

용어 확장을 통한 핀테크 기술 적용가능 산업의 탐색 :네트워크 분석 및 토픽 모델링 접근

Exploring Potential Application Industry for Fintech Technology by Expanding its Terminology: Network Analysis and Topic Modelling Approach

박민규(Mingyu Park)*, 전병민(Byeongmin Jeon)**,
김종우(Jongwoo Kim)***, 금영정(Youngjung Geum)****

초 록

핀테크는 기술주도형 금융혁신을 위한 중요한 사업 영역으로 논의되어 왔다. 핀테크라는 용어는 금융과 기술의 합성어로, 현재 모든 금융 분야와 연계된 ICT 기술을 의미한다. 핀테크 산업에 관한 관심은 시간이 지날수록 크게 높아져 수많은 스타트업에 대한 진폭적인 투자와 지원이 이루어지고 있다. 이에 따라 학문적, 실무적 시점에서 핀테크 동향을 연구하려는 시도가 있어 왔다. 그러나 시장의 높은 관심에도 불구하고 기존 연구는 핀테크 기술에 대한 명확하고 체계적인 정의 및 확장 절차가 부족하였다. 이러한 이유로 핀테크 기술을 바탕으로 적절한 응용 분야를 파악하는 것 역시 한계가 존재하였다. 이에 본 연구는 세 가지 방식을 결합하여 핀테크 관련 용어를 확장하고, 네트워크 분석과 토픽 모델링을 활용해 핀테크 분야 동향을 분석하는 새로운 방법을 제안한다. 이를 위해 새로운 핀테크 용어목록을 생성하고, USPTO로부터 10년간 총 1만 8341건의 특허를 수집하였다. 수집된 특허를 바탕으로 Co-classification 분석과 네트워크 분석을 통해 핀테크 특허의 기술 동향을 파악하였고, 핀테크 도메인을 분석하기 위해 핀테크 트렌드를 파악하는 토픽 모델링을 실시하였다. 이번 연구는 기술주도형 금융 서비스에 참여하고자 하는 경영자와 투자자가 핀테크 기술의 정의와 범위를 정확히 이해하고, 이를 바탕으로 핀테크 기술의 트렌드를 다각적으로 분석함으로써 새로운 핀테크 기술 기회를 포착하는 데 도움이 될 것으로 기대된다.

ABSTRACT

FinTech has been discussed as an important business area towards technology-driven financial innovation. The term fintech is a combination of finance and technology, which

본 연구는 교육부 및 한국연구재단 (보호연구지원사업) 의 지원을 받아 수행된 연구임(NRF- 2020R111A2070429).

* First Author, Master Course of Datascience, Seoul National University of Science and Technology (mgpark@ds.seoultech.ac.kr)

** Co-Author, Undergraduate course of Industrial Information System Engineering, Seoul National University of Science and Technology(dkffkam@naver.com)

*** Co-Author, Undergraduate course of Industrial Information System Engineering, Seoul National University of Science and Technology(Whddn12316@naver.com)

**** Corresponding Author, Associate Professor, Dept. of Industrial & Systems Engineering, Seoul National University of Science and Technology(yjgeum@seoultech.ac.kr)

Received: 2020-09-21, Review completed: 2021-02-10, Accepted: 2021-02-16

means ICT technology currently associated with all finance areas. The popularity of the fintech industry has significantly increased over time, with full investment and support for numerous startups. Therefore, both academia and practice tried to analyze the trend of the fintech area. Despite the fact, however, previous research has limitations in terms of collecting relevant databases for fintech and identifying proper application areas. In response, this study proposed a new method for analyzing the trend of Fintech fields by expanding Fintech's terminology and using network analysis and topic modeling. A new Fintech terminology list was created and a total of 18,341 patents were collected from USPTO for 10 years. The co-classification analysis and network analysis was conducted to identify the technological trends of patent classification. In addition, topic modeling was conducted to identify the trends of fintech in order to analyze the contents of fintech. This study is expected to help both managers and investors who want to be involved in technology-driven financial services seize new FinTech technology opportunities.

키워드 : 핀테크, 기술 동향, 텍스트 마이닝, 네트워크 분석, 토픽 모델링
Fintech, Technology Trend, Text Mining, Network Analysis, Topic Modeling

1. 서 론

핀테크(Fintech)란 Financial과 Technology가 합쳐진 단어로써 금융 문제의 해결책을 제시하기 위해 정보통신(Information and Communication Technology, 이하 ICT) 기술을 사용하는 것을 의미한다[2]. 최근 금융업계는 유망한 ICT 기업들을 인수하여 새롭고 다양한 ICT 기반의 금융서비스를 제공하고 있다. 금융 산업은 모바일 결제뿐만 아니라 보험, 자산관리, 증권, 투자 등의 기존 금융서비스가 기술을 바탕으로 크게 확대되고 있으며, 이러한 중심에 핀테크 기술의 성장이 있다고 알려져 있다[28]. 또한, 인공지능, 블록체인 등의 최신 ICT 기술로 인해 고객과 고객이 직접 금융활동을 수행하는 C2C 금융거래가 일어나고 있을 뿐 아니라, 사물인터넷(Internet of Things, 이하 IoT)과의 결합으로 고객의 일상생활에서 더욱 빠르고 간편한 금융 활동이 가능해졌다.

핀테크 산업은 전 세계적으로 2018년 기준

약 1,276억 달러의 시장 가치를 가지고 있으며, 25%의 연평균 성장률과 함께 2022년에는 약 3,099억 달러의 시장 가치를 가질 것으로 예상된다[12]. 특히 전세계 핀테크 스타트업들은 2018년에 약 390억 달러의 투자를 받아 2017년 대비 그 가치가 120% 상승했으며 새로운 개념의 은행으로서 Monzo, Chime 등과 같은 Digital Challenger Bank가 등장했고, Revolut, Plaid, Nubank, Policy Bazar, Tiger Brokers 등의 신흥 유니콘 기업들이 핀테크 시장의 새로운 강자로 나타나게 되었다[7]. 즉, 핀테크 기술은 전통적 기업들이 혁신의 흐름에 맞춰 ICT 기술과 금융서비스를 접목하고 투자하도록 견인하는 중요한 역할을 해 왔다.

미국은 2018년 기준 전 세계 핀테크 시장의 57%를 차지하고 핀테크 스타트업 투자금 80% 이상을 차지할 정도로 시장규모가 거대하다[13]. 이는 미국 내 소비자의 편의성, 보안성, 단순성, 투명성, 개인화 등 핀테크 혁신으로 활용할 수 있는 핵심 서비스에 투자한 결과라고 볼 수 있다.

그리고 인공지능을 활용한 자동 투자 시스템과 같이 도덕적, 사회적 문제가 되는 기술에 대한 규제 시스템에도 과감한 투자를 아끼지 않고 있다. 한국에서는 최근에 실시된 데이터3법을 통해 오픈뱅킹 서비스가 가능해지면서 핀테크와 은행업계에 큰 전환점이 찾아오게 되었다[23]. 이는 금융 시장에서 다양한 기술이 융합되어 새로운 서비스와 가치를 창출할 기회가 열린 것이다. 따라서 핀테크 기술의 높은 성장 가능성에 초점을 맞추고, 핀테크 기술의 핵심적 특성 및 확장가능성을 면밀히 검토하여, 핀테크를 혁신적 비즈니스로 연계하고 시장을 확대하여 다양한 혁신을 시도하는 것이 요구된다. 이를 위해서는 기술기획의 가장 기본적이고도 가장 핵심적인 활동인 기술동향 분석이 보다 체계적인 프로세스를 통해 수행되는 것이 중요하다.

이러한 이유로 다수 연구에서 핀테크 기술에 대한 기술동향 분석이 수행되어왔다. Seo and Lee[31], Lee and Choi[22], Kim et al.[16]의 연구에서는 핀테크 관련 특허를 수집하여 토픽 모델링 분석으로 기술동향을 살펴본 바 있으며, Li et al.[26]은 한국의 핀테크 온라인 뉴스를 수집하여 텍스트 분석을 통해 서비스와 기술의 동향을 살펴본 바 있다.

다양한 기존 연구에서 핀테크 동향을 성공적으로 연구해 왔으나, 기존 연구들은 다음과 같이 세 가지 한계가 존재한다. 먼저 핀테크 기술의 정의 관점에서 살펴보면, 대부분의 기존 연구에서는 핀테크 기술을 명확히 정의하고 있지 않다. 핀테크는 그 용어에서부터 알 수 있듯이 Finance와 Technology의 결합으로, 금융분야에 사용되는 기술을 뜻한다. 그러나 금융분야에 사용되는 기술 자체는 그 범위가 너무 넓다는

한계가 있다. 시장에서 사용되는 통상적 Fintech의 개념이 존재하기는 하지만 연구자에 따라 핀테크 자체를 정의하는 관점 및 의견이 다를 뿐 아니라, 핀테크 기술의 경계가 뚜렷하지 않아 기술 자체를 정확하게 정의하는 것이 매우 어렵기 때문이다. 이러한 이유로 핀테크 기술동향을 파악하기 위한 핀테크 기술 자체의 정의가 매우 단편적으로 이루어지고 있는 실정이다.

둘째로 핀테크 기술의 수집 관점에서 살펴보자. 핀테크 기술의 수집을 위해 다수 연구들이 핀테크 기술동향 분석을 위해 특허 또는 신문 기사를 활용해 왔다[16, 22, 26, 31]. 그러나 대부분 연구에서 데이터 수집 과정이 체계적이지 않고 키워드, IPC 등 분석가의 관점에 한정하여 데이터가 수집되어왔다. 대부분 연구에서 핀테크라는 키워드 그 자체를 사용해서 관련 기사 및 특허를 수집하거나[22, 26], 특정 USPC나 IPC를 통해 핀테크 관련 기술을 정의하거나[26, 31], 사전 등을 활용하여 자체적 필터링 기법을 이용해 핀테크 기술을 한정하고 있다[6].

세 번째로 핀테크 기술의 적용 산업 관점에서 살펴보자. 기존 연구들은 주로 핀테크 기술 그 자체에 집중하여 분석해 왔으며, 핀테크 기술이 가져올 수 있는 다양한 산업의 파급효과에 대한 심도있는 분석은 등한시되어 왔다. 그러나 핀테크 기술이 금융 관련 기술이기는 하지만, 금융 및 결제 관련 기술은 다양한 산업 활동이 이루어지기 위한 중요한 수단에 관련된 것이다. 따라서 핀테크 기술은 금융산업에만 국한되는 것이 아니라, 다양한 산업에 확대되고 발전할 수 있는 가능성이 큰 기술이며, 이에 핀테크 기술의 다양한 적용 가능성에 대한 집중적 분석을 수행할 필요가 있다.

요약하면 기존 연구들의 경우 핀테크 기술을 분석하기 위한 핀테크 기술 그 자체에 대한 정의가 충분하지 않다는 점, 이러한 이유로 핀테크 기술에 대한 체계적인 수집 과정이 정의되지 않았다는 점, 그리고 핀테크 기술의 분석을 금융산업이 아닌 다양한 산업으로 확장하여 그 확장가능성을 살펴보는 데 소홀하였다는 점 등이 한계로 지적될 수 있다. 이에 본 연구는 핀테크에 대한 정의를 보다 명확하게 하여 핀테크 시장을 충분히 반영할 수 있는 핀테크 특허 수집 방법을 제시하고 그에 따른 기간별 특허 데이터를 가지고 네트워크 분석과 토픽 모델링 분석을 진행하여 최신 핀테크 기술동향을 파악하고 기술의 적용가능 산업을 탐색하고자 한다.

