였다. (Shin and Park, 2003) 연구는 전략경영이론에서 강조되어온 새로운 대안적 역량의 탐색(exploration)과 기존 역량의 활용(exploitation)이라는 조직학습의 두 가지 유형을 신제품개발과정에 적용하여 이 두 가지 서로 다른 유형의 개발활동들이 R&D성과에 어떤 영향을 미치는지를 다중회귀분석을 통해 실증 분석하였다. (Jung,

2014) 연구는 생명공학기업의 R&D혁신역량과 기업성장에 관한 연구에서 기업의 대표적인 성과변수인 수익성과의 관계를 회귀분석 방법을 활용하여 실증 분석하였다. (Shon and Park, 2013) 연구는 각 기업의 혁신(R&D)활동에 의한 산업별 특징 차이와 산업별 혁신(R&D)활동의 비교우위 행태에 관한 분석을 수행하였다. 이 논문에서는 R&D역량을 측정하기 위해서 투입 요소 측면과 결과 측면으로 구분하여 변수를 조작적으로 정의하였으며, 투입요소는 인력, 전담 조직, 경험 및 지식, 재무투자, 기술협력관계에 관한 변수를 정의하였고, 결과변수는 기술도입, 기술인증/성과, 기술상품화 유무에 대한 변수로 정의하여 분산분석(ANOVA)과 다중비교분석(DUNCAN, LSD, TUKEY, SCHEFFE)을 활용하여 분석하였다. (Shim, 2016) 연구는 기술혁신역량이 기업의 생산성에 미치는 영향에 관한 연구에서 기술혁신역량을 R&D역량과 외부네트워크 역량으로 구분하고, 기업의 생산성은 콥-더글라스(Cobb-Douglas) 생산함수 이론을 활용하여 회귀분석을 실시하였다. (Park et al., 2011) 논문은 국내 기업의 하도급 분류에 따른 기술개발역량 차이를 분석하기 위해 산업특성과 기업규모를 통제하면서 기업패널 자료를 통한 분산분석을 수행하여 하도급 분류와 기술개발역량 차이를 분석하였다. (Zhu, 2013) 연구는 기업의 기술혁신역량과 기술사업화역량이 기술성과와 경영성과에 미치는 영향을 회귀분석과 분산분석을 통해 분석하면서, 기술혁신 및 사업화 역량이 기술 성과에 영향을 미치고, 기술성과는 경영성과에 영향을 미친다는 사실을 밝혀냈다. (Hwang and Sung, 2015) 논문은 기술도입 102개 기업을 대상으로 설문조사를 통해 기술사업화역량과 연구개발역량이 혁신성과에 긍정적 영향을 미친다는

사실을 밝혀냈다. 또한 시장변동성은 기술사업화역량과 혁신성과간의 관계를 조절하는 효과로 작용한다는 사실을 추가로 제시하였다. (Park, 2008) 연구에서는 정부출연연구기관들을 대상으로 R&D역량 요인이 기술혁신 성과에 미치는 영향에 관해서 구조방정식 모형을 활용하여 분석하였다. 특히 R&D수행역량을 R&D수주역량, 실행역량, 결과활용역량으로 세분화하여 기술혁신 성과에 미치는 영향을 밝히고자 하였다. (Kim and Yang, 2011) 연구는 국내 게임산업의 CT R&D 성과지표의 타당성을 도출하여 권역별 게임산업 관련 기업들의 경쟁력을 분석하였다. 경쟁력 분석을 위해 기업체수, 종업원수, 매출액에 기초한 입지계수(Location Quotient) 분석과 R&D역량 분석을 권역을 대상으로 수행하였으며, R&D역량의 비교우위 지역을 탐색하면서, 해당 지역에서의 R&D역량 경쟁력이 기업군 네트워크와 기술력을 지원하는 R&D기관의 상호호혜적인 생태계에 기인한다는 사실을 밝혔다.

2.2 데이터마이닝 기반 R&D역량 분석방법

데이터마이닝 기법들 중 대표적인 비지도학습 방법인 군집분석 방법은 동질적인 표본끼리는 동일한 군집에 배치시키고, 이질적인 표본들 간에는 서로 다른 군집에 속하도록 할당한다. 결과적으로 분석결과로 분류되는 각 군집들은 그룹 내에서는 높은 동질성과 유사성을 보이게 된다 (Berry and Linoff, 1997). 이러한 군집분석의 목적은 데이터들 간의 구조적 특성을 파악하기 위하여 주로 사용되며, 군집화를 통하여 복잡한 데이터들을 몇 개의 군집으로 나타낼 수 있는 장점이 있다. 더불어 데이터 집단간의 대표적인 특성이나 유형을 파악하기 용이한 강점이 있다

(Bishop, 2006). (Yoo and Park, 2020) 논문은 시계열 군집분석 방법을 이용하여 디지털 음원의 순위 변화 패턴을 분류하였다. 이 연구에서는 16가지 패턴(분류)을 탐색하여 최종적으로 2가지 유형의 대표적인 패턴을 확인하였다. 결과적으로 군집분석을 활용하여 분류 속성들의 특징을 탐색하고 점차적으로 대표적인 분류 특징들을 축소해 가는 과정을 통해 음원의 순위변화 패턴을 찾아냈다.

R&D역량 분석과 관련하여 국가 R&D과제정보를 활용하여 지역 단위의 R&D역량 강화에 필요한 특성을 텍스트 어넬리티스(text analytics)와 네트워크 분석(network analysis)과 같은 데이터마이닝 기법을 활용하여 탐색하였다(Choi and Lee, 2020). 이 연구에서는 충북지역의 국가 R&D과제 정보를 활용하여 스마트IT부품, 바이오헬스, 수송기계소재부품 관점에서 연구수행주체별 R&D 특성을 파악하였으며, 네트워크 분석을 통해 기술개발 R&D별 국가과학기술분류 간에 상호관련성을 분석하였다.

이처럼 기존 선행연구들은 기술혁신 역량이나 연구개발 역량이 기업의 성과에 미치는 상관성을 파악하기 위한 다양한 시도가 많았으며 분석 방법으로는 회귀분석, 구조방정식 모형, 분산분석, 다중비교분석 등 인과관계를 해석하기 위한 전형적 분석방법을 주도적으로 많이 활용하였다. 특히 R&D역량 특성이 다양한 기업의 성과에 미치는 영향 관계를 분석하기 위한 시도도 있었지만, 관련 연구 역시 전통적인 분석 방법을 활용한 것이 주도적이었다. 최근 선행연구와 같이 국가 연구개발 과제 정보인 증거데이터를 활용하여 권역별 R&D 특성을 탐색하기 위한 시도가 있었으며(Choi and Lee, 2020), 이 연구에서 활용한 방법이 기존의 전통적인 분석 방법이 아

닌 데이터마이닝 기법을 활용한 시도인 점에서 본 연구에서 수행하고자 하는 방법론과 유사성이 높다. 하지만 이 연구의 경우에는 R&D역량 특성을 기업 관점이 아닌 권역별 지역을 중심으로 한 연구로서 본 연구와의 차별성이 존재한다. 특히, 본 연구에서는 R&D역량에 관한 속성 정보를 기술경영능력과 연구개발능력, 그리고 기업 분류 속성에 관한 세부 평가항목들을 구성하고, 이에 대한 최근 10년 동안의 증거자료를 연구개발 중심 기업 단위별로 수집하였으며, 세부 평가항목들이 R&D역량에 미치는 영향을 군집분석 방법을 활용하여 탐색하고자 한 점이 가장 큰 차별점이다.

