

2022년 한국정보시스템학회 추계학술대회

데이터 기반 디지털 금융과 비즈니스 가치 창출



Session C

발표 장소 : 창의관 3층 303호

C1 (13:30~15:10) 데이터활용기술

좌장 : 강재정(제주대)

순서	저자(소속)	논문제목	
1	이현상, 조보근, 오세환, 하성호(경북대)	섬유소재 분야 대용량 텍스트 데이터 활용 기술 동향 분석 : 딥러닝 기반 단어 임베딩 및 시계열 토픽 모델링 기법 적용	159
2	이정민, 이민혁 (부산대)	A Market-Friendly Risk Management Model for DABS Investment : Garnish 'Housing' and 'CMO' Relieves Bitter 'Liquidity Problem' Tastes of DABS Pie	167
3	서창갑(동명대) 정대현(부산대)	빅데이터를 활용한 장애인 콜택시 운영 효율성 제고 방안	170
4	김덕현, 정대울 (경상국립대)	온실가스 배출권 및 신재생에너지 공급인증서 거래 예측모형 개발 및 패턴 분석 : KAU 가격과 REC 가격을 중심으로	178
5	윤혜민, 김상진 (동아대)	컨볼루션 신경망을 이용한 3차원 영상 판독 기술의 비교 연구	189
6	박혜은, 이종태 (서울여대)	기후예측을 통한 미래 과수작물지도 : 바나나, 망고 중심으로	190

섬유소재 분야 대용량 텍스트 데이터 활용 기술 동향 분석: 딥러닝 기반 단어 임베딩 및 시계열 토픽 모델링 기법 적용¹⁾

이현상 (경북대학교 경영학부 박사과정, 주저자 coolwin200@gmail.com)

조보근 (경북대학교 경영학부 박사과정, 공동저자 smchoi@hankuk.ac.kr)

하성호 (경북대학교 경영학부 교수, 교신저자 hsh@knu.ac.kr)

오세환 (경북대학교 경영학부 부교수, 공동저자 sehwan@knu.ac.kr)

··· Abstract ···

In order to analyze technology trends in a specific industry field, existing research uses text analysis such as network analysis and topic modeling. In particular, topic modeling can classify micro topics as well as macro aspects of industry trends. The topic modeling technique has the advantage of being able to perform unsupervised learning without labeling processing on large text data, but there is a limitation in that it cannot analyze the dynamic time-series properties of topics and the topics of keywords that do not appear in the document. In addition, the existing topic modeling methodology may have a problem that topics are not properly clustered depending on the noise of the training data. This paper intends to analyze the technology trends in the textile material field through advanced analysis from the existing topic modeling methodology using the DETM (Dynamic Embedded Topic Model) technique. This proposed technique can use deep learning technology in the topic modeling model structure to set the initial value of the topic probability of a word, and reflect the word embedding and dynamic time series properties in model learning. Through this paper, we intend to empirically analyze the technology trends in the field of textile materials.

Key Words : Textile, Topic Model, DETM, Trend Analysis, Word Embedding

I. 서론

다양한 분야의 기술 연구가 등장하고, 고도화 되기 시작하면서 정성적인 방법론을 기반으로

특정 분야의 기술 동향을 분석하고자 하는 연구의 한계점이 나타나고 있다. 이에 실증적인 방법론으로 대량의 텍스트를 분석하여 주제 및 키워드를 탐색하는 시도가 등장하고 있다

1) 이 논문은 2022년도 산업통상자원부 산업혁신기반구축사업 재원으로 수행된 연구임 (P114000015)

(Amado et al., 2018; Yang et al., 2018; 홍태호 외., 2018; 최병설 외., 2019; 김민수 외., 2020). 빅데이터 텍스트의 전반적인 기술 동향을 분석할 수 있는 연구에는 일반적으로 네트워크 분석(network analysis) 및 토픽 모델링(Topic Modeling) 기법이 활용되고 있다. 네트워크 분석 기법은 텍스트 데이터의 키워드 관계성을 분석할 수 있는 방법론을 의미하며, 토픽 모델링 기법은 문서의 숨겨진 주제를 군집하여 기술 동향을 분석할 수 있다(Scott, 1988; Hofmann, 1999). 그러나 네트워크 분석 기반의 기술 동향 연구의 경우 문서에 전반적으로 나타나는 주제를 포착하기 어렵고, 토픽 모델링 연구의 경우 군집된 키워드만을 분석할 수 있으며 시간적인 정보가 토픽에 반영되지 않는다는 한계점이 존재한다.

따라서 본 연구는 기존 텍스트 동향 분석의 한계점을 극복할 수 있도록 DETM(Dynamic Embedded Topic Model)을 활용하여 섬유소재 분야의 특허 문서 데이터를 분석하고자 한다. DETM 기법은 문서의 시계열 정보를 주제 군집에 반영할 수 있으며, 워드 임베딩(Word Embedding)을 통하여 분석 결과를 개선하면서 주제 안에 포함되지 않는 단어들도 동향을 분석할 수 있다. 본 연구를 통하여 특정 기술을 분석하고자 하는 연구자들과, 섬유소재 분야의 연구자 및 실무자들에게 참고점을 제시하고 해당 분야의 기술적 발전을 기대한다. 또한, 최근 국내 섬유소재 산업 분야에서 빅데이터 플랫폼 구축 사업이 진행되고 있는데, 이에 대한 기술적 동향을 제시함으로써 해당 플랫폼의 활성화를 기대한다(매일신문, 2020).

II. 관련 연구

토픽 모델링 기법이란 문서의 단어 집합에서 숨겨진 주제를 군집하는 방식으로 분석할 수 있는 알고리즘으로, LSA(Latent Semantic Analysis) 연구로 시작되었다(Hofmann, 1999). 토픽 모델링 분석을 통해서 연구자는 의미론적인 주제 기반의 키워드들과 문서당 각 토픽의 확률을 추론할 수 있다. 분석 결과에 따라서 하나의 문서가 어떤 주제들과 얼마나 연관성이 있는지 파악할 수 있다. 최근 토픽 모델링 연구에서는 레이블링 정보에 따른 메타 정보를 주제 군집 프로세스에 반영하거나, 시계열 정보 및 워드 임베딩 벡터값을 군집에 반영하려는 연구가 시도되고 있다(Roberts et al., 2013; Dieng et al., 2019). 토픽 모델링 분석 기법은 텍스트 데이터에 대한 레이블링 처리 없이 수행될 수 있고, 다양한 분석 결과를 정량적으로 제시하기 때문에 대량의 문서 데이터를 분석할 때 유용하게 활용되고 있다(Amado et al., 2018; Yang et al., 2018; 홍태호 외., 2018; 최병설 외., 2019; 김민수 외., 2020).

Amado et al. (2018)은 토픽 모델링을 활용하여 빅데이터 마케팅 분야의 기술 동향을 분석했다. 2010년부터 2015년까지 총 1,560건의 논문 데이터를 분석했으며 빅데이터, 마케팅, 지리적 특성, 제품 및 제품군 등의 주제를 발견했다. Yang et al. (2018)은 스마트팩토리 분야의 기술 동향을 분석하기 위해 LSA 기법을 활용했다. ICT, R&D, 기술 혁신 주제 등에 대해서 국내와 해외 연구 동향을 비교 분석했다. Chen et al. (2020)은 STM(Structural Topic Model) 기법을 통해서 Computer & Education 저널의 논문 데이터 3,963건을 분석했다. 해당 연구는

Computer & Education 투고를 위한 참고점을 제시했다.

Ⅲ. 제안 기법

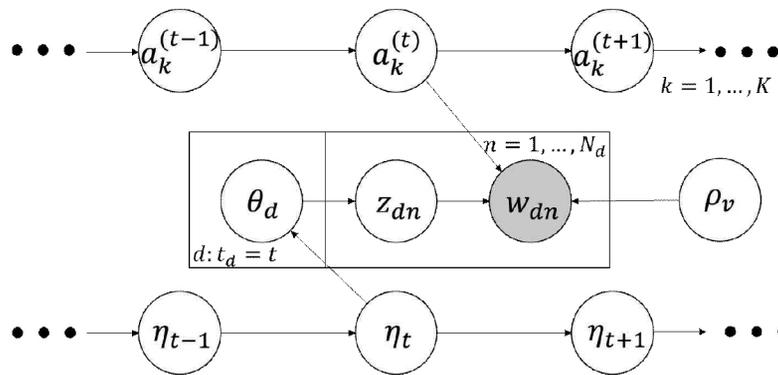
1. DETM

LDA(Latent Dirichlet Allocation) 기법은 문서의 주제를 잠재 변수를 활용하여 확률적으로 추론 후 단어 군집 집합을 생성하는 방법론을 의미한다(Blei et al., 2003). DETM(Dynamic Embedded Topic Model)은 LDA 구조에서 주제 군집 학습 시 3가지 응용 기법을 제시한다(Dieng et al., 2019). 첫 번째, 토픽을 나타내는 다항 분포의 매개 변수에 state-space 기법을 적용하여 토픽별 시계열적인 추세를 분석했다(Blei et al., 2006). 두 번째, 문서의 단어 집합을 워드 임베딩으로 분석하여 벡터값을 모델링에 반

영하는 방식으로 주제 군집 결과를 개선했다. 세 번째, 토픽 모델링 시 확률 초기값을 RNN(Recurrent Neural Network) 기반 구조적 상각 변동 추론(anortized variational inference)을 활용하여 군집 안정성을 보완했다.

〈그림 1〉은 DETM의 모델 학습 프로세스를 표현했다. 단어 토픽 분포 n 은 토픽 임베딩 $a_k^{(t)}$ 와 단어 임베딩 ρ_v 의 내적을 통해서 계산된다. $a_k^{(t)}$ 는 k번째 주제에 대한 임베딩 벡터 공간상에서의 분포 형태로 정의되며 ρ_v 은 워드 임베딩의 벡터값을 의미한다. 워드 임베딩의 벡터값을 통해서 단어의 의미론적 유사도를 파악할 수 있다. w_{dn} 은 단어 및 토픽 임베딩이 일치할 때 증가하기 때문에 유사한 단어가 유사한 주제에 반영된다. η_t 는 시간 t에 따라서 변화하는 잠재 변수를 의미한다. 문서의 토픽 확률 θ_d 는 η_t 로 결정되고, z_{dn} 으로 최종 토픽을 결정한다.

〈그림 1〉 DETM Algorithm



Ⅳ. 분석

1. 데이터

본 연구는 AIHub에서 수집한 2018년부터 2019년까지의 섬유소재 분야 특허 문서 초

록 및 청구항 텍스트 데이터 11,291건을 활용하여 토픽 모델링을 수행했다. 섬유소재 관련 키워드를 통해서 관련 문서들을 수집했으며, 각 데이터는 특허 제목, 등록일자, 텍스트 정보를 가지고 있다. 특허 문서 텍스트 데이터의 명사만을 추출 후 토큰에 대해서 카운

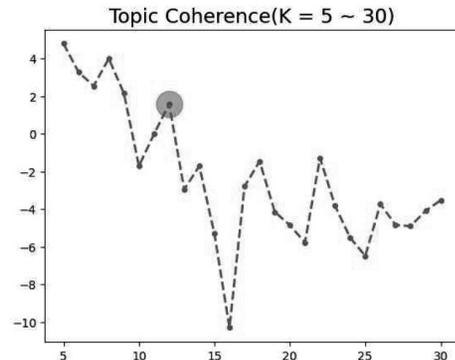
트 벡터화(count vectorization)를 수행했다. 본 연구에서는 단어의 빈도를 기반으로 토픽 모델링을 학습했다. 다음 단계로는 한국어 GloVe 사전학습 모델을 통해서 워드 벡터값을 생성했다. GloVe란 전역적 대수 선형 회귀(global logbilinear regression) 분석을 기반으로 워드 벡터를 생성하는 것을 의미한다(Pennington et al., 2014). 이 워드 벡터값을 통해서 단어들의 의미론적 유사성을 파악할 수 있다. 최종적으로 학습, 검증, 테스트를 위해서 8.5:0.5:1의 비율로 데이터를 분할했다.

2. 모델링

본 연구에서는 데이터 처리 과정에서 생성된 학습, 검증, 테스트 데이터를 DETM 모델에 입력했다. 모델 학습 시 토픽 개수 k 를 설정하기 위해서 토픽 일관성(topic coherence)을 기준으로 최적 토픽 수를 탐색했다. <그림 2>의 결과로 토픽 k 는 12로 설정했다. 토픽 일관성이 내려가는 추세에서 급격하게 증가하는 시점을 선정했다. k 값 설정 후 DETM 모델 학습을 진행했으며, 토픽 혼재성(topic complexity)이 가장 낮은 구간의 학습 모델을 최종 모델로 선정했다. 모델링 결과로 추론된 단어 토픽 확률과 문서 토픽 확률을 사용하여 총 3가지 분석 결과를 도출했다. 첫 번째로, 토픽별 키워드와 점유율, 그리고 트렌드를 분석했다. 해당 결과를 통해서 섬유소재 분야에서 연구되고 있는 주제 및 동향을 알 수 있다. 두 번째, 모델링 결과와 워드 임베딩 벡터값을 활용하여 특정 키워드에 대한 동향을 분석했다. 기존 연구에서는 주제에 군집되지 않는 단어를 분석하기 어렵다는 단점이 존재하지만, 본 연구의 DETM 기법은 유사 단어의 토픽

확률을 계산하는 방식을 활용할 수 있다.

<그림 2> DETM Algorithm



V. 연구 결과

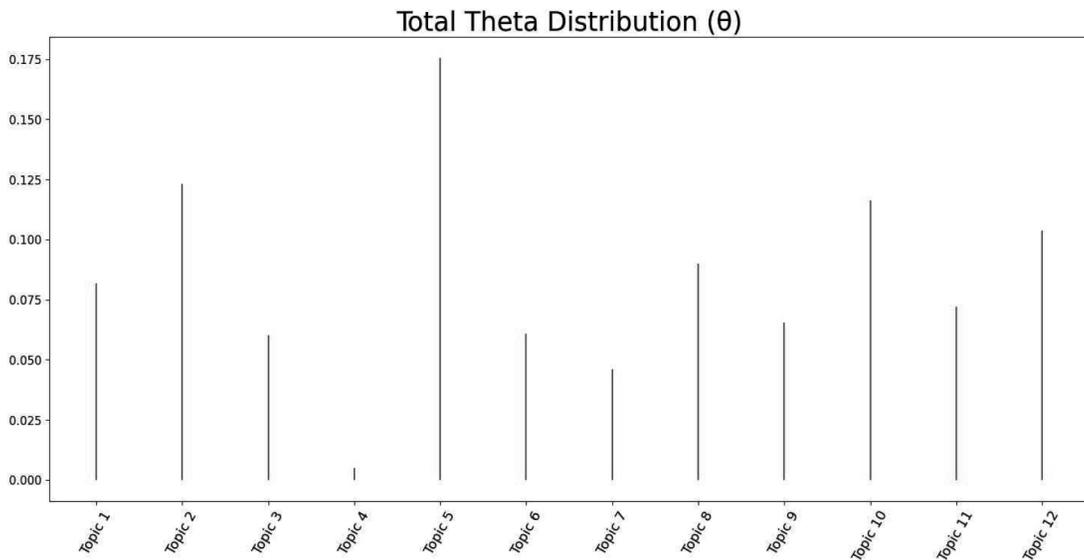
1. 토픽 분석 결과

토픽 분석 결과는 <표 1>과 <그림 3>과 같다. 12개의 토픽명은 기존 연구 및 군집 키워드를 기준으로 부여했다(Raichurkar et al., 2015; Lee et al., 2021). <표 1>의 키워드들은 DETM 분석에 의해서 군집된 각 토픽의 단어 중 가장 포함 확률이 높은 단어들이다. 섬유소재 분야 특히 문서에서 모든 단어들의 토픽 확률을 계산해서 토픽의 점유율을 계산한다. <그림 3>과 같이, 토픽 2, 5, 10의 토픽 점유율이 높게 측정되고, 3, 4, 7은 다소 낮게 나타났다. 토픽 점유율이 낮다는 것은 해당 주제와 관련된 문서 수가 적다는 것을 의미한다. <그림 4>는 토픽 점유율의 추이를 주제별로 시각화한 것이다. 이에 따라 토픽 5, 7, 12 관련 기술, 즉 소재성분, 섬유 강화, 안전과 관련된 특히 기술이 활발하게 연구되고 있다는 사실을 알 수 있다. 반면에 토픽 2, 3, 6, 모듈, 식물, 설계 관련 연구는 등록된 특히 문서가 감소하고 있는 추세가 나타났다.

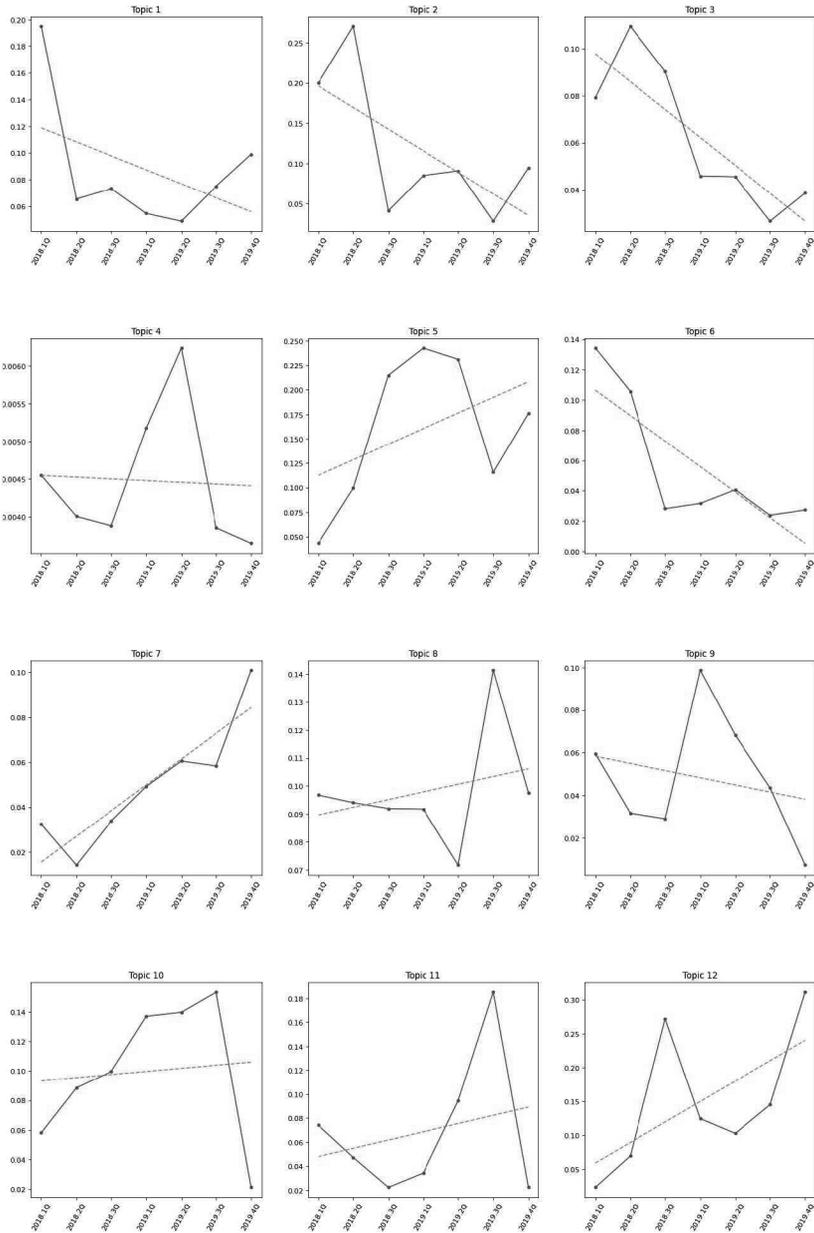
〈표 1〉 토픽 분석 결과

id	토픽명	키워드
Topic 1	패널	패널, 부탁, 부재, 전극, 몰드, 태양, 회동, 전지, 적용, 감응 등
Topic 2	모듈	모듈, 압출, 시트, 조명, 필름, 패턴, 다이오드, 표면, 합성섬유, 마이크 등
Topic 3	직물	위사, 경사, 직물, 제직, 반사, 잔류, 정렬, 홀더, 조직, 레이저 등
Topic 4	가공	분쇄, 발취, 외피, 내구, 고무, 혼방, 라디칼, 절감, 지속, 재생 등
Topic 5	소재성분	탄화수소, 성분, 폴리아미드, 에폭시, 아라미드, 공법, 탄소, 콘크리트, 산화물, 응고, 중합체 등
Topic 6	설계	본체, 상면, 몸체, 용기, 장착, 볼트, 전면, 수납, 스프링, 고정부, 후면, 돌출, 삽입 등
Topic 7	강화	강화, 데이터, 실리콘, 유리, 측정, 코팅, 결정, 계산, 시스템, 검출, 전도성, 모니터링 등
Topic 8	조제	폴리에스테르, 방적사, 조제, 섬도, 단섬유, 향균, 냉감, 증공, 연사, 가연, 핵산, 발현, 복합사 등
Topic 9	천연섬유	폴리아크릴, 천연섬유, 원자, 열풍, 수분, 건강, 이온, 정수, 알킬, 할로젠 등
Topic 10	공기	공기, 터브, 필터, 유입, 유로, 구비, 장치, 순환, 커버, 하우징, 공급, 배출, 밸브 등
Topic 11	화학	활성, 방사선, 화상, 조성물, 카본, 단량체, 화합물, 화학식, 중합, 지방산, 합성, 유기 등
Topic 12	안전	펄스, 냉각, 저항, 피복, 미연, 합성수지, 단계, 방사성, 분자량, 폴리프로필렌, 방사 등

〈그림 3〉 토픽 점유율



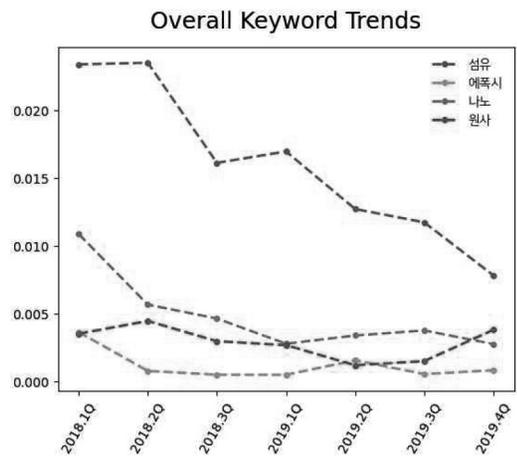
〈그림 4〉 분기별 토픽 트렌드



〈그림 5〉 키워드별 트렌드 분석

2. 키워드 분석 결과

본 연구의 DETM 방법론은 워드 임베딩 기법을 활용하고 있기 때문에 특정 기술 키워드의 트렌드를 복합적으로 분석할 수 있다는 장점이 있다. 〈그림 5〉의 분석 결과는 키워드의 유사 단어 20개를 추출하여 토픽 트렌드를 분석하는 방식으로 해당 분야의 트렌드를 분석했다.