본 연구에서는 이를 위해 CPC를 통한 네트워크 분석을 활용하였다. 특허 CPC는 특허 분류체계 중 하나이며, 특허의 분류체계는 개별 특허들을 큰 틀에서 유사한 산업군/기술군으로 분류하기 위한 체계 중 하나이다. 현존하는 특허 분류체계는 체계적이면서도 section, class, subclass, group, subgroup 으로 구성된 위계적 구조로 구성되어 있으며, 기술적 특성을 가지고 있는 개별 특허의 분류체계이기 때문에 그 자체로 기술적 측면에서의 산업의 특성을 가진다는 특성이 있어 기존 연구에서 산업 분석 도구로 많이 활용되어 왔다[7, 10]. 이러한 CPC 분류체계를 바탕으로 하는 네트워크 분석은 주요 산업 간의 융합을 손쉽게 관찰할 수 있게 해 줌으로써 핀테크 적용가능 산업을 파악하는데 용이하다고 볼 수 있다[9].

또한 토픽모델링의 수행을 통해 각 문서의 토픽을 도출하고, 이를 바탕으로 토픽 군집의 기술적 특성을 파악할 수 있다[27]. 이는 특허 기술에 관한 세부 특징을 살펴볼 수 있을 뿐

아니라 개별 특허에서 나타나는 기술적 특징을 토픽이라는 보다 거시적인 측면에서 살펴봄으로써 적용가능 산업을 더 면밀하게 관찰할 수 있게 한다.

본 연구의 구성은 다음과 같다. 제 2장에서는 본 연구와 관련된 선행 연구들을 살펴본다. 제 3장에서는 본 연구에서 새롭게 제시하는 핀테크 용어 수립 방법과 함께 연구의 전체적인 프로세스를 서술한다. 제 4장에서는 핀테크 특허 데이터를 바탕으로 분석하여 도출된 네트워크와 토픽을 살펴본다. 마지막으로 제 5장에서는 연구의 결론과 의의, 한계점 및 추후 연구 방향에 관해 서술한다.

2. 이론적 배경

2.1 핀테크 정의에 관한 연구

시장의 높은 관심에도 불구하고, 핀테크에 관한 명확한 정의를 탐구한 연구는 많지 않다 [32]. 현재 핀테크의 의미를 깊이 있게 탐구한 Arner et al.[2]는 지금의 핀테크가 진화하기까지의 시대를 세 개의 기간으로 나눠 기술의 진보와 함께 금융의 발전을 설명한다. 핀테크라는 단어는 1980년대 선데이터임이라는 뉴스에서 처음으로 등장했다고 알려졌지만, 많은 벤처 캐피털 회사들이 핀테크에 투자하기 시작한 2007년부터 일반적으로 사용하게 되었다. 하지만 Arner et al.[2]는 핀테크라는 단어가 가진 의미를 고려했을 때 그 역사를 1866년부터 봐야 한다고 주장한다. Fintech 1.0(1866~1967)은 세계 1차 대전을 필두로 개발되는 통신, 정보 기술이 금융 세계화에 큰 영향을 미치는 시기며,

Fintech 2.0(1967~2008)은 아날로그에서 디지털로 전환되는 시점으로서 ATM의 활성화, 금융기관들의 전산화, 온라인뱅킹 등장의 시기이다. 마지막으로 Fintech 3.0(2008-present)은 GFC(Global Finance Crisis) 이후로 금융 시장이 전환 상태를 맞이한 시기이며, 혁신적인 스타트업들이 금융 기술개발에 뛰어들어 모바일, 블록체인, AI, IoT 등의 새로운 시장과 유연한 금융 시스템이 등장하게 된 시기이다. 이러한 핀테크의 역사적 흐름과 현재 시점의 위치를 고려했을 때 핀테크는 다음과 같이 Finance and Investment(Crowdfunding, P2P lending, Robo-advisory services), Operations and Risk Management, Payments and Infrastructure, Data Security and Monetization, Customer Interface의 다섯 가지 유형으로 분류할 수 있다[2].

Schueffel[29]은 핀테크라는 용어가 다양한 비즈니스 맥락에서 적용됐기에 일관성이 부족하고 모호한 개념으로 사용됐다고 주장한다. 지금까지 핀테크에 대한 의미를 함의적으로 정의하려는 시도가 없었기에, 금융서비스의 견고한 과학적 기반을 조성하기 위해 공통적인 의미가 도출되어야 한다고 한다. 보다 명확한 개념 정의를 위해 Schueffel[29]은 서지학적 정보에서 의미분석을 적용하여 핀테크 정의의 유사성을 검토했다. 문서에서 나타난 공통점 중 특별한 점은 핀테크를 새로운 것, 신흥인 것, 혁신적인 것, 교란을 일으키는 것으로 표현한다는 것이다. 그리하여 핀테크란 금융활동을 개선하기 위해 기술을 적용하여 새로운 금융산업이라 정의할 수 있다.

Zavolokina et al.[33]은 핀테크 의미의 모호성이 핀테크 산업의 급속한 성장 때문이라 주장하며 전문 보고서나 과학적 연구뿐만 아니라 미디어에서 발생하는 핀테크 현상까지 참고하여 핀

테크를 정의하고자 하였다. Zavolokina et al.[33]은 핀테크가 어떻게 인식되고 있는지, 시간에 지남에 따라 핀테크에 영향을 미치는 주요 행위자는 누구인지, 언론에서 논의되는 핀테크 주제는 무엇이 있는지에 대해 설명한다. 연구자들은 핀테크는 기술, 조직, 돈의 흐름이 생성, 변화, 개선되거나 파괴되어 새로운 서비스, 제품, 절차, 비즈니스 모델을 만든다고 설명하였으며, 그리고 금융기관, IT 기업, 스타트업, 액셀러레이터, 컨설팅그룹, 정부조직, 교육기관 등 순서의 주요 행위자가 존재한다고 그 특성을 언급하였다.

2.2 핀테크 기술에 관련한 정량적 기술동향 연구

핀테크 기술이 시장 및 학계의 관심을 끌면서, 핀테크 기술에 대한 다양한 기술동향 분석 연구가 시도되어 왔다. 그중 가장 대표적인 연구는 특허를 활용한 동향 분석 연구이다. 특허 문서는 기술적, 상업적 측면에서 풍부한 지식 원천이 되기에 특정 산업의 기술 및 시장 동향을 살펴볼 수 있는 유용한 수단이다[20]. 또한, 기업의 투자와 특허 포트폴리오 사이에 긍정적인 상관관계가 있다고 보고되었다[25]. 이러한 특허의 높은 가치에 따라 핀테크의 기술동향 탐색을 위한 다양한 특허 분석이 이뤄졌다. 특허 분류기준인 USPC의 705/35는 금융과 관련된 데이터 처리 기술로서 핀테크의 핵심기술을 담고 있는 특허를 나타낸다. USPC의 705/35를 기준으로 특허 데이터를 수집하여 토픽 모델링과 네트워크 분석을 활용한 연구는 핀테크 핵심기술의 동향을 제시했다[31]. 하지만 USPC가 2015년 이후로 사용되지 않기에 최신 특허를 활용할 수 없다는 점과 705/35만으로 핀테크 관련

특허를 모두 수집하는 데 한계가 있다.

IPC는 전 세계에서 일반적으로 사용하는 특허 분류 기준으로 특정 도메인에 관한 특허 수집의 지표가 될 수 있다. Kim and Shim[15]은 핀테크와 관련되어 있다고 생각되는 IPC에 따라 특허를 수집하고 중심 국가별 중심 특허를 제시했다. 하지만 IPC 역시 광범위한 핀테크 기술과 시장을 그대로 반영하기에 부족하다는 한계를 갖는다.

특허 분류 기준만으로 핀테크 관련 특허를 수집하는 것이 어렵기에 미디어를 활용해 텍스트 정보를 가지고 핀테크 관련 검색 키워드를 생성하는 연구도 시도되어 왔다. 실제 USPTO에서 Fintech, Finance Technology와 같이 핀테크 자체 키워드로 특허 검색을 했을 때 출력되는 특허 개수는 8개밖에 존재하지 않는다. Fintech라는 단어보다 Fintech 기술을 충분히 설명할 수 있는 별도의 키워드가 필요한 것이다. Lee and Choi[22]는 핀테크 특허 추출을 위해 ‘핀테크’ 키워드를 검색하여 뉴스를 수집하고 빈도분석을 통해 핵심기술 단어목록을 만들었다. 그러나 단순한 빈도분석은 불필요한 단어를 포함할 가능성을 가지며 필요한 단어를 포함하지 않을 가능성 또한 존재하고, 단어수집에 대한 논리적 근거가 다소 부족하다는 한계가 있다.

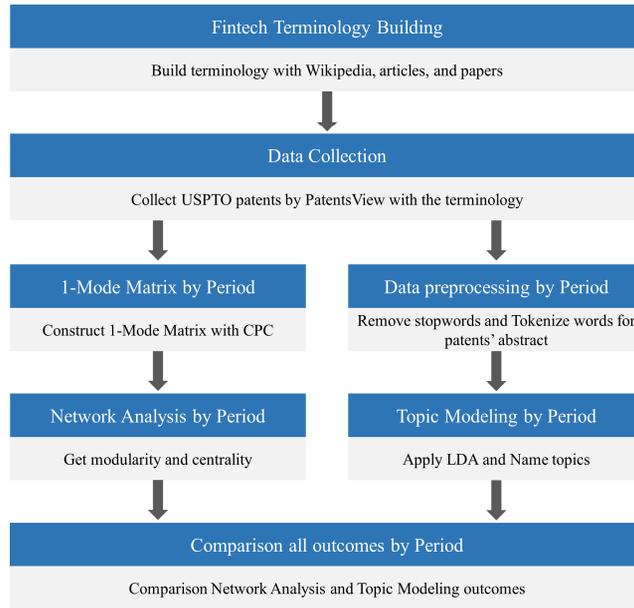
이러한 한계점들을 극복하기 위해 일부 연구자들은 IPC와 서지학적 정보를 결합하여 핀테크 검색 키워드를 생성하기도 하였다. IPC의 가장 상위 분류인 G는 물리, H는 전기에 관한 특허를 나타낸다. Chen et al.[6]는 IPC G와 H를 포함하는 특허는 디지털 컴퓨팅(기술)에 관한 특허라고 가정하여 그에 해당하는 모든 특허를 추출했다. 그 다음 Campbell, Oxford 금융 사전과 최신 단어를 합친 단어목록 중에 하나의 단

어라도 포함하는 특허가 있다면 그 특허는 기술과 금융의 특성을 모두 수반한다고 보았다. 이 논문은 핀테크가 기술과 금융의 합성어라는 개념에 초점을 맞춰 특허를 수집했다는 점에서 기존 연구들보다 한 단계 진일보한 것으로 보이지만, 기술과 금융을 결합하는 과정에서 Oxford 금융 사전과 같은 전통적인 사전을 활용하였기 때문에 금융사전에 포함되지 않는 최신의 기술이 누락될 가능성이 존재한다는 단점이 있다. 또한, 수립된 키워드들이 현저히 많아 불필요한 특허가 수집될 위험이 존재하기에 더 발전된 키워드 생성 방법이 필요한 것으로 보인다.

3. 연구 방법

본 연구의 절차는 <Figure 1>과 같다. 앞서 언급한 것과 같이 핀테크 특허 수집의 정확성을 높이기 위해 본 연구는 세 가지 단계를 활용해 핀테크 관련 용어를 정의하고, 이를 통해 핀테크 기술 관련 특허를 수집한다. 수집된 특허에서 특허의 CPC와 텍스트 정보(Abtract)를 추출하여 핀테크 기술동향을 기간별로 살펴본다.

본 연구에서는 핀테크 기술동향 분석을 크게 두 가지 방법으로 병렬적으로 수행한다. 먼저 특허 CPC에 대한 네트워크 분석을 수행하고 이를 바탕으로 네트워크 클러스터를 도출한다. 일반적으로 한 특허는 여러 개의 CPC 분류체계를 가지고 있으며, 특허 CPC의 동시분류분석을 통해 유사한 CPC를 군집화할 수 있다. 이를 통해 핀테크 기술이 주로 활용되는 산업분야를 도출할 수 있다. 본 연구에서는 이를 위해 기간별 네트워크 분석을 수행하여 모듈성(Modularity) 분석, 중심성(Centrality) 분석을 진행한다.



