

기술 성숙도 및 의존도의 네트워크 분석을 통한 유망 융합 기술 발굴 방법론

최호창

국민대학교 비즈니스IT전문대학원
(hochangchoi@kookmin.ac.kr)

곽기영

국민대학교 경영대학 경영정보학부
(kykwahk@kookmin.ac.kr)

김남규

국민대학교 경영대학 경영정보학부
(ngkim@kookmin.ac.kr)

최근 다양한 분야에서 새로운 기술이 출현하고 있으며, 이들 대부분은 기존 기술들의 융합(Convergence)을 통해 형성되고 있다. 또한 가까운 미래에 출현하게 될 유망한 융합 기술을 예측함으로써 변화하는 기술 지형에 선제적으로 대응하기 위한 수요가 꾸준히 증가하고 있으며, 이러한 수요에 부응하여 많은 기관과 연구자들은 미래 유망 융합 기술 예측을 위한 분석을 수행하고 있다. 하지만 이와 관련한 기존의 많은 연구들은 (i) 고정된 기술 분류 기준을 분석에 사용함으로써 기술 분야의 동적 변화를 반영하지 못했다는 점, (ii) 예측 모형 수립 과정에서 주로 범용성 네트워크 지표를 사용함으로써 기술의 융합이라는 목적에 부합하는 고유 특성을 활용하지 못했다는 점, 그리고 (iii) 유망 분야 예측 모형의 정확성 평가를 위한 객관적 방법을 제시하지 못했다는 점 등에서 한계를 갖고 있다. 이에 본 연구에서는 (i) 토픽 모델링을 통해 기존의 고정된 분류 기준이 아닌 실제 기술 시장의 동적 변화에 따른 새로운 기술군을 도출하고, (ii) 기술 성숙도 및 기술군 간 의존 관계에 따라 각 기술군의 융합적 특성을 반영하는 잠재 성장 중심성(Potential Growth Centrality) 지표를 산출하였으며, (iii) 잠재 성장 중심성에 근거하여 예측한 유망 기술의 성숙도 증가량을 시기별로 추정하여 예측 모형의 정확도를 평가하는 방안을 제시한다. 이와 더불어 제안 방법론의 성능 및 실무 적용 가능성의 평가를 위해 특허 문서 13, 477건에 대한 실험을 수행하였으며, 실험 결과 제안한 잠재 성장 중심성에 따른 예측 모형이 단순히 현재 활용되는 영향도 기반의 예측 모형에 비해 최대 약 2.88배 높은 예측 정확도를 보임을 확인하였다.

주제어 : 기술 융합, 소셜 네트워크 분석, 유망 기술 예측, 잠재 성장 중심성, 토픽 모델링

논문접수일 : 2017년 12월 10일 논문수정일 : 2018년 2월 26일 게재확정일 : 2018년 3월 5일
원고유형 : 일반논문(급행) 교신저자 : 김남규

1. 서론

최근 다양한 분야의 기술은 단일 기술의 고도화 또는 타 기술과의 연계를 통해 여러 모습으로 발전하고 있으며, 특히 둘 이상의 기술군이 상호 작용하여 융합 기술(Convergence Technology)의 형태로 출현하는 경향이 뚜렷이 증가하고 있다. 또한 가까운 미래에 출현하게 될 유망한 융합 기술을 예측함으로써 변화하는 기술 지형에 선제

적으로 대응하기 위한 수요가 꾸준히 증가하고 있으며, 이러한 현상에 부응하여 많은 기관과 연구자들은 미래 유망 융합 기술 예측을 위한 분석을 수행하고 있다. 융합 기술은 생성 원리에 따라 다양한 분야의 특성을 동시에 갖기 때문에 이를 분석하여 유망 기술을 예측하는 것은 매우 어렵지만, 최근 빅데이터 분석 및 소셜 네트워크 분석 기법의 고도화에 따라 기존 기술 지형의 변화에 대한 학습을 통해 미래 유망 기술을 예측하

는 시도에서 일부 성과가 확인되고 있다.

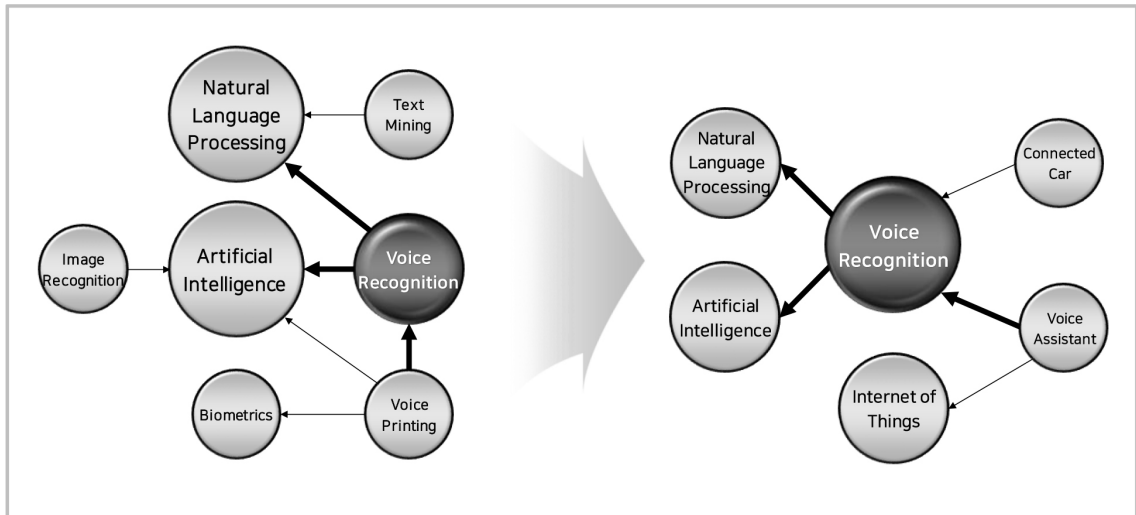
데이터 기반의 융합 기술 분석 연구는 대부분 새로운 융합 기술의 발굴 및 이들의 동향 분석에 초점을 두고 있으며, 기술 관련 연구의 한 분야를 형성할 만큼 활발하게 수행되고 있다. 하지만 유망 융합 기술 발굴을 위한 기존의 분석 방법은 몇 가지 한계를 갖는다. 먼저 대부분의 융합 기술 연구는 사전에 정의된 세부 기술 분류에 기반을 두어 분석을 수행한다. 하지만 최근 출현한 기술 중 상당수가 이미 융합 기술의 특성을 가지므로, 이러한 기술을 기존의 획일화된 분류 기준으로 정의하는 방식은 융합 기술을 모태로 한 융합, 즉 융합의 연쇄 전이를 정확하게 묘사할 수 없다는 한계를 갖는다. 즉 융합을 미래에만 발생하는 현상이 아닌 이미 발생하여 동작하고 있는 현상으로 이해해야 하며, 유망 융합 기술 발굴을 위해서는 현재 이미 태동하여 성장하고 있는 융합 기술에 대한 정확한 이해 및 분석이 선행되어야 함을 의미한다.

이와 더불어 대부분의 기존 분석 방법론은 예측 모형 수립 과정에서 주로 범용성 지표를 활용함으로써, 융합 기술이라는 분야의 특수성을 충분히 활용하지 못한 측면이 있다. 신생 융합 기술은 해당 기술의 모태가 되는 기존 기술에 대한 의존도가 매우 높으며, 의존하고 있는 기술의 변화에 따라 융합 기술이 독립적 분야로 성장하거나 빠른 속도로 사라질 수 있다. 하지만 기존 분석에서는 범용적 관점으로 고안된 지표, 즉 특정 기술에 대한 전통적인 통계 정보 또는 네트워크 지표를 통해 이들의 성장 가능성을 판단한다. 이러한 전통적 분석 지표는 융합의 원리, 즉 충분히 성숙한 두 개 이상의 기술군을 통해 새로운 기술이 등장하고, 이들이 성장하여 또 다른 융합 기술의 생성에 영향을 주는 특성

을 반영하지 않기 때문에 융합 기술의 향후 성장에 대한 정보를 충분히 전달하지 못한다. 보다 구체적으로 기존 분석에서는 현 시점에서는 다수의 주요 기술 분야에 의존하는 정도가 매우 크지만, 향후 성장을 통해 타 기술 분야에 영향력을 발휘할 수 있는 유망 융합 기술을 파악하는 것이 매우 어렵다.

전통적 지표에 의존한 유망 융합 기술 예측의 한계는 <Figure 1>을 통해 설명될 수 있다. 그림에서 노드는 기술 분야를, 링크는 기술 간 영향 관계를 의미한다. 예를 들어 $A \rightarrow B$ 의 그래프는 기술 B가 기술 A에 영향을 미침 또는 기술 A가 기술 B에 의존함을 나타내고 있으며, 두 기술의 관계가 강할수록 링크가 굵게 나타난다. 또한 그림에서는 충분히 성숙하여 다른 기술에 미치는 영향도가 높을수록, 즉 내향 링크의 수가 많거나 전체 강도의 합이 클수록 노드의 크기가 크게 나타난다. 따라서 좌측 그림에서 다른 기술에 대한 영향도가 가장 높은 기술은 “Natural Language Processing”과 “Artificial Intelligence”이며, 가장 많은 기술에 의존하고 있는 기술은 “Voice Printing”임을 알 수 있다. 한편 “Voice Recognition”은 영향도가 높은 두 기술과 강한 관계를 맺고 있음을 알 수 있다.