3. 연구 방법론

3.1 R&D역량 평가지표 구성

최근 국내에서 운용중인 여러 기술신용평가기관(TCB, Tech Credit Bureau)들은 기업이 보유한 기술을 대상으로 기술력평가와 함께 기업신용평가를 수행하고 있다. 다수의 기술평가기관들은 기술력평가를 위한 평가지표로서 기술경영 및 연구개발 능력, 기술성, 시장성, 사업성 평가항목을 공통적으로 활용하고 있다. 기술경영 및 연구개발능력은 기업의 전반적인 관리주체인 경영자와 연구개발인력 등에 관한 수준을 평가하기 위함이다. 이 평가항목을 첫 번째 평가항목으로 사용하는 이유는 동일 기술일지라도 그 기술을 통해 투입요소를 산출물로 전환하는 과정에서 활용주체인 기업의 능력에 따라 상이한 결과나 성과를 도출할 수 있기 때문이다.

따라서 본 연구에서는 R&D역량을 분석하기

위해서 기술평가기관들이 공통적으로 활용하고 있는 기술경영능력, 연구개발능력 관련 평가지표를 참조하여 구성하고, 이외에도 기업 단위의 R&D역량 특성을 탐색해야 하는 분석 목적상 기업분류 속성을 추가하였다. 세 가지 평가항목별 세부 평가지표는 아래 표와 같이 구성하였다.

(Table 1) Evaluation items and detailed evaluation indicators

Evaluation item	Detailed evaluation index
Technology management capability	Representative training in the same industry
	Technical Development Department Operating Years
	Number of patent registrations
	Number of patent applications
	Whether or not utility model is registered
	Whether or not utility model is applied
R&D capability	R&D expense ratio
	Number of technical commercialization achievements
Corporate classification attribute	Number of technology certifications and awards
	Company scale
	Industry classification
	Business type classification
	Employee size

기술경영능력과 연구개발능력에 관한 세부 평가지표 선정 근거는 기술력 평가기관들이 주로적으로 활용하고 있는 소항목 평가지표들을 근거로 정량적 자료획득이 가능한 지표로 일부 수정하여 구성하였다. 기업분류 속성의 경우에는 가장 기본적인 기업분류 프로파일 정보를 고려하여 구성하였다.

3.2 문제의 정의

기본적으로 다양한 특성 요인들에 의해 영향을 받는 결과에 관한 인과관계 분석을 수행하는 전통적 방법론으로 회귀분석, 로지스틱분석, 구조방정식 모형 등 다양하게 존재한다. 본 연구에서는 R&D역량 특성으로 설명될 수 있는 평가항목을 사전에 설정한다. 이후 각각의 평가항목에 따른 과거 자료를 바탕으로, 자료 자체에서 패턴을 찾을 수 있는 정량적 분석 방법인 데이터마이닝을 통해 R&D역량 특성 탐색을 위해서 군집화(clustering)를 수행하고 R&D역량을 예측한다. R&D역량 수준은 기술경영능력과 연구개발능력에 의해 결정될 수 있다는 전제 하에서 평가항목의 세부 평가지표들에 대한 속성값들의 종합 결과로 결정될 수 있다. 이러한 가정을 전제로 기업 단위별 세부평가항목 속성값의 종합점수를 기준으로 개별기업들의 상대적인 R&D역량 수준을 할당할 수 있다. 이러한 개념적 이해를 바탕으로, 개별 기업별 R&D역량 수준을 레이블링(labeling)하고, 이렇게 설정된 R&D역량수준별 그룹과 R&D역량 세부 평가지표에 따른 과거 자료를 바탕으로 군집분석을 수행하여 R&D역량 수준을 예측하고 분류한다. 이때 R&D역량 수준은 5점 척도(매우 낮다-낮다-보통이다-높다-매우 높다)로 구분하여 분석한다. R&D역량 수준은 기술경영능력과 연구개발능력의 세부 평가지표들을 대상으로 min-max 정규화를 통해 도출된 개별 평가지표들의 점수를 합산하여 부여한 후, 모든 기업들의 종합점수들에 대한 순위통계량(order statistics)을 근거로, 20백분위, 40백분위, 60백분위, 80백분위, 100백분위를 5점 척도의 기준으로 활용하였다.

이어서 데이터마이닝 기법을 활용하여 R&D

역량 특성인 세부 평가지표별 정보를 바탕으로 일정한 패턴과 규칙을 탐색한다. 결과적으로 평가지표의 종합 점수에 따른 R&D역량 수준과 군집분석을 통해 도출된 군집들(clusters)의 특성을 상호 비교하여, 종합 점수에 근거한 R&D역량 수준이 군집분석에 의한 군집들(cluster)과 어떠한 유사 특성이 존재하고 있는지를 검토한다.

3.3 데이터마이닝 모형

3.3.1. 데이터

데이터는 최근 10년(2008~2017) 동안의 국내 연구개발 중심 35,398개 기업데이터를 대상으로, <Table 1>에 제시한 기술경영능력 7개, 연구개발 능력 2개, 기업분류 속성 4개 변수를 구성하여 총 13가지 변수를 사용하였다. 해당 변수에 관한 관측 데이터를 모두 포함하고 있는 경우만 선별하여 12,761개 기업의 데이터를 활용하였다. 각각의 변수를 살펴보면 다음과 같다.

실용신안등록 유무와 실용신안출원 유무 변수는 범주형으로 정의하면서 출원 및 등록건수가 없는 경우에 0, 있는 경우에 1로 설정하였다. 기업규모는 범주형으로 정의하면서, 대기업 1, 중기업 2, 소기업 3, 한시성중소기업 4, 중견기업 5, 소상공인 6, 보호대상중견기업 7, 판단제외 8로 설정하였다. 업종분류도 범주형 변수로 정의하면서, 한국표준산업분류(KSIC) 중분류 업종(2자리 기준)도 A01업종에서부터 S96업종까지 68개의 분류를 순차적으로 값을 할당하였다. 기업 유형 분류도 범주형 변수로 정의하면서 개인 1, 일반법인 2, 비영리법인 3, 공공기관 4로 설정하였다.

대표자의 동업종 종사연수, 특허등록건수, 특허출원건수, 기술상용화실적 건수, 기술인증 및

수상실적 보유 건수, 보유 종업원 규모 변수들은 이산형 변수로 정의하면서 해당 변수가 갖을 수 있는 변수값들을 min-max 정규화를 적용하였다.

R&D역량등급은 범주형 변수로 정의하면서 35,398개의 개별기업 데이터를 대상으로 기술경영능력과 연구개발능력 평가항목의 세부지표만을 적용하여 역량수준을 분석하고 레이블링하였다. 레이블링 방법은 5점 척도를 사용하여 매우나쁨 1, 나쁨 2, 보통 3, 좋음 4, 매우좋음 5로 점수를 부여하였다.

3.3.2. 데이터특성

먼저 12,761개의 개별 기업들을 대상으로 R&D역량 탐색을 위한 13가지 속성(attributes)과 R&D역량 수준을 3개/5개의 클래스(class)로 분류하였으며, 각각의 변수의 특성에 따라 범주형과 연속형으로 구분하였다. 연속형 변수는 정규화하여 0 ~ 1의 숫자로 표기하였다.