VI. 결론

본 연구는 섬유소재 분야의 기술 동향을 분석하기 위해서 특히 문서 텍스트 데이터를 활용하여 토픽 모델링을 수행했다. 기존 연구의 한계점을 보완하기 위해서 DETM 방법론을 활용했으며, 12개의 토픽 키워드에 대한 점유율, 트렌드를 분석 결과로써 제시했다. 본 연구에서 적용된 DETM의 워드 임베딩 모델을 활용하여 특정 키워드에 대한 트렌드를 최종적으로 제시했다. 본 연구의 공헌점은 다음과 같다. 첫 번째, 섬유소재 분야의 특히 텍스트 데이터를 활용하여 관련 기술의 동향을 정량적으로 분석했다. 두 번째, 기존 LDA 기반 기술 동향 분석에서 고도화된 DETM 방법론

을 활용하여 한국어 텍스트 데이터에서의 적용 가능성을 실증했다. 세 번째, GloVe 워드 임베딩 모델을 통해서 군집된 단어뿐만 아니라, 특정 키워드의 동향을 분석할 수 있는 방법론을 제시했다. 본 연구의 한계점은 특히 텍스트 데이터 수집의 문제로 문서의 수가 부족하다는 것과 해외 연구를 같이 분석하지 못했다는 것이다. 향후 연구에서는 글로벌 섬유 동향을 국가별로 비교하는 방식으로 고도화가 가능할 것이다. 본 연구를 통해서 섬유 관련 실무자 및 연구자들에게 기술 동향을 제시하고, 특정 분야의 기술 동향을 고도화된 토픽 모델링 기법을 기반으로 분석하는 방법론에 대한 참고점이 될 수 있기를 기대한다.

참고 문헌

김민수, 김희웅. "온라인 해킹 불법 시장 분석: 데이터 마이닝과 소셜 네트워크 분석 활용", 정보시스템연구 29권, 2호, 2020, pp. 221-242.

이현상, 조보근, 오세환, 하성호. "섬유소재 분야 특히 기술 동향 분석: Detm & Stm 텍스트마이닝 방법론 활용", 정보시스템연구 30권, 3호 2021, pp. 201-216.

최병설, 김남규, "감정 딥러닝 필터를 활용한 토픽 모델링 방법론", 정보시스템연구 28권, 4호, 2019, pp. 271-291.

홍태호, 임강, 박지영. "Lda 를 이용한 온라인 리뷰의 다중 토픽별 감성분석-Tripadvisor 사례를 중심으로", 정보시스템연구 27권, 1호, 2018, pp. 89-110.

Amado, A., Paulo C., Paulo R., and Sérgio M., "Research Trends on Big Data in Marketing: A Text Mining and Topic Modeling Based Literature Analysis", European Research on Management and Business Economics, Vol. 24, No. 1, 2018, pp. 1-7.

Blei, D. M., Andrew, Y. Ng., and Michael I J., "Latent Dirichlet Allocation." the Journal of machine Learning research, Vol. 3, 2003, pp. 993–1022.

Chen, X., Di, Z., Gary, C., and Haoran X., "Detecting Latent Topics and Trends in Educational Technologies over Four Decades Using Structural Topic Modeling: A Retrospective of All Volumes of Computers & Education." Computers & Education, Vol. 151, 2020.

Dieng, A. B., Francisco, JR, R., and David, M, B.. "The Dynamic Embedded Topic Model", arXiv preprint arXiv:1907.05545, 2019.

Hofmann, T., "Probabilistic Latent Semantic Indexing", Paper presented at the Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, 1999.

Raichurkar, P., and Manickam, R., "Recent Trends and Developments in Textile Industry in India", International Journal on Textile Engineering & Processes, Vol. 1, no. 4, 2015, pp. 47–50.

Roberts, Margaret, E., Brandon, M, Stewart., Dustin, T., and Edoardo, M, A.. "The Structural Topic Model and Applied Social Science", Paper presented at the Advances in neural information processing systems workshop on topic models: computation, application, and evaluation, 2013.

Scott, J., "Social Network Analysis", Sociology Vol. 22, No. 1, 1988, pp. 109–127.

Yang, H, L., Tai–Woo, C., and Yerim C., "Exploring the Research Trend of Smart Factory with Topic Modeling", Sustainability, Vol. 10, No. 8, 2018.

A Market-Friendly Risk Management Model for DABS Investment : Garnish 'Housing' and 'CMO' Relieves Bitter 'Liquidity Problem' Tastes of DABS Pie

이정민 (부산대학교 경영학과 석사과정, 주저자 leonism@pusan.ac.kr)
이민혁 (부산대학교 경영학과 교수, 교신저자 minhyuk.lee@pusan.ac.kr)

… Extended Abstract …

Publicity of NFT(Non-Fungible Token) and Blockchain is overflowing. According to 'NFT in Real Estate' report released by Upbitcare on April 7, 2022, The influence of these two technologies is rushing to real estate sector through DABS(Digital Asset-Backed securities) contents. These contents generate housing liquidity by cubing housing NFT securitization. Housing NFT securitization means using NFT to convert housing assets into digitization and carve it into a token system. On the other hand, looking at a conclusion of the Upbitcare report, it is said that DABS services possess three risk factors in terms of investment: 1. Liquidity 2. Management risks such as breach of duty, embezzlement, etc. 3. Real estate price fluctuations. These risks require adequate trim to spread the base of DABS investment-related market. This

study penetrates the liquidity problem among these three risks and finds the best way.

As a tool to penetrate the blocked liquidity path, this paper proposes a 'Housing Asset Based Fintech(HABF)' model. This fintech model can wipe out a blot for liquidity problem of DABS investment. The two custom keys are: 1. Reflect housing on underlying assets to transaction object 2. Introduction of a collateralized Mortgage Obligation(CMO) financial model. Normal DABS(N-DABS) model, which does not reflect housing assets, deals only with commercial buildings for underlying assets in certain areas (Seoul, Busan in Korea). This model cannot cover matters leaked out such as local alienation, So, This paper brushes up on one's countermeasures. The proposed model's first kick is

adding an underlying asset target to housing across the country, not just commercial buildings in a specific area. The effects of a diversified product line used as supply targets in real estate-related markets are as follows. [As a certain market's 'supply targets' become more diverse, the 'supply' increases. The increase in supply narrows the gap between 'trading bids'. Changing the trading bids interval increase 'investor options' to choose trading products. Investors are more likely to include the product in their portfolio, and the size of the market related to a specific product increases.] In the end, increasing the number of underlying asset item lines promote amplification of the size of real estate replacement investment-related market. The market's liquidity is not drying up. Next, the second kick is revealed. When underlying assets include housing as well as commercial buildings, the following sections describe a model for reconstructing expanded underlying assets into 'multi-layers' through Pooling and Tranche about DABS. This process consists of three steps. The first step is that a person with a house entrusts real estate to an appraisal corporation. The problem is that in the case of N-DABS, both housing ownership and residency are transferred to a third-party. HABF model supplies participants by creating a platform that guarantees residency

for real estate holders and deals only with ownership. The second step is to reorganize real estate assets through the Pooling and Tranche process. The second step can be prepped in two phases. The first phase, the real estate rights held by the third-party company are pooled into three groups: Prime, Alt-A, and Subprime, depending on asset price. The second phase. Each group is then Tranche and reissued in three grades: Senior, Mezzanine, and Equity. In this process, property securities consist of nine layered sections. The third step is to reduce the real estate layered sections from 9 to 6 grades to make it intuitive. AAA and AA ratings are highly profitable derivatives. BBB, BB, and B ratings are derivatives with uncertain profitability. and final Class C is derivatives with little profitability.

This paper has three contributions. The first is derived-wide-option occurred through risk transfer. This HABF model transfers the risk of bad debts of small and medium-sized city properties to specific derivatives through multi-layers. The HABF model provides investors with derivatives with diverse revenue-risk structures. Investors enjoy the choices to build an optimal portfolio, considering their assets, capital, and liabilities. Beyond simply expanding investor's options, the second contribution is to raise liquidity

of the DABS market. Including housing in underlying asset stretches the whole DABS trading market, where investors can resell derivatives they own. It is relatively easier for sellers to find consumers in a market where supply and demand have increased sufficiently. DABS investment costs, which were considered investment assets among non-current assets, speed up when they are converted into current assets. This liquidity leads to a virtuous cycle structure in which investors' market inflow and supply influx. Eventually, the liquidity of DABS market flows out steadily. The third is to prevent the local alienation problem. If the N-DABS model based on certain large cities is

fully settled in real estate replacement investment market, DABS-based real estate transactions in large cities will be further amplified. It literally means that local investment will be shrunk, the local alienation and centralization problems will grow out of control. The core spirit of blockchain technology is decentralization. It is a serious problem that DABS based on blockchain technology contributes to centralization. Real estate is an important asset for us. To protect the value of local real estate, It is perhaps an obligation, not an option, to add a housing asset to DABS market and apply CMO financial model.

Key Words : DABS, Housing, CMO

빅데이터를 활용한 장애인 콜택시 운영 효율성 제고 방안

서창갑 (동명대학교 경영학과 교수, 주저자 armour@tu.ac.kr)
정대현 (부산대학교 경영학과 강사, 교신저자 jdh@pusan.ac.kr)

… Abstract …

교통 취약계층의 이동수단 편의성에 관한 사회적 배려가 필요하다. 따라서 본 연구는 빅데이터 분석을 활용하여 장애인 콜택시 탑승대기에 영향을 미치는 요인을 분석함으로써 교통약자들의 편의증진을 위한 시사점을 제시하고자 하였다. 부산시의 두리발 운영 관련 빅데이터 7십여만건의 분석을 통하여 현안 문제점 도출과 더불어 개선책 제안이 본 연구의 목적인 것이다. 분석툴은 Excel, R, Power BI, QGIS, SPSS를 활용하여 결과값을 시각화 시켰다. 원시데이터로부터 결측치 제거, 전처리 과정, 빅데이터 분석, 평가에 이르기까지 전체 진행과정을 기술하였다. 주요 제언으로는 고객취소 데이터 관리의 필요성, 다크(dark)데이터 최소화를 위한 보완책 마련, 분산 관리중인 데이터의 통합 데이터베이스화의 검토, 추가분석을 위한 시스템 업그레이드 내용, 인공지능의 머신러닝을 적용한 콜 추이 분석 도입의 필요성, 중요도만족도분석(IPA)을 통한 긴급 조치사항 파악, 증차에 따른 대기시간 변화 분석 등을 제시하였다. 따라서 교통 사회적약자의 현안 문제점을 파악하고 사회적가치를 창출하기 위한 두리발 운영의 효율성을 제고시킬 수 있는 담론을 제시하고자 하였다.

Key Words : 빅데이터, 사회적가치, 교통약자, 대기시간, 장애인 콜택시

I. 서론

사회적가치 확산을 위한 다각도의 노력이 최근 ESG 지속가능경영으로 이어지고 있다. 특히 사회적책임 관점에서 장애인 콜택시 운영의 효율성을 실현함으로써 사회적가치 창출의 필요성이 요구되어 진다. 교통 취약계층의 이동수단 편의성 제공은 공공성 강화와 더불어 공동체적 사회적 배려이기도 하다. 우리나라 총인구 51,638천명 가운데 등록된

장애인 인구는 2,644,700명으로 총인구 대비 5.12%에 해당한다(국가통계포털, 2022). 장애인복지법 제23조(편의시설)에 근거하여 사회적약자들의 이동 편의성을 지원하고 있으나 실질적으로 사용자들이 인식하는 복지의 한계성은 커 보인다(이병화 · 양희택, 2017). 본 연구는 빅데이터 구성요소 5V(volume, velocity, variety, value, variable) 중 가치 창출(정대현 · 박광오, 2021) 관점에서 장애인 콜택시 운영 빅데이터 분석으로 사회적가

치 창출에 필요한 선결 과제가 무엇인지를 파악해 보고자 한다.

교통 취약계층의 이동권은 교통시설 이용, 안전한 이동, 차별성 없는 접근 등 인간으로서의 존엄성, 가치 및 행복추구권의 기본 권리를 의미한다(이병화·양희택, 2017). 따라서 부산시는 ‘두리발’이라는 특별교통수단을 적용함으로써 교통 사회적약자의 교통 편의를 제공하고 있다.

본 연구는 두리발의 운영 효율성을 높일 수 있는 요인 모색을 위하여 탑승 대기시간에 대한 세부 분석을 실시하였다. 두리발 운영으로 축적된 빅데이터 7십여만건의 분석으로 현안 문제점을 파악하고 나아가 개선책과 시사점을 도출하고자 한다. 원시데이터로부터 이상치 제거, 전처리 과정, 빅데이터 분석, 시사점 도출에 이르기까지 전체 진행과정을 기술하게 될 것이다.

II. 연구 개요

1. 두리발 현황

두리발은 교통약자 특별교통수단에 근거 2006년 10대로 출범하여 2022년 8월 기준 총 187대의 차량을 24시간 운영중이다. 운행 요금 규정은 일반 중형택시 요금의 35/100 정도로 책정되어 5km까지는 기본요금 1,800 원에 해당하며, 운행구역은 부산광역시 시계 내로 한정하였으나 양산과 김해는 편도 운행을 허용하고 시외 지역 운행 시 20% 할증요금을 부과중이다. 휠체어 이용고객은 3명, 비휠체어 이용고객은 2명 동승을 허용하고 있으며, 두리발 이용에 제한 지침은 3회 이상 콜센터에 사전 연락없이 배차를 취소하는 경우 1개월 이용 제한이 가능토록 규정하고 있다.

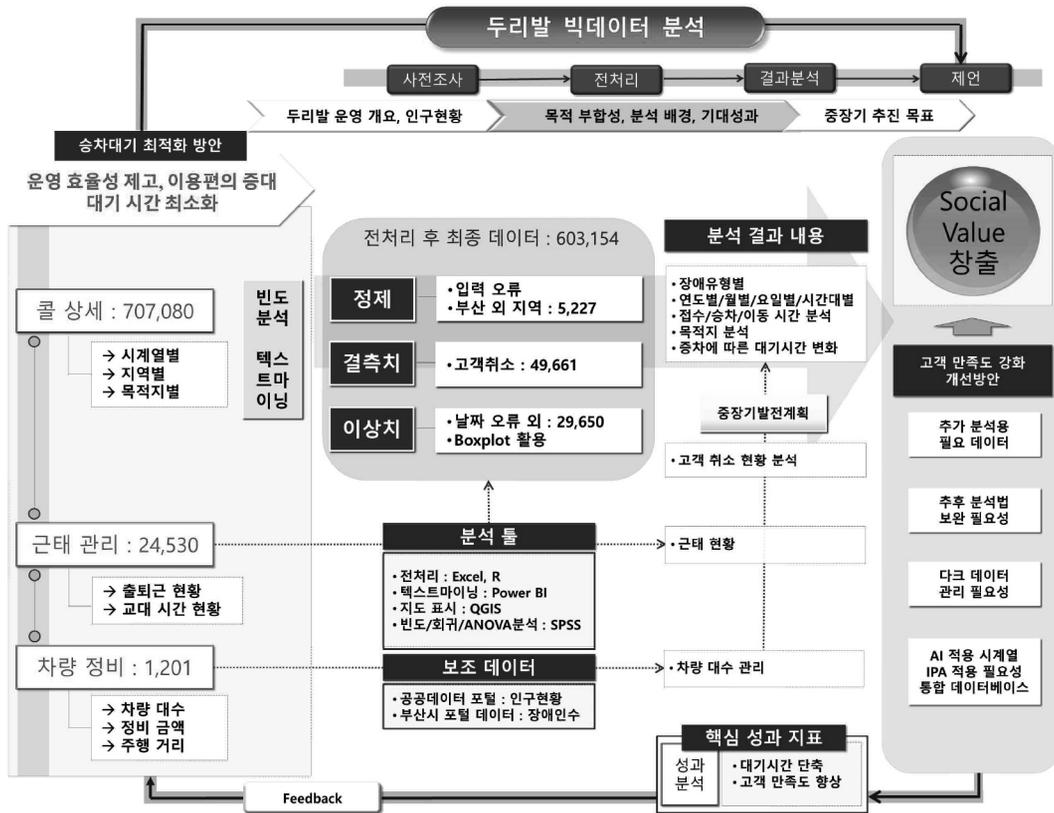
2. 연구 프로세스

전체 연구 프로세스는 <그림 1>과 같이 도식화 하였다. 두리발 콜센터에 축적된 2020년 1월~2022년7월까지의 빅데이터를 분석함으로써 교통약자의 사회적가치를 창출하기 위한 운영 효율성 제고, 이용편의 증대, 대기시간 최소화 등의 목적을 실현하고자 하였다. 본 연구에서는 콜 상세 데이터 70만여건의 시계열별, 지역별, 목적지별 빅데이터 분석을 위하여 빈도분석, 텍스트마이닝 기법을 적용하였다. 무엇보다 분석에 활용할 최종 데이터 선별을 위한 전처리 과정에서 정제, 결측치, 이상치 데이터 식별에 많은 시간이 소요되었다. 분석 툴은 Excel, R, Power BI, QGIS, SPSS를 활용하였으며, 분석 목적에 부합될 수 있도록 결과값을 시각화 시켰다. 그리고 우리나라 전체 인구수와 장애인 현황 파악을 위하여 공공데이터 포털 및 부산시 홈페이지에 공지된 데이터 등을 보조적으로 사용하였다. 분석 결과를 통한 시사점을 파악하고 끝으로 두리발 운영 효율성 및 고객만족도 향상을 위한 제안 내용으로 기술하였다.

3. 인구 현황

KOSIS 2021년 12월 기준으로 <그림 2>와 같이 총인구(51,638,809명)는 감소추세이지만 등록된 장애인 인구(2,644,700명) 비율과 65세이상 노령인구의 증가추세는 이어지고 있다. 향후에도 지속적으로 노령인구의 증가는 이어지기 때문에 교통약자들의 편의를 제공하기 위한 두리발의 정책들이 성과를 달성할 수 있는 빅데이터 분석이 필요하다.

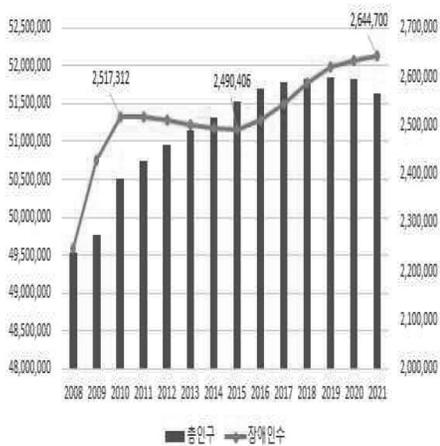
〈그림 1〉 연구 프로세스



Ⅲ. 빅데이터 개요

1. 데이터 전처리

〈그림 2〉 총인구 대비 장애인 증가추세



두리발 콜센터에 축적된 2020년 1월부터 2022년 7월까지의 전체 데이터 707,080건 중에서 603,154건(전체 데이터의 85.3%)의 데이터를 사용하였다. 데이터 전처리 과정에서 주요 결측치로 판단한 내용은 고객취소 데이터 49,661건(7.02%)에 대한 탑승 대기시간 부재로 배제시켰으며 날짜 오류 데이터(하차시간이 탑승시간보다 앞서는 경우 등) 또한 제외시켰다. 그리고 탑승 대기시간 분석 결과에 왜곡 또는 편향된 영향을 미칠 수 있다고 판단된 이상치 29,650건(4.19%)을 분석에서 제외시키고 부산 이외 5,227건을 제외한 최종 분석용 데이터 현황은 〈표 1〉과 같다.

〈표 1〉 최종 분석용 데이터 현황

구분		빈도	비율
소스 데이터		707,080	
결측치	고객취소	49,661	7.02
이상치	날짜오류 외	29,650	4.19
	배차, 탑승 대기시간	19,388	2.74
부산 외	경남, 충북	5,227	0.74
최종 분석용 데이터		603,154	85.30

고객취소 데이터는 별도로 빈도분석을 실시하여 분석 현황 〈표 7〉에서 제시하였으며 이상치는 기사들이 앱을 활용하여 제때 하차시간을 입력하지 못한 다수의 경우로써 실제 하차시간과 상당한 차이를 보이는 경우에 해당하여 분석에서 제외시켰다. 그리고 본 연구에서는 부산지역만 국한시켜 결과값을 제시하였다.

2. 분석 현황

1) 두리발 이용자 분석

두리발 이용자 현황은 장애 32유형별에서 지체1급(21.8%), 뇌병변 1급(16.6%), 65세이상 0급(15.9%)의 사용 비중이 높은 것으로 조사되었다.