현 시점의 기술 영향 관계를 나타낸 <Figure 1>의 좌측 네트워크로부터 향후 유망 기술을 예측하는 경우를 가정하자. 예를 들어 많은 수의 기술에 의존할수록 향후 성장할 가능성이 높다고 가정하면, 좌측 네트워크에서 외향 연결 중심성(Out-degree Centrality)이 가장 높은 “Voice Printing”을 유망 기술 후보로 선정할 것이다. 하지만 이는 융합 기술 분야의 특성, 즉 해당 분야의 모태가 되는 주요 기술의 크기 및 기술 간 관계의 강도를 충분히 반영하고 있다고 할 수 없



〈Figure 1〉 Growth of Convergence Technology

다. 따라서 본 연구에서는 많은 기술에 영향을 주고 있는 성숙 단계의 기술, 즉 내향 연결 중심성(In-degree Centrality)이 높은 기술을 핵심 기술로 식별하고, 여러 핵심 기술과 강한 관계를 맺고 있는 기술을 유망 융합 기술로 선정하는 방법론을 제안하고자 한다. 예를 들어 <Figure 1>의 좌측 네트워크는 “Voice Recognition”이 핵심 기술인 “Natural Language Processing” 및 “Artificial Intelligence”와 강한 관계를 맺고 있으므로 향후 성장 가능성이 높은 융합 분야임을 암시하고 있으며, 우측 네트워크는 실제로 해당 분야가 성장하여 “Connected Car” 및 “Voice Assistant” 분야에 영향을 주고 있음을 나타내고 있다.

이러한 개념을 구현하기 위해 본 연구에서 제안하는 방법론은 다음과 같이 요약된다. 우선 기술 문서에 대한 토픽 모델링을 통해 기존의 고정된 분류 기준이 아닌 실제 기술 시장의 동적 변화에 따른 새로운 기술군을 도출하고자 한

다. 또한 각 문서의 토픽 대응도를 분석하여 기술군 간의 영향 관계를 네트워크로 도식화한다. 다음으로 인접 핵심 기술들과의 연결 강도를 측정하여, 이를 해당 융합 기술의 향후 성장 가능성을 나타내는 잠재 성장 중심성(PGC: Potential Growth Centrality) 지표로 정의하고자 한다. 마지막으로 잠재 성장 중심성이 높은 기술의 영향도, 즉 기술의 성숙도가 시기별로 증가하는 양상을 분석하여 제안 방법론의 성능을 평가하고자 한다.

본 논문의 이후 구성은 다음과 같다. 먼저 2장에서는 본 연구에서 다루는 주요 기법의 간략한 소개 및 융합 기술에 대한 선행 연구의 성과를 요약하고, 3장에서는 본 연구에서 제안하는 방법론을 간단한 예를 통해 소개한다. 또한 4장에서는 제안 방법론을 실제 특허 데이터에 적용한 실험 결과를 제시하고, 마지막 장인 5장에서는 본 연구의 기여 및 한계를 요약한다.

2. 관련 연구

2.1 융합 기술 연구

기술적 융합에 대한 개념이 처음 소개된 이후(Rosenberg, 1963), 융합 기술은 다양한 산업 분야에서 빈번하게 사용되는 단어 중 하나가 되었다. 이러한 융합 기술은 상이한 특성을 갖는 둘 이상의 산업 사이에서 그 경계가 허물어짐에 따라 생성되는, 즉 기존 기술과는 다른 특성을 갖는 새로운 기술로 정의할 수 있으며(Curran and Leker, 2011), 기술의 고도화에 따라 발전 속도가 더딘 기존 기술에 활력을 불어넣는 역할을 수행함으로써 많은 각광을 받고 있다. 이에 따라 융합 기술 분야는 기술의 혁신에 대한 내용을 중점적으로 다루는 기술 관련 연구의 중요 분야로 자리매김하였으며(Lee, 2015), 새로운 융합 기술을 발굴하고 이들의 특성을 파악하기 위한 많은 연구가 다양한 분야에서 수행되고 있다.

융합 기술 연구는 매우 넓은 범위의 기술 분야를 아우르지만 특히 IT 기술 및 산업을 중심으로 활발히 수행되고 있다. 구체적으로 최근 5년간 ICT 분야 융합 기술 동향 파악을 통해 융합 유형을 식별하고 이에 대한 통계적 분석을 수행하여 기술적 융합과 R&D 동맹의 관련성을 분석한 연구(Lee and Zo, 2016), IT 및 BT 분야의 특허 대상 연관성 분석을 통해 다양한 관점에서의 기술적 융합을 파악한 연구(Huang and Huang, 2015), 산업적 상승효과에 대한 평가식 고안을 통해 산업 융합과 기술 진보의 관점에서 산업 분야의 진화를 분석한 연구(Li et al., 2014) 등을 대표적 예로 들 수 있다. 이와 더불어 네트워크 분석을 통해 로봇 기술 분야의 융합 패턴을 찾아내고 QAP 분석을 통해 융합에 영향을 미치는 요인을 식별

한 연구(Lee et al., 2016), 네트워크 분석을 통해 인쇄 전자 분야의 핵심 기술을 식별하고 이를 중심으로 발생하는 기술적 융합의 동적 패턴을 파악한 연구(Kim et al., 2014) 등 고도로 발전된 특정 분야를 중점적으로 다루는 융합 기술 연구 또한 수행된 바 있다.

이러한 융합 기술 연구는 목적에 따라 두 가지의 세부 분야로 나눌 수 있다. 첫 번째는 융합 기술과 관련하여 이미 발생한 현상을 규명하는 것에 초점을 두는 분야로, 앞서 언급된 연구의 대부분이 이에 해당된다고 볼 수 있다. 다음으로 두 번째는 기술 생애주기 관점에서의 접근을 통해 융합 기술의 향후 전망을 파악하는 분야로, 신뢰성 공학에서 활용되는 위험 비율을 근간으로 기존 기술의 대체 기술을 도출하는 방법을 제안한 연구(Jeong et al., 2016)를 예로 들 수 있다. 두 세부 분야를 비교할 때, 이미 발생한 융합 현상을 규명하는 연구에 비해 특정 융합 기술의 향후 전망을 파악하기 위한 연구는 상대적으로 부족한 상황임을 알 수 있다.

2.2 소셜 네트워크 분석

소셜 네트워크 분석은 액터의 구조적 정보를 활용하여 이들의 네트워크 특성을 분석하고, 이를 통해 개별 또는 집단 관점에서 발생하는 인과관계의 작동 원리를 이해하고 설명하는 기법으로 정의할 수 있다(Freeman, 2017; Kwahk, 2017). 네트워크는 기본적으로 액터 및 그들 간 관계로 구성되며, 이때 액터는 노드, 관계는 링크로 나타난다(Marin and Wellman, 2011). 이러한 소셜 네트워크 분석은 분석 대상 네트워크의 패턴을 파악하고 해석하는 것에 목적이 있으며(Park et al., 2013; Park and Kwahk, 2013), 특정 액터가 갖

는 개별적 특성에 비해 전체 네트워크에서 액터가 수행하는 역할에 중점을 둔다. 소셜 네트워크에서의 관계 정보는 그래프 또는 행렬 형태로 표현되며, 네트워크의 구조적 분석 및 시각적 표현을 위해 각 행렬 및 그래프를 활용한다.

네트워크 그래프의 종류는 관계의 방향성 및 강도의 표현 유무에 따라 총 4가지, 즉 방향/이진, 방향/계량, 비방향/이진, 비방향/계량 그래프로 구분할 수 있다. 이 중 비방향/이진 그래프는 단순히 네트워크 관계의 존재 여부만을 다루기 때문에 다양한 구조적 분석이 가능하다는 특징이 있으며, 선행 연구에서 가장 빈번히 사용되고 있다. 한편 방향/계량 그래프는 다소 복잡한 형태로 나타날 수 있지만, 전체 액터 간의 방향성 및 관계 강도를 분석에 반영할 수 있다는 장점이 있다.

분석 대상에 대한 네트워크 정보를 활용하는 대부분의 연구는 다양한 속성의 중심성 분석을 통해 각 액터의 네트워크 특성을 파악하고 있으며, 이 때 중심성은 해당 액터가 네트워크 내에서 얼마나 중요한 위치에 있는지를 나타낸다. 먼저 연결 중심성은 노드의 연결 정도를 기반으로 산출되며, 액터의 활동성을 다룬다. 다음으로 근접 중심성은 네트워크에서 노드가 갖는 간접적 연결까지 고려하며, 액터의 독립성을 강조하는 개념이다. 또한 매개 중심성은 자신과 직접 연결되지 않은 노드 간 관계의 중개 또는 통제 정도를 측정하며, 액터의 통제성을 나타낸다(Freeman, 1978; Scott, 2017; Wasserman and Faust, 1994).

최근에 수행된 관련 예로는 소셜 네트워크 상에서 새로운 유형의 영향력 노드를 찾기 위해 연결 중심성, 근접 중심성, 역 근접 중심성(Reverse-closeness Centrality), 매개 중심성 등을 활용한 연구(Saito et al., 2016), 조직적으로 중요

한 금융 기관을 찾기 위해 연결 중심성, 근접 중심성, 매개 중심성, 보나시치 중심성(Bonacich's Centrality) 등을 활용한 연구(Kuzubaş et al., 2014), 기업에서 이루어지는 돈 세탁(Money Laundering) 방지를 위해 기업 거래 네트워크에 연결 중심성, 근접 중심성, 매개 중심성 등의 분석을 수행한 연구(Colladon and Remondi, 2017) 등이 있으며, 이 외에도 소셜 네트워크 분석에서는 다양한 목적으로 사용될 수 있는 새로운 네트워크 지표가 지속적으로 고안되고 있다.

2.3 네트워크 지표 발굴에 대한 연구

소셜 네트워크 분석의 여러 지표는 다양한 환경에서 공통적으로 발생하는 구조적 정보의 파악을 위해 고안되었으며, 네트워크 관점의 유용한 정보를 제공하고 있다. 하지만 이러한 범용성을 갖는 기존의 네트워크 지표는 특정 상황에서 특수하게 발생할 수 있는 현상을 충분히 설명하지 못하며, 이러한 한계의 극복을 위해 일부 연구에서는 기존 네트워크 지표의 변형 및 새로운 지표의 발굴을 통해 연구 목적에 부합하는 분석을 수행하고 있다.