<Table 2> Data structure

Detailed index	Role	Type	Range of values	Range after normalization
Representative training in the same industry	Property	Discrete	0~65	min-max
Technical Development Department Operating Years	Property	Continuous	-0.503~55.772	min-max
Number of patent registrations	Property	Discrete	0~32137	min-max
Number of patent applications	Property	Discrete	0~114,931	min-max

Detailed index	Role	Type	Range of values	Range after normalization
Whether or not utility model is registered	Property	Categorical	0, 1	0, 1
Whether or not utility model is applied	Property	Categorical	0, 1	0, 1
R&D expense ratio	Property	Continuos	-0.286~110.868	min-max
Number of technical commercialization achievements	Property	Discrete	0 ~ 147,068	min-max
Number of technology certifications and awards	Property	Discrete	0 ~ 100	min-max
Company scale	Property	Categorical	1~8	1~8
Industry classification	Property	Categorical	A01~S96 (68)	A01~S96 (68)
Business type classification	Property	Categorical	1~4	1~4
Employee size	Property	Discrete	1~103,929	min-max
R&D Capability	Class	Categorical	1~5	5-point scale

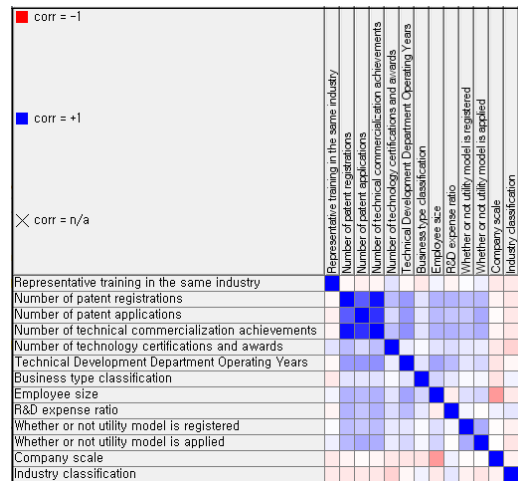
R&D역량 수준의 데이터 분포는 5점 척도의 경우 20백분위, 40백분위, 60백분위, 80백분위, 100백분위를 적용하였기 때문에 클래스 등급간 발생빈도는 기본적으로 균일한 분포(uniform distribution)를 만족한다.

4. 데이터마이닝을 활용한 R&D역량 특성 탐색

데이터마이닝을 이용하여 대표자 동업종 종사

연수, 기술개발전담부서 운영연수 등으로 구성된 총 13가지 변수와 R&D역량수준 간의 패턴과 규칙을 탐색한다. 변수는 기술경영능력, 연구개발능력, 기업분류의 세부 평가지표로, 최근 10년간의 기업정보와 R&D역량 수준에 대한 레이블링 분류 정보이다. 본 연구에서는 분류모형인 군집분석(clustering analysis) 방법을 적용하였다. 분류모형은 비슷하거나 관련된 데이터끼리 그룹으로 묶어주는 것으로 각 군집별로 특성을 파악하여 분류하는 것을 의미한다. 특히 본 연구에서는 군집(cluster)의 수를 사전에 설정하여 분석이 가능한 K-Means 방법을 활용하였다.

기본적으로 속성의 개수와 클래스의 개수, 군집분석 기법의 조건을 달리하여 분석을 수행하였다. 본 연구에서 구성한 13가지 속성에 대해서 상관관계 분석을 통해 획득한 정보를 근거로, R&D역량 수준 클래스와의 관계와 R&D역량 특성 변수들 간의 균형을 고려하여 속성 13가지, 11가지, 10가지를 사용하여 각각 분석을 진행하였다.



(Figure 1) Correlation analysis between 13 attributes

상관관계 분석결과를 참조하여, 특허등록건수와 특허출원건수를 제외한 11가지 속성 시나리오, 그리고 추가적으로 종업원규모를 제외한 10가지 속성 시나리오를 대상으로 추가적인 군집 분석을 수행하였다. 군집의 수는 R&D역량 수준의 범주에 맞춰 각각 5와 3으로 설정하여 분석하였다. 하지만 본 연구결과에서는 역량 수준 5개의 클래스와 13가지의 속성을 기반으로 분석한 결과를 제시하는 것이 보다 많은 정보를 제공할 수 있는 장점이 존재하여 이에 대한 분석결과를 제시하였다.

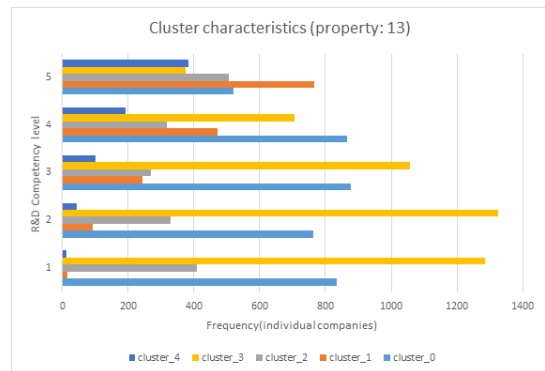
<Table 3> The detail of correlation analysis

Correlation analysis	Representative training in the same industry	Number of patent registrations	Number of technical commercialization achievements	Number of technical commercialization achievements	Number of technology certifications and awards	Technical Development Department Operating Years	Business type classification	Employee size	R&D expense ratio	Whether or not utility model is registered	Whether or not utility model is applied	Company scale	Industry classification
Representative training in the same industry	1.000	-0.027	-0.057	-0.040	0.112	-0.029	-0.087	0.041	-0.045	0.013	0.062	-0.087	-0.087
Number of patent registrations	-0.027	1.000	0.642	0.954	0.268	0.398	0.108	0.302	0.299	0.260	0.280	-0.044	-0.094
Number of patent applications	-0.057	0.642	1.000	0.813	0.153	0.409	0.100	0.272	0.231	0.219	0.335	-0.016	-0.092
Number of technical commercialization	-0.040	0.954	0.813	1.000	0.253	0.436	0.115	0.306	0.305	0.263	0.319	-0.037	-0.095
Number of technology certifications and awards	0.112	0.268	0.153	0.253	1.000	0.074	0.025	0.109	0.136	0.085	0.077	-0.069	-0.175
Technical Development Department Operating Years	-0.029	0.398	0.409	0.436	0.074	1.000	0.110	0.362	0.266	0.104	0.157	-0.102	-0.030
Business type classification	-0.087	0.108	0.100	0.115	0.025	0.110	1.000	0.170	0.090	0.021	0.034	-0.110	0.052
Employee size	0.041	0.302	0.272	0.306	0.109	0.362	0.170	1.000	-0.072	0.124	0.170	-0.400	-0.084
R&D expense ratio	-0.045	0.299	0.231	0.305	0.136	0.266	0.090	-0.072	1.000	0.035	-0.006	0.060	0.099
Whether or not utility model is registered	0.013	0.260	0.219	0.263	0.085	0.104	0.021	0.124	0.035	1.000	0.328	-0.012	-0.054
Whether or not utility model is applied	0.062	0.280	0.335	0.319	0.077	0.157	0.034	0.170	-0.006	0.328	1.000	-0.004	-0.101
Company scale	-0.087	-0.044	-0.016	-0.037	-0.089	-0.102	-0.110	-0.400	0.060	-0.012	-0.004	1.000	-0.053
Industry classification	-0.087	-0.094	-0.092	-0.095	-0.175	-0.030	0.052	-0.084	0.099	-0.054	-0.101	-0.053	1.000

4.1 군집별 속성과 군집의 특성

군집특성을 분석하기 위하여 R&D역량 수준에 따른 군집별 발생빈도를 분석하였으며, <Figure 2>에서처럼 속성이 13개인 경우에, 역량 수준이 매우중음(5) 수준은 5개의 군집이 고르게 분포되는 양상을 보였으며, 그 중에서도 cluster_1이 약간 높은 빈도로 발생하는 양상을 보였다. 역량 수준이 중음(4) 수준은 5개의 군집이 고르게 분포되면서, cluster_0과 cluster_3이 상대적으로 높은 빈도로 나타났으며, 상대적으로