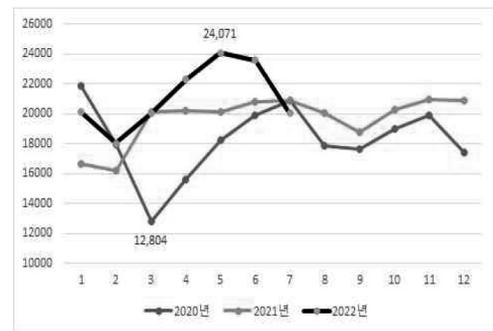
두리발이 담당하고 있는 부산시의 지역별 장애인수는 부산진구(18,474명, 10.52%), 사상구(17,859명, 10.17%), 해운대구(17,611명, 10.03%) 순이며 중구(2,591명, 1.48%)가 최저 분포로 나타났다(부산시 홈페이지, 2022년6월 기준). 이는 지역별 콜 건수와 비교해 보아도 부산진구-사상구-해운대구-북구 순으로 파악되어 장애인수 및 65세 노령인구 거주지와 비례함을 알 수 있다.

〈표 2〉 지역별 콜건수 및 장애인수

지역	콜건수	지역	장애인수
부산진구	73,250	부산진구	18,474
사상구	70,447	사하구	17,859
해운대구	63,376	해운대구	17,611
북구	63,068	북구	16,082
연제구	40,534	사상구	13,379
동래구	38,230	남구	12,914
남구	36,291	동래구	11,850
서구	35,921	금정구	11,619
사하구	33,218	연제구	9,844
금정구	31,658	영도구	8,896
영도구	29,935	기장군	8,493
수영구	22,475	수영구	8,126
기장군	21,771	서구	6,548
동구	21,120	동구	6,103
중구	14,144	강서구	5,241
강서구	7,716	중구	2,591
계	603,154		175,630

연도/월별 빈도분석 결과는 〈그림 3〉과 같이 COVID19의 영향으로 2020년 3월(12,804건)에 급격하게 줄었으나 이후 전반적으로 꾸준히 상승 추세를 보이고 있으며, 요일별로는 수요일이 최다 빈도를 보였다.

〈그림 3〉 연도별 콜건수 추이



시간대별 분석 결과 〈표 3〉과 같이 오전 10시부터 오후 3시 시간대까지 최다 빈도를 보이고 있으며, 오후 8시부터 새벽 5시 시간대까지 낮은 빈도를 보이고 있다. 따라서 콜 집중 시간대에 대한 적절한 배차 및 탑승 대기시간 관리의 중요성을 엿볼 수 있다.

〈표 3〉 시간대별 콜건수 및 총대기시간

시간대	콜건수		대기시간 (분:초)			근무 인원
	건수	비율	배차	탑승	배차+탑승	
0 H	1,232	0.20	5:53	22:36	○ 28:29	2
1 H	787	0.13	3:40	22:26	○ 26:07	
2 H	548	0.09	3:25	22:59	○ 26:25	
3 H	1,280	0.21	4:17	26:58	☾ 31:16	
4 H	2,166	0.36	5:30	25:53	☾ 31:24	
5 H	2,874	0.48	31:23	19:49	● 51:13	
6 H	14,358	2.38	33:19	18:11	● 51:31	44
7 H	35,055	5.81	14:35	20:31	☾ 35:07	
8 H	50,054	8.30	23:21	19:37	☾ 42:59	83
9 H	54,291	9.00	15:05	18:36	☾ 33:43	
10 H	58,064	9.63	10:44	17:27	○ 28:13	
11 H	72,021	11.94	15:48	17:43	☾ 33:32	
12 H	58,180	9.65	23:48	17:49	☾ 41:38	
13 H	46,080	7.64	13:28	18:03	☾ 31:32	
14 H	61,443	10.19	10:01	17:17	○ 27:19	
15 H	56,127	9.31	16:44	19:04	☾ 35:49	
16 H	33,504	5.55	12:43	20:17	☾ 33:01	77
17 H	18,229	3.02	13:08	22:22	☾ 35:31	
18 H	11,280	1.87	14:16	21:43	☾ 36:00	38
19 H	10,332	1.71	10:04	19:42	☾ 29:47	
20 H	8,445	1.40	5:19	18:09	○ 23:29	
21 H	3,284	0.54	24:13	23:38	● 47:52	
22 H	1,966	0.33	17:30	24:35	☾ 42:06	2
23 H	1,554	0.26	9:37	23:52	☾ 33:30	
계/평균	603,154	100%	15:47	18:46	34:34	367

2) 대기시간 분석

배차 대기시간이란 사용자가 콜센터에 접수 후 배차 가능한 기사를 수배하기까지의 소요 시간을 의미한다. 분석 결과 최장 06시대(33분19초) 최단 02시대(3분25초)이다.

탑승 대기시간이란 배차가 완료된 기사가 사용자와 만나기 전까지의 소요시간을 의미한다. 최장 03시대(26분58초) 최단 14시대(17분17초)이다. 따라서 배차+탑승 총대기시간은 오전 6시대(51분31초)가 최장이고 오전 1시대(26분07초)가 최단으로 분석되었다.

분석 결과 근무 교대 시간대인 8H, 12H, 15H, 17H, 21H에서 모두 평균을 상회하고 있다. 9H~12H 시간대의 콜 비율이 40.38%인 반면 현행 근무자 편성내용은 22.62% 배치 중이다. 이를 탄력적으로 콜 건수 비율로 조정해 보면 현행 83명을 146명으로 배치하고 121명을 97명으로 줄이는 검토안을 고려해 볼 필요성이 있다. 물론 현장의 여러 사항

을 고려하여 반영해야겠으나 조정안에 따른 데이터가 축적될 경우 사후 두 모집단의 차이 분석 검토로 대기시간 비교를 통한 성과 분석이 가능할 것이다.

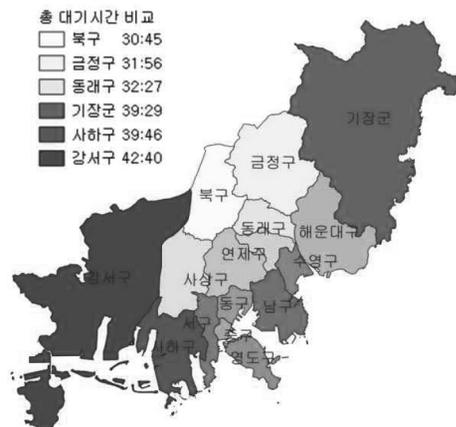
증차에 따른 대기시간의 변화 정도를 분석한 결과는 〈표 4〉와 같다. 2020년1월, 2020년8월, 2021년12월 기준 운행차량 대수는 각각 160대, 181대, 187대이다.

〈표 4〉 차량대수별 대기시간 비교

차량 대수	운영기간		콜건수		운행/대당	대기시간		
	년/월	개월	계	월평균		배차	탑승	계
160	2020/1월~	7	127,287	18,184	114	13:51	18:35	32:26
181	2020/8월~	16	306,863	19,179	106	16:05	18:50	34:55
187	2021/12월~	8	169,004	21,126	113	15:46	18:44	34:30

배차 및 탑승 대기시간과 차량대수에 따른 ANOVA를 이용한 집단비교는 P값 0.000에서 통계적으로 유의한 평균 차이를 보이고 있다. 그러나 대기시간 단축이라는 근본적인 문제점 해결에는 아직까지 큰 영향을 미치지 못하는 것으로 판단 가능하다. 코로나 창궐로 인한 이동제한 등의 영향이 존재하며 콜 건수에 비례하여 배차 및 탑승 대기시간이 연동한다고 볼 수 있기 때문에 추후 추가적인 분석이 필요할 것이다.

〈그림 4〉 총대기시간 지역별 현황



〈그림 4〉의 총대기시간 지역별 분포를 비교해보면 부산외곽 지역인 강서구, 사하구, 기장군 순으로 평균 총대기시간(34분34초)을 상회하고 있어 기사 대기장소와 다소 차이가 발생하는 것으로 예측된다.

3) 목적지 분석

출발지 기준으로 목적지를 분석하기 위하여 사용한 데이터는 총량은 555,431건이다. 대기시간 분석용 데이터 중에서 미분류 데이터 및 언급횟수 100회 이하 단어는 제외시켰으며, 분석을 위하여 원시 데이터를 목적지 및 메모 내용에서 언급된 빈도수가 높은 57개 주제를 선별하여 텍스트마이닝 기법을 활용하였다.

〈표 5〉 목적지 그룹별 현황

대분류	소분류	건수	비율
주거 시설	4 아파트, 주택, 빌라, 맨션	243,070	43.8
의료 기관	10 병원, 의료원, 지압원, 내과, 치과, 외과, 안과, 보건소, 이비인후과, 피부과	197,723	35.6
생활	9 지하철(역), 약국, 빌딩, 미용, 호텔, 해수욕장, 카페, 식당, 애견	28,538	5.1
복지	7 장애, 공원, 스포츠, 회관, 서비스, 박물관, 경로당	23,401	4.2
유통	6 마트, 시장, 백화점, 상가, 농수축협, 유통(기타)	22,802	4.1
종교	3 교회, 성당, 불교	13,679	2.5
관공서	9 시청, 주민센터, 구청, 우체국, 도서관, 공단, 법원, 경찰서, 공사	12,055	2.2
교육 시설	7 어린이집, 초등학교, 대학교, 학원, 중학교, 고등학교, 유치원	7,622	1.4
금융	2 은행, 금고	6,541	1.2

목적지 대분류 건수 비율은 〈표 5〉와 같이 주거시설, 의료기관, 생활, 복지, 유통, 종교, 관공서, 교육시설, 금융 순으로 조사되어 주로 주거시설(43.8%) 및 의료기관(35.6%) 방

문 목적으로 두리발을 사용하고 있음을 알 수 있다.

키워드별 소분류 기분 상위 비율은 〈표 6〉과 같이 아파트(29.8%), 병원(28.5%), 주택(11.7%), 의료원(3.7%) 순으로 파악되었다.

〈표 6〉 소분류 키워드별 횟수

키워드	횟수 (비율)	키워드/횟수
아파트	165,597 29.8	은행 5,597 주민센터 3,069 농수축협 1,173 중학교 834 서비스 309
병원	158,487 28.5	공원 5,067 외과 2,693 유류(기타) 1,168 해수욕장 809 공사 242
주택	64,938 11.7	지압원 4,553 어린이집 2,582 미용 1,143 고등학교 738 박물관 230
의료원	20,362 3.7	내과 4,302 안과 2,168 보건소 1,141 공단 668 불교 170
지하철	19,411 3.5	맨션 4,030 구청 2,134 호텔 1,057 카페 663 경로당 156
장애	12,576 2.3	백화점 3,599 회관 1,933 대학교 1,022 병원 562 유치원 138
교회	11,966 2.2	약국 3,395 성당 1,543 금고 944 이비인후과 499 애견 114
마트	9,800 1.8	시청 3,164 빌딩 1,526 우체국 941 식당 420
빌라	8,505 1.5	치과 3,144 초등학교 1,448 도서관 896 경찰서 379
상가	5,658 1.0	스포츠 3,130 상가 1,404 학원 860 피부과 374

4) 고객취소 현황 분석

사용자들 중 여러 사유로 인한 지역별 취소현황은 〈표 7〉과 같다. 부산진구, 해운대구, 사상구, 북구 순이며 이는 사용자 지역별 콜 건수 순(부산진구, 사상구, 해운대구, 북구) 및 장애인 거주 지역과도 유사한 결과이다.

〈표 7〉 고객 취소 현황_지역별

구분	횟수 (비율)	구분	횟수 (비율)
부산진구	5,899 11.9	금정구	2,544 5.1
해운대구	5,589 11.3	동구	2,311 4.7
사상구	4,628 9.3	영도구	2,242 4.5
북구	4,164 8.4	수영구	2,188 4.4
사하구	3,454 7.0	기장군	1,662 3.3
남구	3,359 6.8	중구	1,269 2.6
동래구	3,285 6.6	강서구	682 1.4
연제구	3,017 6.1	양산시	444 0.9
서구	2,830 5.7	기타	94 0.2
			49,661

세부 분석 내용으로는 6,675명이 취소 경험이 있고 지체1급, 65세이상0급, 뇌병변1급, 시각1급 순으로 조사되었으며, 400회 이상 취소한 경우도 존재하여 정확한 원인 파악이 필요한 실정이다.

또한 고객취소 사유는 〈표 8〉과 같이 다른 교

통수단, 탑승지연, 통화불능, 고객변심 등의 순으로 파악되었다.

〈표 8〉 고객 취소_사유별

구분	2020	2021	2022	계	비율
다른					
교통수단	9,914	12,145	7,684	29,743	59.9
탑승지연	1,557	1,908	1,282	4,747	9.6
통화불능	283	334	204	821	1.7
고객변심	122	300	341	763	1.5
배차오류	430	221	107	758	1.5
차량부족	32	248	18	298	0.6
확인불가	144	280	728	1,152	2.3
기타	2,968	5,151	3,260	11,379	22.9

IV. 결론 및 제언

상기 연구 결과를 바탕으로 한 결론 및 제언은 다음과 같이 요약할 수 있다. 두리발 운영 과정에서 축적된 7십여만건의 빅데이터 분석을 통하여 두리발 운영 효율성 제고, 대기시간 최소화, 이용편의 증대를 목적으로 데이터 클리닝 처리 후 빈도분석, 회귀분석, 분산분석, 텍스트마이닝 분석을 실시하였다.

사회적가치 창출은 사회적 문제점 해결을 통한 공공성 강화에 주안점을 두어야 한다. 교통약자들의 경제적 문화적 배려에 있어서 지속가능한 공동체 이익에 기여함으로써 바람직한 가치를 실현할 수 있어야 한다. 따라서 두리발에 축적된 빅데이터 분석으로 탑승 대기시간의 문제점을 분석하고 사회공동체가 지향하는 공익 가치 실현에 필요한 담론을 제기하고자 한다.

첫째, 고객취소 데이터 분석 결과 철저한 데이터 관리의 필요성이다. 과도한 개인별 취소건수에 대한 조치 사항 강구가 검토되어야 한다. 일정 횟수 이상에 대해서는 한시적 사용제한 등의 조치가 있을 수 있다는 점을 고지함으로써 운영의 효율성을 높일 필요성이 있

다. 기존에 규정은 있으나 취소 상한선에 대한 내부 통제 절차를 강화시켜야 할 것이다. 또한 고객 취소 사유별 분류를 단순 변심, 지연 불만, 대체 수단 이용 등의 유효성 검사의 데이터베이스 구축으로 고객만족도 향상을 위한 분석용으로 활용할 필요성이 있다는 판단이다.

둘째, 다크(dark) 데이터 최소화를 위한 보완책 마련에 주력해야 한다. 분석용 데이터로 활용할 수 없는 다크 데이터의 효율적 관리 방안을 마련해야 할 것이다. 이는 날짜 및 시간 입력에 대한 오류로 판단되어지는 데이터의 수량을 최소화할 수 있도록 교육 및 시스템 보완책이 필요하다고 사료된다. 출퇴근 시간의 앱(app) 기록에 대한 잦은 지연 입력으로 불용데이터로 분류된 경우가 많았다.

셋째, 분산 관리되고 있는 데이터의 통합 데이터베이스화의 검토를 제안한다. 본 연구에서 다룬 데이터인 콜상세, 근태관리, 차량정비 데이터는 분산 관리중이다. 특히 근태관리 및 차량정비 현황은 단순한 엑셀 데이터로 관리하는 실정이라 데이터의 참조무결성에 다수의 오류가 파악되었다. 따라서 현재 콜상세시스템에 기사 근태관리 및 차량정비 데이터를 통합하여 하나의 시스템으로 구축되면 빅데이터 분석 활용도가 높아질 것이다.

넷째, 시스템 업그레이드 필요성이다. 본 연구와 같이 57개 소분류 키워드를 데이터베이스 유효성 검사용으로 사용할 경우 향후 목적지 소분류 항목별 집계를 통한 용이한 분석이 가능할 것이다.

다섯째, 추후 분석법 보완의 필요성이다. 본 연구는 2020년1월~2022년7월까지의 데이터였다. 따라서 2022년12월까지의 데이터 축적 시 인공지능의 머신러닝을 적용한 콜 추이 분석이 가능할 것이다. 단순 추세선 분석을

넘어 콜건수 예측에 따른 적절한 사전 대응 전략 수립에 유익한 정보제공이 뒷받침 될 수 있을 것이다. 분석 방법론으로 Azure ML 활용, ARIMA 모델 및 ETS 지수평활법 등을 적용함으로써 의사결정의 신속성을 높일 수 있을 것이다.

또한 IPA(중요도만족도분석) 분석을 통하여 긴급하게 조치가 필요한 고객의 니즈를 파악함으로써 고객만족도 향상을 꾀할 수 있을 것이다. IPA는 고객이 중요하다고 판단하는 서

비스별 만족도 정도를 파악함으로써 전략적 우선순위를 결정하고 적절한 대응책 마련에 적용 가능할 것이다.

끝으로 추후 연구의 확장을 위하여 전국단위의 장애인 콜택시 비교 분석의 필요성을 제안한다. 진정한 사회적가치 실현을 위해 중앙부처에서 주관하는 빅데이터 분석은 좀 더 의미 있는 교통약자 문제점 해결에 도달할 수 있을 것이다.

(참고문헌 및 기타 세부 분석내용 생략)

온실가스 배출권 및 신재생에너지 공급인증서 거래 예측모형 개발 및 패턴 분석 : KAU 가격과 REC 가격을 중심으로

김덕현 (경상국립대학교 경영정보학과 박사수료, 주저자 kdh294@gnu.ac.kr)
정대을 (경상국립대학교 경영정보학과 교수, 교신저자 dyjeong@gnu.ac.kr)

… Abstract …

과거부터 국제사회는 지구온난화로 인한 심각한 문제가 초래할 것을 논의해왔으나, 근래에 기록적인 자연재해와 이상기후현상이 빈번하게 발생하고 있다. 발생하는 자연재해와 이상기후현상의 빈도는 물론이고, 그 강도가 더욱 강해지고 있으며, 연일 마스크를 통해 세계 곳곳의 이상기후 현상이 보도되고 있는 실정이다. 본 연구진은 탄소중립 추진 방향성에 대하여 ‘온실가스 배출량 감축’ 과 ‘신재생에너지 확대’ 로 인식하고 있다. 그러므로 본 연구에서는 온실가스 배출량을 감축하기 위해 탄생한 거래 시장과 신재생에너지 발전을 권장하고 발전한 에너지를 거래할 수 있도록 탄생한 거래 시장에 대한 다각도 패턴 분석을 진행하였다. 구체적으로 온실가스 배출권 거래와 신재생에너지 공급인증서 거래 관련 시장의 흐름을 파악하기 위해 KAU(할당배출권)과 REC(신재생에너지 공급인증서)의 가격에 대한 기초통계량 분석과 데이터 마이닝을 진행하였다. 수집한 연구자료의 범위는 2015년부터 2022년으로 다양한 웹 데이터베이스로부터 KAU 데이터셋과 REC 데이터셋을 구축하고 예측모형을 개발하였다. 이를 통해 지난 7여 년간 양 거래 시장에 대한 변화 패턴을 KAU와 REC 가격의 변화와 더불어 분석에 투입한 거시경제변수, 거래량, 발전량, 생산량 등으로 파악하고자 하였다.

Key Words : 데이터 마이닝, 탄소중립, 신재생에너지 공급인증서, 온실가스 할당배출권

I. 서 론

전 세계적으로 지구온난화로 인한 자연재해와 이상기후 현상이 빈번하게 발생하고 있다. 2021년 환경부에서 발간한 보고서인 『배출권거래제, 탄소중립으로 가는 길』에서는 현재의 기후변화 상황에 대하여 ‘전 지구적인 비상사태’ 로 정리하고 있다(환경부, 2021).

지구온난화로 인한 폭염 뿐만 아니라 폭우·태풍·가뭄·해수면 상승 등은 대표적인 이상기후 현상이며, 최근 그 빈도와 강도가 강해지고 있다.

언급한 이상기후 현상의 중심에는 지구의 비정상적이고 꾸준한 기온 상승이 거론되고 있으며, 기온 상승의 주된 요인은 온실가스라고 일컬어지는 화석연료에 의한 탄소배출로 판단

하고 있다(탄소중립위원회, 2021). 현재 전 세계는 탄소배출량을 줄이기 위한 노력을 기울이고 있다. 이는 오늘날 자주 언급되고 있는 ‘탄소중립’의 개념으로 널리 통용되고 있다.

본 연구진이 인식하고 있는 탄소중립 추진에 대한 노력의 두 가지 중심축은 다음과 같다. 첫 번째, 탄소배출량에 대한 비용을 산정함으로써 직접적인 탄소배출량을 줄이고자 하는 노력이 시도되고 있다. 두 번째, RE100, 그린뉴딜 등의 개념과 관련된 신·재생에너지이며, 해당 발전원의 확대에 의한 탄소배출량의 감소를 기대하고 있다. 과거 신재생에너지에 대한 논의는 고갈되고 있는 화석연료에 대한 대체 에너지로 초점이 맞추어졌으나, 현재 신재생에너지의 확대는 탄소배출량을 줄이기 위한 획기적인 방안으로 인지되고 있다(환경부, 2020; 한국에너지공단, 2021).

본 연구에서는 탄소중립이라는 목표 달성을 위한 ‘탄소 배출에 대한 감축’과 ‘신재생에너지 확대’라는 두 가지 중심축에 대하여, 각각의 거래시장에서 매매가 일어나고 있는 권리에 대한 가격 추이와 패턴 탐색을 수행하였다.

분석에 활용한 기법으로 기초통계량 분석과 데이터 마이닝에 의한 분석 결과를 다루고자 한다. 개발된 예측모형의 성능 평가와 분석 결과 기반 시사점 도출을 중점적으로 수행하였다.