최근 수행된 연구의 예로는 문서와 저자로 구성되는 이분 네트워크에서 문서의 인용 정보 및 저자의 사회적 인지도를 통해 새로운 지표를 고안하고 이를 통해 학술 논문의 수상을 예측한 연구(Kinouchi et al., 2018), 특정 액터의 부재시 발생하는 단일 액터 정보의 제거를 통해 새로운 관점에서 영향력 있는 액터를 식별한 연구(Wang et al., 2017), 네트워크의 매개 중심성, 내향 연결 중심성, 지역 중심성을 근간으로 핵심 단백질 식별을 위한 새로운 지표를 고안한 연구(Qin et al., 2016), 핵심 단백질 예측의 성과 향상을 위해 가

증화된 연결 중심성을 새로운 지표로 제안한 연구(Tang et al., 2014) 등을 들 수 있으며, 이렇게 발굴된 새로운 지표는 기존 지표를 통해 파악하기 어려운 특수한 상황에 대한 정보를 제공하고 있다.

융합 기술은 인접 기술 간 영향 관계를 통해 생성되고 성장하기 때문에, 일반적 기술에 비해 매우 복합적인 특성을 갖는 것으로 알려져 있다. 이에 따라 해당 분야의 연구에서는 특정 기술의 향후 전망 및 미래 상황을 예측함에 있어 많은 어려움을 겪고 있으며, 이를 목적으로 하는 연구는 단순히 융합 기술의 과거 현상을 분석, 파악하는 연구에 비해 활성화 되지 못하고 있다. 일부 연구에서 특허 문서에 대한 정량적 분석을 통해 융합 기술의 향후 전망을 예측하는 시도가 이루어진 바 있으나(Altuntas et al., 2015), 이와 같은 방식은 전술한 바와 같이 융합 기술이 갖는 특수성을 충분히 활용하지 못했다는 한계를 갖는다. 다시 말해서 단순히 데이터에서 얻을 수 있는 정보만을 활용한 연구에서는 융합 기술의 향후 변화를 제대로 파악하기 어렵다는 것을 의미한다. 이러한 맥락에서 특정 융합 기술에 대한 향후 전망 파악을 목적으로 하는 융합 기술 연구에서는 융합의 원리를 충분히 반영하는 새로운 지표의 발굴을 통해 네트워크 분석을 수행할 필요가 있다.

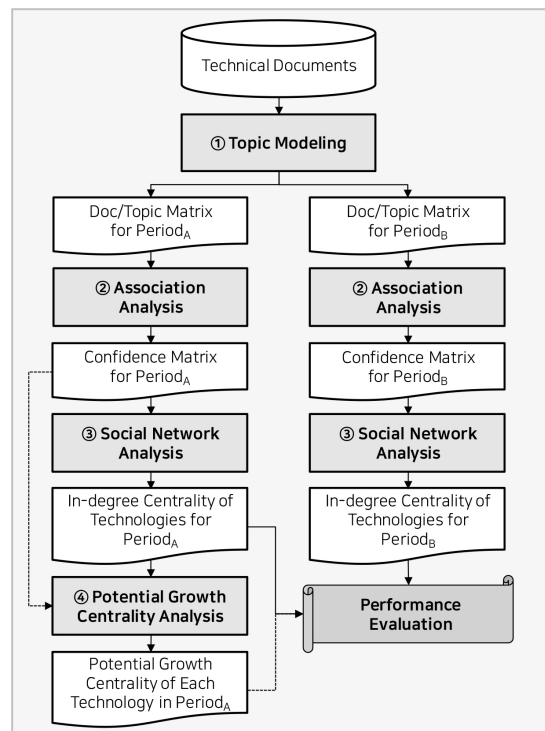
3. 제안 방법론

3.1 연구 모형

본 절에서는 이미 출현한 융합 기술을 식별하고, 이들의 잠재 성장 가능성을 파악하기 위한

방안을 제시한다. 구체적으로 기술적 내용의 문서에 대한 토픽 모델링을 통해 융합 기술을 식별하고, 이들이 갖는 연관 관계에 대한 네트워크 분석을 통해 특정 기술의 잠재 성장 가능성을 파악할 수 있는 잠재 성장 중심성 지표를 새롭게 제안한다. 제안 방법론의 전체적 개요는 <Figure 2>와 같으며, 아래의 단계로 구성된다.

먼저 기술적 내용의 문서에 대한 토픽 모델링을 통해 주요 기술을 식별하고, 문서별 토픽 가중치 정보를 현재 시점(Period A)과 미래 시점(Period B)으로 구분하여 추출한다(①). 다음으로 각 시점별로 기술 간 공유 문서 수에 기반을 둔 연관성 분석을 수행하여, 그 결과로 기술 간 의존 관계 네트워크를 구성한다(②). 이후 네트워



<Figure 2> Research Overview

크에 대한 소셜 네트워크 분석을 통해 각 기술의 내향 연결 중심성을 산출하고, 이 값을 해당 기술의 성숙도, 즉 해당 기술이 다른 기술에 영향을 미치는 정도로 파악한다(③). 마지막으로 각 기술의 인접 기술 중 성숙도가 큰 핵심 기술을 식별하고, 해당 기술과 핵심 기술 간 관계의 강도를 반영하여 해당 기술의 잠재 성장 중심성을 산출한다(④). 이와 더불어 현재 시점 기준으로 산출한 지표값에 따라 유망 기술을 선정하고, 이들 유망 기술의 성장 여부를 측정하여 제안 방법론의 성능을 평가한다. 본 장에서는 제안 방법론의 이해를 돕기 위해 자동차 산업 분야에 대해 가상으로 도입한 예를 설명하며, 실제 데이터를 분석한 실험 결과는 4장에서 제시한다.

3.2 융합 기술 식별 및 기술 간 의존 관계 파악

본 절은 <Figure 2>의 ①과 ②에 해당하는 과정, 즉 기술적 내용의 문서에 토픽 모델링을 적용하여 융합 기술을 식별하고, 이를 통해 도출한 문서별 토픽 가중치를 시기별로 구분하여 추출하는 과정과 이를 토대로 기술 간 의존 관계를 파악하는 과정을 소개한다. 우선 분석을 위해 수집된 문서는 분석 대상 산업군에 해당하는 기술의 정보를 충분히 다루고 있어야 한다. 구체적으로 기술적 내용의 문서는 새롭게 개발된 개별 기술의 상세 정보를 설명하는 특허 문서뿐만 아니라, 특정 기술에 대한 전문적 연구 성과를 나타내는 학술 논문 등을 포함할 수 있다.

본 방법론에서는 고정된 기술 분류 기준을 사용하는 대신 기술 문서에 대한 토픽 모델링을 통해 주요 기술 토픽을 도출하므로, <Figure 3>의 좌측 상단 표와 같이 기존에 존재하지 않던 새로

운 융합 기술군의 정의가 가능하다. 도출된 토픽 중 Topic 1과 Topic 2는 각각 자동차 분야의 전통적 영역인 조향 및 파워트레인 관련 기술이며, Topic 3은 IT 관련 기술 토픽으로 파악된다. 또한 Topic 4와 Topic 5는 모두 자동차 분야와 타 산업의 융합을 통해 나타난 토픽으로 각각 자율주행 및 친환경 차량 관련 기술을 나타낸다.

토픽 모델링은 이미 많은 텍스트 분석 연구에서 상세히 다루고 있으므로(Blei et al., 2003; Choi and Kim, 2017; Deerwester et al., 1990; Kim et al., 2017), 본 연구에서는 토픽 모델링에 대한 자세한 개념 소개는 생략하고 제안 방법론에 활용된 주요 과정만을 설명한다. 토픽 모델링의 산출물은 크게 토픽에 대한 용어의 참여 정보를 나타내는 용어/토픽 행렬(Term/Topic Matrix) 및 토픽에 대한 문서의 참여 정보를 나타내는 문서/토픽 행렬(Document/Topic Matrix)로 나뉜다. 또한 각 행렬은 특정 토픽에 대한 용어 및 문서의 부합 정도를 나타내는 용어 가중치(Term Weight) 및 문서 가중치(Document Weight)를 값으로 가지며, 문서/토픽 행렬의 경우 일반적으로 문서 가중치의 ‘평균 + 10’의 문서 임계값(Document Cutoff)을 적용하여 특정 토픽에 대한 문서의 참여 여부를 <Figure 3>의 좌측 하단 도표와 같은 형태로 나타낸다.

분석에 사용된 각 문서는 공개 일시와 같은 시기 정보를 포함한다. 이를 통해 전체 문서/기술 행렬은 현재 시점을 가정하는 지표 생성 시기와 미래 시점을 가정하는 지표 검증 시기로 구분하여 추출될 수 있다. 본 예는 <Figure 3>의 좌측 하단 도표와 같이 지표 생성 시기에 해당하는 문서 650건과 지표 검증 시기에 해당하는 문서 550건이 전체 분석 데이터를 구성하는 경우를 나타낸다. 또한 이후 설명에서는 지표 생성 시기를

Topic Information							
1	스티어링휠, 토크, 기어, 컨트롤러, 센서						
2	엔진, 연료, 펌프, 모터, 제어						
3	통신, 신호, 감지, 서버, 센서						
4	위치, 속도, 공간, 센서, 제어						
5	전기, 배터리, 충전, 모터, 제어						

Total	Topic1	Topic2	Topic3	Topic4	Topic5
Doc.A-1	1	0	1	0	0
Doc.A-2	1	0	0	1	0
Doc.A-3	1	0	1	0	0
Doc.A-4	0	1	0	0	1
Doc.A-5	0	1	1	1	1
⋮					
Doc.A-650	0	1	0	1	1

Period A	Topic1	Topic2	Topic3	Topic4	Topic5
Doc.A-1	1	0	1	0	0
Doc.A-2	1	0	0	1	0
Doc.A-3	1	0	1	0	0
Doc.A-4	0	1	0	0	1
Doc.A-5	0	1	1	1	1
⋮					
Doc.A-650	0	1	0	1	1

Period B	Topic1	Topic2	Topic3	Topic4	Topic5
Doc.B-1	1	1	1	1	0
Doc.B-2	0	1	1	1	1
Doc.B-3	1	0	0	1	0
Doc.B-4	1	0	1	1	0
Doc.B-5	1	0	0	1	1
⋮					
Doc.B-550	0	1	1	1	0

〈Figure 3〉 Topic Information and Document/Topic Matrix for Each Period

‘Period A’, 지표 검증 시기를 ‘Period B’로 지칭하며, <Figure 3>의 우측 도표와 같은 방식으로 구분하여 사용한다.