<Figure 1> Overall Process of the Research

두 번째로, 토픽 모델링을 활용하여 기술동향을 분석한다. 핀테크 기술동향을 풍부하게 살펴보기 위해서는 기술 분류체계, 기술 자체의 내용 등 다양한 측면을 살펴봐야 한다. 앞에서 수행한 특허 분류체계에 대한 분석은 거시적 측면에서 특허의 유형을 다룬다는 점에서 특허 그 자체의 내용을 반영하지는 못한다는 단점이 있다. 이러한 단점을 극복하기 위해 본 연구에서는 각 특허의 텍스트 데이터를 통해 기간별로 토픽 그룹을 추출하고 각 토픽의 특성을 규정한다.

본 연구에서는 두 가지 방법을 병렬적으로

진행함으로써 핀테크 기술의 산업분야를 종합적으로 살펴볼 수 있다. 마지막 단계로 두 가지 결과의 비슷한 점과 다른 점을 파악하여 최종적으로 핀테크 기술동향을 살펴본다.

3.1 핀테크 용어 수립

먼저 핀테크 기술동향을 분석하기 위한 용어를 수립한다. 이를 위해 다음 <Table 1>과 같이 3가지 방법을 활용해 핀테크 용어 목록을 수립한다.

<Table 1> Terminology Building Method

| Approach method | Database | Goal |
|---------------------------------------|---------------------------------|--|
| Arner et al.[2] Classification System | Wikipedia | Reflect the definition of FinTech-related technologies and the latest technologies |
| Literature Term | Papers and professional reports | Reflect the characteristics of FinTech technology used in academia or in the field |
| Finance+Technology | Word combination | Reflect scalability to various industries of FinTech technology |

첫 번째로 핀테크 기술 관련해 많이 인용되고 있는 Arner et al.[2]의 분류체계를 활용한다. 해당 연구는 핀테크 개념의 역사적 측면과 실용적 측면을 심도 있게 살펴봤기 때문에 핀테크 기술의 용어를 수립하는 데 도움을 줄 수 있다. 해당 논문에서 정의된 핀테크 분류 용어를 위키피디아에 검색하여 도출된 하이퍼링크 단어를 수집하였다. 위키피디아는 광범위한 주제를 다루고, 최신 정보를 반영하며, 도메인의 세밀하고 구체적인 지식을 제공하고, 풍부한 의미론적 정보를 담고 있다[14, 21]. 이러한 이점 때문에 위키피디아는 토픽 모델링[1], 형태 분석[18]과 같은 자연어처리 관련 연구에 사용된 바 있으며, 이에 기술동향을 분석하기 위한 목적에도 적합한 것으로 보인다. 위키피디아의 활용은 Oxford Dictionary 등을 활용한 기존 연구[6]에 비해 훨씬 유연한 백과사전이라는 점, 실시간으로 기술의 변화를 반영할 수 있다는 점에서 큰 장점을 가진다. 첫 번째 방식으로 <Table 2>와 같이 50개의 핀테크 키워드가 도출되었다.

두 번째로 핀테크의 학문적 특성과 시장의 특성을 반영하기 위해 핀테크 정의와 개념 등을 연구한 논문과 핀테크 보고서 등을 참고해 용어 목록을 수립한다. 이때 문서에서 수립된 단어들은 핀테크와 직접적으로 연관되어 있거나 실제 시장에서 수행되는 핀테크 관련 기술 혹은 서비스에 관한 것이다. 구체적으로 용어 목록은 Arner et al.[2], Schueffel[29], 그리고 Lee[23]의 논문에서 정의하는 핀테크 핵심 분야, 세부 분야의 단어들(money transfer, robo-advisory service, mortgage, online banking)을 포함한다. 또한, 핀테크 전문 잡지에서 핀테크에

자주 쓰이는 유망한 최신 기술이라고 주장하는 단어들(cryptocurrency, crowdfunding, wallet, payment, insurance, loan, lending, P2P, digital asset)을 포함했다[11, 12]. The business research company의 fintech report를 살펴보면 핀테크가 전통적 방식에서 crowdfundering(crowd funding)과 같은 새로운 방식으로 변화했고, 2020년에 블록체인(cryptocurrency)과 같은 기술을 결제 기술에 적용하는 기업들이 많아진 것을 확인할 수 있다[12]. 종합하여 <Table 2>와 같이 13개의 키워드가 핀테크 문헌으로부터 도출되어 추가되었다.

마지막으로, 핀테크가 금융 산업 뿐 아니라 다양한 산업에서 혁신의 수단으로 사용될 수 있다는 점을 고려해, 논문 및 보고서 등에 직접적으로 등장하지 않더라도 핀테크 기술과 관련된 다양한 활동을 고려할 필요가 있다. 이에 본 연구에서는 finance와 technology의 합성어로 이루어진 핀테크(fintech)의 정의를 고려하여, Technology, finance와 관련된 단어의 조합으로 단어를 생성한다. 예를 들어, Technology와 관련된 단어 mobile과 finance와 관련된 단어 identification을 조합하면 Mobile identification이라는 핀테크 용어가 생성된다. Technology를 설명하는 단어는 위키피디아에 ICT(Information and communications technology)를 검색했을 때 나오는 문서에서 수기로 추출된다. Finance도 마찬가지로 위키피디아에 finance를 검색했을 때 나오는 단어가 추출된다. 추출된 두 집합의 조합이 핀테크 의미를 반영하도록 최종 Technology+Finance 목록을 구성한다. 이러한 방식으로 98개의 키워드가 추가되었다.

<Table 2> Fintech Terminology

| Five major areas' words in Wikipedia(50) | Literature terms (13) | Technology +Finance Combination(98) |
|---|---|--|
| 1. Finance and Investment: capital marketing, open banking, investment management, robo-advisor, robo-advisory, stock market, deal sourcing, deal origination, crowdfunding, private equity, microfinance, assurance contract, alternative finance, fundraising, person-to-person lending, crowdlending, credit model, alternative financial service, debt collection, loan portfolios, digital advice, exchange-traded fund, closed-end fund, open-end fund, net asset value 2. Operations and Risk management: financial risk management, financial risk, credit risk, foreign exchange risk, cash management, financial engineering, investment risk, margin at risk 3. Payments and Infrastructure: currency exchange, micropayment, underbanked, card-based, carrier billing, billing, account transfer, credit card, one-time-password 4. Data security and Monetization: data security, disk encryption, transparent encryption data recovery, smart card, security token 5. Customer Interface: customer spending habits, consumer banking, consumer transaction | cryptocurrency, wallet, money transfer, payment, insurance, loan, lending, mortgage, digital asset, P2P(peer to per), crowdfunding, robo-advisory service, online banking | Technology(7) digital, internet, online, mobile, p2p, peer-to-peer, electronic + Finance(14) lending, insurance, finance, identification, asset, account, mortgage, loan, investing, investment, borrowing, trading, estate, money |

3.2 데이터 수집

위 단계에서 수립된 핀테크 용어 목록을 가지고 USPTO 특허를 수집한다. Patents-View에서 제공하는 query API를 사용하여 핀테크 용어 목록의 단어 중 하나라도 제목이나 초록에 포함이 되면 특허를 수집하도록 파이썬 크롤러를 개발하였다. 수집하는 특허는 특허번호(Patent Number), 특허이름(Patent Title), 특허등록일(Patent Date), 특허타입(Patent Type), 특허초록(Patent Abstract), CPC의 정보를 담고 있다.

3.3 기술 CPC 네트워크 분석

연결망 분석(Network Analysis)이란 연결망 데이터를 활용하여 연구하고자 하는 대상의 연결망과 대상의 구조 등을 분석하는 방법을 의미한다[30]. 네트워크 분석에서 필수로 사용되는 노드(node)는 분석하고자 하는 하나의 대상을 의미하고, 엣지(Edge)는 노드 간의 연결을 의미한다. 본 연구에서는 연구 대상(노드)를 CPC로 설정했는데, CPC 코드에서 사용되는 정보는 섹션(Section), 클래스(Class), 서브클래스(Subclass), 메인그룹(Main Group)이다.

예를 들어 G05B2219 /1184라는 CPC에서 서브 그룹(Sub Group)인 1184을 제외한 G05B2219만 사용한다. 여기서 노드와 엣지 정보를 얻기 위해 각 특허에 대한 CPC를 가지고 특허-CPC 행렬(2-Mode Matrix)를 만든 후 CPC-CPC 행렬(1-Mode Matrix)를 생성한다. 특허-CPC 행렬은 행에 각 특허를 나열하고, 모든 특허에서 나온 CPC를 중복제거 하여 고유 CPC만 열에 나열한다. 그리고 각 특허에서 각 CPC가 등장한 횟수를 행렬에 삽입한다. CPC-CPC 행렬은 식 (1)과 같이 특허-CPC 행렬의 전치행렬과 특허-CPC 행렬의 곱으로 나타난다.

$$c_{ij} = \sum_k x_{ik}x_{jk} \text{ or } C = X^T X \quad (1)$$

X = 특허-CPC 행렬

C = CPC-CPC 행렬

다음으로 R 프로그램의 igraph 모듈을 사용하여 CPC-CPC 행렬로 네트워크를 만든다. 그리고 Gephi 프로그램의 모듈성 분석을 수행하여 네트워크의 커뮤니티 구조를 탐지한다. 모듈성 분석은 연결 밀도가 높은 네트워크끼리 묶어서 군집화하는 방법을 의미한다. Gephi 프로그램에서 수행되는 모듈성 분석은 Louvain 알고리즘을 사용한다[5]. Louvain 알고리즘은 Phase 1과 Phase 2로 나뉘는데, Phase 1에서 한 노드를 원래의 커뮤니티에서 빼내어 인접한 커뮤니티에 재배치하고 모듈성을 측정한다. 만약 모듈성이 높다면 재배치된 커뮤니티로 이동하고 낮다면 원래의 커뮤니티로 돌아간다. Phase 2에서는 Phase 1에서 도출된 커뮤니티를 모아 새로운 커뮤니티를 만드는 방식을 사용한다. 이 두 과정은 모듈성의 변화가 없을 때까지 반복해서 진행된다. 이를 도식화하면 식 (2)와 같다.

$$\Delta Q = \left[\frac{\sum_{in} + 2k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \quad (2)$$

ΔQ : 모듈성 변화량

\sum_{in} : 커뮤니티 C의 내부 엣지 weight의 합

\sum_{tot} : 커뮤니티 C에서 발생하는 모든 엣지 weight의 합

k_i : 노드 i에서 발생한 엣지 weight의 합

$k_{i,in}$: 노드 i로 연결되는 엣지 weight의 합

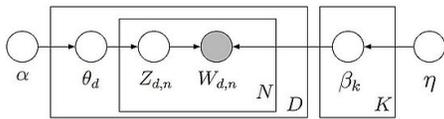
m : 네트워크 안의 모든 엣지 weight의 합

모듈성 분석으로 각 군집을 살펴본 뒤 군집들의 융합도를 확인하기 위해 군집별 연결비율을 계산한다. A라는 군집에 속한 노드가 B라는 군집의 노드에 연결된 횟수를 A의 노드가 다른 모든 군집의 노드에 연결된 횟수로 나눠주어 융합도를 계산한다.