II. 선행 연구

1. 온실가스 배출권 거래제

탄소가격제에는 대표적으로 탄소세와 배출권 거래제가 있다. 탄소세는 석탄, 석유, 천연가

스 등의 이산화탄소를 배출하는 화석연료를 사용한 배출자에게 배출량만큼 부과하는 세금을 의미한다(온실가스종합정보센터, 2021).

탄소세를 화석연료 공급업체에게 직접 부과하면 이것이 차례대로 화석연료를 통해 생산한 제품과 전력, 더 나아가 일반 소비자들에게 제공되는 제품 및 서비스의 가격이 높아지는 방식으로 세금 부담이 전가된다.

화석연료를 사용한 배출업체에게 배출량만큼 세금을 부가하게 되면, 무분별한 에너지 사용 절감과 재생에너지 등 대체에너지에 대한 투자와 개발을 촉진할 수 있다. 특히 발전사(전력 생산자)들은 전력 생산에 저탄소 연료를 사용하게 되는 효과가 있다.

온실가스 배출권 거래제는 정부가 온실가스를 배출하는 사업장을 대상으로 연 단위 배출권을 할당하여 할당 범위 내에서 배출 행위를 할 수 있도록 허용하고, 할당된 배출권 중 여분 또는 부족분에 대해서는 시장의 거래를 통해 확보할 수 있도록 하는 제도를 의미한다. 2015년부터 본격적으로 거래제도가 시행되어 지속되고 있다(온실가스종합정보센터, 2018).

현 단계에서는 의무적 할당량을 충족 목적으로 주요 할당 기업들은 배출권을 거래하고 있으나, 향후 기업들이 잘 활용하면 탄소 배출량을 줄일 수 있을 뿐만 아니라 추가적인 이익까지 창출할 수 있는 제도로 전망받고 있다. 특히 청정개발체제(CDM), P-CDM 등을 활용하여 온실가스 감축 활동을 수행하여, 환경부로부터 인증 받은 상쇄배출권을 통해 해당 사업장의 감축량으로 인정받는 제도는 온실가스 의무감축량을 할당받은 기업들의 감축 목표 달성의 유연성을 확보해주고자 하는 제도로 정착되어 있다.

국내에서는 배출 총량을 정부가 계획하고 설

정하기 때문에 줄이고자 하는 목표량을 확실히 알 수 있다는 장점이 있으나, 배출권 자체의 가격 변동성은 존재하고 있는 실정이다.

국내 할당배출권(Korea Allowance Unit, 이하 KAU)의 가격은 2019년 12월까지 크게 상승하였다가, 코로나19의 영향으로 큰 폭으로 하락하였다. 2021년 이후의 추이는 증가와 하락을 반복하고 있으며, 최근 에너지 물가의 상승과 더불어 탄소중립 추진에 대한 고도화와 맞물려 다시 가격 상승세를 보이고 있다.

2. 신재생에너지 공급인증서 거래제

신·재생에너지 보급 확대와 관련 사업 육성에 관련하여 과거로부터 꾸준히 추진되어 오고 있는 RPS 제도를 들 수 있다.

일정 규모 이상의 발전사업자를 공급의무자로 정의하고, 총 발전량의 일정 비율 이상을 신·재생에너지로 공급하도록 의무화하는 제도를 의미한다.

RPS 제도로 인하여, 공급의무자와 발전사업자의 신재생에너지 공급인증서(Renewable Energy Certificate, 이하 REC)가 거래될 수 있는 환경이 조성되었다.

신재생에너지 공급인증서 거래제는 신재생에너지 공급의무화 제도 시행 이후 2012년부터 본격적으로 시행되었다(이철용, 2015).

REC는 발전사업자가 신재생에너지 설비를 이용하여 전기를 생산·공급했음을 증명하는 인증서의 단위로, 공급의무자는 공급인증서를 구매하여 의무공급량에 대한 의무이행 실적에 활용할 수 있다.

현재 한국형 RE100 (K-RE100) 도입 및 참여기업들의 확산을 위해 플랫폼 형태의 REC 거래시스템이 시범운영 중에 있다. REC 구매 실적을 통해 온실가스 감축에 대한 기업 차원

의 기여를 인정하겠다는 규정 하에서 녹색 프리미엄제도와 같은 기존 정책적 지원과 연계를 진행 중에 있다(한국에너지공단, 2021).

REC의 가격은 KAU보다 변동성이 더 큰 편인데, 2016년 3월 태양광-비태양광 통합시장이 형성되면서 가격이 크게 상승하였으나, 공급 과잉의 문제로 인하여 2018년 이후 10만원 선이 무너지며 큰 폭으로 하락하였다. 최근 에너지 물가의 상승과 국제사회에서의 굴지 기업들의 신재생에너지 100% 사용이라는 RE100 등으로 인한 가격 상승세를 보이고 있다.

Ⅲ. 연구 모형과 분석기법

1. 연구 진행과정

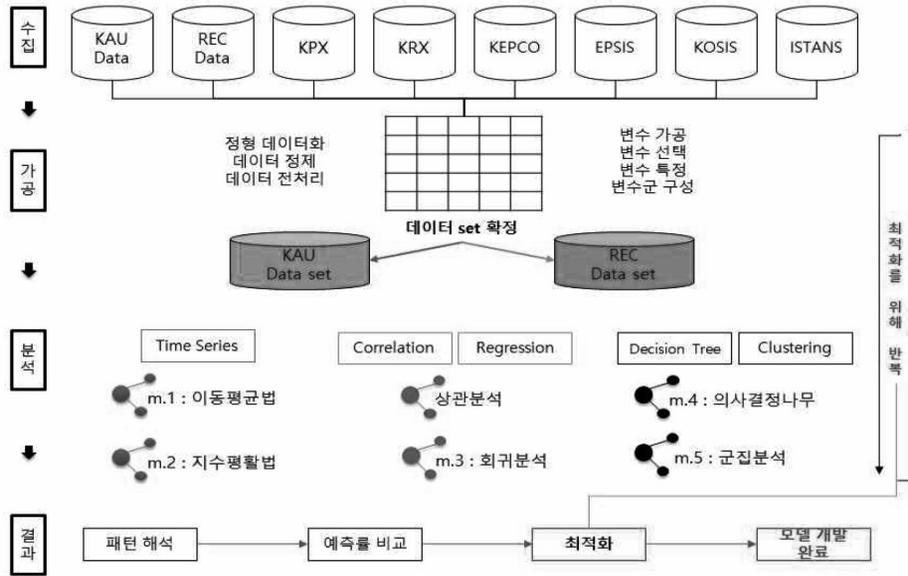
본 연구는 KAU와 REC에 대한 가격 예측모형을 <그림 1>과 같은 진행과정에 의해 개발하였다.

연구의 전체적 흐름에서 데이터 분석의 기본 프로세스인 ‘수집→가공(전처리)→분석→결과 도출’의 과정을 거쳐 예측모형을 개발하였다. 데이터 수집 과정에서 목표변수인 KAU와 REC에 대한 가격 데이터를 수집하였으며, 가격과 함께 수집될 수 있는 체결량, 거래량, 거래대금 등의 데이터를 수집하였다.

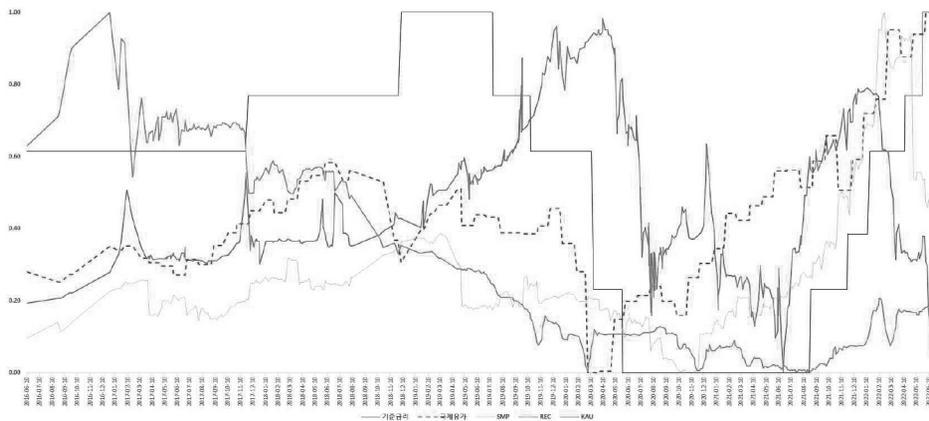
분석에 적합한 데이터로 전처리하는 가공 단계에서 비정형데이터의 정형데이터화, 데이터 정제, 변수 가공 및 파생 등을 수행하였다. 가공 단계를 거친 데이터들은 한 가지의 목표변수를 예측하기 위한 데이터셋으로 종합하였다.

온실가스 배출권 중 할당배출권의 가격을 예측하기 위한 데이터 집합은 ‘KAU 데이터셋’이며, 신재생에너지 공급인증서의 가격

〈그림 1〉 연구 프레임워크



〈그림 2〉 주요 변수의 정규화 그래프('16.06 ~ '22.07)



을 예측하기 위한 데이터 집합은 'REC 데이터셋'이다.

분석 과정에서 시계열 분석, 통계적 분석, 데이터 마이닝을 진행하였고 예측모형의 성능을 향상하기 위해 변수 추가·제거, 파라미터 수정 과정을 거쳤다.

앞서 언급한 바와 같이, 본 고에서는 통계적 분석 중 기초통계량 분석과 데이터 마이닝의 의사결정나무에 대한 결과를 해석하고, 패턴 도출만을 다루었다.

2. 주요 변수와 분석 기법

1) 목표변수와 주요 독립변수

(1) 목표변수

본 연구에서의 목표변수는 KAU(할당배출권)과 REC(신재생에너지 공급인증서)로 구분된다. 두 목표변수에 대한 연구자료의 범위는 2015년 08월부터 2022년 07월까지 총 2,506일이며, 거래대금·체결건수 등의 부가

자료를 통해 해당일에 거래가 일어난 경우만을 고려하였다.

거래 여부를 중심으로 한 KAU에 관한 데이터셋의 총 인스턴스의 수는 1,386개이며, REC에 관한 데이터셋의 총 인스턴스의 수는 588개이다.

〈그림 2〉는 거래대금이 존재했던 KAU 거래일에 대한 정규화 그래프로, 연구 자료의 기간 중 KAU와 REC 가격과 기준금리, 국제유가, SMP 가격의 증감을 파악할 수 있다.

정규화(Normalization)는 0과 1 사이에 값으로 기존 데이터를 치환하여 변수 간 간접적인 비교와 시간의 흐름에 대한 변동 추이를 파악하기 용이하다.

REC 가격은 17년도까지 매우 높은 가격대를 기록하다가 최근까지 크게 하락한 가격대가 유지되고 있으며, KAU 가격은 코로나19 팬데믹 전후로 변동성이 큰 양상을 보이고 있다.

가격 변동 패턴을 예측하기 위하여, 목표변수의 범주화도 함께 진행하였다. 특정일을 제외하면 매일 거래가 발생하는 주식시장과 달리, 두 시장 모두 전일 대비 개념이 아닌 전 거래 대비 개념을 도입하였다.

KAU 가격 변동의 경우, 전 거래일에 비해 가격이 초과한 경우(등)가 579개, 가격이 같거나 하락한 이하의 경우(락)가 807개로 범주형 목표변수가 하위집단을 가진다.

REC 가격 변동의 경우, 전 거래 대비 초과(등)가 330개, 이하(락)가 258개로 범주형 목표변수가 하위집단을 가진다.

(2) 주요변수

KAU와 REC 가격을 예측하기 위한 모형 개발에 투입된 주요변수는 다음과 같다. 시계열 데이터이기 때문에, 해당일자를 기준으로 변

수를 파생한 해당연도·주·분기·월·요일을 독립변수로 활용하였다.

거시경제변수로 환율, 기준금리, 시장금리(국고채 3년), 국제유가, 전체 생산자물가지수, 전력·수도·가스 생산자물가지수를 활용하였다.

전력 거래량·생산량(태양광·풍력·수력·화력·전체), SMP 가격 등 전력에 관한 데이터 또한 변수화하여 활용하였다. 특히, RPS 제도의 변동 여부를 반영하기 위하여 RPS의 무비율도 독립변수로 활용하였다.

〈그림 2〉에서 거시경제변수인 기준금리, 에너지 물가를 상징하는 국제유가, 전력 거래 가격의 기초가 되는 계통한계가격(SMP)에 대한 증감을 정규화를 통해 일목요연하게 도식화하였다.

2) 데이터 마이닝 기법

본 연구에서는 데이터 마이닝의 분류분석 기법 중 하나인 의사결정나무를 활용하여 예측모형을 개발하였다. 의사결정나무는 분석결과가 나무 구조의 규칙들로 표현되기 때문에, 연구자가 분석결과를 쉽게 이해하고 설명할 수 있는 장점이 있다(최종후, 서두성, 1999). 즉, 목표변수에 영향을 주는 주요 요인들을 중심으로 규칙들이 블랙박스 아닌 화이트박스의 형태로 도출되기 때문에, 규칙들을 토대로 핵심 요인을 탐색할 수 있으며, 요인들의 조합에 따라 규칙의 결과물이 달라질 수 있기 때문에 준거점을 파악하기 용이하다.

의사결정나무 중에서도 수치·범주형 데이터를 예측하기 위한 랜덤 포레스트 알고리즘과 범주형 데이터를 예측하기 위한 C4.5 알고리즘을 활용하였다(Quinlan, 1993; Breiman, 2001).

〈표 1〉 목표변수의 기초통계량 분석

구분	단위	KAU		REC	
		전체 기간	최근 1년	전체 기간	최근 1년
산술평균	원	25,140.01	26,564.90	73,966.48	42,742.16
표준편차	원	6,571.45	6,092.76	37,158.59	9,721.23
최대값	원	42,000	35,400	174,044	58,600
최소값	원	22,550	28,000	29,200	29,200
인스턴스 (n)	거래	1,386	255	588	102

IV. 분석 결과

1. 기초통계량 분석과 가격 추이

〈표 1〉를 통해 목표변수에 대한 전체기간과 최근 1년에 대한 기초통계량 분석 결과를 요약하였다.

목표변수는 KAU와 REC 가격이며, 전체기간은 연구 자료의 전체범위인 2015년 08월부터 2022년 07월까지로, KAU 거래는 1,386회, REC 거래는 588회 이루어졌다. 최근 1년은 2021년 07월부터 2022년 07월까지로 KAU 거래는 255회, REC 거래는 102회 이루어졌다.

먼저, KAU에 관한 기초통계량에 대한 해석은 다음과 같다. 전체기간 대비 최근 1년간 평균 가격은 상승하였고 편차는 줄어들었다. 또한 최대값과 최소값을 기반으로 전체기간 중 최근 1년이 아닌 기간에 최고가와 최저가가 포함되는 것으로 해석된다.

KAU의 전체기간과 최근 1년에 대한 기초통계량이 주는 함의는 형성된 가격 자체가 일부 상승되었다는 점과 변동성이 최근 1년 내에 줄어들었다는 점으로 요약할 수 있다.

REC에 관한 기초통계량에 대한 해석은 다음과 같다. 전체기간 대비 최근 1년간 평균가격은 하락하였고 편차는 크게 줄어들었다. 최고가는 17만원으로 전체기간 내에서 기록되었고, 최저가는 약 2만9천원으로 최근 1년 사이

에 기록된 것으로 판단된다.

REC의 전체기간과 최근 1년에 대한 기초통계량이 주는 함의는 형성된 가격 자체가 크게 하락되었다는 점과 변동성도 크게 줄어들었다는 점이다.

또한 자료 기간 내에 최고가가 17만원, 전체 평균이 약 7만4천원인데 반해 현재는 가격이 크게 하락한 상황임을 알 수 있다.

끝으로 지난 7년간 온실가스 배출권보다 신재생에너지 공급인증서의 가격 변동폭이 매우 컸다는 점을 파악할 수 있고, 과거에 비해 최근 1년 간 REC 가격에 대한 편차가 크게 줄어든 점은 변동성이 비교적 약화된 것으로 해석된다.

하위 절에서는 데이터 마이닝을 통해 분석된 결과를 제시하고, 패턴과 시사점을 도출하고자 한다. 분석도구로 Weka 3.8.5 버전을 활용하였다.

2. 의사결정나무 기반 예측모형

1) 목표변수와 주요 독립변수

본 연구에서는 KAU 데이터셋과 REC 데이터셋에 대하여 랜덤 포레스트 기법알고리즘과 C4.5 알고리즘을 통해 의사결정나무 분석을 시행하였다. 분석 결과에 대한 성능 평가를 위해 〈표 2〉와 〈표 3〉으로 요약하였다.

학습 방법과 규칙 노드 안에 포함되어야 하는

최소 인스턴스 수 등을 통해 최적의 모형을 개발하고자 하였으며, 선정된 예측모형은 진하게 표시하였다.

KAU와 REC에 대한 의사결정나무 분석 결과는 크게 3가지로 정리할 수 있다. 첫째, 랜덤 포레스트 알고리즘이 C4.5보다 성능이 대체적으로 우수한 것으로 보인다. 둘째, 7:3 분할 학습보다 10 fold 교차 검증(Cross validation)을 도입하였을 때 예측률 성능이 일부 향상되었다. 셋째, 도식화가 기능이 제공되는 C4.5의 경우 파라미터 세팅을 통해 랜덤 포레스트 수준의 성능까지 예측률이 향상

되었다.

양상불 기법이 가미된 랜덤 포레스트가 성능이 대체적으로 좋은 편이나, 이에 준하는 예측률을 보이는 C4.5 기반의 예측모형을 최종 모형으로 선택하였다. 그 이유는 도표를 통한 시각화가 가능하며, 시각화된 도표는 의사결정나무의 장점인 준거점 파악에 용이하기 때문이다.

그러므로 KAU와 REC 가격 예측모형은 ‘C4.5+10 fold 교차 검증+최소 인스턴스 수 설정’의 조합에 의해 도출된 모형들(각각 5번)을 최종적으로 선택하였다.

〈표 2〉 KAU 의사결정나무 분석 결과

구분	알고리즘	파라미터 세팅	예측률	MAE	기타 내용
1	Random F.	70:30 분할 학습	61.29 %	0.446	없음
2	Random F.	10 fold 교차 검증	63.49 %	0.439	KAU 모형 중 예측률이 가장 높음
3	C4.5	70:30 분할 학습	56.25 %	0.488	규칙의 개수가 1개로, 시각화 불가능 가장 낮은 성능을 보이고 있음
4	C4.5	10 fold 교차 검증	58.22 %	0.486	규칙의 개수가 1로, 시각화 불가능
5	C4.5	10 fold 교차 검증	62.69 %	0.445	의사결정나무 도표를 통한 시각화 가능 최소 인스턴스 수 : 15개 (전체 n의 0.01)

〈표 3〉 REC 의사결정나무 분석 결과

구분	알고리즘	파라미터 세팅	예측률	MAE	기타 내용
1	Random F.	70:30 분할 학습	63.63 %	0.443	없음
2	Random F.	10 fold 교차 검증	64.79 %	0.439	REC 모형 중 예측률이 가장 높음
3	C4.5	70:30 분할 학습	61.36 %	0.422	규칙의 개수가 94개로, 전지가 필요함
4	C4.5	10 fold 교차 검증	59.35 %	0.427	규칙의 개수가 94개로, 전지가 필요함 가장 낮은 예측률을 기록함
5	C4.5	10 fold 교차 검증	61.90 %	0.435	의사결정나무 도표를 통한 시각화 가능 최소 인스턴스 수 : 10개 (전체 n의 0.02)

2) 의사결정나무 주요 규칙

〈부록 1〉과 〈부록 2〉는 KAU와 REC의 의사결정나무 도표이며, 전체 예측률은 각각 62.69%와 61.90%이다. 분지 기준이 되는 독립변수들의 조합에 의해 세부 규칙의 적중률(hit ratio)이 각기 달라지는데, 전체 예측률 대비 세부 규칙의 적중률이 높은 규칙들을 주

요 규칙으로 선별하였다.

주요 규칙을 선별한 이유는 전체 예측률 자체가 70% 미만을 보이고 있다는 점과 더불어, 전체 예측률보다 낮은 적중률을 보이는 규칙은 패턴으로 고려하기 어려운 부분이 존재한다고 판단하였다.

전체 예측률보다 70% 이상 적중률을 기록한 세부 주요 규칙들은 각각 〈표 4〉와 〈표 5〉를

통해 정리하였다.

의사결정나무 주요 규칙에 대한 표를 해석하는 방법은 다음과 같다. 규칙(리프노드)이란 예측모형을 개발할 때, 투입된 독립변수의 조합에 의해 생성되는 것을 의미한다.

내용 상에 언급된 독립변수의 조합에 의해 ‘전 거래 대비 가격이 상승했는지, 하락했는

지’에 대한 규칙이 정해진다.

해당 인스턴스란 독립변수의 조합 내에 포함되는 인스턴스의 수를 의미하고, 오분류는 독립변수의 조합 내에 포함되지만 규칙이 틀림으로 인해 오분류가 된 인스턴스의 수를 의미한다.