다음으로 이렇게 생성된 정보에 대한 기술 단위의 연관성 분석을 수행함으로써, 시기별 기술 간 의존 관계 네트워크를 형성할 수 있다. 산업 기술을 분석 대상으로 하는 연구에서는 대부분 단일 기술의 동시 분류 정보를 통해 기술 간 연관 관계를 분석한다. 하지만 융합 기술은 다른 일반 기술에 비해 그 속성 자체가 빠르게 변화하는 경향을 보이며, 이로 인해 시간의 흐름에 따라 해당 기술이 속한 산업군 자체가 변화할 여지가 있다. 따라서 본 연구에서는 각 기술에 대해 고정된 분류 체계가 아닌, 각 기술이 포함하고 있는 공통 주제에 기반을 두어 연관성 분석을 수

행한다.

연관성 분석은 데이터의 동시 발생 빈도 분석을 통해 이들 간의 유의미한 관계를 발견하여 규칙으로 나타내는 기법으로, 소셜 네트워크 분석 및 텍스트 분석에서 다양한 목적으로 활용된 바 있다(Cho et al., 2012; Kim et al., 2016). 이 기법은 도출된 규칙의 평가를 위해 주로 지지도(Support), 신뢰도(Confidence), 향상도(Lift)의 세 가지 흥미성 척도를 사용하고 있다. 예를 들어 규칙 $A \rightarrow B$ 가 있을 때, 지지도는 전체 트랜잭션 수에 대한 A, B의 동시 출현 수의 비율로, 신뢰도는 A의 출현 수에 대한 A, B의 동시 출현 수의 비율로 정의된다. 마지막으로 향상도는 A와 B의 상관관계를 뜻하며, 값이 1이면 독립, 값이 1보다 크면 양의 상관관계, 값이 1보다 작으면 음의

상관관계를 나타낸다. 이들 중 신뢰도는 기술 문서 분석에서 기술 A에 속한 문서가 기술 A와 B에 동시에 속할 조건부 확률을 나타낼 수 있다. 예를 들어 기술 A에 속한 문서의 대부분이 기술 B에도 속해 있다면, 기술 A가 기술 B에 의존하는 정도, 즉 기술 B가 기술 A에 영향을 미치는 정도가 매우 크다고 할 수 있다. 따라서 기술 의존도 네트워크에서 노드 간 링크로 나타내는 신뢰도는 특정 기술이 타 기술에 의존하는 정도로 이해할 수 있다.

본 연구에서는 연관성 규칙의 신뢰도를 활용하여 기술 간 의존 관계를 파악하였다. 이와 관

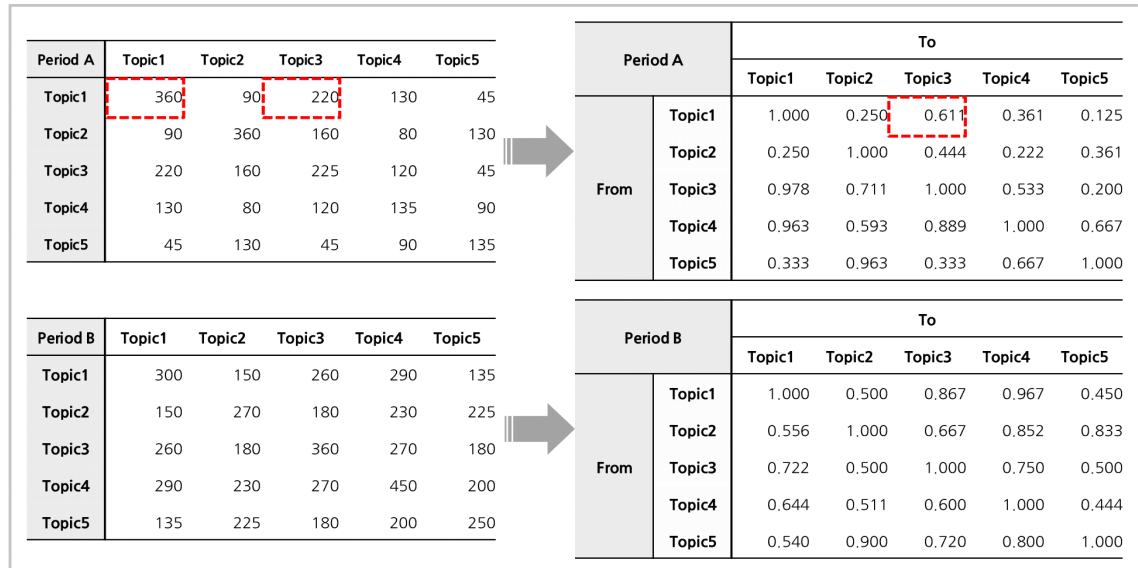
련한 표기는 <Table 1>에 제시하였으며, 값 산출을 위한 수식은 다음과 같다.

$$Conf.(T_i, T_j) = \frac{p(T_i, T_j)}{p(T_i)}$$

토픽 간 신뢰도의 산출을 위해서는 각 토픽별 문서 수 및 각 토픽 쌍에 동시에 출현하는 문서 수의 집계가 필요하다. 이 값은 <Figure 3>의 우측 도표로부터 산출할 수 있으며, 그 결과는 <Figure 4>의 좌측 도표와 같은 형태로 나타난다. 이를 통해 <Figure 4>의 좌측 도표로부터 모

<Table 1> Notations for Confidence between Topics

Notation	Description
$p(T_i)$	토픽 “Topic i”에 참여하는 문서의 수
$p(T_i, T_j)$	토픽 “Topic i” 및 토픽 “Topic j”에 동시에 참여하는 문서의 수
$Conf.(T_i, T_j)$	토픽 “Topic i”의 기술이 토픽 “Topic j”의 기술에 대해 갖는 의존도



<Figure 4> Confidence between Each Pair of Topics

든 토픽 쌍에 대한 신뢰도를 산출할 수 있으며, 그 결과가 <Figure 4>의 우측 도표로 나타나있다. 예를 들어 Period A에서 토픽 Topic 3에 대한 토픽 Topic 1의 신뢰도는 다음과 같이 산출된다.

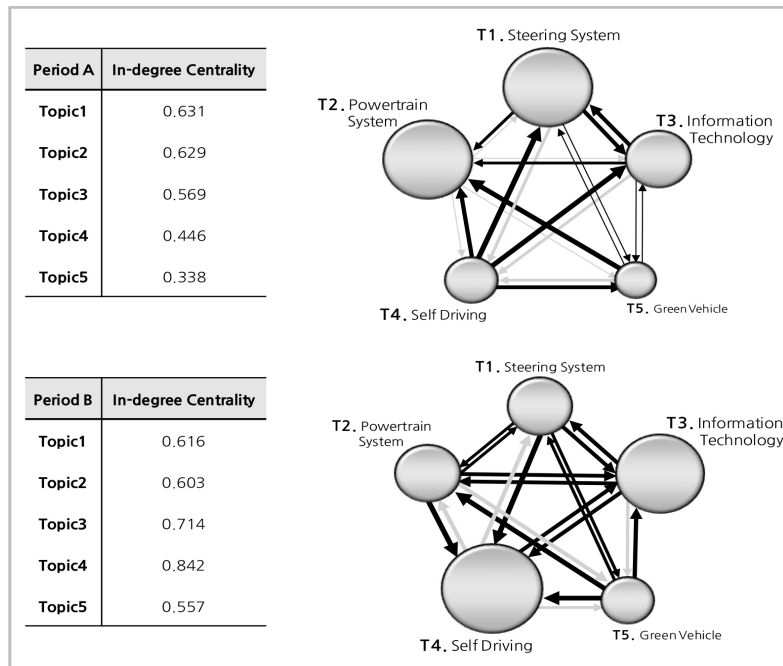
$$Conf.(T_1, T_3) = \frac{220}{360} = 0.611$$

3.3 융합 기술의 기술 성숙도 파악

본 절은 <Figure 2>의 ③에 해당하는 과정, 즉 각 융합 기술 및 이들 간 의존도를 활용하여 융합 기술 네트워크를 구성하고, 이에 대한 소셜 네트워크 분석을 통해 각 기술의 성숙도를 파악하는 과정을 소개한다. 앞서 설명한 바와 같이, 전체 융합 기술은 각 기술 및 기술 간 의존도를 노드와 링크로 가지는 방향/계량 네트워크 형태

로 구성된다. 이에 따라 융합 기술 간 의존 관계에 기반한 소셜 네트워크 분석이 가능하며, 여러 중심성 지표를 통해 각 기술의 구조적 정보를 파악할 수 있다. 이 중 연결 중심성은 각 액터가 갖는 연결 정도 또는 강도에 기반하여 이들의 위치적 우위를 판단하며, 특히 내향 연결 중심성은 특정 액터가 타 액터로부터 지목 받은 정도를 통해 산출되기 때문에 중요한 또는 인기있는 액터를 판단하는 역할을 수행한다. 이에 따라 본 연구는 각 토픽 간 신뢰도, 즉 융합 기술 간 의존 관계를 통해 각 융합 기술에 대한 내향 연결 중심성을 산출하고, 이를 해당 기술의 성숙도를 나타내는 지표로 선정한다.

<Figure 5>의 좌측 도표는 Period A, B의 각 기술에 대한 내향 연결 중심성을 분석한 결과이며, 이 정보를 바탕으로 시기별 기술 성숙도 및 기술



<Figure 5> In-degree Centrality and Topic Network

간 의존도를 노드의 크기와 링크의 강도로 표현하는 <Figure 5>의 우측 네트워크를 형성할 수 있다. 우선 <Figure 5>의 우측 상단에 위치한 Period A의 네트워크를 살펴보면, Topic 1, Topic 2 그리고 Topic 3의 기술이 다른 기술군에 비해 상대적으로 성숙하였음을 파악할 수 있으며, Topic 4 기술의 경우 Topic 1 및 Topic 3 기술에 대해 강한 의존 관계를 갖는 것을 확인할 수 있다. 또한 <Figure 5>의 우측 하단에 위치한 Period B의 네트워크에서는 Topic 3 및 Topic 4 기술이 갖는 기술 성숙도 및 타 기술군에 대한 영향력이 두드러지며, Period A에 비해 Period B에서 Topic 4 기술이 매우 성숙한 분야로 성장하였음을 알 수 있다.