마지막으로 네트워크의 연결 정도 중심성(Degree Centrality), 고유벡터 중심성(Eigenvector Centrality), 매개 중심성(Betweenness Centrality)을 살펴보고 기간별로 어떤 노드(CPC)가 상대적 우위성을 갖는지 확인한다. 중심성이 높은 노드일수록 네트워크 관계에서 해당 노드의 관여 정도가 증가한다고 볼 수 있다[3, 19]. 연결 정도 중심성은 중심성 척도 중 가장 간단한 척도로서 한 노드에 연결된 모든 엣지의 개수이다. 고유벡터 중심성은 CPC-CPC 행렬에 대한 고유값 분해를 하여 고유벡터를 구해 각 CPC에 대한 벡터값을 도출한다. 매개 중심성은 특정 노드 b에서 노드 c로 가는 최단 경로에 노드 a가 포함되어 있다면 노드 a에 매개 중심성을 부여하는 개념이다. 다른 모든 노드의 최단 경로에 노드 a가 포함된 확률을 구해 각 노드에 대한 매개 중심성을 계산한다.

3.4 토픽 모델링

토픽 모델링은 문서에서 단어들에 대한 추상적인 주제를 파악하기 위해 사용되는 통계 모델의 한 유형이다. LDA(Latent Dirichlet Allocation)는 토픽 모델링 방법 중 가장 대표적인 모델인데 문서 내에 잠재되어있는 토픽을 찾아내는 알고리즘이다. 문서는 여러 개의 토픽을 혼합한 것이라고 가정하고 문서 안의 단어들은 각 토픽에 속할 확률값을 가지게 된다. 그래서 설명 변수인 문서와 단어를 가지고 잠재 변수를 추론하여 문서에 속한 토픽들의 확률과 단어에 속한 토픽들의 확률을 구할 수 있다[4].



<Figure 2> LDA Model Architecture

- K : 토픽의 개수
- D : 전체 문서 개수
- N : d번째 문서의 단어 개수
- α : 문서별 토픽 k의 Dirichlet prior weight, θ 값을 결정하는 파라미터
- μ : 문서별 토픽 w의 Dirichlet prior weight, β 값을 결정하는 파라미터
- θ_d : 문서별 토픽의 비율
- β_k : 토픽별 단어 w의 생성확률
- $Z_{d,n}$: 문서 d의 n번째 단어의 토픽(index)
- $W_{d,n}$: 문서 d의 n번째 단어(문서에 관측되는 변수, index)

<Figure 2>에서 사용자가 설정할 수 있는 변수는 d번째 문서의 n번째 단어인 $W_{d,n}$ 가 유일하다. 이 변수를 가지고 α 와 β 를 제외한 모든 잠재 변수를 추정해야 한다. 본 연구에서 d는

특히 초록을 나타내고, $Z_{d,n}$ 는 핀테크에 관한 기술 주제를 의미한다. α 에 따라 θ 가 분포하게 될 Dirichlet 분포의 모양이 결정되고, μ 에 따라 β 가 분포하게 될 Dirichlet 분포의 모양이 결정된다. 그리고 θ 에 따라 특히 문서 내에 존재하는 단어들의 토픽인 $Z_{d,n}$ 가 결정된다. 각 단어의 토픽들을 나타내는 $Z_{d,n}$ 와 각 단어에 대한 전체 토픽에 대한 비율 β 에 따라 단어 $W_{d,n}$ 가 결정된다. 모델 결과는 토픽에 속하는 단어의 집합으로 도출된다.

본 연구는 기간별로 특히초록에서 stopwords를 제거하고 tokenize하여 LDA를 적용한다. LDA를 통해 도출된 단어 집합들을 살펴보면 단어 집합이 나타내는 주제의 이름을 생성한다. 마지막으로 기간별로 핀테크의 주제들을 살펴봄으로써 어떠한 기술적 변화가 있었는지 확인하고, 핀테크 기술동향을 파악한다.

4. 연구 결과

4.1 핀테크 용어 수립

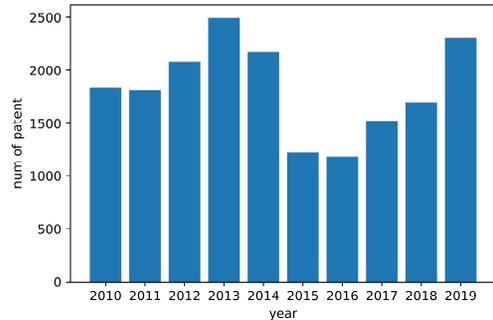
본 연구에서 생성된 핀테크 용어 목록에서 몇 개의 단어는 다의어의 특징을 가지고 있기에 핀테크 특히뿐만 아니라 핀테크와 관련 없는 특히도 수집하게 된다. 그래서 account, saving, card, shape risk, liquidity risk, self-checkout, near sound data transfer, airdrop, investor, identification(11개)은 제거되고 용어 목록이 수립되었다.

앞서 언급한 첫 번째 용어 수립 방법에 따라 위키피디아에 Crowdfunding, P2P lending, robo-advisory services, risk management, payment, data security를 검색하여 총 50개의 핀

테크 관련 단어를 수집했다. 두 번째 방법에 따라 서지학적 정보를 활용하여 13개의 단어를 수집했으며, 마지막 방법에 따라 기술+금융 합성어는 총 98개가 나왔다. 세 가지 결과를 종합하였고 <Table 2>와 같이 총 161개의 핀테크 용어를 수립하였다.

4.2 데이터 수집

<Table 2>의 핀테크 용어 목록으로 수집된 특허의 총 개수는 18,341개이다. 본 연구는 핀테크의 최신 기술동향을 살펴보기 위해 특허 등록일 기준으로 데이터 수집 기간을 2010년 1월부터 2020년 1월로서 10년으로 설정했으며 기간은 총 세 개의 기간으로 나누었다. 기간 1은 2010년부터 2013년까지 4년으로 8,223개의 특허, 기간 2는 2014년부터 2016년까지 3년으로 4,588개의 특허, 기간 3은 2017년부터 2019까지 3년으로 5,529개의 데이터를 얻었다(<Figure 3> 참조). 핀테크 관련 특허는 2012년부터 2014년까지 많이 등록되었고, 2015년부터 특허 등록수가 현저히 감소 하는 것을 확인할 수 있다. 2019년이 돼서야 2012년부터 2014년까지의 수준에 이르렀다. 이는 2014년까지 새로운 핀테크 기술이 지속해서 등장하다가 2015년부터 신기술 개발 정체 시기가 찾아오고, 실제로 핀테크 투자 금액이 최고를 찍었던 2019년에 기술 기반 스타트업이 대거 등장한 것으로 보인다.



<Figure 3> Number of Patent by Year

4.3 핀테크 기술 CPC분석 : 네트워크 분석

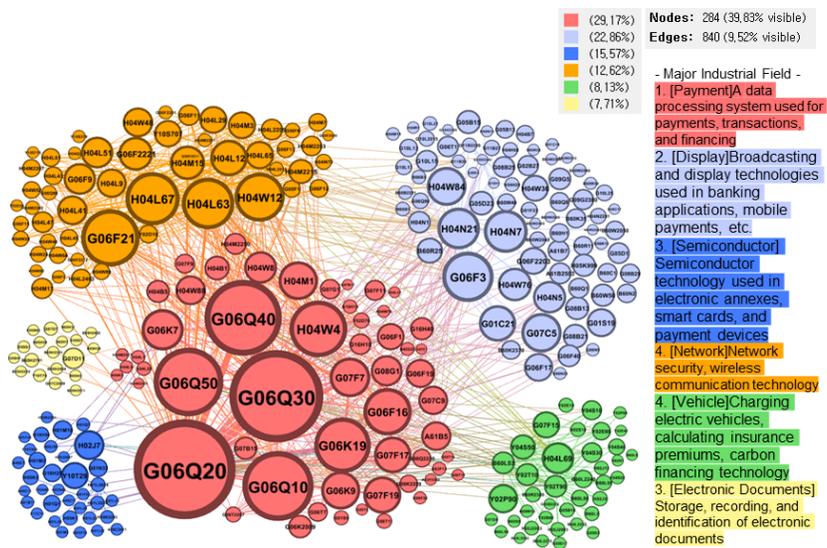
기간별로 CPC에 관한 네트워크의 기본정보는 <Table 3>과 같다. 기간 1이 4년인 것을 고려하여 기간이 지난해에 따른 네트워크 지표 증감률을 계산했다. 기간 1에서 기간 2로 갈 때는 특허 수가 26%나 감소하다가 기간 2에서 기간 3으로 갈 때는 특허 수가 21% 증가한다. 기간 1에서 기간 2로 갈 때 특허 수가 상당히 감소했지만, 노드 수와 엣지 수가 42%, 74%로 엄청나게 증가한 것으로 보아 핀테크 관련 기술이 다양해졌고 서로 많은 연결성이 나타난 것으로 보인다. 반면 모듈성의 군집 개수는 13% 낮아져 산업의 기술 세분화가 감소했다고 볼 수 있다. 기간 2에서 기간 3으로 갈 때 노드는 19% 증가했고 엣지의 경우 1% 감소해 연결성의 변화는 미비하지만, 군집 개수는 46% 증가하여 산업의 기술 세분화가 이루어진 것으로 보인다.

<Table 3> Basic Information of the Network by Period

| | Period 1 (4 years) | Period 2 (3 years) | Period 3 (3 years) | Period 1 → Period 2 | Period 2 → Period 3 |
|------------------|-----------------------|-----------------------|-----------------------|---------------------|---------------------|
| Number of Patent | 8,223 | 4,588 | 5,529 | 26% ↓ | 21% ↑ |
| Node | 669 | 713 | 846 | 42% ↑ | 19% ↑ |
| Edge | 6,747 | 8,822 | 8,766 | 74% ↑ | 1% ↓ |
| Modularity | 20 | 13 | 19 | 13% ↓ | 46% ↑ |

<Table 4> Convergence of Top Clusters for Period1

| % | Payment | Network | Electronic Documents | Vehicle | Semiconductor | Display |
|----------------------|---------|---------|----------------------|---------|---------------|---------|
| Payment | 35 | 31 | 38 | 31 | 31 | 28 |
| Network | 30 | 52 | 8 | 12 | 7 | 39 |
| Electronic Documents | 11 | 2 | 44 | 0 | 8 | 6 |
| Vehicle | 9 | 4 | 0 | 54 | 2 | 1 |
| Semiconductor | 5 | 1 | 5 | 1 | 50 | 1 |
| Display | 7 | 10 | 5 | 0 | 1 | 25 |



<Figure 5> Network for Period2

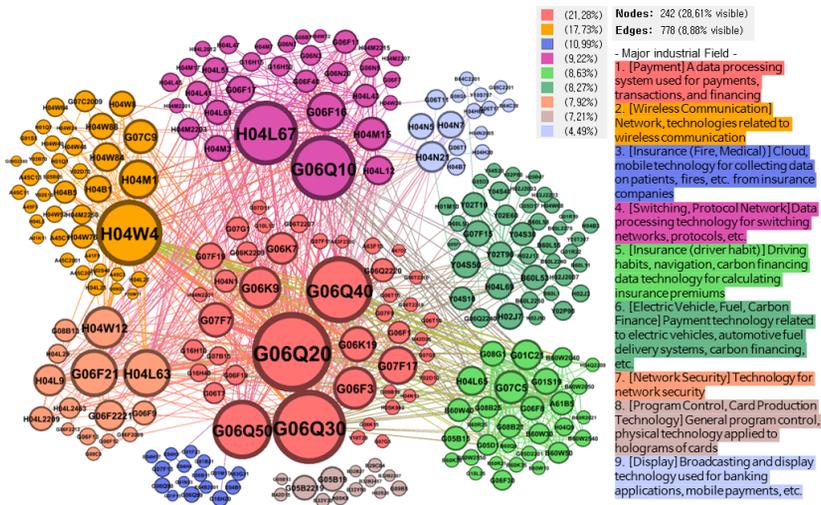
<Table 4>는 기간 1의 상위 6개 군집의 각 융합도 비율을 열별로 나타낸다. 자기 자신과 강한 연결을 보이는 군집은 네트워크, 차량, 전자문서이다. 핀테크 핵심 기술인 결제 관련 CPC는 자기 자신보다 다른 군집들과 모두 강한 연결을 보여 다른 산업 분야와의 융합이 활발히 이뤄짐을 알 수 있다. 반도체는 결제에 사용되는 물리적 제품과 관련이 깊다. 일반적인 신용카드나 RFID 등과 같은 결제 시스템에 사용되기에 결제와 높은 융합도를 보인다. 그리고 네트워크 관련 기술과 디스플레이 관련 기술이 다른 군집들보다 높은 융합도를 보이는데 기간 1에는 은행 방송 시스템 기술과

네트워크 통신과의 융합기술이 많이 나타나 두 군집의 기술적 연관도가 필수적임을 유추할 수 있다. 마지막으로 자동차 같은 경우 공유차량이나 전기차는 자동결제 시스템이 요구될 수 있고, 자동차의 디스플레이를 통해 상품이나 서비스를 결제하는 등의 기술이 사용된다. 또한, 운전자의 습관을 수집하는 네트워크 기술을 활용해 보험료 산정과 같은 금융서비스가 제시된다.