낮은 발생빈도로 cluster_1, cluster_2, cluster_4 순으로 발생되었다. 역량 수준 보통(3)은 상대적으로 상당히 높은 빈도로 cluster_3, cluster_0 순으로 발생되었고, 이에 비해 상당히 낮은 빈도로 cluster_2, cluster_1, cluster_4가 순으로 발생되었다. 역량 수준 나쁨(2)과 매우나쁨(1)은 상대적으로 상당히 높은 빈도로 cluster_3, cluster_0순으로 발생되었고, 이에 비해 상대적으로 상당히 낮은 발생빈도로 cluster_2, cluster_1, cluster_4 순으로 발생되었다. 역량 수준 나쁨(2)과 매우나쁨(1)의 차별점은 cluster_1과 cluster_4가 확연히 상대적으로 낮게 발생되고 있다는 점이다. 속성의 가지수가 각각 11개, 10개의 경우에도 비슷한 양상으로 분석되었다. 종합해보면, 역량 수준이 나쁠수록 발생빈도가 감소하고 좋을수록 증가하는 군집은 cluster_1과 cluster_4로 분석되었다. 이와는 반대로 cluster_3의 경우에는 역량 수준이 높을수록 감소하고, 낮을수록 증가하는 경향을 띠는 것으로 분석되었다. cluster_2의 발생빈도는 역량 수준이 보통(3)에서 상대적으로 가장 낮은 발생빈도를 보여주고 있으며, 이를 기준으로 매우나쁨과 매우중음 단계로 갈수록 점진적으로 증가하고 있는 대칭적 구조의 발생빈도 양상을 보여

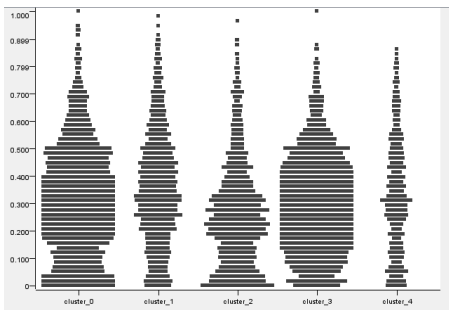


<Figure 2> Frequency of occurrence by cluster

줬다. 마지막으로 cluster_0은 역량 수준이 매우 나쁨(1), 나쁨(2), 보통(3), 좋음(4)에서 거의 비슷한 수준의 발생빈도를 보였고, 매우좋은 수준에서 다른 역량 수준들의 절반 정도수준에 해당되는 낮은 발생빈도를 보였다.

R&D역량 수준 5개 범주와 비교분석을 위하여 5개 군집(cluster_0 ~ cluster4)을 생성하여 분석한 결과는 다음과 같으며, 군집분석 결과인 다음의 그림들에서 x축은 분석된 군집을 나타내고, y축은 세부 평가지표들에 대한 min-max 정규화 점수를 나타내며, 군집별 너비는 정규화 점수의 발생빈도를 표현한다.

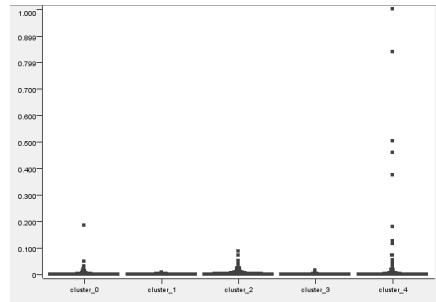
먼저 5개 군집과 대표자 동업종 종사연수에 대한 비교결과, 모든 군집들에서 안정적인 발생구조 패턴을 띄고 있는 것으로 나타났다. 특히 cluster_0과 cluster_3이 상대적으로 오랜 기간 동안 안정적으로 동업종에서 종사하고 있는 특성을 보였으며, 이와는 상대적으로 cluster_4와 cluster_1이 대표자의 교체가 빈번하게 교체되고 있는 양상을 보였다.



〈Figure 3〉 CEO, same industry work experience

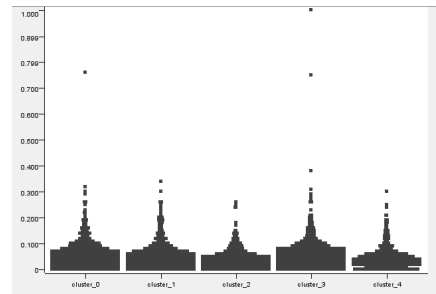
군집별 특허등록건수에 대한 비교분석결과, cluster_1과 cluster_3은 상당히 낮은 출원등록건

수의 양상을 보였고, cluster_0과 cluster_2는 비슷한 수준에서 cluster_1 및 cluster_3 보다 약간 상회하는 수준의 출원등록건수를 보였다. 이와는 반대로 cluster_4는 높은 출원등록건수를 보유한 소수의 기업들이 분포하고 있는 것으로 나타났다.



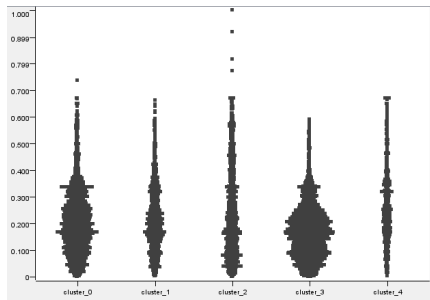
〈Figure 4〉 Number of patent registrations

군집별 특허출원건수와 기술상용화건수의 발생양상도 상기 특허등록건수의 발생양상과 거의 비슷한 수준으로 나타났다. 기술인증건수의 경우에는 모든 군집에서 전반적으로 작은 건수의 기술인증 실적을 보였으며, 상대적으로 소수이긴 하지만 많은 건수의 기술인증실적을 보유한 기업들이 존재하고 있는 군집은 cluster_3으로 나타났다. 다음으로 cluster_0, cluster1, cluster_4, cluster_2 순으로 나타났다.



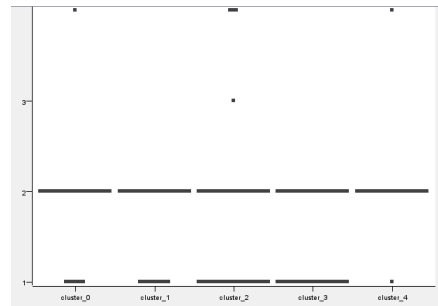
〈Figure 5〉 Number of technology certifications

기업부설연구소나 연구개발전담부서의 운영 유지기간의 경우에 cluster_2는 짧은기간에서부터 오랜 기간에 이르기 까지 비교적 발생빈도가 낮게 나타났다. 나머지 군집들에서는 가장 오랜 기간의 70% 수준 정도에서 연구개발 전담부서를 유지하고 있는 것으로 나타나고 있으며, 다만 cluster_3, cluster_4는 대략 35백분위 까지는 상대적으로 높은 발생빈도로 짧은 기간동안의 연구개발 전담부서를 유지해오고 있는 것으로 나타났다.