3.4 잠재 성장 중심성 산출

본 절은 <Figure 2>의 ④에 해당하는 과정, 즉 이전 과정에서 도출한 융합 기술의 정보를 활용하여 각 기술의 잠재 성장 중심성을 산출하는 과정을 소개한다. 본 절의 목적은 본 연구의 핵심인 융합의 원리를 반영하는 새로운 네트워크 중심성, 즉 잠재 성장 중심성의 발굴 과정을 설명하는 것이다. 다시 말해서 현재는 해당 기술의 성숙도가 낮고 타 기술에 대한 의존도가 높지만, 충분히 성숙한 여러 기술과의 강한 관계를 갖고 있기 때문에 향후 더욱 성장할 가능성이 높은 기술군을 식별하기 위한 새로운 지표를 고안하고자 한다.

잠재 성장 중심성은 융합 기술에서 나타나는 세 가지 특성을 반영한다. 먼저 첫 번째 속성은 기술의 융합적 성장 가능성을 뜻하는 잠재성(Potential)이다. 또한 두 번째 속성은 성숙 기술에 대한 융합 기술의 밀접한 관계를 나타내는 인

접성(Adjacency)이다. 마지막으로 세 번째 속성은 융합이 발생하는 원리를 반영하는 융합성(Convergence)이다. 초기 단계의 융합 기술은 해당 기술과 강한 관계를 갖는 성숙 단계 기술의 정보를 받아들임으로써 성장하며, 일정 수준 이상으로 성장한 융합 기술은 독자적인 기술 개발을 통해 성숙 단계의 기술로 발전하는 패턴을 보인다. 또한 융합 기술은 융합의 발생 원리에 따라 둘 이상 기술군의 영향을 받아 출현하고 성장한다는 특성을 갖는다. 이에 따라 새 지표의 산출 과정에는 특정 기술이 여전히 융합을 통해 성장할 가능성이 있는지를 판단하고, 이들과 밀접한 관계를 갖는 성숙 기술의 정보를 반영함과 함께 단일 인접 기술이 아닌 여러 개의 인접 기술에 대한 정보를 동시에 반영할 필요가 있다.

잠재 성장 중심성의 산출 과정을 구체적으로 설명하면 다음과 같다. 먼저 분석 대상 시점의 모든 토픽에 대한 내향 연결 중심성을 파악하고, 임계값(Threshold) 이상의 중심성을 갖는 토픽을 충분히 성숙한 분야의 기술군으로 선별하여 이를 토큰(Token)으로 지정한다. 토큰 기술은 향후 타 기술의 성장에 도움을 주는 핵심 기술로 동작한다. 이러한 토큰 기술을 활용하여 각 관심 기술의 잠재 성장 중심성을 산출하는 식은 다음과 같으며, 이와 관련된 주요 용어의 정의는 <Table 2>와 같다.

$$PGC_N(T_i) = \frac{\sum_{j=1}^N Conf.(T_i, TK_j) \times IC.(TK_j)}{N \times IC.(T_i)}$$

예를 들어 Period A 시점에서 Topic 4 기술의 잠재 성장 중심성을 산출하는 것이 분석의 목적이라고 하자. 이를 위한 첫 번째 과정은 토큰 토픽을 식별하는 것이다. 본 예에서는 성숙 단계의

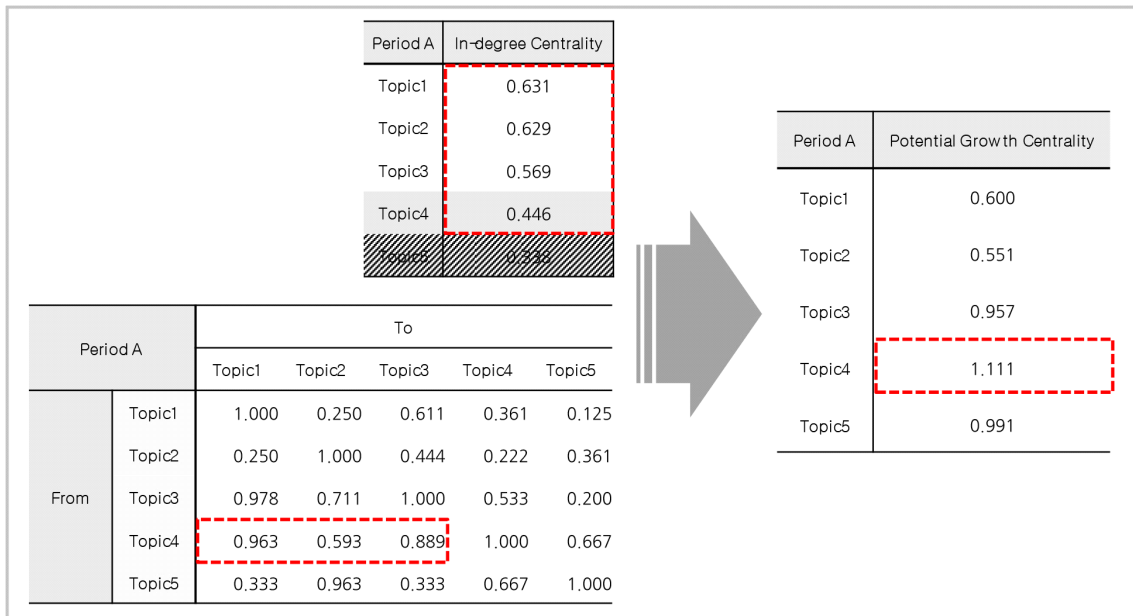
〈Table 2〉 Notations for Calculating Potential Growth Centrality of Topic

Notation	Description
$Conf.(T_i, TK_j)$	“Token Topic j”에 대한 융합 기술 “Topic i”의 의존도 (= 신뢰도)
$IC.(TK_j)$	“Token Topic j”의 내향 연결 중심성
$IC.(T_i)$	융합 기술 “Topic i”의 내향 연결 중심성
$PGC_N(T_i)$	상위 토큰 N개를 기준으로 한 융합 기술 “Topic i”의 잠재 성장 중심성

기술을 판단하는 임계값을 ‘0.5’ 이상의 내향 연결 중심성으로 가정하였으며, 이에 따라 <Figure 5>의 좌측 도표에 의해 Topic 1, Topic 2, Topic 3의 기술이 토큰으로 지정된다. 이때 상위 토큰

3개에 대한 융합 기술 Topic 4의 잠재 성장 중심성은 다음과 같이 계산되며, 동일한 방식으로 모든 융합 기술에 대한 잠재 성장 중심성을 산출할 수 있다(Figure 6).

$$PGC_N(T_i) = \frac{(0.963 \times 0.631) + (0.593 \times 0.629) + (0.889 \times 0.569)}{3 \times 0.446} = \mathbf{1.111}$$



〈Figure 6〉 Potential Growth Centrality of Topics for Period A

3.5 제안 방법론의 우수성 평가 방안

본 절에서는 본 연구에서 제안한 잠재 성장 중심성 기반 유망 기술 예측 방법론의 우수성 평가 방안을 소개한다. 제안 방법론의 검증은 잠재 성장 중심성을 근거로 유망 기술의 순위를 부여하고, 유망성이 높은 기술의 성숙도 증가폭과 유망성이 낮은 기술의 성숙도 증가폭의 차이를 측정하는 방식으로 수행하고자 한다. 이때 성숙도 증가폭이 클수록 제안 방법론의 예측력이 우수한 것으로 이해할 수 있다. 또한 기존의 다양한 지표를 기준으로 동일한 방법의 평가를 수행하여 기존 지표와 제안 지표의 예측력 성능을 비교하고자 한다.

<Figure 7>에서 Period A의 잠재 성장 중심성 (PGC)은 Topic 4 > Topic 5 > Topic 3 > Topic 1 > Topic 2의 순으로 높게 나타남을 알 수 있다. 또한 이들 토픽의 내향 연결 중심성의 증가량 역시 <Figure 7>의 좌측 도표에서 확인할 수 있다. 이를 바탕으로 PGC 기준 상위 두 토픽인 Topic 4와 Topic 5의 내향 연결 중심성 증가량의 평균은 ‘0.308’, PGC 기준 하위 두 토픽인 Topic 1과 Topic 2의 내향 연결 중심성 증가량의 평균은 ‘-0.021’임을 <Figure 7>의 우측 도표와 같이 확인할 수 있다. 따라서 상위 토픽과 하위 토픽의

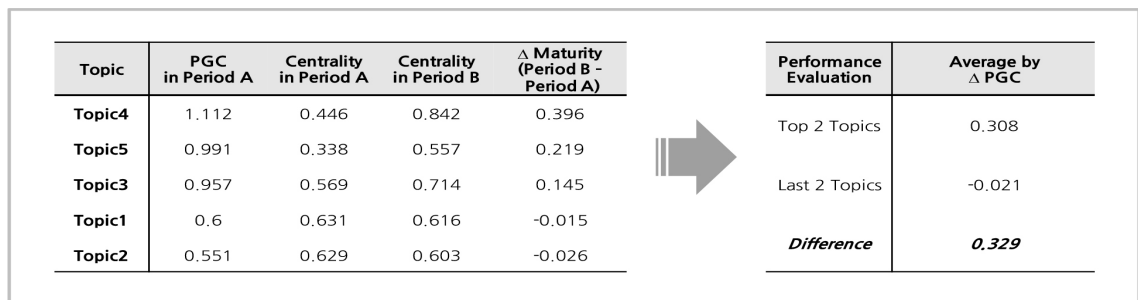
내향 연결 중심성 증가량 차이는 ‘0.329’로 나타나며, 이는 PGC에 따라 부여한 순위가 향후 해당 기술의 성숙도 예측에 유의하게 사용될 수 있음을 암시한다.

본 절에서는 앞에서 다룬 간단한 예를 통해 제안 모델의 우수성 평가 방안을 소개하였으며, 실제 실험 결과는 다음 장에서 소개한다. 실제 실험의 분석에서는 제안 지표 및 다양한 기존 지표를 활용하여 위와 같은 내향 연결 중심성 증가량 평균의 차이뿐만 아니라 각 지표의 순위에 따른 성숙도 증가량의 변화 양상을 파악하고, 이를 누적 향상도 도표(Cumulative Lift Chart)를 통해 보다 자세히 제시한다.