<Figure 5>는 기간 2의 네트워크를 나타낸다. Cut-off 값은 기간 1과 똑같은 값으로 적용됐으며 노드는 전체 713개 중 284개로서 39.83%를 나타낸다. 엣지는 전체 8,822개 중 840개로서

<Table 5> Convergence of top clusters for Period2

| % | Payment | Display | Semiconductor | Network | Vehicle | Electronic Documents |
|----------------------|---------|---------|---------------|---------|---------|----------------------|
| Payment | 33 | 33 | 38 | 33 | 27 | 36 |
| Display | 26 | 39 | 8 | 28 | 13 | 3 |
| Semiconductor | 9 | 1 | 44 | 2 | 7 | 1 |
| Network | 21 | 22 | 0 | 32 | 10 | 1 |
| Vehicle | 7 | 4 | 5 | 4 | 43 | 0 |
| Electronic Documents | 4 | 0 | 5 | 0 | 0 | 58 |



<Figure 6> Network for Period3

9.52%를 나타낸다. 기간 2의 그래프 또한 5%를 넘는 상위 군집들이 6개로 전체의 96.06%를 나타낸다. 기간 2의 군집들은 상위 순서에 따라 다음과 같다. ‘지불, 거래, 금융에 사용되는 데이터 처리 시스템’, ‘은행응용프로그램, 모바일 결제 등에 쓰이는 방송, 디스플레이 기술’, ‘전자식별기, 스마트 카드, 결제 기기에 쓰이는 배터리, 의료, 물리적 기술’, ‘네트워크 보안, 무선 통신 관련 기술’, ‘전기 자동차 요금 부과, 보험료 산정의 네비게이션, 탄소금융 기술’, ‘전자문서의 보관, 기록 및 식별’. 기간 2에서는 기간 1과 비슷한 군집들을 확인할 수 있었고, 디스플레이 관련 CPC가 상당히 증가한 것으로 보아 키오스크와

같이 금융과 디스플레이가 결합된 기술이 발전했다고 여겨진다.

기간 2의 상위 6개의 군집별 융합도를 살펴보면 기간 1과 비슷한 양상을 띠는 것을 확인할 수 있다(<Table 5> 참조). ‘전기 자동차 요금 부과, 보험료 산정의 네비게이션, 탄소금융 기술’ 군집은 전자문서 관련 기술을 제외한 다른 군집들과 고루 분포된 융합도를 나타낸다. 차량 관련 기술에 네트워크, 디스플레이, 배터리, 결제 등이 융합되어 더 우수한 금융기술이 등장한 것으로 보인다.

마지막으로 <Figure 6>은 기간 3의 네트워크를 나타낸다. Cut-off 값은 기간 1, 기간 2와 똑같은

값으로 적용되어 노드는 전체 846개 중 242개로
서 28.61%를 나타낸다. 옛지는 전체 8,766개 중
778개로서 8.88%를 나타낸다. 기간 1, 기간 2에
등장한 군집을 포함하여 상위 군집은 총 9개로
전체의 95.74%를 나타낸다. 기간 3의 군집들은
상위 순서에 따라 다음과 같다. ‘지불, 거래, 금융
에 사용되는 데이터 처리 시스템’, ‘네트워크, 무
선 통신 관련 기술’, ‘보험사에서 환자, 화재 등에
관한 데이터를 수집하기 위한 클라우드, 모바일
기술’, ‘스위칭 네트워크, 프로토콜 등에 관한 데
이터 처리 기술’, ‘보험료 산정을 위한 운전 습관,
네비게이션, 탄소금융 데이터 기술’, ‘전기 자동
차, 자동차 연료 공급 시스템, 탄소금융 등에
관련된 결제 기술’, ‘네트워크 보안을 위한 기술’,
‘일반적 프로그램 제어, 카드의 홀로그래프에 적용
되는 물리적 기술’, ‘은행응용프로그램, 모바일
결제 등에 쓰이는 방송, 디스플레이 기술’. 기간

3 모듈 분석에서는 네트워크 군집이 스위칭, 프로
토콜; 네트워크 보안; 무선 통신으로 분리되면서
산업 내 중요도가 상승한 것으로 보인다. 또한,
결제 관련 node들이 군집에서 분리되어 네트워크
노드와 섞여 새로운 군집을 생성했다. 그리고 보
험료 산정을 위한 다양한 분야에서의 데이터
수집 기술이 발전하였다(화재, 의료, 운전 등).

<Table 6>은 기간 3의 군집별 융합도를 나
타낸다. 대부분의 네트워크 관련 군집은 서로
강한 연결을 보였다. 보험(화재의료) 군집은 자
기 자신과 강한 연결을 보이고 결제와의 융합
도가 보인다. 이는 금융서비스가 보험(운전습
관) 산업과 융합하는 것처럼 보험과 관련된 다
양한 분야에 결제 혹은 금융 서비스가 결합했
다고 볼 수 있다. 보험(운전습관), 디스플레이
관련 CPC가 스위칭 네트워크, 프로토콜과 융
합도가 큰 것을 확인하였다.

<Table 6> Convergence of Top Clusters for Period 3

| % | Payment | Wireless Commu- nication | Insu- rance | Swit- ching, Protocol | Insu- rance | Electric Vehicle, Fuel, | Network Security | Program Control, | Display |
|---|---------|--------------------------------|----------------|-----------------------------|----------------|-------------------------------|---------------------|---------------------|---------|
| Payment | 31 | 22 | 23 | 25 | 23 | 18 | 26 | 19 | 30 |
| Wireless Communication | 14 | 34 | 5 | 14 | 11 | 13 | 15 | 7 | 11 |
| Insurance | 6 | 2 | 54 | 2 | 5 | 1 | 1 | 0 | 1 |
| Switching, Protocol Network | 14 | 13 | 6 | 23 | 13 | 8 | 20 | 6 | 15 |
| Insurance (driver habit) | 9 | 7 | 8 | 10 | 34 | 5 | 5 | 4 | 8 |
| Electric Vehicle, Fuel, Carbon Finance | 7 | 8 | 2 | 5 | 5 | 46 | 5 | 3 | 2 |
| Network Security | 10 | 9 | 1 | 13 | 5 | 5 | 21 | 3 | 9 |
| Program Control, Card Production Technology | 3 | 2 | 0 | 2 | 2 | 1 | 1 | 54 | 4 |
| Display | 6 | 4 | 1 | 5 | 4 | 1 | 5 | 5 | 18 |

4.4 핀테크 기술 내용 분석 : LDA 분석

핀테크 기술의 CPC분석에 이어, 핀테크 기술의 내용 자체를 분석하기 위해 LDA분석을 수행하였다. 이를 위해 기간별로 8223, 4588, 5529개의 문서에 대한 초록을 전처리하여 분석을 수행하였다.

LDA 분석을 효과적으로 수행하기 위해서는 LDA에 사용되는 하이퍼 파라미터에 대한 알맞은 조정이 필요하다. 하이퍼 파라미터에는 Topic (토픽의 수)과 Passes(학습 빈도)가 있다. 먼저 최적의 Passes를 구하기 위해 Topic을 30으로 설정하고 Perplexity와 Coherence를 살펴보았다. Perplexity는 특정 확률 모델이 실제로 관측되는 값을 얼마나 잘 예측하는지를 의미하며, Coherence는 주제의 일관성을 측정한다. Perplexity가 낮을수록, Coherence가 높을수록 적당한 Topic과 Passes를 가진다. 모든 기간의 최적 Passes는 30으로 동일하고 Topic의 수는 13, 16, 13으로 나타났다. 각 기간의 LDA 결과는 토픽들과 그 안의 단어들로 나타나는데 단어들은 토픽에 속할 확률을 가지고 있다. 만약 토픽 내의 단어들이 하나의 주제를 나타내지 못한다고 판단되면 그 토픽은 제외하였다.

<Table 7>은 기간 1에서 선별된 토픽의 상위 15개의 단어를 나타낸다. 기간 1에서는 13개의 토픽 중 10개의 토픽이 선별됐고 지불은 12%, 네트워크, 통신은 10.2%, 계좌, 송금은 8.9%, 금융상품관리는 8.9%, 보안은 8.3%, 자산은 7%, 카드, 인식은 7%, 보험은 5.7%, 세일즈는 5.2%, 문서, 기록은 2.4%로서 각 토픽 비율을 나타낸다. 위 10개의 토픽은 전체의 77.2%를 나타낸다. 각 토픽 밑에 구성된 단어들에 해당하는 비율은 토픽 내에서 단어가 나타나는 확률을 나타낸다.

지불 토픽을 살펴보면 payment가 8.8%, transaction이 6.5%로서 결제, 지불과 관한 토픽임을 알 수 있다. LDA 특성상 토픽에 속하는 모든 단어가 지불과 직접적으로 연결되어 있지는 않지만, 최상위 단어들이 나타내는 공통적인 의미가 지불을 나타낸다고 할 수 있다. 예를 들어 지불에 속해 있는 service는 결제서비스를 나타내고, request는 결제를 요청하고, network는 결제 네트워크에 관한 단어라고 볼 수 있다. 금융상품관리 같은 경우 financial, risk, instrument 등 상위 단어들이 금융상품관리에 해당하는 특허의 요약에 자주 등장하기에 금융상품관리라고 명명할 수 있다.

기간 2에서는 16개의 토픽 중 11개의 토픽이 선별됐고 지불은 22.1%, 스마트 카드는 10.9%, 보험은 9.9%, 금융상품관리시스템은 7.9%, 네트워크, 통신은 6.4%, 전자 신호, 보안은 3.8%, 보안은 3.3%, 보안(블록체인)은 3%, 거래 요청은 2.6%, 가상, 이미지 시스템은 2.6%, 앱 보안은 2.2%로서 각 토픽 비율을 나타낸다. 위 11개의 토픽은 전체의 74.7%를 나타낸다. 기간 2에서 지불은 기간 1과 비교하여 다소 큰 비율을 차지하고 있고, 지불의 큰 비율에도 불구하고 보험 토픽 비율이 크게 증가한 것으로 보아 보험과의 융합이 활발히 이뤄졌다고 이야기할 수 있다. 블록체인과 그에 필요한 보안 토픽이 등장한 것과 앱 보안에 관한 토픽이 등장했기에 보안에 대한 중요성이 대두됐다고 볼 수 있다.

마지막으로 기간 3에서는 13개의 토픽 중 10개의 토픽이 선별됐고 지불은 18.1%, 거래 통신은 12.2%, 스마트 카드는 8.9%, 네트워크 서버는 8.7%, 보안, 접근은 7.8%, 금융상품관리시스템은 7.1%, 신용카드는 6.3%, 보험은 5.3%, 모바일은 5.2%, 자동차는 4.8%로서 각 토픽 비율을 나타낸다.