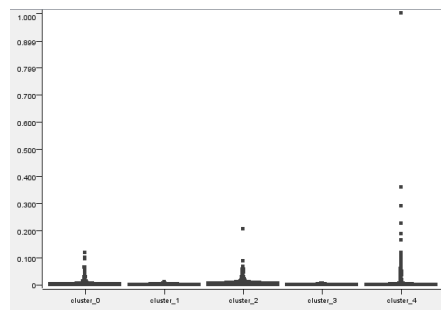
〈Figure 6〉 Operation period of R&D department

기업유형분류는 일반법인(2)이 모든 군집들에서 거의 동일한 수준의 발생빈도를 보이고 있다. 개인기업(1)은 cluster_4에서 가장 낮은 발생빈도를 보이고 있으며, cluster_2와 cluster_3에서는 상대적으로 높은 발생빈도를 보이고 있다. 이에 반해 cluster_0과 cluster_1은 상대적으로 낮은 발생빈도를 보이고 있는 것으로 나타났다. 비영리법인(3)은 cluster_2에서 아주 낮은 빈도를 보이고, 나머지 cluster들에서는 발생되지 않았다. 공공기관(4)은 cluster_0, cluster_4, cluster_2에서 상대적으로 아주 낮은 발생빈도를 보이고 있으며, cluster_1과 cluster_3에서는 발생되지 않았다.



〈Figure 7〉 Enterprise type classification

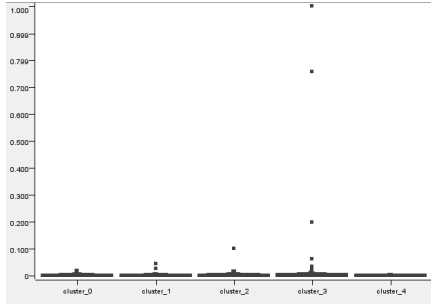
보유 종업원 규모는 cluster_4가 상대적으로 많은 규모를 보유하고 있는 소수의 개별기업들이 존재하고 있는 것으로 나타났으며, cluster_1과 cluster_3은 상당히 작은 규모의 종업원을 보유하고 있는 개별기업들이 다수 존재하고 있는 것으로 나타났다. cluster_0은 cluster_1과 cluster_3을 약간 상회하는 수준에서의 종업원을 보유하고 있는 소수의 기업들이 존재하고 있는 것으로 나타났다.



〈Figure 8〉 Employ holding scale

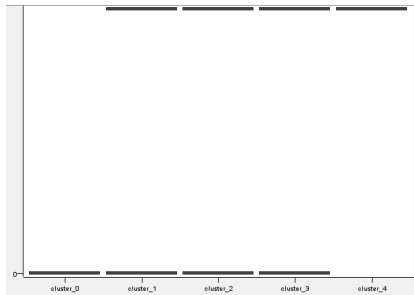
연평균 연구개발비 비율은 모든 군집들에서 전반적으로 낮은 비율을 나타내고 있었으며, cluster_3이 발생빈도는 높지 않지만 연구개발비 비율이 큰값을 보유하고 있는 소수의 기업들이

존재하고 있는 것으로 분석되었다.



〈Figure 9〉 Average R&D cost ratio

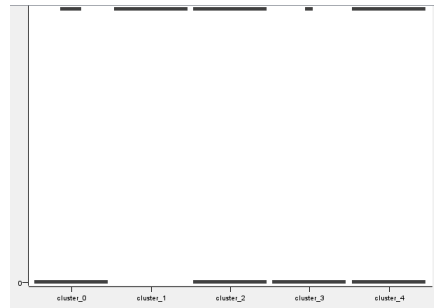
실용신안등록 경험 유무는 cluster_1, cluster_2, cluster_3에서 모두 거의 동일 비율로 나타나고 있으며, 다만 cluster_0은 등록한 경우는 존재하지 않고 등록하지 않은 경우만 있었으며, cluster_4는 등록하지 않은 경우는 존재하지 않고 등록한 경우만 나타났다.



〈Figure 10〉 Existence of utility model registration experience

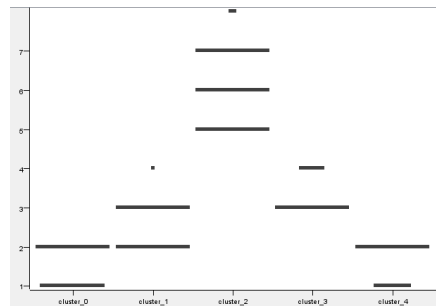
실용신안출원 경험 유무는 cluster_2와 cluster_4가 거의 비슷한 비율로 출원 경험과 출원 무경험 양상을 보였으며, cluster_1은 단지 출원 경험만 일정수준 발생빈도로 나타났고 출원 무경험은 존재하지 않았으며, cluster_0과 cluster_3은 상호

비슷한 양상을 보이면서 cluster_0이 출원 경험 대비 출원무경험의 비율이 2배 이상으로 나타났다. 비슷하게 cluster_3은 출원 경험 대비 출원 무경험의 비율이 5배 이상 훨씬 높은 수준으로 나타났다.



〈Figure 11〉 Existence of utilit model application experience

기업규모는 cluster_0은 거의 비슷한 비율로 대기업(1)과 중기업(2) 유형을 나타내고 있으며, cluster_1은 중기업(2)과 소기업(3)이 비슷한 비율로 나타나면서 한시성중소기업(4)도 아주 작은 규모로 발생되고 있는 양상을 보였다. cluster_3은 중견기업(5), 소상공인(6), 보호대상중견기업(7) 일정수준 이상의 규모로서 거의 비슷한 비율로 발생되고 있다. cluster_3은 소기업(3)과 한시



〈Figure 12〉 Type of company scale

성중소기업(4)이 약간 상이한 발생비율로 나타났으며, cluster_4는 cluster_1과 비슷하게 대기업(1)과 중기업(2) 근소한 차이의 발생비율로 일정 수준 존재하고 있는 것으로 나타났다. 결과적으로 기업규모는 cluster_2를 기준으로 대칭적 구조의 양상을 보였다.

4.2 군집별 특성과 R&D역량 특성 비교

기술경영능력과 연구개발능력 중심의 증거데이터를 활용한 R&D역량수준 레이블링 결과와 기업분류 속성을 추가하여 구성된 13개의 R&D역량 속성의 특성들을 비교분석하기 위하여 비지도학습(unsupervised learning) 방법으로 데이터를 분석하였다. 데이터마이닝 방법 중에서 K-Means clustering 방법을 적용하여 개별 기업들이 보유하고 있는 특성으로 분류하여 분석하였다.

기본적으로 R&D역량 수준이 5가지 범주임을 고려하여 클래스는 5개로 설정하여 분석을 실시하였고, 5개의 클래스로 구분하여 분석하였을 때 군집별 크기를 보면 37.22%(cluster_3), 30.23%(cluster_0), 14.35%(cluster_2), 12.45%(cluster_1), 5.75%(cluster_4)로 나타났다. 5개의 클래스로 구분한 R&D역량 속성들의 분류를 비교해 보면 다음과 같다.

<Table 4> Ratio comparison of K-Means cluster analysis and R&D capability level

K-Means clustering			Labeling for R&D competency level		
cluster	count	ratio	rank	count	ratio
cluster_1	1,589	12.45%	1	1,475	11.56%
cluster_0	3,857	30.23%	2	3,919	30.71%
cluster_3	4,749	37.22%	3	4,001	31.35%
cluster_2	1,831	14.35%	4	2,757	21.60%
cluster_4	734	5.75%	5	609	4.77%
Total	12,760	100.00%	Total	12,761	100.00%

등급별 비율은 K-Means 군집분석에서 결과로 도출된 비율과 R&D역량 수준 등급을 부여한 비율이 유사함을 확인할 수 있다. 등급 3인 R&D역량 보유기업들이 가장 많고 다른 수준의 역량을 보유한 기업들은 상대적으로 작은 것으로 나타났다. 이와 유사하게 군집분석 결과도 cluster_3을 기준으로 좌우 대칭적 구조의 비율양상을 보였다.