4. 실험

4.1 실험 개요

본 절에서는 실험 환경 및 데이터에 대해 간략히 소개한다. 실험의 토픽 모델링은 SAS Enterprise Miner Workstation 14.1을 통해 수행하였으며, 불용어 사전(Stop List)을 활용하여 결과를 정제하였다. 또한 실험의 소셜 네트워크 분석은 NetMiner 4.0을 통해 수행하였다. 실험 데이



<Figure 7> Example of Performance Evaluation

터는 2011년 1월부터 2016년 12월까지의 기간 동안 국내에 등록된 특허 중 2017년 10월 기준 권리가 소멸되지 않은 특허 문서이며, 특허 요약 문에 “자동차”, “차량” 또는 “차”의 용어가 포함된 문서만을 분석 대상으로 하였다. 데이터는 특허 DB검색 사이트인 WISDOMAIN을 통해 수집하였으며, 수집된 전체 특허의 수는 총 13,477건이다.

4.2 제안 방법론의 적용 및 평가

4.2.1 융합 기술 간 의존 관계 및 기술 성숙도 파악

본 부절에서는 3.2 ~ 3.3절에서 소개된 과정에 따라 실제 특허 문서를 분석한 결과를 제시한다. 먼저 융합 기술의 식별을 위해 전체 13,477건의 특허 문서에서 20개의 토픽과 토픽별 5개의 핵

Topic Information		Topic Information	
1	영상, 카메라, 촬영, 인식, 이미지	11	엔진, 연료, 밸브, 가스, 펌프
2	프레임, 상부, 하부, 본체, 레일	12	센서, 정보, 주행, 위치, 속도
3	배터리, 충전, 전기, 전력, 전압	13	주차, 공간, 위치, 센서, 관리
4	수지, 조성물, 중량, 폴리프로필렌, 열가소성	14	운전자, 정보, 주행, 운전, 사고
5	정보, 통신, 단말기, 서버, 시스템	15	부재, 회전, 하우징, 방향, 샤프트
6	엔진, 토크, 모터, 제어, 클러치	16	제조, 냉연강판, 불순물, 강판, 중량
7	공조, 온도, 케이스, 실내, 증발기	17	브레이크, 제동, 압력, 제어, 페달
8	신호, 센서, 감지, 제어부, 제어	18	하우징, 램프, 모듈, 내부, 운전자
9	충돌, 충격, 흡수, 에어백, 사고	19	자동차, 센서, 조향, 커버, 상부
10	시트, 요소, 위치, 등받이, 쿠션	20	도어, 유닛, 패널, 프레임, 위치

〈Figure 8〉 Topic Information

Period 1	Topic1	Topic2	Topic3	Topic4	Topic5	...	Topic20
Topic1	208	5	5	0	63		6
Topic2	5	415	15	2	4		72
Topic3	5	15	252	0	19	...	11
Topic4	0	2	0	199	0		2
Topic5	63	4	19	0	363		24
⋮			⋮				
Topic20	6	72	11	2	24		325

〈Figure 9〉 Count of Shared Documents for Each Pair of Topics in Period 1 (Part)

심 용어를 도출하였으며, 분석 결과는 <Figure 8>과 같다. 이와 더불어 각 토픽쌍 별 공유 문서의 수를 Period 1(2011~2012년), Period 2(2013~2014년), Period 3(2015~2016년)으로 구분하였으며, 결과의 일부를 <Figure 9>에 제시하였다.

다음으로 <Figure 9>를 토대로 토픽 간 신뢰도, 즉 기술 간 의존도를 산출하여 그 결과의 일

부를 <Figure 10>에 제시하였다. 이후 <Figure 10>의 토픽을 노드로, 의존도를 링크 가중치로 부여하여 시기별 네트워크를 구성하였으며, 이에 대한 소셜 네트워크 분석을 통해 각 토픽이 갖는 시기별 내향 연결 중심성, 즉 기술 의존 관계에 따른 기술 성숙도를 산출하였다(Figure 11).

Period 1		To						
		Topic1	Topic2	Topic3	Topic4	Topic5	...	Topic20
From	Topic1	1.00000	0.02404	0.02404	0.00000	0.30288		0.02885
	Topic2	0.01205	1.00000	0.03614	0.00482	0.00964		0.17349
	Topic3	0.01984	0.05952	1.00000	0.00000	0.07540	...	0.04365
	Topic4	0.00000	0.01005	0.00000	1.00000	0.00000		0.01005
	Topic5	0.17355	0.01102	0.05234	0.00000	1.00000		0.06612
	⋮			⋮				
	Topic20	0.01846	0.22154	0.03385	0.00615	0.07385		1.00000

〈Figure 10〉 Dependency between Each Pair of Topics in Period 1 (Part)

Topic	Maturity (In-degree Centrality)		
	Period 1	Period 2	Period 3
Topic1	0.06505	0.06631	0.07823
Topic2	0.11867	0.10212	0.09409
Topic3	0.06932	0.06471	0.06286
Topic4	0.02678	0.03178	0.02959
Topic5	0.10514	0.09267	0.11586
Topic6	0.10242	0.10251	0.09853
Topic7	0.05864	0.06408	0.05102
Topic8	0.14618	0.13604	0.12990
Topic9	0.08564	0.08987	0.07795
Topic10	0.05161	0.06286	0.06655

Topic	Maturity (In-degree Centrality)		
	Period 1	Period 2	Period 3
Topic11	0.09413	0.09230	0.07513
Topic12	0.16194	0.16779	0.16126
Topic13	0.05599	0.06129	0.05217
Topic14	0.12413	0.12930	0.13365
Topic15	0.13372	0.13035	0.10882
Topic16	0.04151	0.03333	0.03623
Topic17	0.09964	0.10721	0.09542
Topic18	0.13243	0.10779	0.12656
Topic19	0.11991	0.12534	0.10880
Topic20	0.11981	0.12554	0.11294

〈Figure 11〉 Maturity(In-degree Centrality) of Topics

4.2.2 잠재 성장 중심성 산출

본 부절에서는 3.4절에 소개된 과정에 따라, 실제 특허 문서를 분석하여 잠재 성장 중심성을 산출한 결과를 제시한다. 잠재 성장 중심성은 현재 시점에 특정 기술과 인접 기술 간 의존 관계를 통해 산출되며, 이를 통해 해당 기술의 향후 성장 가능성을 예측한다. 우선 타 기술의 성장에 영향을 끼치는 핵심 기술(토큰)로 인정받기 위한 노드의 내향 연결 중심성 임계값을 ‘0.1’로 설정하였다. 또한 3.4절에 소개된 방법에 따라 전체 토픽 20개에 대한 PGC 값을 산출하였으며, Period 1과 Period 2의 PGC 값이 <Figure 12>에 나타나있다.

4.2.3 잠재 성장 중심성에 따른 유망 기술 예측의 성능 평가

본 부절에서는 잠재 성장 중심성을 근거로 유망 기술의 순위를 부여하고, 유망성이 높은 기술의 성숙도 증가폭과 유망성이 낮은 기술의 성

숙도 증가폭 차이를 측정하는 방식으로 제안 방법론의 성능을 평가한다. 또한 평가에서는 현재 시점과 미래 시점을 (Period 1, Period 2), (Period 2, Period 3), 그리고 (Period 1, Period 3)의 세 가지 경우로 설정하였으며, 실험에 따른 20개의 토픽이 갖는 성숙도 증가폭(= 미래 시점의 성숙도 - 현재 시점의 성숙도)은 <Figure 13>에 요약하였다.

우선 잠재 성장 중심성에 따른 유망 기술과 비유망 기술의 성숙도 증가량을 비교하기 위해, 상위/하위 토픽의 개수를 각 5개로 지정하였다. 또한 각 융합 기술이 영향을 받고 있는 인접 핵심 기술의 수, 즉 토큰의 수는 3개로 한정하였다. 이때, <Figure 7>의 우측 도표와 동일한 방식으로 산출한 상위 토픽과 하위 토픽의 내향 연결 중심성 증가량 평균의 차이를 <Table 3>에 요약하였다. <Table 3>는 제안 방법론에 의한 값(PGC)뿐만 아니라 단순한 내향 연결 중심성에 근거하여 예측한 값(IC), 그리고 아이젠벡터 중심성에 근거하여 예측한 값(EC)을 함께 비교하여 제시하

Topic	Potential Growth Centrality	
	Period 1	Period 2
Topic1	0.61301	0.59447
Topic2	0.25642	0.28659
Topic3	0.41717	0.33915
Topic4	0.25108	0.27996
Topic5	0.41090	0.50669
Topic6	0.31340	0.27998
Topic7	0.46540	0.34437
Topic8	0.24424	0.25508
Topic9	0.32287	0.27279
Topic10	0.62707	0.43596

Topic	Potential Growth Centrality	
	Period 1	Period 2
Topic11	0.21039	0.26098
Topic12	0.23942	0.20981
Topic13	0.71543	0.59489
Topic14	0.41776	0.33987
Topic15	0.21550	0.18558
Topic16	0.25077	0.34468
Topic17	0.36684	0.30525
Topic18	0.23498	0.22709
Topic19	0.23849	0.18088
Topic20	0.24631	0.22360

<Figure 12> Potential Growth Centrality of Topics in Period 1 and Period 2

고 있다. <Table 3>에서 보는 바와 같이, 제안 방법론의 PGC에 따른 유망 기술 예측 모형이 내향 연결 중심성 또는 아이겐벡터 중심성에 기반을 둔 모형에 비해 유망 기술의 성숙도 성장을 더 정확하게 예측함을 알 수 있다.