〈Table 7〉 10 Topics with Top 15 Words for Period 1

| Payment[2] 1.2% | Network, Communication [3] 10.2% | | Account, Remittance[4] 8.9% | | Financial instrument management[5] 8.9% | | Security[6] 8.3% | | Asset[7] 7% | | Card, Identification[8] 7% | | Insurance[10] 5.7% | | Sales[11] 5.2% | | Document, Record[13] 2.4% | |
|--------------------|-------------------------------------|-----|--------------------------------|-----|--|-----|---------------------|-----|--------------------|-----|-------------------------------|-----|-----------------------|------|-------------------|-----|------------------------------|-----|
| | words | % | words | % | words | % | words | % | words | % | words | % | words | % | words | % | words | % |
| payment | 8.8 | 3.3 | account | 9.8 | financial | 7.4 | security | 4.8 | value | 5.5 | card | 8.6 | insurance | 11.2 | product | 5.4 | item | 8 |
| transaction | 6.5 | 2.7 | card | 6.7 | risk | 2.8 | token | 2.5 | amount | 3 | machine | 4.5 | policy | 5 | consumer | 4.8 | document | 5.8 |
| service | 3.8 | 2.7 | credit | 6.1 | instrument | 2 | server | 2.2 | time | 2.5 | banking | 2.3 | claim | 3.3 | purchase | 3.1 | member | 3.5 |
| request | 2.9 | 2.5 | customer | 3.9 | parameter | 1.5 | application | 2.2 | period | 2.1 | reader | 2.1 | loan | 2.2 | offer | 2.2 | electronic | 3.4 |
| provider | 2.3 | 2.1 | transaction | 2.9 | model | 1.4 | access | 2.1 | deposit | 1.8 | check | 1.7 | benefit | 1.7 | price | 2 | primary | 1.8 |
| merchant | 1.9 | 2.1 | financial | 2.5 | component | 1.3 | key | 1.9 | rate | 1.7 | read | 1.6 | insured | 1.7 | buyer | 1.9 | image | 1.7 |
| processing | 1.6 | 1.6 | fund | 1.5 | plurality | 1.3 | card | 1.9 | investment | 1.6 | automated | 1.6 | plan | 1.7 | seller | 1.6 | form | 1.3 |
| party | 1.5 | 1.4 | institution | 1.4 | product | 1.2 | authenti- cation | 1.8 | asset | 1.3 | cash | 1.6 | coverage | 1.6 | sale | 1.5 | container | 1.2 |
| receiving | 1.1 | 1.4 | transfer | 1.4 | portfolio | 1.1 | computer | 1.4 | property | 1.3 | transaction | 1.5 | premium | 1.6 | currency | 1.4 | package | 1.2 |
| received | 1 | 1.4 | number | 1.3 | used | 1 | smart | 1.3 | fund | 1.3 | magnetic | 1.3 | vehicle | 1.3 | gift | 1.1 | travel | 0.9 |
| authoriza- tion | 0.9 | 1.3 | money | 1.1 | individual | 1 | secure | 1.2 | pre- determined | 1 | financial | 1.2 | life | 1.2 | good | 1.1 | paper | 0.9 |
| network | 0.9 | 1.2 | bank | 1.1 | rating | 0.9 | client | 1.1 | interest | 1 | responsive | 0.9 | loss | 1.1 | exchange | 1.1 | energy | 0.8 |
| customer | 0.7 | 1.1 | amount | 1 | score | 0.9 | remote | 1 | plurality | 1 | layer | 0.9 | risk | 1.1 | bid | 1 | project | 0.8 |
| vendor | 0.7 | 0.9 | purchase | 0.9 | factor | 0.8 | memory | 0.9 | portion | 0.9 | display | 0.9 | mortgage | 0.9 | purchaser | 1 | etc | 0.7 |
| electronic | 0.7 | 0.8 | balance | 0.9 | market | 0.8 | request | 0.9 | determining | 0.8 | operative | 0.8 | health | 0.9 | item | 0.9 | signature | 0.7 |

〈Table 8〉 11 Topics with Top 15 Words for Period 2

| Topic | Smart card[2] 10.9% | | Insurance[3] 9.9% | | Financial instrument management system[4] 7.9% | | Network, Communication [6] 6.4% | | Electronic signal, Security[10] 3.8% | | Security[11] 3.3% | | Security (Blockchain) [12] 3% | | Transaction request[14] 2.6% | | Virtual, Image system[15] 2.6% | | App security [16] 2.2% | | | |
|---------------------|------------------------|-----|----------------------|------|---|-----|------------------------------------|-----|---|-----|----------------------|-----|----------------------------------|------|---------------------------------|------|-----------------------------------|-----|---------------------------|-----|-------------|-----|
| | words | % | words | % | words | % | words | % | words | % | words | % | words | % | words | % | words | % | words | % | | |
| Payment[1] 22.1% | payment | 6.7 | card | 12.7 | insurance | 3.4 | financial | 8.4 | communication | 7 | secure | 4.9 | token | 11.7 | security | 10.1 | message | 9.7 | image | 7.9 | application | 5 |
| | transaction | 6.3 | smart | 3.1 | object | 1.3 | management | 1.7 | mobile | 4.6 | digital | 4.9 | security | 9.9 | token | 9.8 | medium | 6.4 | virtual | 7.5 | trust | 3.1 |
| | account | 2.2 | reader | 1.8 | business | 1.2 | instrument | 1.3 | wireless | 4.4 | module | 4.3 | service | 6.3 | key | 5.2 | program | 4.1 | machine | 6.7 | identity | 3 |
| | merchant | 1.7 | payment | 1.6 | invention | 1.1 | used | 1.2 | terminal | 2.4 | circuitry | 3.2 | access | 5.5 | authentication | 5.2 | content | 3.5 | file | 5.4 | group | 2.5 |
| | request | 1.4 | credit | 1.5 | policy | 1 | loan | 1.1 | unit | 2.2 | integrated | 2.8 | application | 2.8 | server | 2.5 | feature | 3.5 | server | 3.7 | token | 2.2 |
| | consumer | 1.3 | display | 1.3 | product | 1 | and/or | 1.1 | proximity | 1.8 | plugin | 2.3 | client | 2 | domain | 2.4 | security | 2.9 | host | 3.2 | disclosure | 1.8 |
| | financial | 1.2 | input | 1.2 | claim | 1 | activity | 1.1 | storage | 1.4 | storage | 2.1 | authentication | 2 | client | 2.3 | onetime | 2.3 | computer | 2.9 | security | 1.5 |
| | mobile | 1.1 | electronic | 1.2 | benefit | 0.8 | database | 1 | dci | 1.3 | dci | 1.9 | computer | 1.9 | secret | 1.7 | file | 2.3 | game | 1.9 | present | 1.4 |
| | receiving | 1 | read | 1.1 | present | 0.7 | credit | 1 | ced | 1.2 | ced | 1.6 | request | 1.8 | public | 1.7 | token | 2.2 | used | 1.5 | aspect | 1.4 |
| | amount | 1 | configured | 1.1 | company | 0.7 | element | 0.9 | visual | 1.1 | visual | 1.6 | identity | 1.6 | communication | 1.4 | processing | 1.8 | distributed | 1.5 | mobile | 1.3 |
| | processing | 1 | magnetic | 1 | condition | 0.7 | institution | 0.9 | browser | 1.1 | browser | 1.4 | server | 1.5 | provides | 1.3 | computer | 1.8 | player | 1.3 | full | 1.3 |
| | purchase | 0.9 | order | 1 | entity | 0.6 | process | 0.9 | content | 1.1 | content | 1.4 | privacy | 1.2 | identifier | 1.2 | stored | 1.6 | gaming | 1.3 | low | 1.2 |
| | electronic | 0.9 | processor | 0.8 | type | 0.6 | various | 0.7 | custody | 1 | custody | 1.4 | software | 1.2 | remote | 1.2 | tokencode | 1.5 | automated | 1.3 | arrangement | 1.2 |
| | identifier | 0.9 | memory | 0.7 | model | 0.6 | control | 0.7 | document | 1 | document | 1.3 | resource | 1.2 | secure | 1.1 | value | 1.5 | procedure | 1.2 | dynamic | 1.2 |
| | network | 0.9 | interface | 0.7 | interface | 0.6 | report | 0.7 | processor | 0.9 | processor | 1.3 | refresh | 0.9 | unblock | 1 | instruction | 1.4 | disk | 1 | federation | 1.1 |

〈Table 9〉 10 Topics with Top 15 Words for Period 3

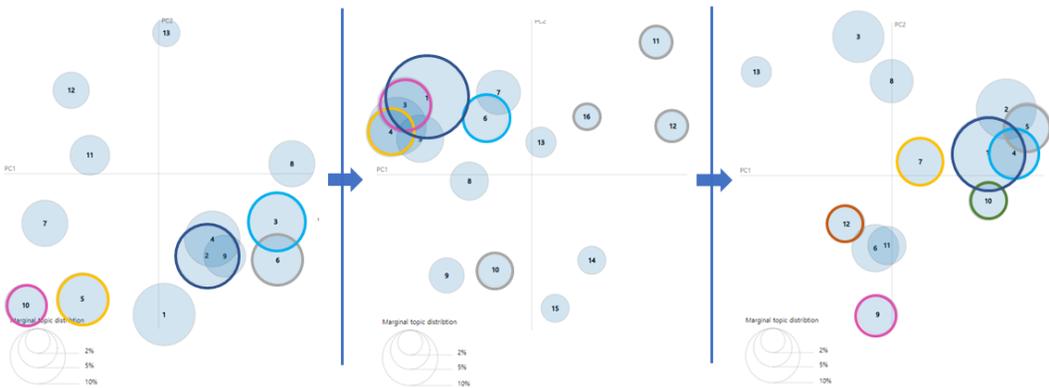
| Payment[1] 18.1% | Transactional communication [2] 12.2% | | Smart card[3] 8.9% | | Network server [4] 8.7% | | Security, Access[5] 7.8% | | Financial instrument Management System[7] 7.1% | | Credit card[8] 6.3% | | Insurance[9] 5.3% | | Mobile[10] 5.2% | | Vehicle[12] 4.8% | |
|---------------------|--|-------|-----------------------|-------|----------------------------|-------|-----------------------------|-------|---|-------|------------------------|-------|----------------------|-------|--------------------|-------|---------------------|-------|
| | % | words | % | words | % | words | % | words | % | words | % | words | % | words | % | words | % | words |
| 12.5 | payment | 3.1 | message | 5.9 | card | 8 | service | 7.7 | security | 6.4 | financial | 11.7 | insurance | 9.5 | mobile | 6.6 | value | 2.7 |
| 7.7 | transaction | 2.4 | financial | 3.4 | smart | 4.7 | network | 4.5 | token | 4.4 | transaction | 7.9 | vehicle | 6.6 | digital | 5.1 | vehicle | 2.4 |
| 3.6 | merchant | 2.4 | request | 2 | module | 2.9 | provider | 3.4 | key | 1.6 | record | 5.7 | policy | 3.8 | asset | 3.8 | storage | 1.6 |
| 1.8 | account | 2.3 | server | 1.5 | reader | 2 | billing | 2.5 | application | 1.3 | source | 3 | risk | 3.8 | wallet | 3 | driver | 0.9 |
| 1.8 | mobile | 2.1 | communication | 1.4 | component | 1.9 | client | 2.4 | access | 1.2 | machine | 2.8 | property | 1.8 | content | 2.8 | location | 0.9 |
| 1.5 | purchase | 2 | processor | 1.2 | signal | 1.5 | entity | 1.7 | secure | 1.2 | cash | 1.6 | claim | 1.8 | computing | 2.8 | time | 0.9 |
| 1.5 | consumer | 1.6 | receiving | 1.6 | power | 1.4 | server | 1.5 | authentication | 1.1 | profile | 1.5 | determine | 1.5 | location | 1.4 | individual | 0.8 |
| 1.4 | identifier | 1.6 | configured | 1.2 | chip | 1.2 | application | 1.2 | encrypted | 1 | management | 1.2 | feature | 1.3 | medium | 1.2 | telematics | 0.7 |
| 1.4 | request | 1.4 | interface | 1.1 | communication | 1.2 | communication | 1.1 | server | 0.9 | transfer | 1.1 | operation | 1.3 | target | 1.2 | medium | 0.7 |
| 1.3 | processing | 1.4 | terminal | 1 | terminal | 1.1 | web | 1 | request | 0.9 | electronic | 1.1 | and/or | 1.3 | node | 1.1 | member | 0.7 |
| 1.1 | customer | 1.4 | computer | 1 | electronic | 1 | access | 1 | code | 0.9 | instrument | 0.9 | insured | 1.2 | resource | 0.9 | current | 0.7 |
| 1.1 | authorization | 1.3 | response | 1 | control | 1 | medical | 0.9 | storage | 0.9 | banking | 0.9 | autonomous | 1.2 | crypto-currency | 0.8 | speed | 0.7 |
| 1 | item | 1.2 | electronic | 0.9 | wireless | 0.9 | wireless | 0.9 | stored | 0.8 | automated | 0.8 | determined | 1.1 | provisioning | 0.7 | life | 0.7 |
| 1 | receiving | 1.1 | document | 0.7 | configured | 0.8 | platform | 0.9 | encryption | 0.8 | payee | 0.8 | level | 1.1 | tag | 0.7 | window | 0.7 |
| 0.8 | code | 1 | receive | 0.7 | machine | 0.8 | management | 0.9 | computer | 0.7 | new | 0.8 | coverage | 1 | stored | 0.6 | cost | 0.6 |