<Table 4>에서 제시한 K-Means 군집과 R&D역량 수준 등급부여 결과와의 비교는 R&D역량 수준의 클래스를 고려하여 분석한 결과가 아니기 때문에 크게 의미가 없다고 볼 수 있다. 또한 이 연구에서 주요한 부분은 어떤 형태의 개별 기업들이 어떤 기술경영능력, 연구개발능력, 기업분류 속성을 보이고 있는지에 대한 것에 초점이 맞추어져 있는 것을 감안하면 두 가지의 분류에 대한 등급간의 정확한 비교 보다는 R&D역량의 특성 파악에 중점을 두는 것이 적절해 보인다. 이러한 특성 비교는 아래와 같은 결과로 확인해 볼 수 있다.

<Table 5> Comparison of R&D competency level characteristics by cluster

Level	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
Level 1	833	14	408	1,284	13
Level 2	762	93	330	1,326	44
Level 3	876	244	270	1,057	102
Level 4	865	473	318	706	191
Level 5	521	765	505	376	384
Total	3,857	1,589	1,831	4,749	734
High level	35.9%	77.9%	44.9%	22.8%	78.3%
Low level	41.4%	6.7%	40.3%	55.0%	7.8%

<Table 5>의 결과를 보면, cluster_1과 cluster_4의 형태를 보면, R&D역량 수준이 높은 기업들이 각각 77.9%, 78.3%, 역량 수준이 낮은 기업들은

각각 6.7%, 7.8%로 나타났다. 반면에 cluster_3에 은 기업들이 55.0%로 나타났다. 이외에 cluster_0 서는 R&D역량 수준이 높은 기업들이 22.8%, 낮 과 cluster_2는 R&D역량 수준이 비슷한 수준으

<Table 6> Characteristics of each cluster and properties of R&D capabilities

Cluster	Cluster features	Characteristics of competency attributes
cluster_0	The ratio of high and low competence level is similar (the ratio of low is slightly higher)	<ul style="list-style-type: none"> - The representative's training in the same industry has a stable occurrence structure and a tendency to work for a long period of time. - The number of technical certifications is generally small - The R&D department's maintenance period is relatively long, and the frequency of occurrence is low. - The frequency of occurrence of private companies is relatively low - Low incidence of small-scale employees - The annual average R&D expense ratio is low - No experience in utility model registration and application - There are types of large and medium-sized companies with similar ratios
cluster_1	High competency level	<ul style="list-style-type: none"> - The frequency of occurrence is relatively low when the representative's years of employment in the same industry are relatively stable. - The number of patent applications and technology commercialization is relatively small. - Low incidence of private companies - No incidence of non-profit corporation type - No incidence of public institution type. - High incidence of small-scale employees - Many types of small and medium enterprises
cluster_2	The ratio of the high and low competence level is similar (the high ratio is slightly higher)	<ul style="list-style-type: none"> - The representative's training in the same industry has a stable occurrence structure and a tendency to work for a long period of time. - The number of patent registrations tends to be low, similar to cluster 0 - Technology certifications tend to be generally low - The R&D department's maintenance period is short or long - High incidence rate for private and general corporation types at similar rates - The number of employees held is similar to cluster 0 and is low - The annual average R&D expense ratio is low - The ratio of the availability of utility model registration and application experience is similar - Medium-sized enterprises, small business owners, and types of protected medium-sized enterprises at a similar rate
cluster_3	Low competency level	<ul style="list-style-type: none"> - The frequency of occurrence is relatively high when the representative's years of employment in the same industry are relatively stable. - The number of technology certifications is relatively high, and the frequency of occurrence is also high. - High incidence of private enterprise type - No incidence of non-profit corporation type - No incidence of public institution type. - High incidence of small-scale employees - There are a few companies with a large annual average R&D expenditure ratio - More than 5 times the ratio of no experience compared to the experience of applying for a utility model - Many types of small and temporary small and medium enterprises
cluster_4	High competency level	<ul style="list-style-type: none"> - The frequency of occurrence is relatively low when the representative's years of employment in the same industry are relatively stable. - The number of patent applications and technology commercialization is relatively high - Low incidence of private enterprise type - No incidence of non-profit corporation type - Low incidence of public institution types - The incidence of small-sized employees is high, and the incidence of large-sized employees is low, but it exists. - Only having experience in registering a utility model exists - Many types of large and medium enterprises

로 양분되는 양상을 보였다. 군집들은 각 형태의 속성들이 비슷한 그룹으로 묶인 기업들인데, 이들의 특성이 R&D역량 수준이 높은 군집과 낮은 군집, 그리고 높고 낮음이 비슷한 수준인 군집으로 구별하여 R&D역량 수준의 특성을 분류할 수 있다.

<Table 6>에서 제시하는 바와 같이 R&D역량 수준이 높은 군집특성을 살펴보면, 기업 대표의 동업종 종사연수가 비교적 안정적인 경우의 발생빈도가 상대적으로 낮은 경향을 보이고 있으며, 이것은 일반적인 기대치와는 상이한 양상을 보여준다. 이 같은 결과는 기업 대표의 교체주기가 짧을수록 역량 수준에 기여할 가능성이 높다는 사실을 보여준다. 일반적인 예상과 동일하게 역량 수준이 높은 경우에 개인기업의 발생빈도가 낮게 발생되고 있다는 점이다. 이와 함께 비영리법인 유형의 발생빈도도 상대적으로 낮게 나타나고 있으며, 공공기관 유형의 발생빈도가 낮거나 아예 존재하지 않은 경우도 있는 것으로 나타났다. 또한 역량 수준이 높은 경우, 소규모의 종업원을 보유하고 있는 경향이 뚜렷하다는 점이다. 물론 낮은 발생빈도로 큰 규모의 종업원을 보유하고 있는 경우도 존재하고 있다. 마지막으로 역량 수준이 높으면 실용신안등록 경험이 반드시 존재하고 있는 것으로 나타났다.

이와는 반대로 R&D역량 수준이 낮은 군집특성을 살펴보면, 대표이사의 동업종 종사연수가 안정적인 경우의 발생빈도가 상대적으로 높다는 점이다. 이 또한 일반적 기대치와는 다른 양상을 보여주고 있다. 전반적으로 기술인증 실적이 낮은 하지만 상대적으로 보다 많은 실적을 보유하고 있는 기업일수록 역량 수준은 낮게 나타나고 있다. 또한 역량 수준이 낮은 경우에 비영리 법인과 공공기관 유형의 발생빈도가 존재하지 않

은 점이다. 이외에도 소규모의 종업원 보유기업의 발생빈도가 높다는 점이다. 이러한 특징은 역량수준이 높은 경우에도 동일 특성을 보인다. 그리고 연구개발비 비율이 큰 소수의 기업들이 존재하고 있는 특징이 있으며, 실용신안출원 무경험 비율이 유경험 대비 5배 이상 높다는 점이다. 기업규모 측면에서는 소기업과 한시성중소기업 유형이 다수 존재하고 있다.

이외에도 역량 수준이 낮고 높은 비율이 유사하게 발생되고 있는 군집이 분석되었으며, 두 군집의 차이점은 높은 비율이 약간 상회하거나 낮은 비율이 약간 상회하는 경향으로 구분된다. 두 군집의 경우에 역량 수준의 발생점유 비율 측면에서의 유사한 특징을 보임으로 인해 역량 속성 측면에서 일부 속성들에 대해서는 동질성을 보이는 경향이 존재하였다. 다만 기업유형의 발생양상이나 실용신안등록 및 출원경험에 대한 발생형태가 약간 차이가 있었으며, 기업규모의 유형이 상이한 특징을 보였다.