다음으로 제안 방법론의 성능을 보다 엄밀히 비교하기 위해 각 기술의 유망도에 따른 성숙도 증가량을 <Figure 14>의 누적 향상도 도표로 나타내었으며, 이때 그래프는 각 모형에서 예측한 상위 유망 기술에서 시작되는 성숙도 증가량의

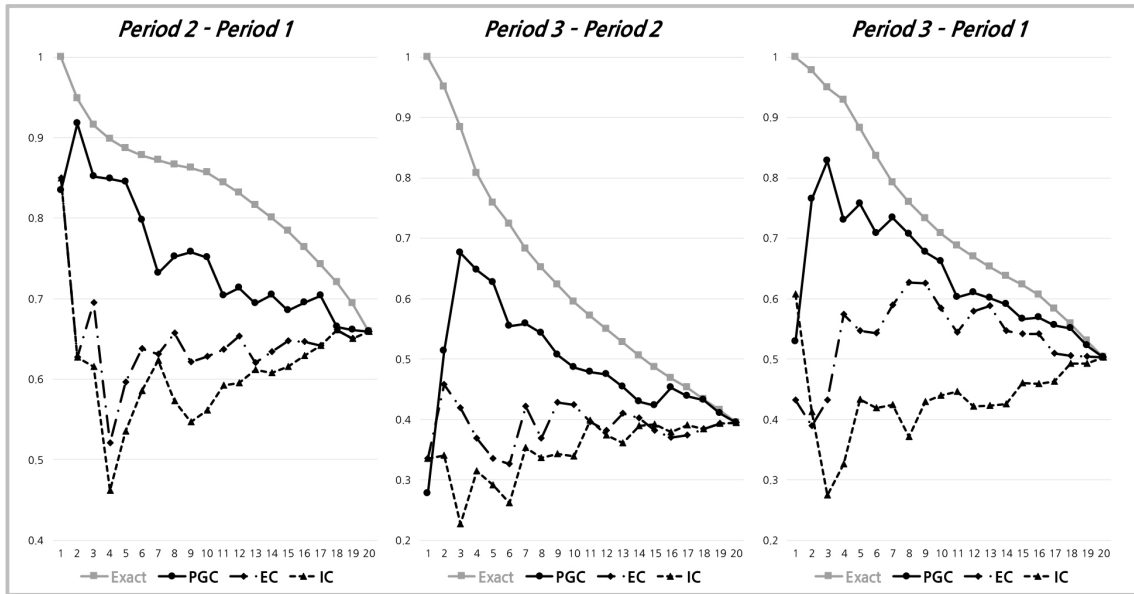
누적 평균으로 도출된다. 본 분석에서는 각 기술의 성숙도 증가량을 0에서 1 사이의 값으로 정규화 하였으며, <Table 3>에서 활용된 PGC, IC, 그리고 EC뿐만 아니라, 모든 성숙도가 알려진 상태에서 동작하는 이상적 모형의 수치(Exact)를 비교에 함께 포함하였다. 분석 결과 예측 모형의 우수성은 $PGC > EC > IC$ 순에 따라 높은 것으로 파악되었으며, 제안 모델인 PGC의 경우 이상적 모형인 Exact와 비교적 근접하게 나타남을 알 수 있다.

Topic	Maturity (In-degree Centrality)			Topic	Maturity (In-degree Centrality)		
	Period 2 - Period 1	Period 3 - Period 2	Period 3 - Period 1		Period 2 - Period 1	Period 3 - Period 2	Period 3 - Period 1
Topic1	0.00125	0.01192	-0.01901	Topic11	-0.00184	-0.01717	-0.00389
Topic2	-0.01655	-0.00803	0.00282	Topic12	0.00584	-0.00652	-0.00769
Topic3	-0.00461	-0.00185	-0.02490	Topic13	0.00530	-0.00912	0.01073
Topic4	0.00501	-0.00219	-0.00527	Topic14	0.00517	0.00435	-0.00422
Topic5	-0.01247	0.02320	-0.00068	Topic15	-0.00337	-0.02153	0.00952
Topic6	0.00009	-0.00398	-0.01628	Topic16	-0.00818	0.00291	-0.00646
Topic7	0.00544	-0.01305	-0.00587	Topic17	0.00758	-0.01180	-0.00762
Topic8	-0.01014	-0.00614	-0.01111	Topic18	-0.02465	0.01878	0.01494
Topic9	0.00423	-0.01192	-0.02459	Topic19	0.00542	-0.01653	0.01317
Topic10	0.01126	0.00369	-0.00687	Topic20	0.00573	-0.01260	-0.00382

〈Figure 13〉 Maturity Growth of 20 Topics

〈Table 3〉 Performance Comparison among PGC and Traditional Network Measures

Model	Maturity (In-degree Centrality)		
	Period 2 - Period 1	Period 3 - Period 2	Period 3 - Period 1
PGC	0.00940	0.01420	0.01755
IC (In-degree Centrality)	0.00919	0.00493	0.00785
EC (Eigenvector Centrality)	0.00357	0.00444	0.00634



〈Figure 14〉 Cumulative Lift Chart for PGC and Traditional Measures

본 장에서는 제안 방법론의 실무 적용 가능성 평가를 위해 제안 방법론을 실제 특허 문서 분석에 적용한 실험 결과를 제시하였다. 실험 결과 융합 기술의 원리를 반영하는 지표, 즉 잠재 성장 중심성 기반의 예측 모형이 전통적인 네트워크 지표에 기반을 둔 예측 모형에 비해 유망 기술의 성숙도 성장을 보다 정확하게 예측함을 확인하였다.

5. 결론

최근 단일 기술의 고도화에 따라 많은 기술이 융합을 통해 생성 및 발전하고 있으며, 이러한 융합 기술의 연구가 다방면에서 활발하게 수행되고 있다. 특히 데이터 분석을 통해 융합 기술의 성장을 사전에 예측하기 위한 수요가 점차 증

가하고 있다. 이에 본 연구에서는 토픽 모델링을 통해 융합 기술을 식별하고, 연관 분석을 통해 기술별 의존 관계를 파악하고, 네트워크 분석을 통해 기술 간 의존도와 기술의 성숙도를 도출한 뒤, 이러한 지표에 융합의 원리를 반영하여 유망 기술 예측을 위한 새로운 지표인 잠재 성장 중심성을 도출하는 방안을 제시하였다. 또한 실제 특허 문서 13,477건을 활용한 실험을 통해 제안 지표가 유사 목적에 사용되는 기존 네트워크 지표에 비해 유망 융합 기술의 성장을 더욱 정확하게 예측할 수 있음을 확인하였다.

본 연구의 학술 및 실무적 기여는 다음과 같다. 먼저 학술적 측면에서 제안 방법론의 가장 큰 기여는 융합의 원리를 반영하는 새로운 네트워크 지표를 제안하였다는 점이다. 즉 충분히 성숙한 인접 기술의 영향을 받아 생성 및 성장하는 융합 기술의 특성을 지표 산출 과정에 반영함으

로써, 융합 기술의 향후 성장을 예측하기 위한 연구의 토대를 마련하였다. 또한 기술적 내용의 문서에 토픽 모델링을 적용하여 융합 기술을 식별하고, 이에 대한 연관성 분석을 통해 기술 간 의존 관계를 파악하는 방법 역시 텍스트 분석을 활용한 융합 기술 연구의 새로운 방향을 제시하였다는 점에서 기여를 인정받을 수 있다. 이와 더불어 실무적 측면에서는 제안 방법론에 따라 유망 기술을 예측함으로써, 여러 기관 및 연구자들이 빠르게 변화하는 기술 지형에 선제적으로 대응할 수 있을 것으로 기대한다.

그럼에도 불구하고 본 연구는 향후 다음의 측면에서 보완이 필요하다. 먼저 분석을 통해 식별된 융합 기술군의 계층(Hierarchy)을 구분할 필요가 있다. 본 연구에서는 문서의 토픽 모델링을 통해 나타난 융합 기술을 하나의 기술군으로 파악하여 분석을 수행하였지만, 이들 간의 계층 정보를 추가적으로 활용한다면 더욱 유의미한 통찰을 얻을 수 있을 것으로 판단된다. 또한 잠재 성장 중심성 지표의 산출 과정이 더욱 정교화 될 필요가 있다. 물론 제안 방법론에서의 검증 및 평가 과정을 통해 기존 지표 대비 제안 지표의 우수성을 확인하였지만, 해당 지표는 견고한 이론적 토대를 근거로 도출된 것이 아니라 방대한 양의 반복 실험을 통해 경험적으로 고안되었다는 한계를 갖는다. 따라서 향후 후속 연구에서는 융합의 원리를 더욱 체계적으로 반영하기 위한 수학적 보완이 이루어져야 한다. 이와 더불어 본 연구에서는 제안 방법론의 성능 분석이 전통적 네트워크 지표에 기반을 둔 모형과의 비교를 통해 이루어졌지만, 향후 다양한 분야에서 유사 목적으로 고안된 예측 모형과의 성능 비교를 통해 제안 방법론의 우수성을 보다 엄밀히 검증할 필요가 있다. 마지막으로 본 연구에서 새롭게 제안

된 방안이 현업의 상황을 제대로 반영할 수 있는지에 대한 실무적 측면에서의 고찰이 필요하다.

참고문헌(References)

- Altuntas, S., T. Dereli and A. Kusiak, "Forecasting Technology Success based on Patent Data," *Technological Forecasting and Social Change*, Vol.96, (2015), 202~214.
- Blei, D. M., A. Y. Ng and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, Vol.3, (2003), 993~1022.
- Cho, I., N. Kim and K. -Y. Kwahk, "Recommending Core Research Keywords using Social Network and Data Mining Analysis," *Entrue Journal of Information Technology*, Vol.11, No.1(2012), 87~99.
- Choi, H. and N. Kim, "Efficient Topic Modeling by Mapping Global and Local Topics," *Journal of Intelligence and Information Systems*, Vol.23, No.3(2017), 69~94.
- Colladon, A. F. and E. Remondi, "Using Social Network Analysis to prevent Money Laundering," *Expert Systems with Applications*, Vol.67, (2017), 49~58.
- Curran, C.-S. and J. Leker, "Patent Indicators for Monitoring Convergence - Examples from NFF and ICT," *Technological Forecasting and Social Change*, Vol.78, No.2(2011), 256~273.
- Deerwester, S., S. T. Dumais, G. W. Furnas, T. K. Landauer and R. Harshman, "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science*, Vol.41, No.6(1990), 391~407.