기간 3에서는 지불 토픽의 비율이 줄어들면서 다른 토픽들의 중요도가 골고루 나뉘진 것을 확인할 수 있다. 또한, 스마트 카드, 신용카드 토픽이 동시에 등장한 것을 보아 결제 수단, 방식에 대한 새로운 기술들이 개발됐다고 볼 수 있다. 보안 토픽은 기간 2와 비교하여 오히려 하나로 뭉쳐졌고, 모바일 시장의 성장과 전기자동차와의 융합을 엿볼 수 있다. gensim 라이브러리의 LDAvis 패키지를 통해 LDA 토픽에 대해 시각화를 진행했다. LDAvis는 토픽에서 단어가 발생할 확률 벡터인 β_k 에 Principal Component Analysis (PCA)를 적용하여 2차원의 벡터로 압축한 후 시각화한다. 2차원 벡터로 표현된 토픽들이 가까이 위치한다면 유사하다고 판단한다. <Figure 7>은 기간별 LDA 토픽들의 유사도를 시각화하여 나타낸다. <Table 7>, <Table 8>, <Table 9>에서 각 토픽 옆에 있는 번호와 <Figure 7>의 번호는 동일한 토픽을 나타낸다. 번호 순서대로 원(토픽)의 우위성을 나타내고, 원이 서로 붙어있을수록 유사한 단어들을 공유한다. 기간이 지날수록 탐색 원과 하늘색 원이 가까워지는 것을 보아 지불과 네트워크, 통신의 결합이 증가한다고 볼 수 있다. 기간

1에서부터 금융상품관리시스템(노란색 원)이 나타났고, 기간 2에서 보안(회색 원)이 세분화되어 여러 토픽으로 나뉘었다. 기간 3에서는 모바일 결제와 모바일 지갑(초록색 원), 차량(갈색 원) 관련 토픽이 나타났고 보험(분홍색 원) 관련 토픽에서 자동차와 관련된 단어가 중요해짐을 발견했다.

4.5 네트워크 분석 결과와 토픽 모델링 결과 비교

두 분석 방법은 같은 데이터를 사용하지만, 데이터가 가진 속성의 차이에 따라 그 결과가 다르게 나타난다. CPC는 이미 갖춰진 특허 분류체계에서 기술의 구조적인 측면을 살펴볼 수 있지만, 특허의 초록에서 추출된 단어들은 특정 기술의 세부적인 내용을 살펴볼 수 있는 점에 있어서 다르다. 이러한 두 결과를 비교한다면 네트워크 분석에서 거시적으로 살펴봤던 산업에 대해 토픽 모델링을 통해 미시적 관점에서 살펴볼 기회가 제공된다. 또한, 중첩되지 않는 산업 주제를 통해 적용가능 산업을 더욱 확장하여 살펴볼 수 있다.



<Figure 7> Inter Topic Distance Map by Period

네트워크 분석 결과와 토픽 모델링의 결과를 기간별로 비교해보자. 우선 기간 1에서 네트워크 분석에서는 탄소 금융, 전기자동차에 관한 군집이 있지만, 토픽 모델링에서는 금융상품관리시스템과 세일즈에 관한 토픽이 나타났다. 이는 텍스트 정보에서 금융상품의 관리나 판매에 대한 전통적 기술이 나타나 아직 탄소 금융이나 차량 관련 핀테크 기술이 덜 발전한 것으로 해석할 수 있다. 기간 2의 네트워크 분석에서 디스플레이에 대한 중요도, 차량 관련 기술이 발전했지만, 토픽 모델링에서는 여전히 금융상품관리시스템이 높은 비율을 차지했고 보안의 중요성이 블록체인과 앱과 같은 세분화된 분야에서 강조되었다. 네트워크 보안 관련 CPC가 텍스트로 설명되어 구체적인 분야를 파악할 수 있었다. 마지막으로 기간 3의 네트워크 분석에서 의료, 화재, 운전 등의 토픽이 등장, 보험 관련 분야에 금융서비스가 결합한 것을 알 수 있다. 토픽 모델링에서도 차량과 관련된 보험 토픽이 중요해짐을 확인할 수 있다. 특히 차량 관련 토픽에서는 value, storage, location, time의 단어들이 등장했는데, 이는 자동차에 관한 클라우드 서비스, 시간 단축 알고리즘과 같은 기술이 핀테크에 적용이 가능할 것으로 보인다.

두 분석 모델을 전체적으로 비교, 분석해보면 네트워크 분석의 경우 기간 1부터 탄소금융, 전기자동차와 보험 등의 주제가 등장하여 자동차와 핀테크 융합의 중요성이 일찍이 대두되었고, 이에 관한 주제가 기간이 지남에 따라 보험, 자동차 결제기술로 세분화되었다. 반면 토픽 모델링은 꾸준히 높은 토픽의 비율을 유지한 금융상품관리시스템의 중요성이 지속적으로 강조되었고, 기간 3이 되어서야 자동차와 보험 관련 중요성이 나타났지만, 보안에 관련된

중요성은 조금 더 이른 기간 2부터 나타났다. 또한, 네트워크 분석에서 볼 수 없었던 모바일 결제와 모바일 지갑의 중요성을 알 수 있었다.

그 외에 전체적으로 토픽 모델링 분석과 네트워크 분석의 군집들은 유사한 주제를 가지는 것을 확인할 수 있었다. 종합적으로 기간이 지남에 따라 결제, 송금 분야와 네트워크 통신 분야가 융합되고 있는 점, 보안에 관련된 주제가 꾸준히 강조되고 있다는 점이다. 따라서 네트워크 통신과 보안에 관한 연구와 투자는 지속적으로 요구될 것이다.

또한, 두 모델에서 공통으로 전기차에 대한 수요가 급증함에 따라 핀테크와 융합하는 모습을 발견할 수 있었다. 이는 mobile-banking이 car-banking으로 적용될 수 있음을 암시한다. 구체적인 예로, 운전자 핸들의 지문센서를 통해 금융거래를 시도할 수 있다. 보험, 의료 산업과 핀테크의 융합으로 소비자는 합리적인 금융거래를, 기업은 효율적인 금융거래를 할 수 있다. 예를 들어 운전자 습관을 데이터화 하여 보험 및 의료 산업에 적용하거나, 웨어러블 디바이스로 의료비용을 결제하는 서비스를 제공할 수 있다.

4.6 기존 연구와의 비교 및 고찰

본 연구는 핀테크 관련 데이터를 수집하는데 있어 보다 체계적인 키워드 및 데이터 수집 과정을 제안하고, 이를 바탕으로 핀테크 기술의 적용가능 산업을 폭넓게 파악하고자 하는데 그 독창성이 있다. 따라서 본 연구에서 제안하는 데이터 수집 과정과 기존 연구에서 사용되어 온 데이터 수집 과정을 비교해 볼 필요가 있다.

기존 연구는 크게 두가지로 분류되며, Oxford 등의 금융사전을 사용하거나 뉴스를 활용해 용어 목록을 수립하는 방법과 USPC를 사용하는 BM특허(705/35)를 가지고 분석한 방법으로 나눌 수 있다. 전자의 경우, Lee and Choi[22]는 핀테크 관련 뉴스를 가지고 빈도분석을 적용해 <Table 10>과 같이 핀테크 관련 용어를 수립하였다. 본 연구의 용어 목록과 비교해보면 용어들이 다소 일반적이거나 핀테크의 세부 기술을 나타내기에 부족한 것을 확인할 수 있다. 후자의 경우, BM특허를 이용한 첫 번째 방법과 본 연구의 방법을 비교해 보고자 한다.

먼저 기존 USPC 705/35를 활용하여 2010년부터 2015년까지(USPC는 2015년부터 지원 종료)의 핀테크 특허를 수집하면 4,385개가 수집된다. 이는 본 연구 기간 1(2010~2013)의 특허수 8,223개와 비교해 보면 본 연구 기간이 USPC 연구 방법의 기간보다 짧음에도 불구하고 약 2배 정도 차이 나는 것을 확인할 수 있다. 본 연구와의 정확한 비교를 위해 uspc 방법으로 수집된 특허를 가지고 네트워크 분석을 진행한 결과 노드는 245개, 엣지는 2,330개였다. 이는 본 연구의 네트워크 지표가 노드, 엣지 모두 2배 이상 큰 것을 보여준다. <Table 11>은 USPC를 활용한 핀테크 네트워크가 핀테크 산업을 표현하기에 상대적으로 작음을 보여준다. USPC를 활용한 네트워크에 대해 모듈성 분석을 수행하면 11개의 군집이 나타나지만 본 연구 period1의 네트워크와 동일하게 5% cut-off를 적용하면 3개의 군집만을 얻을 수 있었다. 3개의 군집은 ‘지불, 거래, 금융에 사용되는 데이터 처리 시스템’, ‘네트워크 보안, 무선 통신 관련 기술’ 그리고 ‘전기 자동차, 네비게이션 관련 기술’로 나타났다. 물론 네트워크 크기가 크다고 해서 무조건

그 산업을 더 잘 반영한다고 단정하기는 어렵지만, 모듈성 분석 결과를 확인했을 때 본 연구의 방법이 USPC 705/35보다 더 다양한 핀테크 산업을 나타내기에 새롭게 핀테크 용어를 수립하는 것이 충분히 효과적임을 확인할 수 있다.

USPC 705/35 특허를 가지고 본 연구의 방법과 동일하게 최적 토픽 개수와 passes를 구하여 토픽 모델링을 구현한 결과, 15개 중 8개의 유효한 토픽을 얻었다. 8개의 토픽은 ‘지불’, ‘포트폴리오’, ‘컴퓨터’, ‘문서, 기록’, ‘카드’, ‘보험’, ‘대출’ 그리고 ‘트레이딩’이다. 본 연구의 토픽 모델링 결과와 비교했을 때, 핀테크 산업 중 중요하다고 여겨지는 ‘보안’과 ‘금융상품관리’에 관한 토픽 결과를 찾기 어려웠다. 이는 네트워크 비교 분석 결과와 비슷하게 USPC 705/35 특허가 핀테크를 충분히 표현하기에 비교적 어려운 것을 보여준다.