〈표 4〉 KAU 의사결정나무 주요 규칙

번호	내용 (독립변수의 조합)	가격 등락여부	해당 인스턴스	오분류	적중률
1	해당연도 <= 2018	이하(락)	438 개	123 개	71.92%
2	해당연도 > 2018 & 해당 주 <= 28 & 월별 풍력 거래량 <= 132 Gwh & 환율 > 1,132.2 & KAU 당일거래량 <= 17,106	초과(등)	17 개	4 개	76.47%
3	해당연도 > 2018 & 해당 주 > 28 & 해당연도 <= 2019	초과(등)	161 개	44 개	72.67%
4	해당연도 > 2018 & 해당 주 > 28 & 해당연도 > 2019 & KAU19 & REC 가격 <= 45,500	이하(락)	24 개	5 개	79.17%
5	해당연도 > 2018 & 해당 주 > 28 & 해당연도 > 2019 & KAU20 & 금요일	초과(등)	24 개	6 개	75.00%
6	해당연도 > 2018 & 해당 주 > 28 & 해당연도 > 2019 & KAU21 & KAU 당일거래량 <= 7,450	이하(락)	31 개	8 개	74.19%
7	해당연도 > 2018 & 해당 주 > 28 & 해당연도 > 2019 & KAU21 & KAU 당일거래량 > 7,450	초과(등)	71 개	21 개	70.42%

〈표 5〉 REC 의사결정나무 주요 규칙

번호	내용 (독립변수의 조합)	가격 등락여부	해당 인스턴스	오분류	적중률
1	육지체결량 <= 960 & 해당연도 > 2018	이하(락)	231 개	61 개	73.59%
2	육지체결량 <= 960 & 해당연도 <= 2018 & 육지거래량 <= 6,245	이하(락)	14 개	0 개	100.0%
3	육지체결량 <= 960 & 해당연도 <= 2018 & 육지거래량 > 6,245 & 해당연도 <= 2017 & 기준금리 > 1.3	초과(등)	22 개	5 개	77.27%
4	육지체결량 <= 960 & 해당연도 <= 2018 & 육지거래량 > 6,245 & 해당연도 <= 2017 & 기준금리 <= 1.3 & 환율 <= 1,105.7	이하(락)	14 개	3 개	78.57%
5	육지체결량 <= 960 & 해당연도 <= 2018 & 육지거래량 > 6,245 & 해당연도 > 2017 & 화력 거래량 <= 28,124 Gwh	이하(락)	15 개	0 개	100.0%
6	육지체결량 <= 960 & 해당연도 <= 2018 & 육지거래량 > 6,245 & 해당연도 > 2017 & 화력 거래량 > 28,124 Gwh & 육지거래량 <= 265	이하(락)	10 개	0 개	100.0%
7	육지체결량 <= 960 & 해당연도 <= 2018 & 육지거래량 > 6,245 & 해당연도 > 2017 & 화력 거래량 > 28,124 Gwh & 육지거래량 > 265 & 수력 거래량 <= 176 Gwh	이하(락)	12 개	3 개	75.00%

KAU와 REC 의사결정나무의 주요 규칙들은 다음과 같다.

KAU 관련 주요 규칙은 전 거래 대비 가격 상승 규칙이 4개, 하락 규칙이 3개로 고르게 나타났다. REC 관련 주요 규칙은 전 거래 대비 가격 하락 규칙이 6개, 상승 규칙이 1개로 나타났다.

KAU 관련 주요 규칙 도출에 관한 분지 기준들은 연도, 주, 풍력 거래량, KAU 종목, REC 가격, 요일 등이 있다. REC에 관한 분지 기준들은 육지체결량, 해당연도, 육지거래량, 기준금리, 화력 거래량, 수력 거래량 등이 있다. 현재까지 개발된 모형의 패턴들은 주로 해당연도, 요일, 주, 기준금리 등 시계열 관련 변

수들이 규칙의 준거로 등장하였다. 예를 들어, <표 5>의 4번 규칙은 연도가 분지의 기준으로 두 번 나타나는데, 2017년 이전에 관한 패턴임을 알 수 있다. 또한 3번 규칙과 달리 기준금리가 1.3 이하이기 때문에 2016년 06월 이후에 일어난 거래들에 관한 규칙임을 추론할 수 있다.

시계열 변수에 의한 규칙 추론이 갖는 함의는 해당 시기에 어떤 정책 변화가 있었는지, 기준금리의 변동이나 탄소배출권이나 REC라는 권리가 얼마만큼 거래되고, 얼마에 거래되었는지 등을 재추적할 수 있는 근거로 활용될 수 있다.

해당 시기나 분지의 기준으로 나타나는 주요 요인들의 변동이나 구체적 수치를 참조함으로써, 단순 가격에 의한 의사결정이 아닌 KAU와 REC의 수요자들은 구매 전략을 계획하고, 공급자들은 판매 전략을 계획이 가능하다.

하지만 의사결정나무는 독립변수와 목표변수의 상관적 관계를 파악할 수 있기 때문에 지나친 인과론적 해석은 지양할 필요가 있다.

V. 결 론

탄소중립에 대한 전 세계적인 관심은 현재 진행형이다. 국내에서도 마찬가지로 탄소배출권 거래제와 신재생에너지 공급의무화 제도가 시행되어 왔고, 탄소저감을 위한 여러 전략이 추진되고 있다. 직접적인 탄소배출의 원인인 화석연료를 대체하기 위한 새로운 에너지원에 대한 확산도 각계에서 주문되고 있는 실정이다.

본 연구는 지난 7년 동안 탄소배출권 거래시장과 신재생에너지 공급인증서 거래시장에 대한 가격 추이와 주요 변수들의 패턴을 파악하고자 집중하였다.

그 결과 KAU와 REC가 거래되고 있는 시장 모두에서 큰 변동이 파악되었고, 일부는 통제가 어려운 외부환경적 요소에 의한 것들과 국내 정책과 시장 변동에 의한 것으로 확인되었다.

본 연구의 근본적인 목적은 분석 결과에서 살펴본 바와 같이, 현행 제도 내에서 거래되고 있는 권리들의 가격이 여전히 불안정한 상태에 있기 때문에, 이를 보완할 수 있는 데이터 기반의 기초자료를 생성하는데 있다.

각각의 시장에서 거래되고 있는 권리들의 큰 가격 변동성은 제도의 안착을 저해하고, 시장 참여자들에게 상호호혜적인 구조를 만들어내지 못하게 하는 환경을 조성한다.

변동성이 큰 시장에 대한 가격예측은 보통 발전사로 구성되어 있는 현재 수요자(의무공급자·의무할당자)에게 구매 의사결정을 지원할 수 있는 근거로 활용될 수 있다. 반대로 현재 공급자(민간 신재생에너지 발전사, 상쇄배출권이나 할당배출권을 판매하고자 하는 기업)에게는 판매 의사결정을 지원할 수 있는 근거로도 활용될 수 있다.

가격예측과 패턴을 통해 수요자와 공급자에게 정보의 비대칭성을 해소함으로써, 이른바 ‘win-win’ 구조를 형성하게 된다면 시장의 가격 안정화를 도모할 수 있을 것으로 기대한다.

시장 안정화는 정부 및 탄소중립 유관기관의 적극적인 정책 추진을 통해, 의무에 의한 고정적인 수량 확보보다 자율적 참여에 따른 수요와 공급을 유도할 수 있다.

결과적으로 탄소 감축과 에너지 대체로 인한 순탄소배출량을 대폭 감소시켜 현재 논의되고 있는 탄소중립을 실현할 수 있을 것이다.

참 고 문 헌

강형창, 김철수 “데이터마이닝에서의 군집분석 응용”, 기초과학연구, 제18권, 제2호, 2005, pp. 121-127.

이철용, “신재생에너지 공급인증서(REC) 가격 예측 방법론 개발 및 운용.” 에너지경제연구원 기본연구보고서, 정책 이슈페이퍼 16-06, 2015.

최종후, 서두성 “데이터마이닝 의사결정나무의 응용”, 통계분석연구, 제4권, 제1호, 1999, pp. 61-83.

온실가스 종합정보센터, 제1차 계획기간 배출권거래제 운영결과보고서, 2018, 환경부.

온실가스종합정보센터, 2019 배출권거래제 운영결과보고서, 2021, 환경부.

탄소중립위원회, 2050 탄소중립 시나리오, 2021, 탄소중립위원회 사무처.

한국에너지공단, 한국형 RE100 (K-RE100) 도입방안 및 인증서 거래시장 시범사업 안내, 2021, 한국에너지공단 신·재생에너지센터.

환경부, 온실가스 배출권거래제 제 3차 계획기간 국가 배출권 할당계획안, 2020.

환경부, 배출권거래제, 2021, 환경부 기후경제과.

Breiman, L., "Random Forest," Machine Learning, Vol. 45, 2001, pp. 5-32.

Quinlan, J. R., C4.5: Programs for Machine Learning, 1993, Morgan Kaufmann Publisher.

컨볼루션 신경망을 이용한 3차원 영상 판독 기술의 비교 연구

윤혜민 (동아대학교 경영정보학과 박사과정, 주저자 yoonlea1205@donga.ac.kr)
김상진 (동아대학교 경영정보학과 교수, 교신저자 skim10@dau.ac.kr)

… Extended Abstract …

최근 산업 전반에서 디지털 트랜스포메이션으로 인해 변화하는 환경에 대응하고 미래 가치 창출을 위한 다양한 서비스가 현실화되고 있다. 의료 및 헬스케어 분야에서도 자동화된 의료환경 및 인공지능 기술을 접목한 서비스들이 이미 현장에서 운영되고 있다. 본 연구에서는 정밀 의료 및 원격 진료를 위한 기술 중 하나인 인공지능 기반의 자동 영상 판독 시스템의 비교를 위하여 컨볼루션 신경망을 이용한 3차원 영상 판독 모델을 구현하고 두 가지 방법으로 성능과 해석 가능성 측면에서 비교하였다. 우리는 정상 그룹과 파킨슨병으로 구성된 그룹의 질병분류를 위해 FPCIT 3차원 영상을 이용하여 2D 및 3D 컨볼루션 신경망을 구현하였다. 2D 영상 분석을 위하여

단순화된 rank pooling 방법을 이용한 영상 표현 기법을 사용하였고, 두 모델을 Layer-wise Relevance Propagation을 이용하여 해석 가능성을 평가하였다. 3D에서 2D로 요약된 영상은 병변의 차이를 보이는 선조체 영상의 국소적 차이를 잘 나타내 주었다. 두 방법을 비교한 결과, 2D 및 3D 영상 분류기의 성능은 서로 유의한 차이가 나지 않았다. 따라서, 2D 영상 분석은 영상 스크리닝 및 영상 검색 시스템과 같이 방대한 데이터 처리에서 효율적일 수 있고, 3D 영상 분석은 서비스 대상 고객의 이해와 데이터의 해석이 필요한 시스템에서 효과적일 수 있다.

Key Words : 3D영상, 컨볼루션신경망

기후예측을 통한 미래 과수작물지도 : 바나나, 망고 중심으로

박혜은 (서울여자대학교 컴퓨터학과 석사과정 asphe@swu.ac.kr)
이종태 (서울여자대학교 경영학과 조교수 light4u@swu.ac.kr)

… Abstract …

지속된 온난화와 기후변화의 심화로 한국의 생산 농작물과 경작지형이 크게 변화하고 있다. 제주의 대표 작물인 귤 생산지는 2019년 경기도까지 북상하였으며 제주지역엔 망고, 바나나, 올리브 등의 열대 및 아열대 작물이 자라고 있다. 이에 따라 기후변화에 대비한 과수작물 변화연구의 필요성이 커지고 있다.

이러한 배경을 바탕으로 본 연구는 열대 및 아열대 과수작물 재배 가능지역의 확대 경향을 예측하고 기후 변화에 단계별로 대응하기 위한 딥러닝 알고리즘 기반 데이터 분석 연구를 진행했다. 연구의 수행을 위하여 현재 바나나와 망고가 생산되고 있는 내륙 지방인 산청군과 완도군의 기온변화를 기상청 제공 오픈 데이터를 통해 수집, 예측했다. 산청군의 경우 1990년~2019년까지 10,950일, 완도군은 2000년~2019년까지 7,300일의 데이터를 분석했으며, 오픈소스 기반 데이터 마이닝툴인 Orange Data Mining Software를 활용하여 수행하였다. 대표적인 예측 알고리즘인 Linear Regression, Support Vector Machine, Neural Network 알고리즘의 예측력을 비교한 결과, 기대와 달리 Linear Regression Model의 예측력이 가장 우수한 것으로 나타났으며, 이러한 원인은 계절의 변화와 이에 따른 기온과 강수량 변화, 일조량 변화가 뚜렷한 한국의 기후 특성과 같은 요인이 영향을 준 것으로 사료된다.

본 연구의 결과는 우리나라 농업환경에 적합한 열대 및 아열대 과수작물 재배 인프라를 갖추고, 장기적인 기후변화에 대응한 농업 경영 및 관련 정책 수립을 위한 근거로 활용할 수 있는 것으로 기대한다.

Key Words : LSTM, RNN, Climate Prediction, SVM, 딥러닝 예측

I. 서 론

최근 지속적인 기후변화와 평균기온 상승으로 인해 우리나라의 농작물 재배 지역 및 재배 품목이 크게 변화하고 있다. 전남지역에서 제주도의 대표 작물인 감귤이 생산되고 있으며 한라봉이 정읍에서 사과는 강원도 고성에서 생산되고 있다[1]. 이러한 변화에 대응하여

농촌진흥청은 우리나라의 6대 과수작물인 사과, 배, 복숭아, 포도, 단감, 감귤의 총 재배 가능지를 2090년까지 10년 단위로 예측한 결과를 발표했다. 해당 보고서에 따르면 사과, 복숭아는 2000년대 후반에는 강원도 일부 지역에서만 재배할 수 있을 것으로 예상되고, 감귤은 남해안과 강원도 해안지역까지 재배 한계선이 확대될 것으로 예상했다. 농업은 기

상 의존도가 매우 높은 산업분야로서 기후변화에 따른 다양한 작물과 수목에 대한 복합적인 반응이나 그 대책을 예측하고 대응하기 위한 연구와 노력들이 나타나고 있다[2]. 2020년 기상청에서 발표한 SSP5-8.5 기후변화시나리오에 따르면 금세기 말(2081~2100)에는 연평균 기온이 7.0°C 증가할 것으로 예상되고 있다. 실제로 한반도가 아열대 기후로 변화하면서 망고, 용과 등의 아열대 과일뿐 아니라 울금, 여주, 아스파라거스 등 열대·아열대 채소재배도 증가하고 있으며(김창영 외, 2019), 아열대 과수의 총 재배지역은 2010년 33.9 ha에서 2014년 58 ha, 2020년에는 171.3 ha까지 늘었다(방경원, 2022). 금세기 후반기엔 우리나라의 대부분 지역이 아열대 기후로 변화할 것으로 예측되는 만큼, 기존의 과수작물 재배 지역 변화뿐 아니라 과수작물 품종 변화 연구의 필요성 또한 증가하고 있다.

따라서 본 연구에서는 우리나라에 확대 도입할 수 있도록 바나나와 망고 생산에 적합한 생육환경을 조사하고, 현재 바나나와 망고가 생산되고 있는 산청과 영광, 완도지역 중심으로 평균기온 변화를 예측을 통해 해당 작물의 재배에 적합한 지역을 알아보하고자 한다.

II. 배경이론

1. 바나나의 생육환경

온도는 바나나 생육에 가장 중요한 환경 요인으로 작용하기 때문에 바나나 재배의 성패는 온도 관리의 조건에 달려있다. 열대 과수인 바나나는 고온을 좋아하는 식물로 세계적인 주산지의 연평균 기온은 27°C 내외이며 기온차가 심하지 않다.[3]. 바나나는 고온을 요구

하면서도 최고온도는 27°C이고 이보다 더 높은 온도에서는 일소장해를 받기 쉽고 21°C 이하의 온도에서는 생육이 정지된다[4]. 따라서 우리나라의 6~9월과 같은 고온에서는 가온 없이도 바나나 생육이 적합하나, 주로 제주도와 경남 산청, 포항, 전남 해남 등의 남부지방과 같은 일부 지역을 제외하고는 충청권 이상 지역에선 재배할 수 없다. 바나나가격은 수입산 바나나가격보다 2배가 높은 고소득 작목으로 여러 아열대 작물 중에서도 특히 인기가 높다. 이러한 수요와 농업기술의 발전을 바탕으로 시설 하우스를 통해서도 바나나 재배가 이루어지고 있지만, 우리나라 기후가 아열대로 변해가는 만큼 바나나 재배 가능지역 또한 점차 확대될 것으로 기대한다.

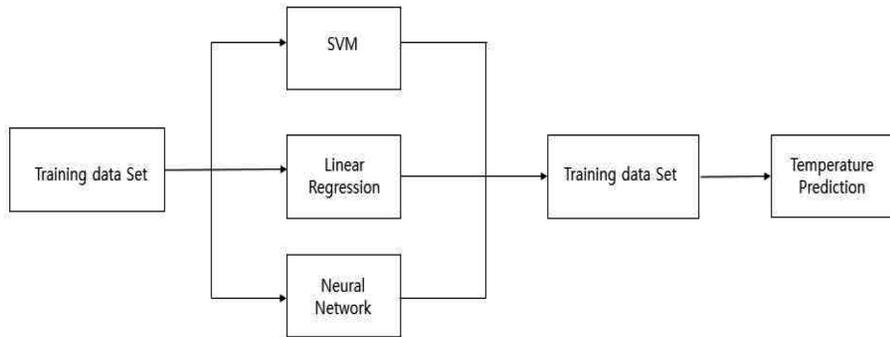
2. 망고의 생육환경

망고는 아열대성 과수로서 기본적으로 따뜻한 기후를 좋아하며 생육적온은 24~27°C로 알려져 있다. 개화기나 유과기시 온도가 5°C 이하의 짧은 노출도 심각한 해를 입을 수 있으며 37°C 이상의 고온에서도 해를 입는다. 또한, 망고는 호광성 과수로 충분한 광을 필요로 하며, 광이 충분한 곳에서 개화가 빠르고 신초의 생장, 결실, 과실착색이 좋아진다[5]. 농촌진흥청에서 실시한 아열대 작물 재배 실태 분석에 따르면 현재 우리나라의 망고 주산지는 제주도에서 점차 북상하고 있다. 제주도의 망고 재배지역은 크게 변화가 없는 반면 전남지역의 영광, 완도, 고흥을 비롯한 경남의 김해, 함안 지역의 망고 재배면적은 확대되고 있다. 그러나 망고는 전적으로 노지재배가 불가하여 난방비가 소요되는 시설재배 및 용기재배가 여러 지역에서 적용되고 있으나 재배법이 정립되지 않았다[6]. 우리나라는 망

고를 겨울철에 가온을 할 수 있는 시설 하우스에서 재배되고 있으며 재배역사가 짧고 재배 관련 연구가 미흡하여 아직까지 시설재배에 대한 명확한 표준 재배 방법이 없는 실정

이다[7]. 따라서 기온예측을 통한 망고 재배 가능지역의 추론 및 기후변화가 재배지 변화에 미치는 영향에 대해 밝히고자 한다.

〈그림 1〉 워크플로우



Ⅲ. 연구방법

이와 같은 연구목적을 달성하기 그림1과 같은 워크플로우로 연구를 진행했다. 첫 번째로 기상청의 기상자료 개방포털(data.kma.go.k)에서 종관기상관측(ASOS) 일자료 데이터 중 현재 바나나와 망고를 생산지역인 산청군 1990년~2019년까지 30년간의 평균, 최고, 최저기온 데이터와 완도군의 2000~2019년까지 20년치 데이터를 활용했다. 이후 Orange Data Mining Software를 활용하여 Linear Regression, Support Vector Machine, Neural Network 알고리즘의 예측력을 비교했다. Orange Data Mining Software는 오픈소스 크로스 플랫폼 소프트웨어로서 위젯 기반의 시각적인 인터페이스를 이용해서 손쉽게 인공지능 기능을 구현할 수 있을 뿐만 아니라, 데이터 분석을 위해서도 유용하게 사용할 수 있다. 또한 탐색적 데이터 분석 및 시각화에 사용할 수 있으며 실험선택, 예측 모델링 및 추천 시스템을 위한 플랫폼을 제공하며 계층리서치, 바이오의학,

바이오정보 및 교육분야에서 활용할 수 있는 오픈소스이다[9]. 따라서 본 연구에서는 Orange Data Mining Software을 활용하여 두 번째 단계인 알고리즘별 예측력을 비교했다.

1. Data

산청군의 1990년~2019년 30년치 기온데이터, 완도군의 2000년~2019년 20년치 기온데이터 중 평균기온을 target으로 설정했다.