- Freeman, L. C. (Ed.), *Research Methods in Social Network Analysis*, Routledge, Oxfordshire, 2017.
- Freeman, L. C., "Centrality in Social Networks Conceptual Clarification," *Social Networks*, Vol.1, No.3(1978), 215~239.
- Huang, B. and L. Huang, "Patent-based Association Analysis for Technological Convergence of IT and BT," *Proceedings of 2015 International Conference on Logistics, Informatics and Service Sciences (LISS)*, IEEE, (2015), 1~6.
- Jeong, Y., I. Park and B. Yoon, "Forecasting Technology Substitution based on Hazard Function," *Technological Forecasting and Social Change*, Vol.104, (2016), 259~272.
- Kim, E., Y. Cho and W. Kim, "Dynamic Patterns of Technological Convergence in Printed Electronics Technologies: Patent Citation Network," *Scientometrics*, Vol.98, No.2(2014), 975~998.
- Kim, J. -E., Y. Hyun and Y. -J. Choi, "Discovery of Market Convergence Opportunity combining Text Mining and Social Network Analysis: Evidence from Large-scale Product Databases," *Journal of Intelligence and Information Systems*, Vol.22, No.4(2016), 87~107.
- Kim, N., D. Lee, H. Choi and W. X. S. Wong, "Investigations on Techniques and Applications of Text Analytics," *The Journal of Korean Institute of Communications and Information Sciences*, Vol.42, No.2(2017), 471~492.
- Kinouchi, O., L. D. H. Soares and G. C. Cardoso, "A Simple Centrality Index for Scientific Social Recognition," *Physica A: Statistical Mechanics and its Applications*, Vol.491, No.1(2018), 632~640.
- Kuzubaş, T. U., I. Ömercikoğlu and B. Saltoğlu, "Network Centrality Measures and Systemic Risk: An Application to the Turkish Financial Crisis," *Physica A: Statistical Mechanics and its Applications*, Vol.405, (2014), 203~215.
- Kwahk, K. -Y., *Social Network Analysis*, 2nd ed., Cheongram, Seoul, 2017.
- Lee, H. and H. Zo, "R&D Allies: How they impact Technology Convergence in the Area of ICT," *Proceedings of 2016 International Conference on Information and Communication Technology Convergence (ICTC)*, IEEE, (2016), 340~343.
- Lee, K., "Toward a New Paradigm of Technological Innovation: Convergence Innovation," *Asian Journal of Technology Innovation*, Vol.23, Sup.1(2015), 1~8.
- Lee, W. J., W. K. Lee and S. Y. Sohn, "Patent Network Analysis and Quadratic Assignment Procedures to Identify the Convergence of Robot Technologies," *PloS one*, Vol.11, No.10(2016), e0165091.
- Li, Y., Y. Li, Y. Zhao and F. Wang, "Which Factor Dominates the Industry Evolution? A Synergy Analysis Based on China's ICT Industry," arXiv preprint arXiv:1403.4305, (2014). Available at <https://arxiv.org/abs/1403.4305> (Accessed 10 November, 2017).
- Marin, A. and B. Wellman, "Social Network Analysis: An Introduction," *The SAGE Handbook of Social Network Analysis*, (2011), 11~25.
- Park, J. H., K. -Y. Kwahk, H. Han and Y. Kim, "A Study on the Determinants of Patent Citation Relationships among Companies:

- MR-QAP Analysis,” *Journal of Intelligence and Information Systems*, Vol.19, No.4(2013), 21~37.
- Park, J. H. and K. -Y. Kwahk, “The Effect of Patent Citation Relationship on Business Performance: A Social Network Analysis Perspective,” *Journal of Intelligence and Information Systems*, Vol.19, No.3(2013), 127~139.
- Qin, C., Y. Sun and Y. Dong, “A New Method for Identifying Essential Proteins based on Network Topology Properties and Protein Complexes,” *PloS one*, Vol.11, No.8(2016), e0161042.
- Rosenberg, N., “Technological Change in the Machine Tool Industry, 1840-1910,” *The Journal of Economic History*, Vol.23, No.4 (1963), 414~443.
- Saito, K., M. Kimura, K. Ohara and H. Motoda, “Super Mediator-A New Centrality Measure of Node Importance for Information Diffusion over Social Network,” *Information Sciences*, Vol.329, (2016), 985~1000.
- Scott, J., *Social Network Analysis*, 4th ed., Sage, London, 2017.
- Tang, X., J. Wang, J. Zhong and Y. Pan, “Predicting Essential Proteins based on Weighted Degree Centrality,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, Vol.11, No.2(2014), 407~418.
- Wang, S., Y. Du and Y. Deng, “A New Measure of Identifying Influential Nodes: Efficiency Centrality,” *Communications in Nonlinear Science and Numerical Simulation*, Vol.47, (2017), 151~163.
- Wasserman, S. and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, 1994.

Abstract

Discovering Promising Convergence Technologies Using Network Analysis of Maturity and Dependency of Technology

Hochang Choi* · Kee-Young Kwahk** · Namgyu Kim***

Recently, most of the technologies have been developed in various forms through the advancement of single technology or interaction with other technologies. Particularly, these technologies have the characteristic of the convergence caused by the interaction between two or more techniques. In addition, efforts in responding to technological changes by advance are continuously increasing through forecasting promising convergence technologies that will emerge in the near future. According to this phenomenon, many researchers are attempting to perform various analyses about forecasting promising convergence technologies. A convergence technology has characteristics of various technologies according to the principle of generation. Therefore, forecasting promising convergence technologies is much more difficult than forecasting general technologies with high growth potential. Nevertheless, some achievements have been confirmed in an attempt to forecasting promising technologies using big data analysis and social network analysis.

Studies of convergence technology through data analysis are actively conducted with the theme of discovering new convergence technologies and analyzing their trends. According that, information about new convergence technologies is being provided more abundantly than in the past. However, existing methods in analyzing convergence technology have some limitations. Firstly, most studies deal with convergence technology analyze data through predefined technology classifications. The technologies appearing recently tend to have characteristics of convergence and thus consist of technologies from various fields. In other words, the new convergence technologies may not belong to the defined classification. Therefore, the existing method does not properly reflect the dynamic change of the convergence phenomenon.

Secondly, in order to forecast the promising convergence technologies, most of the existing analysis method use the general purpose indicators in process. This method does not fully utilize the specificity of

* Graduate School of Business IT, Kookmin University

** School of MIS, Kookmin University

*** Corresponding Author: Namgyu Kim

School of MIS, Kookmin University

77 Jeongneung-ro, Seongbuk-gu, Seoul 136-702, Korea

Tel: +82-2-910-5425, Fax: +82-2-910-5209, E-mail: ngkim@kookmin.ac.kr

convergence phenomenon. The new convergence technology is highly dependent on the existing technology, which is the origin of that technology. Based on that, it can grow into the independent field or disappear rapidly, according to the change of the dependent technology. In the existing analysis, the potential growth of convergence technology is judged through the traditional indicators designed from the general purpose. However, these indicators do not reflect the principle of convergence. In other words, these indicators do not reflect the characteristics of convergence technology, which brings the meaning of new technologies emerge through two or more mature technologies and grown technologies affect the creation of another technology.

Thirdly, previous studies do not provide objective methods for evaluating the accuracy of models in forecasting promising convergence technologies. In the studies of convergence technology, the subject of forecasting promising technologies was relatively insufficient due to the complexity of the field. Therefore, it is difficult to find a method to evaluate the accuracy of the model that forecasting promising convergence technologies. In order to activate the field of forecasting promising convergence technology, it is important to establish a method for objectively verifying and evaluating the accuracy of the model proposed by each study.

To overcome these limitations, we propose a new method for analysis of convergence technologies. First of all, through topic modeling, we derive a new technology classification in terms of text content. It reflects the dynamic change of the actual technology market, not the existing fixed classification standard. In addition, we identify the influence relationships between technologies through the topic correspondence weights of each document, and structuralize them into a network. In addition, we devise a centrality indicator (PGC, potential growth centrality) to forecast the future growth of technology by utilizing the centrality information of each technology. It reflects the convergence characteristics of each technology, according to technology maturity and interdependence between technologies. Along with this, we propose a method to evaluate the accuracy of forecasting model by measuring the growth rate of promising technology. It is based on the variation of potential growth centrality by period.

In this paper, we conduct experiments with 13,477 patent documents dealing with technical contents to evaluate the performance and practical applicability of the proposed method. As a result, it is confirmed that the forecast model based on a centrality indicator of the proposed method has a maximum forecast accuracy of about 2.88 times higher than the accuracy of the forecast model based on the currently used network indicators.

Key Words : Technology Convergence, Social Network Analysis, Promising Technology Forecast, Potential Growth Centrality, Topic Modeling

Received : December 10, 2017 Revised : February 26, 2018 Accepted : March 5, 2018

Publication Type : Regular Paper(Fast-track) Corresponding Author : Namgyu Kim

저 자 소 개



최 호 창

현재 국민대학교 비즈니스IT전문대학원 석사과정에 재학 중이다. 국민대학교 경영학부에서 학사 학위를 취득하였으며, 주요 관심분야는 텍스트 마이닝, 데이터 마이닝 및 소셜네트워크분석 등이다.



광 기 영

현재 국민대학교 경영대학과 비즈니스IT전문대학원 교수로 재직 중이다. 서울대학교 경영대학을 졸업하고 KAIST 경영과학과와 테크노경영대학원에서 석사 및 박사학위를 취득하였다. 주요 연구관심분야는 Social network analysis and its application, R-based data analytics, Social communication ecology, IT-enabled organizational agility, Knowledge management 등이다.



김 남 규

현재 국민대학교 경영정보학부 부교수로 재직 중이다. 서울대학교 컴퓨터공학과에서 학사 학위를 취득하고, KAIST 테크노경영대학원에서 Database와 MIS를 전공하여 경영공학 석사 및 박사학위를 취득하였다. 한국지능정보시스템학회 부회장, 한국정보기술응용학회 부회장, 한국경영학회 상임이사, 한국경영정보학회 이사, 한국인터넷정보학회 이사, 한국CRM학회 이사를 역임하였다. 주요 관심분야는 Text Mining, Data Mining, Data Modeling 등이다.