5. 결론

글로벌 경영환경 변화가 연구개발 중심의 기업들에게 기업경영의 효율성(efficiency) 및 효과성(effectiveness)의 제고를 요구하고 있다. 특히 R&D중심 기업들은 연구개발 활동에 대한 지속적인 투자가 필수불가결한 사항이며, 연구개발 경쟁력을 제고시키기 위해서 투자를 강화시켜야만 한다. 하지만 설비나 연구개발 투자 요소는 기업의 입장에서 투자자본을 회수하지 못할 가능성이 존재하는 미래의 불확실성에 고스란히 직면하게 된다.

본 연구에서는 데이터마이닝 기법중 하나인

군집분석 방법을 활용하여 기업들의 연구개발 역량에 영향을 주는 특성을 기술경영능력, 연구개발능력, 기업분류 속성 관점에서 탐색하였다. 이 같은 평가항목에 근거한 개별적인 세부 요인들이 국내 12,761개의 개별 기업의 R&D역량 수준이 높고 낮은 경우에 각각 어떤 특징들을 보여주는지 최근 10년 동안의 증거데이터를 기반으로 탐색을 시도하였다.

결과적으로 R&D역량 수준이 높은 경우의 대표적인 특징들은 첫째, 기업 대표의 동업종 종사연수가 비교적 안정적인 경우의 발생빈도가 상대적으로 낮은 경향을 보인다는 점이다. 이것이 의미하는 바는 기업 대표의 교체주기가 빠를수록 역량수준에 기여할 가능성이 높다는 것을 의미한다. 둘째, 기업분류 속성 중에서 개인기업의 발생빈도가 낮게 발생되고 있다는 점이다. 이것은 개인기업인 경우에 R&D역량 수준이 낮을 가능성이 높다는 것을 의미한다. 셋째, 공공기관 유형의 발생빈도가 낮거나 아예 존재하지 않은 경우도 있으며, 가장 두드러진 특징 중 하나인 소규모 종업원을 보유하고 있는 경향이 뚜렷하다는 점이다. 상기의 사실 중에서 일반적으로 기대했던 결과는 상이한 결과가 도출된 것을 본 연구를 통해 알 수 있었다.

이와 반대로 R&D역량 수준이 낮은 경우의 주요 특징들은 첫째, 대표이사의 동업종 종사연수가 안정적인 경우의 발생빈도가 상대적으로 높다는 점이다. 이것이 의미하는 바는 오랜 기간 동안 대표이사직을 수행하고 있는 기업의 경우에는 R&D역량 수준에 긍정적 영향을 미치지 못하고 있다는 사실이다. 둘째, 기술인증 실적이 모든 군집들에 있어서 전반적으로 낮은 상황이지만, 그 중에서도 일부 실적을 보유하고 있는 것이 R&D역량 수준에 오히려 부정적인 영향을

미치고 있는 것으로 나타났다. 셋째, 비영리법인과 공공기관 유형의 발생빈도가 존재하지 않은 점이다. 특히 공공기관 유형의 발생빈도가 낮은 경우는 R&D역량 수준이 높은 경우에도 나타나는 양상이다. 마지막으로 연구개발비 비율이 큰 소수의 기업들이 R&D역량 수준이 낮은 것으로 나타났으며, 실용신안출원 무경험 비율이 유경험 대비 5배 이상 높다는 사실을 확인하였다.

본 연구 결과는 국내 연구개발 기업 전체를 대상으로 증거데이터에 기반한 군집분석 및 실험 결과를 제시하였다. 특히 본 연구의 결과가 제시하고 있는 시사점은 기술변화 속도와 시장수요의 변화에 효과적으로 대응할 수 있는 전문 경영자의 교체주기가 빠를수록 역량제고에 기여할 가능성이 높다는 사실이다. 개인기업의 경우에 법인기업으로의 전환을 통해 연구개발 인력들의 기업에 대한 소속감을 제고시킴으로써 연구개발 역량의 투입강도를 높일 필요성이 있다는 점이다. 이외에도 조직적 측면에서 팀단위의 조직구성을 통해 책임과 권한의 정확성을 제공할 필요가 있다는 점이다. 기술상용화실적건수나 기술인증건수는 역량제고에 기여하는 경우와 그렇지 않은 경우가 모두 발생되고 있어, 경영자 입장에서 연구개발 역량제고를 위한 중요 인자로 검토하는데 한계가 있는 것으로 확인되었다. 마지막으로 실용신안출원의 경험 여부는 연구개발 역량에 중요한 영향을 미치는 요인으로 파악되어, R&D역량 제고를 위해서는 실용신안출원 장려를 위한 동기부여를 제공할 필요가 있다는 점이다. 이처럼 본 연구의 결과는 연구개발 중심 개별기업들의 R&D역량 제고를 위한 거시적인 기업경영 전략의 중요한 시사점을 제공할 수 있을 것으로 기대된다.

참고문헌(References)

- Berry, M., Linoff, G., *Data mining techniques: For marketing*, (1997), Sales and Marketing Support.
- Bishop, C. M., *Pattern recognition and machine learning*, (2006), springer.
- Bobe, B., Bobe, A.C., Benchmarking Innovation Practices of European Firms, Joint Research Centre, European Commission (EC). Brussels, Luxembourg: ECSC-EEC-EAEC., (1998)
- Burgelman, R.A., Christensen, C.M., Wheelwright, S.C., *Strategic Management of Technology and Innovation.*, McGraw-Hill/ Irwin., (2004)
- Choi, T.M., “A Study on the Relationship between Marketing Differentiation, R&D Capabilities and Performance in Information Technology Based Start-up Companies,” Ph.D. dissertation, Gyeongsang National University, Jinju, 2012.
- Choi, W.S. and J.H. Lee, “A Study on the Characteristics of R&D for Technology Development in Chungbuk Area: Focused on text analysis and network analysis,” *Journal of Industrial Innovation*, Vol. 36, No. 2 (2020), 1-19.
- Christensen, J.F., Asset profiles for technological innovation. *Research Policy*, 24: 727-745., (1995)
- Cooper, R.G., Overhauling the new product process. *Industrial Marketing Management*, 25, 465-482., (1996)
- Hwang, K.Y. and E.H. Sung, “The Relationships between Technology Commercialization Competence, R&D Capacity, Innovation and Export Performance: In the Firms Introduced the Technology from Government-funded Research Institutes in Daedeok Innopolis,” *Journal of Korea Trade*, Vol. 40, No. 1(2015), 285-309.
- Jeong, D.Y., “A Study on the Effect of R&D Capability and Technology Commercialization Capability of SMEs in Collaboration with Government Funded Research Institutes on Their Innovation Performance,” Masters dissertation, Korea University, Seoul, 2015.
- Jung, S.Y., “The Study on Relation between R&D Innovation Capability and Performance in Biotechnology Firms,” Ph.D. dissertation, Konkuk University Seoul, 2014.
- Kim, D.W., and D.H. Kim, “A Study on the R&D Capabilities of ICT SMEs in Yangje District,” *Journal of the Korea Convergence Society*, Vol. 8, No. 7(2017), 265-272.
- Kim, M.S., S.J. Kim and K.H. Kyung, “The Empirical Study on Relation between R&D Innovation Capability and Performance in Knowledge-Based Service Firms,” *Journal of the Korean Society for Quality Management*, Vol. 40, No. 4(2012), 631-640.
- Kim, Y.J., and D.W. Yang, “The comparative analysis of game capability by region-focus on the analysis of LQ index and CT R&D competence level-,” *The Journal of digital policy & management*, Vol. 9, No. 3(2011), 133-144.
- Park, J.P., “Analysis of R&D Competence Factors on Technology Innovation in Government-funded Research Institutes,” Masters dissertation, Chungnam National University, Daejeon, 2008.
- Park, M.S., H.J. Shon and H.H. Lee, “Classification of the Subcontracting Companies Accompanied by R&D Capabilities Compare and Research,”