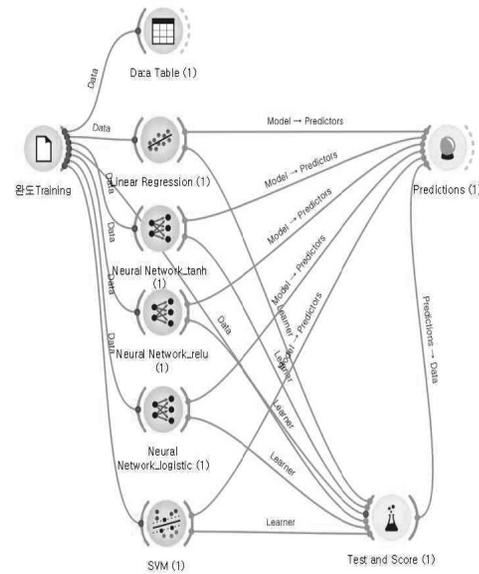
〈그림 2〉 산청군 기온데이터 예시

	A	B	C	D	E
1	지점	일시	평균기온(°C)	최저기온(°C)	최고기온(°C)
2	산청	1990-01-01	0.3	-5.8	7.7
3	산청	1990-01-02	0.2	-1.7	3.6
4	산청	1990-01-03	-0.8	-3.5	2.9
5	산청	1990-01-04	-1.8	-7.8	5.1
6	산청	1990-01-05	1.9	-3.3	7.7
7	산청	1990-01-06	1.7	-2.1	5.6
8	산청	1990-01-07	-1.8	-7.9	7.7
9	산청	1990-01-08	-1.1	-8.5	6.2
10	산청	1990-01-09	5.1	1.9	9
11	산청	1990-01-10	7.2	3.1	10.3
12	산청	1990-01-11	3.3	-2.2	9.4
13	산청	1990-01-12	3.4	0	6.1
14	산청	1990-01-13	-0.6	-5.5	3.9

〈그림 3〉 완도군 기온데이터 예시

	A	B	C	D	E
1	지점	일시	평균기온(°C)	최저기온(°C)	최고기온(°C)
2	완도	2000-01-01	7.8	2.5	13.6
3	완도	2000-01-02	6.9	4.3	11.5
4	완도	2000-01-03	2.9	-1.3	6.3
5	완도	2000-01-04	4.4	-1.9	11.2
6	완도	2000-01-05	9.3	5.1	12
7	완도	2000-01-06	6.9	2.2	11.4
8	완도	2000-01-07	0	-1.3	2.2
9	완도	2000-01-08	1.8	-6	3.6
10	완도	2000-01-09	2.7	1	4
11	완도	2000-01-10	3.4	-1	6.8
12	완도	2000-01-11	4.2	-2.7	9.6
13	완도	2000-01-12	7	5.3	8.8

〈그림 5〉 Orange Data Mining 완도군 실행

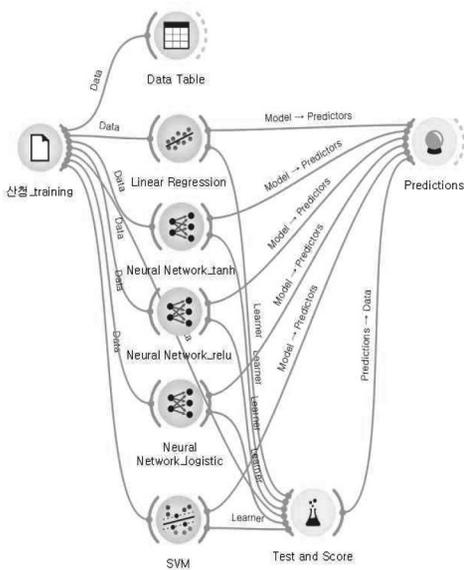


2. Algorithms

Orange Data Mining Software에서 제공하는 알고리즘 중 Support Vector Machine, Linear Regression, Neural Network 알고리즘을 이용하여 각각 예측력을 비교하였다. 데이터는 Training data set과 Test data set은 8:2의 비율로 설정했다.

3. 결과

〈그림 4〉 Orange Data Mining 산천군 실행



〈그림 6〉 산천군 실험결과

Model	MSE	RMSE	MAE	R2
Neural Network	71.805	8.474	7.420	-0.010
Linear Regression (1)	0.326	0.571	0.438	0.995
SVM (1)	69.878	8.359	7.345	0.017
Neural Network	71.193	8.438	7.335	-0.001
Neural Network	71.526	8.457	7.405	-0.006

〈그림 7〉 완도군 실험결과

Model	MSE	RMSE	MAE	R2
Neural Network	81.774	9.043	7.923	-0.010
Linear Regression	0.519	0.720	0.564	0.994
SVM	79.488	8.916	7.828	0.018
Neural Network	795250.487	891.768	888.762	-9823.388
Neural Network	82.077	9.060	7.939	-0.014

완도군의 경우 데이터 양이 산천군에 비해 10년치가 부족한만큼 Neural Network분석을 위한 데이터양이 충분하지 못했던 것으로 사료된다. 특히 최근 다양한 수치 수치 예측에서 활용되고 있는 RNN(Recurrent Neural Network) 알고리즘의 일종인 LSTM(Long

Short-Term Memory Models) 알고리즘의 예측력이 높을것으로 기대한 것과는 반대로 Linear Regression의 예측력이 가장 높은 것으로 나타났다. 계절의 변화와 이에 따른 기온과 강수량 변화, 일조량 변화가 뚜렷한 한국의 기후 특성과 같은 요인이 영향을 준 것으로 사료된다. 또한 LSTM모형은 과거시점의 수치 데이터 뿐만아니라 외부 변수를 학습을 통해 예측 정확도 역시 높일 수 있는 것으로 알려져 있는점을 감안한다면 평균데이터 뿐 아니라 기온에 영향을 미치는 다양한 변수 학습을 추가한다면 예측력이 개선될 것으로 예상된다.

IV. 결론

본 연구에서는 현재 우리나라에서 바나나와 망고를 생산하고 있는 산청군과 완도군 기온데이터를 활용하여 SVM, Linear Regression, Neural Network 알고리즘의 예측력을 비교한 결과 Linear Regression 알고리즘의 예측력이 가장 높았음을 확인했다. 향후 추가 연구를 통해 계절감이 있는 수치데이터에서 LSTM의 예측력을 높일 수 있는 방안과 더불어 해당 지역을 중심으로 열대·아열대 과수작물 재배지역 확산을 시간대별로 예측하고자 한다. 더 나아가서 바나나와 망고 재배가 가능한 지역을 추천하기 위한 작물지도까지 제작할 예정이다.

참고 문헌¹⁾

- [1] 조성주, 김화년, “기후변화에 따른 작물 재배면적 할당 및 이동 변화 분석” Korean Journal of Agricultural Management and Policy, 제49권, 제2호, 2012, pp. 191-211.
- [2] 광태식 외, “지구 온난화에 따른 국내 과수작물 재배지 변화에 대한 GIS 예측 모형 연구 - 여섯 가지 열대 및 아열대 과수를 중심으로 -” 한국공간정보시스템학회논문지, 제10권, 제3호, 2008, pp. 93-106.
- [3] 농촌진흥청 농사로(<http://www.nongsaro.go.kr>)
- [4] 제주농업기술원 (<https://agri.jeju.go.kr/>)
- [5] 제주농업기술원 (<https://agri.jeju.go.kr/>)
- [6] 조혜성 외, “망고 용기재배에서 환경 및 생육과 토양수분의 관계” , 한국원예학회 원예과학기술지, 제 39권, 별호 II [초록집], 2021, pp136-137

[7] 김윤미 외, "인공지능과 데이터 과학을 위한 Orange 플랫폼의 교육적 이용 사례", 한국컴퓨터교육학회, 제26권, 제1호, 2022, pp. 213-216

[8] 임찬규 외 "망고 시설재배 표준화를 위한 생육단계, 수체생육 및 온도조사", 한국원예학회, 2021,

[9] Kodati, Vivekanandam, "Analysis of Heart Disease using in Data Mining Tools Orange and Weka " Software & Data Engineering Quarterly, Vol. 18 No. 1, 2018

[10] 김기태 외 "Algorithms vs. Data? LSTM model based Demand Forecasting Using External Weather" 한국지능정보시스템학회, 2020, pp. 24-25

1) 참고문헌은 본문이 끝난 뒤 기재한다. 기재순서는 국문문헌, 외국어문헌의 순서로 하며 모두 저자명에 의거하여 정렬한다. 참고문헌이 정기간행물인 경우 저자 명, "논문 제목", 잡지 명(이탈릭체), 권호, 면, 발행 년도, 페이지 순서로 기입한다.

2022년 한국정보시스템학회 추계학술대회

데이터 기반 디지털 금융과 비즈니스 가치 창출



Session C

발표 장소 : 창의관 3층 302호

C2 (15:20~17:00) 디지털서비스와 교육

좌장 : 이웅규(대구대)

순서	저자(소속)	논문제목	
1	정종호 (대구가톨릭대)	Research Report: A Summary of the Studies of the Effects of Various Performance Graphs on Computer-Based Idea Generation	198
2	정종호 (대구가톨릭대)	The effect of alternate row shading when displaying comments on computer-based idea generation	203
3	김종기, 왕젠보 (부산대)	The Privacy Paradox Regarding the Use of Location-protective Settings in the Context of COVID-19 Contact Tracing Applications	208
4	권희석, 이민혁 (부산대)	웨이블릿 변환과 멀티헤드 어텐션 기법을 적용한 주가 예측	212
5	LIU TINGTING(전남대) 이웅규(대구대), 고준(전남대)	소비자의 디지털 그림자 노동 반응이 지속사용의도에 미치는 영향 : 만족지연의 조절효과를 중심으로	214
6	이웅규, 조아라 (대구대)	휴대용 디지털 기기 액세서리 구매과정에서의 디지털그림자노동 역할 : 아이패드 사례를 중심으로	222

Research Report : A Summary of the Studies of the Effects of Various Performance Graphs on Computer-Based Idea Generation

정종호 (대구가톨릭대학교 소프트웨어융합학과 교수, 주저자 jjung@cu.ac.kr)

... Abstract ...

Information Graphics (Infographics) presents quantifiable data and information in various graphical formats quickly and clearly. Since human brains are adjusted to catch images faster than texts or numbers, infographics in the form of “Information Dashboard” has already been widely used in business. However, Groupware has long overlooked the adoption and use of infographics, in particular, in the idea generation process. Jung and colleagues have conducted a series of studies in an effort to improve the performance of idea generation by applying the notion of infographics in the form of real-time performance feedback. In sum, we summarize the results of our various studies and conclude its implications.

Key Words : Idea Generation, Performance, Infographics

I . Introduction

The value of creativity to come up with novel and useful ideas is critical in today’s fast-paced business environment. An individual-based (e.g., Nominal) or a group-based (e.g., Face-to-Face and Electronic Brainstorming) idea generation has been widely recognized and used in many successful companies as a means to facilitate creativity and innovation.

Given that both Nominal and Electronic Brainstorming (EBS) consistently outperformed Face-to-Face, the productivity of EBS when compared to that of Nominal did not yield noticeable statistical differences regardless of group sizes and other factors. There is certainly no harm in utilizing EBS because they are no less productive than are baseline groups (i.e., Face-to-Face). However, given that paper-and-pencil-based nominal groups’ performance

is no worse than that of EBS and the cost of setting up and maintaining EBS is high, the net benefit of EBS to the companies is questionable. Thus, over the years, we have pursued a series of performance feedback (outcome-based and process-based) studies in an effort to significantly increase the performance of EBS.

II. Experiment

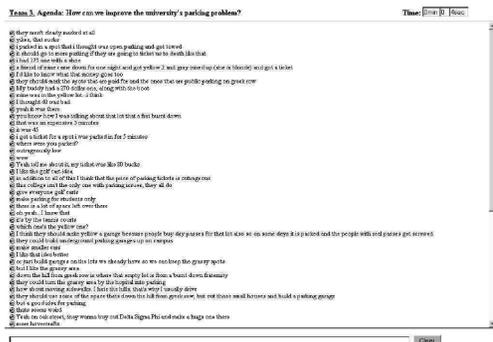
1. Research Design

Business students from a large university served as subjects. Participants were asked to generate ideas on “How can we improve the university’s parking problem?” Following prior studies, this task was chosen for its high relevance to the subjects because it stimulates participants to draw on their personal knowledge and experience.

2. Outcome-based Feedback

(Jung et al., 2010; Jung, 2013, 2018)

(Figure 1) Traditional EBS Environment



(Figure 2) A Bar-based Performance Graph

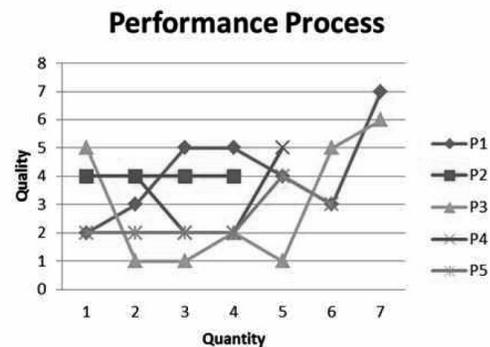


We developed and deployed a real-time electronic visual representation of individual performances in the form of a bar or a dot chart (See Figure 2). The chart summarizes cumulative contributions each subject makes and allows the performance comparison with others in an assigned group and is redrawn every ten seconds based on their accumulated contributions.

3. Process-based Feedback

(Jung, 2014, 2017)

(Figure 3) A Process-based Performance Graph



A process chart allows participants to monitor process variation by comparing new performance data to past performance data. The chart

objectively shows the pattern of event (i.e., individual performances) in the temporal process. This feedback visually displays individuals' performances two-dimensionally (quality for each idea vertically and quantity of ideas horizontally).

Two trained experts who had been intensively exposed to idea generation task for at least two years were instructed to rate comments on the dashboard real-time. Participants in the feedback conditions were told that a built-in quality algorithm would rate comments based on a certain probability. This further controls any potential confounding effects in the experimental setting.

4. Procedure

Subjects were told that their contributions would be anonymous, but would be able to evaluate each group member's performance. All subjects were allowed to become familiar with the operation of the system prior to the main task by first working on a practice task. The experimenter then read aloud the experimental instructions while the participants followed along on their own copies. Specifically, participants were given a version of Osborn's [1957] brainstorming rules and were instructed to follow them. The rules directed

students to generate as many ideas as possible, to withhold criticism, to include wild ideas, and to build on the ideas of others. They were also told that their contributions would be used to improve the campus-parking problem. The system was programmed to stop automatically after 10 minutes, after which the subjects completed a brief questionnaire and were then released.

5. Dependent Variables

The dependent variables were both quantity (efficiency) and quality (effectiveness) of ideas after removing redundant ideas. The manner by which these performance measures were operationalized is consistent with many prior studies. A master list of ideas (generated in previous studies) was rated by expert raters. Any ideas generated that were not on the master list were also evaluated for quality by the same three expert raters and included in each subject's final performance scores.

III. Discussion

1. Outcome-based Approach

The rationale behind the use of a performance feedback mechanism is to heighten self-awareness and in turn to increase self-attention. According to the notion of social identity, humans

have desire to maintain a positive self-concept in a group. The self-concept is personal and social in nature. This triggers self-evaluation process to take place and objective information such as performance feedback that is diagnostic concerning one's and others' abilities is essential (Jung et al., 2010). When such information is available, individuals can observe and compare himself (herself) to others who are outperforming the individual. Self-improvement motives are, then, likely to emerge. As a result, individuals tend to be motivated to improve their image (i.e., self-improvement) by enhancing their performance.

2. Process-based Approach

As the goal of brainstorming is “the generation of as many quality ideas as possible,” the task has quantity and quality dimensions. Thus, both quantity and quality of performance feedback are equally important to achieve the goal. Austin and Bobko (1985) note that “if quality were a crucial outcome [as in brainstorming], then not including a quality measure of performance leads to criterion deficiency” (p. 291). Prior studies also suggest high correlation between quantity and quality (i.e., quantity breeds quality) (Jung, 2014, 2017). Thus, both dimensions (quantity and quality) of performance feedback

seem equally important to achieve the goal.

IV. Conclusion

Like an adage “a picture is worth a thousand words,” human brains process images better and more quickly than sets of words or numbers. In addition, “to the user, the system is the interface.” The computer screen is the only communication channel to individuals in computer-based idea generation. Referring to figure 1, in a traditional EBS environment, social comparison and matching leads to performance equalization. To be specific, high performers lower their effort to low performers and low performers vice versa. However, table 1 clearly shows that regardless of an outcome-based or process-based approach, performance feedback induces upward social comparison (i.e., the matching of one's own performance to that of better performing group members). Thus, providing a visual performance feedback mechanism should help to improve the productivity of idea generation.

〈Table 1〉 Means and Standard Deviations for Idea Generation Performance

Quantity			Quality		
Outcome-based	M	43.30	Outcome-based	M	127.67
	SD	10.70		SD	23.58
Process-based	M	59.20	Process-based	M	135.84
	SD	9.86		SD	29.24
Baseline	M	31.10	Baseline	M	91.41
	SD	9.72		SD	27.34

References

Austin, J. T., and Bobko, P.(1985), ‘Goal setting theory: Unexplored areas and future research needs,’ *Journal of Occupational Psychology*, 58, 289–308.

Jung, J. H., Schneider, C., and Valacich, J.(2010), ‘Enhancing the Motivational Affordance of Information Systems: The Effects of Real-Time Performance Feedback and Goal Setting in Group Collaboration Environments,’ *Management Science*, 56, 724–742.

Jung, J. H.(2013), ‘The Effect of Total-Identification on the Performance of Computer-Mediated Group Idea Generation,’ *Journal of Business Research*, 28, 27–50.

Jung, J. H.(2014), ‘The Effect of Real-Time Individual Process Performance Feedback on Computer-based Group Idea Generation,’ *The Journal of Information Systems*, 23, 91–107.

Jung, J. H.(2017), ‘Process Performance Feedback and Quality Goal Setting as Sources of Process Restrictiveness and Behavior Guidance in Electronic Brainstorming,’ *The Journal of Information Systems*, 26, 1–15.

Jung, J. H.(2018), ‘The Effect of Graphical Formats on Computer-Based Idea Generation Performance,’ *The Journal of Information Systems*, 153–169.

The effect of alternate row shading when displaying comments on computer-based idea generation

정종호 (대구가톨릭대학교 경영학부 교수, 주저자 jjung@cu.ac.kr)

… Abstract …

Computer-based idea generation when compared to other techniques has been a favorable choice because of its built-in wide range of options. However, the user interface, which facilitates divergence and convergence of creative ideas, has never been changed since the beginning. In particular, the way comments are displayed on the computer screen has not given an another thought. Given that comments are randomly displayed and accumulated rapidly on the screen, recent studies point out that such practice causes cognitive overload, hindering the performance of idea generation. As a plausible solution, this study explores a possibility of the alternate row shading technique when displaying comments on the computer screen.

Key Words : Idea Generation, Productivity, Alternate Row Shading

I. 서론

Despite the rich structural features of Electronic Brainstorming (EBS) including anonymity, group memory, and parallel input, prior studies suggest that less than expected performance of EBS is due to remaining crucial process losses that may hinder cognitive stimulation (Jung, 2020). Examples include junk comments, production time underutilization, attention blocking to stimuli, dual-

task interference, information overload, incomplete use of information, and social comparison and matching. Among various process losses discussed above, this study discusses a plausible way to reduce information overload in the process of idea generation.

II. Theoretical Framework

Cognitive stimulation and cognitive

interference are both sides of the same coin in the idea generation process (Jung, 2020). They both occur by reading the contributions of others in computer-interacting groups. The former occurs when participants are stimulated by task-relevant ideas generated by others. On the other hand, the latter occurs when the contributions of others interfere with one's train of thought. Given that many prior studies have pointed out a much less significant influence of cognitive stimulation in computer-based groups (Jung et al., 2010), quickly accumulating a high volume of comments are one such contributor to cognitive interference.

A typical outcome of group idea generation is composed of ideas and frivolous comments (i.e., junk comments). Unlike in face-to-face groups, where the spoken words quickly disappear soon after they are verbalized, all contributions including frivolous comments in computer-based groups are captured in group memory and are constantly displayed on the computer screen throughout the idea generation session.

With the current practice of a random-like display (first come, first served) of all contributions on the computer screen (Jung et al., 2012), Hilmer and Dennis (2001) points out that individuals need

to exert extra effort to discern and process stimulating ideas for cognitive stimulation if information is presented in a large unorganized pool of information (as in computer-based groups). Thus, such practice may significantly interrupt or distract individuals' flow of cognition (Jung et al., 2012) – refers to the intrapsychic processes that enable the acquisition, storage, transformation, and use of knowledge (Jung et al., 2012) – on the primary task (i.e., to generate as many quality ideas as possible).

III. Experiment

1. Research Design

Business students from a large university will serve as subjects. Participants will be asked to generate ideas on “How can we improve the university's parking problem?” Following prior studies, this task was chosen for its high relevance to the subjects because it stimulates participants to draw on their personal knowledge and experience.

2. Row Shading Technique

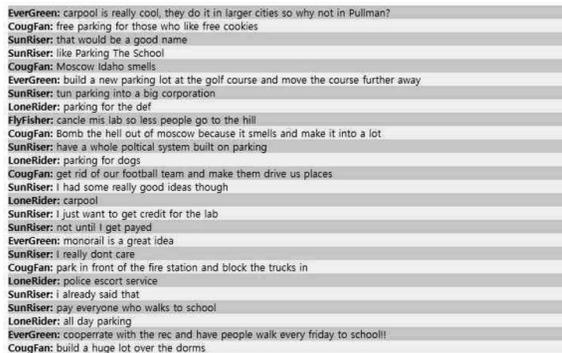
In the perspective of human-computer interaction, WYSIWYG (what you see is what you get) is a widely adopted notion of information processing. We mentioned above that the interface of EBS displays

others' ideas randome—like in a text format and those ideas quickly build up on the computer screen as an ideation session progresses (see figure 1). In such an environment, most individuals perceive overload (Jung et al., 2010) and exert additional time and effort to find stimulating ideas for cognitive stimulation. Girotra et al. (2010) even suggest that “building on others’ ideas is counterproductive ... such buildup neither create more ideas, nor are the ideas that build on previous ideas better” (p. 591).

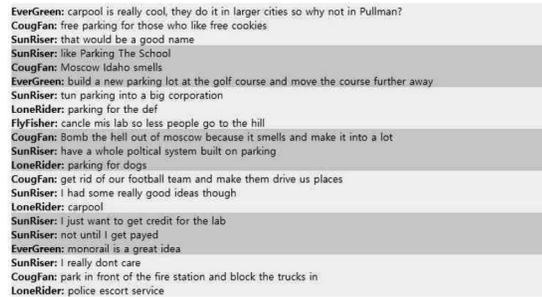
(Figure 1) Traditional EBS Environment



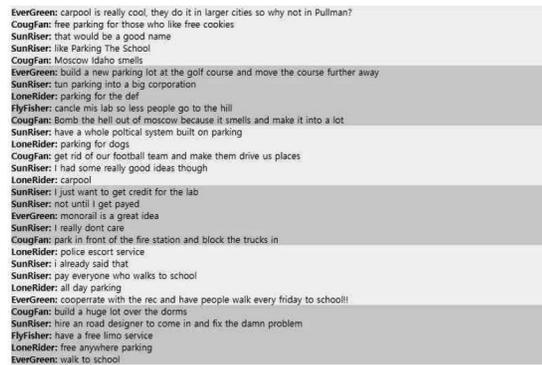
(Figure 2) An Example of Every other row shading



(Figure 3) Shading groups of three rows



(Figure 4) Shading groups of five rows



We develop and deploy various real-time electronic visual representation of individual performances based on a row shading technique (See Figure 2, 3, 4). To differentiate rows from each other, a different shade is used as background color for every second row (or groups of three and five). Adding a color to alternate rows can make the comments easier to scan, providing a visual aid to follow a row from the left to the right and back again – without confusing one row with another.