<Table 10> Previous Research's Terms

| Lee and Choi(2015)'s fintech terms | |
|---|--|
| finance, payment, bank, electronic, investment, trading, organization, foreign exchange, stock, security, card, identification, biometrics, token, hacking, detection, malignant, defense, generator, service, client, communication, experience, platform, speed, infra, tool, easy, voice | |

<Table 11> Comparison with Previous Research and this Research

| | fintech network (USPC 705/35) (2010~2015) | period 1 network (2010~2013) |
|------------------|---|------------------------------|
| Number of Patent | 4,385 | 8,223 |
| Node | 245 | 669 |
| Edge | 2,330 | 6,747 |
| Modularity | 11 | 20 |

5. 결 론

본 연구는 핀테크 기술동향을 정확히 파악하기 위해 서지학적 정보를 활용하여 3가지 종합적 방법을 사용하고 핀테크 용어 목록을 생성하였다. 그리고 생성된 단어들로 핀테크 관련 특허를 얻었다.

다음으로 수집된 특허들로 네트워크 분석과 토픽 모델링을 진행했다. CPC의 네트워크를 기간별로 살펴본 결과 지불에 관한 기술은 계속 압도적이었으며 시간이 지날수록 무선 네트워크, 통신 관련 CPC와 결합한 형태를 띠었다. 무선 네트워크, 통신 관련 CPC는 점점 세분되어 나타났으며, 중심성 분석에서도 우위를 점하였다. 토픽 모델링 분석 결과 네트워크 분석과 비슷한 양상을 보여 무선 네트워크, 통신, 보험, 자동차, 보안 등이 중요 토픽임을 확인할 수 있었다. 네트워크 분석과는 달리 금융상품 관리시스템과 모바일, 앱 관련 토픽이 등장하여 CPC로는 찾아볼 수 없던 주제를 텍스트를 통해 확인할 수 있었다.

이에 따라 기업은 무선 네트워크, 통신의 속도, 효율성, 보안, 접근성 등에 투자하여 더 경쟁력 있는 핀테크 기술을 가질 것으로 기대한다. 또한, 전기 자동차와 관련된 CPC가 기간 3에 대거 등장한 것으로 보아 자동차와 핀테크를 활용하는 기술들이 계속 증가할 것으로 예상된다. 보험사가 정확한 보험료 산정을 위해 운전자의 습관을 데이터화 하는 것처럼 핀테크 기술은 단순한 모바일 지불, 결제뿐만 아니라 IOT가 적용되는 물체와의 융합이 일어날 것으로 보인다.

본 연구의 결과는 핀테크 기술을 바탕으로 새로운 금융서비스에 참여하고자 하는 경영자

및 투자자가 핀테크 기술의 정의와 특징을 명확히 이해하고, 핀테크 기술의 동향과 적용산업의 변화를 분석할 수 있는 틀을 제공할 수 있다는 데 그 의의가 있다.

본 연구의 한계점은 다음과 같다. 먼저 핀테크의 적용 범위가 방대하기에 본 연구에서 제시한 특허 수집방법이 반영하지 못하는 범위가 존재할 수 있다는 점이다. 핀테크 관련 특허를 정확하게 분류할 수 있는 연구가 추가로 진행되어 정보의 손실을 최소화하기 위한 추가 연구가 필요하다. 둘째, 텍스트 데이터에 대하여 클러스터 개수를 정하는 문제나 클러스터를 고르고 이름을 짓는 문제에서 주관적인 해석과 판단을 배제하기 어려울 수 있다. 이는 클러스터 분석, 토픽 모델링 분석에서 겪는 고질적인 문제로서 최신 알고리즘을 적용하여 주관적 해석을 최소화할 수 있을 것이다. 셋째, 변화 속도가 빠른 핀테크 기술 특성상 10년이란 기간이 최신 기술동향을 살펴보기에 길다고 느껴질 수 있다. 더 짧은 기간에 따라 특허를 수집하여 추가적인 분석을 진행한다면 더욱 자세한 핀테크 기술동향을 파악할 수 있을 것으로 생각된다. 마지막으로, 본 연구에서는 기간을 구분하여 별도로 토픽 모델링을 실시하였으나, 보다 효과적인 동태적 분석을 위해 dynamic topic modelling 등과 같은 다른 방법론을 종합적으로 고려할 수 있을 것이다.

References

- [1] Allahyari, M. and Kochut, K., "Semantic tagging using topic models exploiting Wiki-

- pedia category network,” In 2016 IEEE Tenth International Conference on Semantic Computing(ICSC), Vol. 44, No. 3, pp. 63-70, 2016.
- [2] Arner, W., Barberis, J., and Buckley, R. P., “The evolution of Fintech: A new post-crisis paradigm,” University of Hong Kong Faculty of Law Research Paper, 2015.
- [3] Bavelas, A., “Communication Patterns in Task Oriented Groups,” Journal of the Acoustical Society of America, Vol. 22, pp. 271-282, 1950.
- [4] Blei, D. M., Ng, A. Y., and Jordan, M. I., “Latent dirichlet allocation,” Journal of machine Learning Research, Vol. 3, pp. 993-1022, 2003.
- [5] Blondel, V. D., Guillaume, J. L., Lambiotte, R., and Lefebvre, E., “Fast unfolding of communities in large networks,” Journal of Statistical Mechanics: Theory and Experiment, No. 10, pp. P10008, 2008.
- [6] Chen, M. A., Wu, Q., and Yang, B., “How valuable is FinTech innovation?,” The Review of Financial Studies, Vol. 32, No. 5, pp. 2062-2106, 2009.
- [7] Curran, C. S. and Leker, J., “Patent indicators for monitoring convergence-examples from NFF and ICT,” Technological Forecasting and Social Change, Vol. 78, No. 2, pp. 256-273, 2011.
- [8] Davis, L., Fintech Trends to Watch, Retrieved from <https://www.cbinsights.com/research/report/fintech-trends-2019>, 2019.
- [9] Devezas, T. C., “Evolutionary theory of technological change: state-of-the-art and new approaches,” Technol. Technological Forecasting and Social Change, Vol. 72, No. 9, pp. 1137-1152, 2005.
- [10] Geum, Y., Kim, C., Lee, S., and Kim, M. S., “Technological convergence of IT and BT: evidence from patent analysis,” Etri Journal, Vol. 34, No. 3, pp. 439-449, 2012.
- [11] Ghosh, I., Visualizing the current landscape of the fintech industry, Retrieved from <https://www.visualcapitalist.com/current-fintech-industry/>, 2020.
- [12] Gianchandani, N., The business research company, Retrieved from <https://www.thebusinessresearchcompany.com/report/fintech-market>, 2019.
- [13] Hong, M. and Jeong, B., “A study on the Activation Plan according to the Trend of FinTech Industry in Major Countries,” The e-Business Studies, Vol. 19, No. 6, pp. 369-383, 2018.
- [14] Joorabchi, A., English, M., and Mahdi, A. E., “Automatic mapping of user tags to wikipedia concepts: The case of a Q&A website-stackoverflow,” Journal of Information Science, Vol. 41, No. 5, pp. 570-583, 2015.
- [15] Kim, G. and Shim, M., “Study on the Mobile FinTech Vacant Technology using Patent Analysis,” International Journal of u-and e-Service, Science and Technology, Vol. 9, No. 9, pp. 55-64, 2016.
- [16] Kim, T., Choi, H., and Lee H., “A Study on the Research Trends in Fintech using

- Topic Modeling,” *Journal of the Korea Academia-Industrial cooperation Society*, Vol. 17, No. 11, pp. 670-681, 2016.
- [17] Kwon, H., Park, Y., and Geum, Y., “Toward data-driven idea generation: Application of Wikipedia to morphological analysis,” *Technological Forecasting and Social Change*, Vol. 132, pp. 56-80, 2018.
- [18] Kwon, S., *Reviewing the Patent Issues of BM Inventions Focused on Fintech (Master degree)*, Graduate School of Legal Studies, Korea University, Seoul, 2018.
- [19] Leavitt, H. J., “Some effects of certain communication patterns on group performance,” *The Journal of Abnormal and Social Psychology*, Vol. 46, No. 1, pp. 38-50, 1951.
- [20] Lee, S., Yoon, B., and Park, Y., “An approach to discovering new technology opportunities: Keyword-based patent map approach,” *Technovation*, Vol. 29, No. 6-7, pp. 481-497, 2009.
- [21] Lee, G. and Kim, H., “Automated Development of Rank-Based Concept Hierarchical Structures using Wikipedia Links,” *The Journal of Society for e-Business Studies*, Vol. 20, No. 4, pp. 61-76, 2015.
- [22] Lee, H. and Choi, E., “Essential Technical Patent Extraction Method Associated with Fintech Based on Text Mining,” In *Proceedings of the Korea Information Processing Society Conference*, pp. 1219-1222, 2015.
- [23] Lee, Y., “Analysis of ‘Data 3 Laws’ and future tasks,” *The Korean Association of Comparative Private Law*, Vol. 27, pp. 423-465, 2020.
- [24] Lee, I. and Shin, Y. J., “Fintech: Ecosystem, business models, investment decisions, and challenges,” *Business Horizons*, Vol. 61, No. 1, pp. 35-46, 2018.
- [25] Lerner, J., “The importance of patent scope: an empirical analysis,” *The RAND Journal of Economics*, pp. 319-333, 1994.
- [26] Li, G., Dai, J. S., Park, E., and Park, S., “A study on the service and trend of Fintech security based on text-mining: Focused on the data of Korean online news,” *Journal of Computer Virology and Hacking Techniques*, Vol. 13, No. 4, pp. 249-255, 2017.
- [27] Momeni, A. and Rost, K., “Identification and monitoring of possible disruptive technologies by patent-development paths and topic modeling,” *Technological Forecasting and Social Change*, Vol. 104, pp. 16-29, 2016.
- [28] Park, B., Choi, S., Kim, D., and Guahk, J., “Analysis of domestic and foreign Fintech service and policy trends,” *The Journal of Korean Institute of Communications and Information Sciences*, Vol. 34, No. 3, pp. 3-10, 2017.
- [29] Schueffel, P., “Taming the beast: a scientific definition of fintech,” *Journal of Innovation Management*, Vol. 4, No. 4, pp. 32-54, 2016.
- [30] Scott, J., “Social network analysis,” So-

- ciology, Vol. 22, No. 1, pp. 109-127, 1988.
- [31] Seo, S. and Lee, H., "Fintech trend analysis using topic modeling of BM patents," Seoul National University of Science and Technology, Seoul, 2015.
- [32] Shim, Y. and Shin, D. H., "Analyzing China's fintech industry from the perspective of actor-network theory," Telecommunications Policy, Vol. 40, No. 2-3, pp. 168-181, 2016.
- [33] Zavolokina, L., Dolata, M., and Schwabe, G., "FinTech-What's in a Name?," Thirty Seventh International Conference on Information Systems, Dublin, 2016.

저 자 소 개



박민규 (E-mail: mgpark@ds.seoultech.ac.kr)
2014년 3월 한국방송통신대학교 경영학과 (학사)
2019년 9월 ~ 현재 서울과학기술대학교 일반대학원 데이터사이언스학과 (석사과정)
관심분야 Data mining, Machine Learning



전병민 (E-mail: dkffkam@naver.com)
2015년 3월 ~ 현재 서울과학기술대학교 산업공학과 산업정보시스템전공 (학석사연계과정)
관심분야 Data mining



김종우 (E-mail: whddn12316@naver.com)
2016년 3월 ~ 현재 서울과학기술대학교 산업공학과 산업정보시스템전공 (학사과정)
관심분야 Data mining, Natural language



김영정 (E-mail: yjgeum@seoultech.ac.kr)
2004년 KAIST 산업공학 (학사)
2013년 서울대학교 산업공학 (박사)
2014년 ~ 2020년 3월 서울과학기술대학교 산업공학과 조교수
2020년 4월 ~ 현재 서울과학기술대학교 산업공학과 부교수
관심분야 Technology Management, Data-driven innovation, Service Engineering, Business Model Innovation