- Journal of Information Technology and Architecture, Vol. 8, No. 2(2011), 121-135.
- Shim, K.S., “The Effects of Technological Innovation Competencies, Entrepreneurial Characteristics and Technological Innovation Performance on Productivity of Venture Business,” Masters dissertation, Yeonsei University, Seoul, 2016.
- Shin, D.Y. and S.C. Park, “The Effects of Exploration and Exploitation on R&D Performances,” Journal of Strategic Management, Vol. 6, No. 1(2003), 93-126.
- Shon, H.J. and M.S. Park, “Analysis of Corporate R&D Capability with Industrial’s Innovation Trend,” Journal of Information Technology and Architecture, Vol. 10, No. 1(2013), 47-62.
- Song, M.J., “A study on the effect of internal competence on R&D performance,” Masters dissertation, Yeungnam University, Gyeongsan, 2018.
- Souitaris, V., “Firm-specific Competencies Determining Technological Innovation : A Surveying Greece,” *R&D Management*, Vol. 32, No. 1, (2002), 61-77.
- Wilks, D. S., *Statistical methods in the atmospheric sciences*, Vol.100, (2011), Academic press.
- Yam, R.C.M., Guan, J.C., Pun, K.F., Tang, E.P.Y., An audit of technological innovation capabilities in chinese firms: some empirical findings in Beijing, China., *Research Policy*, 33:, (2004), 1123-1140.
- Yoo, I.J. and Park, D.H., “Derivation of Digital Music’s Ranking Change Through Time Series Clustering,” Journal of intelligence and information systems, Vol. 26, No. 3(2020), 171-191.
- Yoon, H., A. Hong and S. Jung, “The effects of R&Ds, technology innovation capability and the innovation support system of small- and medium-sized business on the company performance,” *Innovation Studies*, Vol. 13, No. 2(2018), 209-238.
- Zhu, S.G., “A study on the influence of technological innovation capability and technology commercialization capability on technological innovation performance and management performance,” Masters dissertation, Gachon University, Geonggi-do, 2013.

Abstract

A Study on the Characteristics of Enterprise R&D Capabilities Using Data Mining

Sang-Gook Kim* · Jung-Sun Lim** · Wan Park***

As the global business environment changes, uncertainties in technology development and market needs increase, and competition among companies intensifies, interests and demands for R&D activities of individual companies are increasing. In order to cope with these environmental changes, R&D companies are strengthening R&D investment as one of the means to enhance the qualitative competitiveness of R&D while paying more attention to facility investment. As a result, facilities or R&D investment elements are inevitably a burden for R&D companies to bear future uncertainties. It is true that the management strategy of increasing investment in R&D as a means of enhancing R&D capability is highly uncertain in terms of corporate performance. In this study, the structural factors that influence the R&D capabilities of companies are explored in terms of technology management capabilities, R&D capabilities, and corporate classification attributes by utilizing data mining techniques, and the characteristics these individual factors present according to the level of R&D capabilities are analyzed. This study also showed cluster analysis and experimental results based on evidence data for all domestic R&D companies, and is expected to provide important implications for corporate management strategies to enhance R&D capabilities of individual companies. For each of the three viewpoints, detailed evaluation indexes were composed of 7, 2, and 4, respectively, to quantitatively measure individual levels in the corresponding area. In the case of technology management capability and R&D capability, the sub-item evaluation indexes that are being used by current domestic technology evaluation agencies were referenced, and the final detailed evaluation index was newly constructed in consideration of whether data could be obtained quantitatively. In the case of corporate classification attributes, the most basic corporate classification profile information is

* Corresponding author: Sang-Gook Kim

Data Analysis Platform Center, Korea Institute of Science and Technology Information
66 Hoegi-ro, Dongdaemun-gu, Seoul 02456, South Korea
Tel: +82-2-3299-6294, Fax: +82-2-3299-6041, E-mail: sgkim@kisti.re.kr

** R&D Investment Analysis Center, Korea Institute of Science and Technology Information
66 Hoegi-ro, Dongdaemun-gu, Seoul 02456, South Korea

*** Science and Technology Management Policy, University of Science A Technology,
217 Gajeong-ro, Yuseong-ga, Daejeon Metropolitan City, South Korea.

considered. In particular, in order to grasp the homogeneity of the R&D competency level, a comprehensive score for each company was given using detailed evaluation indicators of technology management capability and R&D capability, and the competency level was classified into five grades and compared with the cluster analysis results. In order to give the meaning according to the comparative evaluation between the analyzed cluster and the competency level grade, the clusters with high and low trends in R&D competency level were searched for each cluster. Afterwards, characteristics according to detailed evaluation indicators were analyzed in the cluster. Through this method of conducting research, two groups with high R&D competency and one with low level of R&D competency were analyzed, and the remaining two clusters were similar with almost high incidence. As a result, in this study, individual characteristics according to detailed evaluation indexes were analyzed for two clusters with high competency level and one cluster with low competency level. The implications of the results of this study are that the faster the replacement cycle of professional managers who can effectively respond to changes in technology and market demand, the more likely they will contribute to enhancing R&D capabilities. In the case of a private company, it is necessary to increase the intensity of input of R&D capabilities by enhancing the sense of belonging of R&D personnel to the company through conversion to a corporate company, and to provide the accuracy of responsibility and authority through the organization of the team unit. Since the number of technical commercialization achievements and technology certifications are occurring both in the case of contributing to capacity improvement and in case of not, it was confirmed that there is a limit in reviewing it as an important factor for enhancing R&D capacity from the perspective of management. Lastly, the experience of utility model filing was identified as a factor that has an important influence on R&D capability, and it was confirmed the need to provide motivation to encourage utility model filings in order to enhance R&D capability. As such, the results of this study are expected to provide important implications for corporate management strategies to enhance individual companies' R&D capabilities.

Key Words : Data mining, cluster analysis, R&D capability, technology management capability, enterprise classification attribute

Received : November 4, 2020 Revised : January 28, 2021 Accepted : February 2, 2021

Corresponding Author : Sang-Gook Kim

저자 소개



김 상 국

서울대학교에서 산업공학으로 석사학위를 취득하고, 美 플로리다공과대학에서 경영과학(Operations Research)으로 박사학위를 취득했다. 현재 한국과학기술정보연구원(KISTI)에서 책임연구원으로 재직 중이며, 동국대학교 경영학과 겸임교수를 역임하고 있다. 전공분야는 확률과정 및 대기이론이며, 최근 관심분야는 빅데이터 기반 기계학습 및 인공지능을 활용한 과학기술정보의 정보계량 연구이다. 현재 기술가치 및 기술력 평가 모형개발과 국가 R&D 투자의사결정을 위한 정량분석 방법에 관한 연구를 수행하고 있다.



임 정 선

KAIST 생명화학공학과에서 석사학위를, Purdue 대학에서 화학공학 박사학위를 취득하였다. 현재 한국과학기술정보연구원(KISTI)에서 선임연구원으로 재직 중이다. 주요 연구분야는 자기조립, 나노기술정책, 과학기술 정보계량 연구이다.



박 완

KAIST 경영대학원에서 경영학 석사학위를 취득하였고, 한국과학기술연구원 부설 녹색기술센터(GTC)에서 연구원으로 재직하였다. 현재 UST 과학기술경영정책 박사과정에 재학 중이며, 주요 연구 및 관심 분야는 과학기술투자의 효율성 분석, 과학기술정책, 기술가치평가, 기후변화정책 등이다.