3. 연구가설

H1: Individuals or groups in the row shading conditions will perform better

than the traditional condition

4. Procedure

Subjects were told that their contributions would be anonymous, but would be able to evaluate each group member's performance. All subjects were allowed to become familiar with the operation of the system prior to the main task by first working on a practice task. The experimenter then read aloud the experimental instructions while the participants followed along on their own copies. Specifically, participants were given a version of Osborn's (1957) brainstorming rules and were instructed to follow them. The rules directed students to generate as many ideas as possible, to withhold criticism, to include wild ideas, and to build on the ideas of others. They were also told that their contributions would be used to improve the campus-parking problem. The system was programmed to stop automatically after 15 minutes, after which the subjects completed a brief questionnaire and were then released.

5. Dependent Variables

The dependent variables were both quantity and quality of ideas after removing redundant ideas. The manner by which these performance measures were operationalized is consistent with

many prior studies (e.g., Jung et al., 2010). A master list of ideas (generated in previous studies) was rated by expert raters. Any ideas generated that were not on the master list were also evaluated for quality by the same three expert raters and included in each subject's final performance scores.

IV. Discussion

A design of the group interface when displaying the comments needs to be reconsidered. Since the beginning of EBS, the user interface, which displayed all comments first-come first-serve based, has not been changed with the assumption that a random-like display facilitates divergent thinking. However, we argued that such a practice (1) induces cognitive loading to discern stimulating ideas and (2) interferes with a train of thought, hindering cognitive stimulation. Nijstad et al. (2003) suggested one solution to maintain stimulation and to prevent interference simultaneously, i.e., a clustering of semantically related ideas. The current EBS that supports categorization of ideas only after idea generation session ends needs to consolidate categorization of ideas into the idea generation process. In this way, similar ideas are clustered, reducing attention blocking and cognitive interference via information overload.

참 고 문 헌

Girotra, K., Terwiesch, C., and Ulrich, K. T.(2010), ‘Idea Feneration and the Quality of the Best Idea,’ *Management Science*, 56, 591–605.

Hilmer, K. M., and Dennis, A. R.(2001), ‘Stimulating Thinking: Cultivating Better Decisions with Groupware Through Categorization,’ *Journal of Management Information Systems*, 17, 93–114.

Jung, J. H.(2020), ‘An interaction between cognitive ability and personality on the performance of computer–based group idea generation,’ *The Journal of Information Systems*, 29, 265–286.

Jung, J. H., Schneider, C., and Valacich, J.(2010), ‘Enhancing the Motivational Affordance of Information Systems: The Effects of Real–Time Performance Feedback and Goal Setting in Group Collaboration Environments,’ *Management Science*, 56, 724–742.

Jung, J. H., Lee, Y., and Karsten, R.(2012), ‘The Moderating Effect of Extraversion–Introversion Differences on Group Idea Generation Performance,’ *Small Group Research*, 43, 30–49.

Nijstad, B. A., Stroebe, W., and Lodewijkx, H. F. M.(1999), ‘Persistence of Brainstorming Groups: How Do People Know When to Stop?,’ *Journal of Experimental Social Psychology*, 32, 165–185.

Osborn, A. F., *Applied Imagination* (Rev. ed.), Scribner, 1957.

The Privacy Paradox Regarding the Use of Location-protective Settings in the Context of COVID-19 Contact Tracing Applications

김종기 (부산대학교 경영학과 교수, 주저자 jkkim1@pusan.ac.kr)
왕젠보 (부산대학교 경영학과 박사수료생, 교신저자 luckyblair@pusan.ac.kr)

… Abstract …

As one of the most efficient and effective strategies to mitigate the spread of the coronavirus disease 2019 (COVID-19), contact tracing applications have also caused serious public concerns about privacy all around the world. Till now, the ongoing COVID-19 pandemic has tragically caused millions of deaths and totally devastated the world economy. In response to the pandemic, many countries have developed and adopted contact tracing applications to help reduce the spread of the virus. However, such widespread use of contact tracing applications has increased the collection of users' personal information, which puts personal privacy at risk of being misused or leaked. Even though contact tracing applications have gained more academic attention, it is still at an early stage and thus it lacks of empirical approaches. Also, the majority of existing studies have focused on the positive perspectives regarding the use of these applications, there are few empirical studies addressing the potential negative consequences and corresponding solutions. What is more, accumulated evidence shows that individuals express their concerns but do not take actual actions to cope with the concerns and protect their own privacy, which refers to a discrepancy between individuals' attitudes towards privacy and their actual behaviors, i.e., the privacy paradox. Therefore, this study seeks to examine the privacy paradox regarding the use of location-protective settings in the context of COVID-19 contact tracing applications. In specific, based on Information-Motivation-Behavioral (IMB) Skills model and implementation intention, we aim to identify the existence of privacy paradox and investigate the factors that determine users' location-protective behaviors.

Key Words : Privacy Paradox, Location Settings, COVID-19 Contact Tracing Application, Implementation Intention, Information-Motivation-Behavioral (IMB) Skills model

I . Introduction

According to the World Health

Organization (WHO), as of October 7, 2022, the coronavirus disease 2019 (COVID-19) has caused millions of

illnesses and more than 6.5 million deaths globally. The rapidly spreading and long-lasting COVID-19 pandemic has affected and even permanently changed our daily lives in many ways. In addition to the adverse effects on human health and life, COVID-19 also affect economic activities (Feyisa, 2020). It is not an exaggeration to say that COVID-19 has almost brought down the existing world economy and forced it to form a new pattern. Because of the business depression caused by the pandemic, unemployment rises and people are becoming poorer than ever. Even now, the global economy is still recovering from the repetition of such a vicious cycle. Therefore, governments and health authorities come up with strategies to fight the pandemic and bring normal daily life back to people. Apparently, contact tracing applications have done some excellent work in this and earned their credits. However, the increasing use of these applications calls for more collection of users' private information, which raises potential privacy threats and risks.

So far, prior studies on COVID-19 contact tracing applications have focused on the positive aspects that these applications help us prevent further spread of coronavirus. There has been little literature in terms of the potential negative consequences due to the extensive use of COVID-19 contact

tracing applications, not to mention the empirical approaches. According to accumulated academic evidence and polls, people always claim that they are concerned about their privacy being violated, but they do not take any actual actions to protect it, which is called the privacy paradox.

In this study, we focus on COVID-19 contact tracing application users and their use of location-protective settings. Based on the information-motivation-behavioral (IMB) skills model and implementation intention theory, we seek to investigate the factors that influence the actual use of location-protective settings, and we attempt to examine whether the privacy paradox exists in our research context. In addition, we hope to emphasize the mediating effects of implementation intention on the relationship between intention and behavior, as well as the moderating effects of privacy self-efficacy on the relationship between implementation intention and behavior.

II . Literature Review

1. Information-Motivation-Behavioral (IMB) Skills Model

The IMB skills model, first proposed by Fisher and Fisher, was developed to explain influences on performing health

behaviors at the individual level (Rubens et al., 2015). The model was designed on the basis of the assumption that information about specific techniques associated with improving or preventing a health condition, along with the motivation to take an action, will operate through necessary behavioral skills that influence a health behavior (Fisher & Fisher, 2000). The IMB skills model has mostly been applied in the behavioral health and social psychology literatures to study and understand health-related behaviors. While in the information systems literature, the IMB skills model has been suggested as a theoretical foundation to explore security and privacy behaviors. The IMB model posits that performing a behavior is a function of the extent to which someone is well informed about the behavior, motivated to perform the behavior (e.g., has positive personal beliefs and attitudes towards the behavior or outcome), and has the requisite skills to execute the behavior across various situations (Fisher et al., 2003). That is to say, individuals who are well-informed (information), motivated to act (motivation), and have

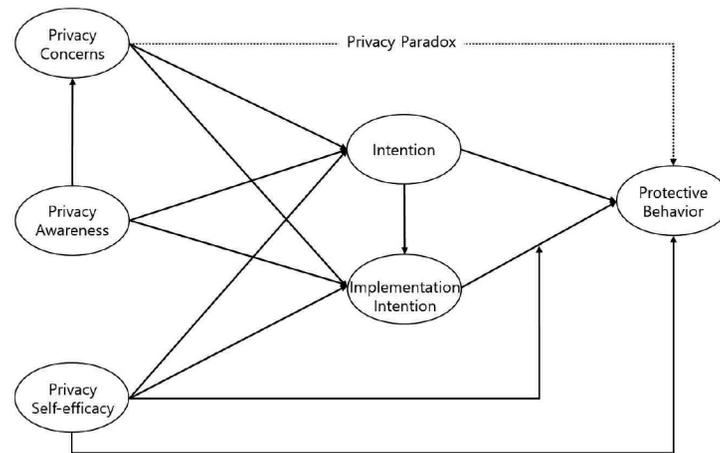
the necessary behavioral skills to act are more likely to act on the behavior.

III. Research Model and Hypotheses

1. Research Model

Targeted at COVID-19 contact tracing application users, this study aims to investigate these users' behaviors to protect their privacy through location settings. In specific, based on the information-motivation-behavioral (IMB) skills model, we designed a model (See <Figure 1>) to identify the factors that affect the actual use of location-protective settings and to confirm the existence of the privacy paradox in our research context. In addition, we hope to highlight the mediating role of implementation intention playing regarding the relationship between intention and behavior and to verify the moderating effects of privacy self-efficacy on the relationship between implementation intention and behavior in our research.

〈Figure 1〉 Research Model



References

Feyisa, H. L. (2020). The World Economy at COVID-19 quarantine: contemporary review. *International Journal of Economics, Finance and Management Sciences*, 8(2), 63–74.

Fisher, J. D., & Fisher, W. A. (2000). Theoretical approaches to individual-level change in HIV risk behavior. In *Handbook of HIV prevention* (pp. 3–55). Springer, Boston, MA.

Fisher, W. A., Fisher, J. D., & Harman, J. (2003). The information-motivation-behavioral skills model: A general social psychological approach to understanding and promoting health behavior. *Social psychological foundations of health and illness*, 22(4), 82–106.

Seto, E., Challa, P., & Ware, P. (2021). Adoption of COVID-19 contact tracing apps: a balance between privacy and effectiveness. *Journal of medical Internet research*, 23(3), e25726.

웨이블릿 변환과 멀티헤드 어텐션 기법을 적용한 주가 예측

권희석 (부산대학교 경영학과 석사과정, 주저자 mjs1995@naver.com)
이민혁 (부산대학교 경영학과 조교수, 교신저자 minhyuk.lee@pusan.ac.kr)

… Extended Abstract …

본 연구는 웨이블릿 변환으로 주가 데이터의 잡음을 제거하고 멀티헤드 어텐션(Multi-Head Attention) 기반의 LSTM(Long Short-Term Memory Model)을 이용하여 종가를 예측하는 WMA-LSTM 모델을 제안한다. 선행 연구에서는 웨이블릿을 통해 잡음이 제거된 주가 데이터를 가지고 LSTM에 Attention을 결합한 모델을 제안했고 LSTM 모델보다 좋은 성능을 보였다. 하지만 단일 Attention 층만을 사용하면 놓치는 데이터가 있을 수 있다. 이를 극복하기 위해 여러 Attention 층을 이용하면 학습에 사용할 데이터 감쇄를 최소화할 수 있고 예측 성능의 향상도 기대해 볼 수 있다. LSTM은 기존 RNN과 달리 이전 셀의 정보가 시그모이드 함수를 거치지 않고 다음 셀로 전달된다. 이전 상태 정보의 소실이 비교적 적게 일어나기 때문에 장기 의존성 문제를 개선한 모델이다. Attention 기법은 인간 시각의 주의 과정을 차용하여 전체 데이터에서 중요하다고 판단하는 부분을 중점적으로 학습한다. 데이터에서 중요한 부분을 판별하는 방법은 소프트맥스(Softmax) 함수이고 이것의 결과인 확률값이

은닉상태와 행렬곱의 연산을 거친다. Multi-Head Attention 은 여러 Attention 층이 병렬적으로 학습을 수행하고 결과를 병합(Concatenate)한다. 웨이블릿 변환(Wavelet Transform)은 주가 데이터의 잡음을 제거하기 위해 사용했다. 웨이블릿 분석은 영상 및 신호처리 분야에서 뛰어난 성과를 보였는데 주가 데이터도 일종의 신호로서 웨이블릿을 적용할 수 있다. 기존 푸리에 분석의 단점을 보완한 방법이고 다양한 시간 및 주파수 영역에서 금융 시계열 데이터를 분해하고 재구성할 수 있는 이점이 있다.

비교 모델은 웨이블릿 변환을 적용한 LSTM(이하 W-LSTM), 웨이블릿 변환과 Attention을 적용한 LSTM(이하 WA-LSTM)이 있다. W-LSTM은 웨이블릿 변환된 데이터를 단방향 LSTM의 입력값으로 한 모델이다. WA-LSTM은 웨이블릿 변환된 데이터를 입력값으로 하는 단방향 LSTM과 단일 Attention 층의 결합 모델이다. 제안 모델은 번역에 사용되는 Transformer의 일부인 Multi-Head Attention을 단방향 LSTM

과 연결된 모형이다. 여기서 4개의 Attention Head를 사용했고 각 Attention이 학습을 진행하고 병합하는 과정이 포함된다.

데이터는 Python의 yahoo_fin 라이브러리를 통해 yahoo finance에서 가져왔고 KOSPI, S&P500, HSI(INDEXHANGSENG, 항생지수) 세 가지 지수 데이터를 사용했다. 해당 데이터의 기간은 2012년 1월 2일부터 2022년 6월 30일까지 총 10년 7개월치를 사용했다. 연구에 사용한 환경과 언어는 각각 Jupyter hub, tensorflow 2.10.0과 Python이고 딥러닝을 구성하는데 keras 라이브러리를 사용하였다. 성능 지표는 MSE(Mean Squared Error), RMSE(Root Mean Squared Error), MAE(Mean Absolute Error)이며 값이 적을수록 뛰어난 성능을 의미한다. 연구의 진행 순서는 다음과 같다. 첫 번째, 전체 가격 데이터 중에서 예측할 종가를 제외하고 웨이블릿 변환을 통해 1%의 잡음을 제거했다. 웨이블릿 변환에 사용한 함수는 coif N 중

coif 3를 사용했다. 두 번째, 학습, 검증, 실험 데이터를 8:1:1의 비율로 순서대로 나눴다. 이때, 검증과 실험 데이터는 전체의 10%에 과거 20일이 더해진 형태이다. 그 이유는 과거 20일의 시가, 종가, 저가, 고가, 수정 종가, 거래량을 사용하여 그다음 날의 종가를 예측하는데 20일 치의 데이터 손실을 막기 위함이다. 세 번째, 입력 변수, 출력 변수의 범위를 조정하기 위하여 -1과 1 사이로 정규화를 시킨다. 네 번째, 전처리 과정을 모두 마친 후에 입력 변수를 제안 모형과 비교 모형에 넣어 학습시키고 도출된 결과 지표를 확인하여 성능을 검증한다.

실험 결과, 제안한 WMA-LSTM 모형이 비교 모형의 예측 결과보다 높은 예측 성과를 보였다. 본 연구는 기존 시계열 연구의 한계를 다중 Attention을 구성하여 극복할 수 있고, 실무에서도 같은 데이터를 서로 다르게 볼 수 있는 주안점들을 결합함으로써 더 나은 예측을 할 수 있음을 시사한다.

소비자의 디지털 그림자 노동 반응이 지속사용의도에 미치는 영향: 만족지연의 조절효과를 중심으로

LIU TINGTING (전남대학교 디지털미래융합서비스협동과정 박사, 주저자 qiuwentoto@naver.com)
이웅규 (대구대학교 경영학과 교수, 공저자 woong3041@empal.com)
고준 (전남대학교 경영학과 교수, 교신저자 kjooh@chonnam.ac.kr)

… Abstract …

Digital transformation and the COVID-19 pandemic have facilitated the rapid development and dissemination of non-face-to-face technologies such as SST(Self-service technology). According to the latest research, consumers should use digital shadow works to do self-service technology. Therefore, in connection with the actual situation, this study first examines whether consumers actively or passively select self-service technology on their response to digital shadow works. Whether changes in consumers' responses to digital shadow works according to SST usage conditions make a difference in consumers' continuous SST usage behaviors are examined. In particular, it was verified whether an individual's ability to delay gratification had a moderating effect in this process. This study verifies with SPSS,24 after data collection is completed using a survey method. The scope of research on digital shadow works has been expanded by empirically analyzing and clearly explaining the effect on consumers' responses to digital shadow works according to the SST usage situation of consumers. In practice, it can contribute to providing improvement measures to SST companies by examining the effects of differences in consumer reaction and individual satisfaction delay ability from the perspective of digital shadow works according to the situation of SST usage.

Key Words : Digital Shadow Works, Self-Service Technology, Satisfaction Delay Ability

I. 서론

셀프서비스기술(Self-service technology; SST)은 서비스 직원이 직접 참여하지 않고 고객이 기술적 화면을 통해 생산하는 서비스의

핵심 수단이다(Meuter et al., 2000). SST는 기업의 자원을 절약할 수 있는데 고객의 입장에서 SST 사용이 항상 이점만을 가져다주는 것은 아니다(박상철 외, 2020). 디지털 환경에서의 그림자노동은 키오스크와 같은 기

술을 바탕으로 소비자인 사용자가 직접 생산, 판매, 결제를 하며 이때 소비자가 보상 없이도 꼭 해야 하는 노동이다(이용규, 2019). 최근 SST에 대한 연구의 주요한 이슈는 SST 시스템 특징, 채택, 수용을 둘러싼 문제들이었다. 기존 연구들(전효아 외, 2013; 이지원, 2016)에 따르면, SST에 대한 능동적인 사용 과정에서 관계적 혜택, 자기 효능감은 소비자의 반응에 영향을 끼친다. 특히 코로나 시대에서는 SST를 통해 기업과 사회의 피해를 줄이고 고객 만족을 이끌어낸 것은 사실이다. 예를 들어, COVID-19이 자신의 안전을 위협하는 사회 불안요인 속에서, 직원의 능력이 부족한 상태에서 서비스가 실패하면 소비자는 회사에 대해 강한 부정적 반응을 보일 수 있다(김상희, 2012).

한편, 디지털 그림자노동과 관련해서 SST에 의한 소비자의 소비행위의 변화, 노동의 전가 등에 대한 연구는 일부 있었으나, 디지털 그림자노동의 긍정적 영향에 대한 연구는 전무하였다.

심리학 연구에서 만족지연능력은 나중에 더 큰 보상을 받기 위해 작은 보상을 즉시 포기하는 능력이며(Fujita, 2011; Michelle et al., 2011), 자기 조절적 적응력으로 정의된다. 본 연구는 만족지연능력이 본 연구의 맥락에서 디지털 그림자노동과 소비자의 지속적인 SST 사용행동에 영향을 미칠 가능성이 높다고 본다. 예를 들어, 소비자들은 앞으로 디지털 사회에 더 잘 적응하기 위해 현재의 직원 서비스를 포기하고 SST 사용을 배우는 것을 선택할 수 있다.

본 연구는 SST와 디지털 그림자노동 관련 연구를 바탕으로, 소비자들은 SST의 사용 상황에 따라(능동적, 수동적) 디지털 그림자노동에 대한 소비자의 반응 변화를 고찰하고자 한

다. 또한 디지털 그림자노동과 SST 지속사용 의도의 관계에 대해, 디지털 그림자노동의 부정적인 영향에만 초점 둔 기존 연구와 달리, 본 연구는 심리학에서 제시한 만족지연능력 변수를 도입하여 디지털 그림자노동이 가져올 수 있는 긍정적 현상에 대해 살펴볼 것이다. 본 연구는 다음의 연구 목적을 가진다.

1. 셀프서비스기술 사용에 대한 소비자의 능동적과 수동적 선택이 소비자의 디지털 그림자노동에 반응에 어떻게 영향을 미치는지를 고찰한다.
2. 디지털 그림자노동에 대한 소비자의 반응이 의사결정 행동에 미치는 영향을 분석한다.
3. 소비자의 만족지연 능력이 디지털그림자노동 반응과 지속사용의도 간의 관계를 조절하는지를 실증한다.

II. 문헌고찰과 가설

1. DSW 반응에 대한 영향요인

그림자노동은 오스트리아 철학자Ivan(1981)에 의해 최초로 제안된 개념으로, 보상을 받을 수 없는 노동(unpaid labor)'으로 정의하였다. Lambert(2015)는 사회현상에 대해 그림자노동의 개념을 확장하여 최근 기술의 발전에 따라 인간이 떠맡게 되는 다양한 종류의 추가적인 노동을 설명하고자 하였다. 디지털 기술에 발전을 따라 그림자노동은 SST와 밀접한 관련이 있다. 이용규(2021)는 디지털 환경에서 제공되는 아무런 대가없는 서비스에 수반되는 그림자노동을 디지털그림자노동(digital shadow work, 이하 DSW)로 명명하였다.

DSW는 SST를 포함해 현대 사회에서의 자동화 시스템 도입에 따른 기업의 소비자에 대한

노동 전가 현상을 설명하는 개념이다. 박상철 외(2020)는 맥도날드제이션의 관점에서 SST 사용자의 그림자 노동의 형성과정에 대해 설명하면서 디지털 기술 환경에서 그림자노동이 발생하고 전개되는 과정을 규명하였다. 유티팅과 고준(2021)은 무인마트 환경에서 소비자의 그림자노동에 대한 부정적 태도에 영향을 미치는 요인을 연구하였다.

하지만, 기존 연구들은 다양한 맥락에서의 DSW와 행동적 반응에 대하여 세밀하게 연구하지 못한 한계가 있다.

본 연구는 기존 연구의 한계점을 극복하기 위해 소비자의 SST 사용 상황에 대한 포괄적 고찰을 통하여 SST 사용 상황 관련 영향요인들을 도출하고 이들 변수간의 관계를 규명하고자 한다.

Peter(1980)는 지각된 혜택이란 제품을 사용하면서 발생하는 고객들이 원하는 주관적인 보상 또는 기대라 설명하였다. 전효아 외(2013)는 관계혜택(정보적·심리적·고객화혜택) 중 정보적·심리적 혜택이 기업 태도에 긍정적인 영향을 미친다고 보았다. 유티팅과 고준(2021)의 연구에 따르면, 심리와 경제적 혜택은 소비자들의 그림자노동에 대한 부정적 반응에 영향을 미칠 수 있다. 따라서 소비자들은 SST 이용했을 때 지각된 혜택을 인지하면 SST를 적극 선택할 것이고 DSW에 대한 지각이 감소될 것이다. 이에 가설 H1을 도출하였다.

H1: 지각된 혜택은 디지털 그림자노동에 대한 반응에 부(-)의 영향을 미칠 것이다.

사회인지 이론에서는 자기효능감은 인간의 행동을 결정하는 주요 인지적 개념 중 하나로 간주한다(Bandura, 1999). 자기효능감은 업

무 수행 능력에 대한 개인의 믿음으로 정의할 수 있는데(Gist, 1987), 역할수행에 있어 핵심 요인으로 작용한다. SST 관련 선행연구에서도, 자기효능감은 SST의 수용에 밀접하게 관련된 소비자의 개인적 특성 중의 하나로 밝혀졌다(Blut et al., 2016). SST 환경에서 자기효능감이 높은 소비자일수록, SST의 유용성, 용이성, 행동통제력이 높다고 인식하였으며 이는 결과적으로 사용의도와 사용행동을 증가시킨 것으로 나타났다(Blut et al., 2016). 따라서 소비자들은 자기효능감 높을 때 SST를 능동적으로 사용하며 DSW를 수행하므로 다음의 가설 H2를 설정하였다.

H2: 자기효능감은 디지털 그림자노동에 대한 반응에 부(-)의 영향을 미칠 것이다.

고객 의사결정과정에서 외적 요인과 내적 요인은 행동의도에 영향을 미치는 중요한 요인으로(Moore et al., 1991), 제품이나 서비스에 대한 고객만족과 평가도 의사결정에 영향을 미칠 것이다(Fornell et al., 1996). 따라서 유쾌하고 즐거운 서비스접점의 창출은 고객반응에 결정적인 역할을 한다(Gabbott et al., 2000). 이처럼 서비스 접점이 고객반응의 형성에 매우 중요한 역할을 하지만 실제 서비스접점에서 서비스 제공자는 항상 성공적인 서비스만을 제공할 수는 없다. 즉 서비스 실패가 유발될 수도 있다. 이러한 실패는 고객의 감정적 반응에 영향을 미치고 고객 이탈, 부정적 구전 등과 같은 행동적 반응을 유발한다(Tax et al., 1998). 따라서 소비자들은 외부요인에 의해 SST를 사용하다면 부정적인 반응이 생길 수 있으므로 다음의 가설 H3을 제시한다.

H3: 판매원의 서비스 실패는 디지털 그림자 노동에 대한 반응에 부(-)의 영향을 미칠 것이다.

사회 불안이란 사회적 상호작용에서의 공포 및 회피가 역기능적인 신념에서 비롯된다(Clark et al., 1995). 사회불안은 대면 서비스에서 느끼는 혼잡과 같은 상황적 환경에서도 발생할 수 있다. 소비자는 SST 환경에서 혼잡을 감지하면 사회불안과 맞서기 위해 어쩔 수 없이 반응한다. 최근에 COVID-19의 발생으로 사회적 제한과 봉쇄 조치가 실시되었다. 사회불안은 사람들로 하여금 사회적 상황을 피하거나 안전한 행동을 하도록 동기를 부여한다(Narendra et al., 2022). 본 연구는 외부환경에 의한 사회불안으로 인하여 SST를 사용한다고 보고 다음의 가설 H4를 제시한다.

H4: 사회불안은 디지털 그림자노동에 대한 반응에 부(-)의 영향을 미칠 것이다.

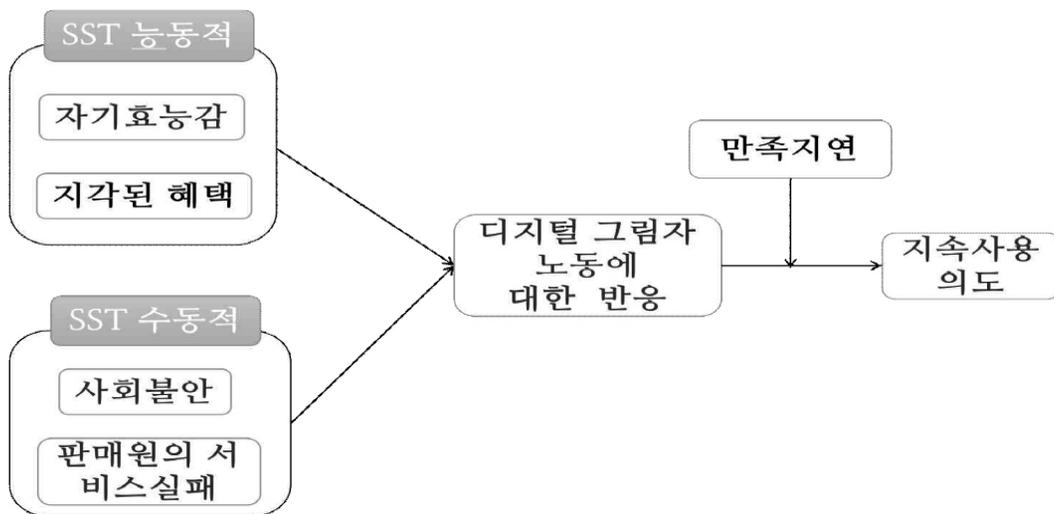
2. DSW 반응과 지속사용의도

지속사용의도는 정보시스템 분야에서 매우 중요한 의미가 가지고 있다. 기술이나 시스템이 아무리 우수하더라도 소비자들이 사용하지 않으면 기술이나 시스템이 효과가 없다. 다시 말해 소비자가 기술이나 시스템에 대해 만족하여 지속적으로 사용하여야 정보시스템이 지속적으로 발전할 수 있다(Bhattacharjee, 2001).

이용자 만족과 마찬가지로 지속사용 의도는 이용자와 기업 간의 지속적인 관계를 유지하기 위한 핵심 개념으로 받아들여지고 있다(서우중, 2009). 소비자들은 이전의 경험으로 기대 목표 이득을 얻게 되면 향후 기대에 의해서 이 서비스 사용의도가 높을 것이다. 따라서 다음의 가설5를 제시한다.

H5: 디지털그림자노동에 대한 반응은 셀프서비스기술 지속사용의도에 부(-)의 영향을 미칠 것이다.

〈그림 1〉 연구 모형



3. 만족지연의 조절효과

만족지연능력은 나중에 더 큰 보상을 받기 위해 현재의 작은 보상을 포기할 수 있는 능력과 성향(Fujita, 2011)이다. Mischel(1981)은 만족지연 행동과 사회적 책임성의 연구에서 만족지연능력은 사회적 책임성의 기본요소가 된다고 하였다. Funder 외(1983)도 이러한 주장을 뒷받침하고 있는데, 만족지연능력과 환경에 융통성 있게 적응할 수 있는 자아탄성, 자아통제 능력과의 관련성을 종단적으로 연구하였다. 한편, 자아통제능력에서의 결핍은 범죄행위나 반사회적인 행위로 이어질 수 있다. Rose(1991)는 만족지연능력의 결핍은 개인에게 실패 경험을 계속 겪게 하고, 반사회적 행동, 정신병리, 미성숙 등 사회적 적응에 문제를 가져올 수 있다고 주장하였다. (Mischel, 1981). 따라서 만족지연능력에 따라서 DSW에 대한 반응의 지속사용의도는 달라질 것이다.

H6: 디지털그림자노동에 대한 반응의 지속사용의도에 미치는 영향은 만족지연능력에 따라서 달라질 것이다.

〈그림 1〉은 본 연구의 연구모형이다.

Ⅲ. 실증 분석

본 연구는 최근 SST의 실제적 사용경험을 있고 DSW 인식한 사람을 대상으로 온라인 설문조사를 실시하였고, 수집된 469부 중에서 450부가 분석에 이용되었다. 각 항목별 질문은 리커트 5점 척도로 구성되었다(1: 전혀 그렇지 않다, 2: 그렇지 않다, 3: 보통이다, 4: 그렇다, 5: 매우 그렇다). 변수에 대한 측정항

목들은 〈부록 1〉에 제시하였다.

우선 측정변수에 대한 신뢰도 분석을 실시하였다. Cronbach's α 값이 모두 0.7보다 크게 나타나 측정도구의 신뢰성이 확보되었다(부록 1 참조). 본 연구에 사용된 변수에 대한 측정의 구조성은 탐색적 요인 구현의 정확성과 적절성을 가지고 측정하였다. 탐색적 요인 분석에서는 정보 손실을 최소화하기 위하여 주성분 분석을 수행하며, 분석 결과는 도출된 7개 요인에 따른 기존 변수들에 대한 전체 설명력은 74.24%로 나타났다. 한편, 자기응답방식의 설문에서는 동일방법편의(ommon method bias; CMB) 문제가 발생할 수 있는데(Podaskoff and Organ, 1986), Harman's 단일요인 분석(one factor test)을 실시한 결과, CMB 문제는 없는 것으로 나타났다. 독립변수와 종속변수의 관계를 실증하고자 회귀분석을 실시하였다. 가설 검증 결과는 다음의 〈표 1〉에 제시되어 있는데, 가설 1과 가설4는 기각되었고 가설2와 가설3은 지지되었다. 또한 가설5은 지지되었고, 조절효과 가설인 가설6도 지지되었다. 가설6의 경우 상호작용효과(디지털 그림자노동에 대한 반응x만족지연) 계수는 0.192로 나타났고($p < 0.01$), 수정된 R 제곱값은 0.112로 상호작용계수 추가에 따른 모형의 증가된 설명력이 유의하여 가설6이 채택되었다.

IV. 연구결과와 토의

본 연구는 SST 열풍이 거세지고 있는 상황에서 효과적인 소비자 행동을 파악하기 위해 SST 사용상황의 영향요인에 따른 DSW에 대한 반응을 실증하였다는 점에서 의미를 지니고 있다. 본 연구결과는 소비자의 의사결정에 있어서 환경의 영향보다는 SST가 소비자

자신한테 가져다줄 수 있는 혜택이 무엇인지를 가장 중요시 생각하고 있음을 시사한다. 즉 자신한테 이득이라고 생각하면 적극적으로 수용하는 반면에 이득이 되지 않는 경우에 받아들이지 않을 것이다. 코로나와 같은 환경적 변화가 SST의 발전에 영향을 미칠 수 있지만, 소비자의 SST 사용에 대한 설명에는 SST가 소비자에게 줄 수 있는 가치나 혜택이 가장 중요하다. 또한 본 연구에서는 만족지연능력의 조절효과를 검증하였다.

연구결과에 따르면, DSW에 대한 반응이 SST 지속사용의도에 미치는 영향은 소비자의 만족지연능력에 따라 달라질 수 있었다. 즉 만족지연능력은 DSW의 부정적 효과를 완화시킬 수 있다.

소비자가 능동적인 상황에서 지각된 혜택이 DSW에 부(-)의 영향을 미치는 것으로 나타

난 본 연구의 결과는 선행연구 결과와 일치하였다. 수동적인 상황에서 서비스 직원의 서비스 실패도 소비자의 DSW에 대한 반응에 유의하였다. 따라서 기업은 서비스 인력의 보완 노력을 간과해서는 안 될 것임을 시사한다. 학술적 관점에서 본 연구는 디지털 정보시스템 분야의 새로운 개념으로서 DSW를 만족지연능력과 같은 심리학 변수와 결합하여 SST와 DSW의 관계에 대한 연구영역을 넓힌 기여가 있다.

향후에는 설문조사가 아닌 실험을 통해 연구결과를 도출할 필요가 있으며, 표본 수집의 타당성이 확보되어야 할 것이다. 또한 DSW의 유형에 따른 효과를 살펴보는 것도 의미가 있을 것이다. 또한 만족지연능력의 조절효과에 대한 이론적 보완이 필요하다.

〈표 1〉 회귀 분석 결과

종속변수	독립변수	비표준화 계수		표준 계수	t	유의 확률
		B	표준 오차	베타		
디지털 그림자 노동에 대한 반응	(상수)	4.825	0.279		17.324	0.000
	자기 효능감	-0.048	0.045	-0.047	-1.066	0.287
	지각된 혜택	-0.415	0.059	-0.36	-7.064	0.000
	판매원의서비스실패	-0.159	0.049	-0.157	-3.235	0.001
	사회불안	0.05	0.054	0.038	0.921	0.358
R제곱 = .235				F = 34.582		
수정 R제곱 = .228				유의확률 = 0.000		
지속사용의도	(상수)	4.378	0.118		37.017	0.000
	디지털그림자노동에 대한반응	-0.215	0.039	-0.252	-5.541	0.000
R제곱 = .063				F = 30.700		
수정된 R제곱 = .061				유의확률 = 0.000		

〈부록 1〉 변수의 측정항목

변수	조작적 정의와 측정항목	관련 문헌	Cronbach's α
지각된 혜택	나는 셀프서비스기술이 편리하다고 생각한다.	Peter, 1980 ; 전효아 외, 2013	0.873
	나는 셀프서비스기술의 이용이 직원 서비스보다 기다린 시간을 줄여 준다고 생각한다.		
	나는 셀프서비스기술의 안전성이 높다고 생각한다.		
	나는 셀프서비스기술의 접근성이 좋다고 생각한다.		
	나는 셀프서비스기술이 무차별 서비스를 받을 수 있다고 생각한다.		
자기 효능감	나는 셀프서비스기술을 사용하고 이해하는 능력이 있다.	Bandura, 1999; Gist, 1987; Blut et al., 2016	0.926
	나는 셀프서비스기술 사용법을 다른 사람에게 전달할 수 있는 능력이 있다.		
	셀프서비스기술 화면에서 나는 필요로 하는 것을 찾을 수 있는 능력이 있다.		
	나는 전반적으로 셀프서비스기술을 이용하는 방법에 대해 알고 있다.		
판매원의 서비스 실패	판매원의 기술이 부족했기 때문에 나는 셀프서비스기술의 이용을 선택한다.	Moore et al., 1991; Fornell et al., 1996; Gabbott et al., 2000; Tax et al., 1998	0.892
	판매원이 집중하지 않았기 때문에 나는 셀프서비스기술을 사용하지 않게 된다.		
	판매원이 나에게 제공한 서비스 품질이 낮기 때문에 셀프서비스기술을 사용한다.		
사회 불안	나는 코로나를 피하기 위해 자주 사람들로부터 멀리 떨어져 있고 싶어진다.	Clark et al., 1995; Narendra et al., 2022	0.774
	나는 평소에 접하지 않았던 코로나 상황에서 느끼는 회피적 대응과 불안감이 있다.		
	외부환경이 나에게 대해 내리는 부정적 위험에 대하여 두려워한다.		
디지털 그림자 노동에 대한 반응	기업 또는 직원을 대신해서 내가 하는 귀찮은 일이다.	Ivan, 1981; Lambert, 2015; 이웅규, 2021; 박상철 외, 2020; 유틀링과 고준, 2021	0.944
	나는 디지털 그림자 노동하는 것이 개인적으로 많은 시간을 점유하고 있다고 생각한다.		
	나는 디지털 그림자 노동을 하는 것이 부담된다.		
	디지털 그림자 노동을 하는 것은 나에게 추가적인 노력과 시간을 요구한다.		
만족 지연	당장 셀프서비스기술을 사용하기 불편하더라도 장기적 입장에서 인내심을 가지고 셀프서비스기술을 사용한다.	Fujita, 2011; Mischel, 1981; Funder et al., 1983; Rose, 1991	0.901
	셀프서비스기술을 잘 사용을 위해서 더 많은 노력이 필요하다더라도 열심히 배울 용의가 있다.		
	셀프서비스기술이 사용에 스트레스 받더라도 이를 견디고 계속 배울 의지가 있다.		
	셀프서비스기술을 사용을 직원 서비스보다 불편하더라도 더 큰 이익을 위하여 현재 불편을 참고 계속 셀프서비스기술을 사용한다.		
	현재 셀프서비스기술 사용은 힘들어도 장기적인 편의를 위해 참아 다.		
지속 사용 의도	나는 셀프서비스기술을 중단하지 않고 반복적으로 사용할 것이다.	Bhattacharjee, 2001; 서우중, 2009	0.935
	나는 앞으로도 셀프서비스기술을 계속 이용할 의향이 있다.		
	가능하다면 나는 셀프서비스기술을 지속적으로 사용할 것이다.		
	나는 다른 방법보다 셀프서비스기술을 지속적으로 사용할 마음이 있다.		
	나는 셀프서비스기술을 포기하지 않고 잘 사용할 의지가 있다.		

참고 문헌

김상희, “모든 고객은 판매원의 서비스 실패에 대해 분노하는가? 판매원의 서비스 실패에 대한 고객귀인이 고객의 감정적 반응과 행동적 반응에 미치는 영향,” 마케팅연구, 제27권, 제2호, 2012, pp. 27-52.

박상철, 이웅규, 고준, 류성열, “디지털 기술 환경에서의 그림자 노동 메커니즘 규명 연구”, 경영학연구, 제49권, 제1호, 2020, pp. 31-50.

박상철, 류성열, “맥도날디제이션 관점에서 본 SST 사용자의 그림자 노동”, 인터넷전자상거래연구, 제20권, 제1호, 2020, pp. 99-115.

박상철, “그림자노동 차원에서 본 IT 사용자의 비밀번호 변경행동 연구”, 정보시스템연구, 제28권, 제2호, 2020, pp. 93-107.

이반 일리치 지음, 박홍규 옮김, “그림자 노동”, 2005, 서울 미토.

*기타 참고문헌리스트는 지면관계상 생략하였습니다.

휴대용 디지털 기기 액세서리 구매과정에서의 디지털그림자노동 역할 : 아이패드 사례를 중심으로

이웅규 (대구대학교 경영학부 교수 주저자 woong3041@empal.com)
조아라 (대구대학교 경영학부 박사과정 교신저자 ara2147483647@gmail.com)

… Abstract …

Digital device accessories are a kind of optional supplementary material necessary for management rather than directly affecting the actual function or operation or performance improvement. Nevertheless, as the use of portable digital devices increases, the demand for various accessories such as covers, pouches, and liquid crystal films is steadily increasing. This study looks at the purchasing process of these digital device accessories from the perspective of digital shadow work. By adopting grounded theory methodology, total 24 iPad users were interviewed. Based on open coding and axis coding, we derived 20 codes and 7 classes and suggested a model for digital device accessory purchasing process. This study found that purchasing digital device accessories is outsourcing to avoid digital shadow work and triggered another digital shadow work.

Key Words : Digital Shadow Work, Digital Device Accessory, iPad users

I. 서론

휴대용 디지털기기가 보편화되면서 해당 기기의 액세서리 사용도 더불어 늘어나고 있다. 경우에 따라서는 디지털기기 못지 않게 액세서리 구매에 관심을 사용자가 늘어가고 있고 액세서리 구매에 적지 않은 투자를 하고 있다.

실제 디지털기기 액세서리 시장은 꾸준히 성장할 것이라는 것이 일반적인 견해다. 가령, Mordor Intelligence 같은 컨설팅 회사에서는 2027년까지 모바일장비 커버시장의 연평균 성장률을 6.13%로 보고 있고 그 가운데에서도 아시아태평양 지역은 가장 크면서도 가장 빠르게 성장하는 시장으로 보고 있다.¹⁾ 따라서 액세서리의 수용 또는 구매 과정의 이

1) 참조; <https://www.mordorintelligence.com/industry-reports/mobile-protective-cases-market>

해는 오늘날 정보기술 사용의 기본이 되는 태블릿이나 스마트폰의 사용 행태를 이해하는 새로운 단초가 될 수 있을 것이다. 하지만 그림에도 불구하고 액세서리에 대한 정보기술 관점에서의 연구나 조사는 그리 많지 않다. 디지털기기의 액세서리는 다음과 같은 두가지 특성을 가지고 있다. 첫째, 필수적이지는 않지만 디지털기기의 기능을 보완하거나 강화시킨다. 액세서리가 없다고 해서 디지털기기의 성능이나 기능이 현격하게 달라지지는 않지만 액세서리가 주어지면 사용 또는 관리가 좀 더 용이해진다. 태블릿의 커버가 주어지면 태블릿의 안전관리가 용이해질 것이고 키보드가 주어지면 입력이 상대적으로 편리해진다. 다시 말해 액세서리는 디지털기기 사용자들의 디지털그림자노동(digital shadow work, 이하 DSW)을 경감시키는 역할을 한다. DSW는 디지털서비스 사용자가 아무런 댓가 없이 수행하고 있는 보이지 않는 작업이다(이웅규, 2021). 가령 커버나 파우치가 없다면 사용자는 항상 기기의 파손을 신경써야 하고 키보드가 없다면 입력의 불편을 감수해야 한다. 둘째, 액세서리 구매를 위해서 쉽지 않은 의사결정이 필요할 수 있다. 비록 보완적 기능을 가지고 있는 동일한 액세서리라 하여도 다양한 모델이 있고 서로 다른 다양한 기능을 가지고 있다. 가령, 어떤 커버는 단순히 기기를 보호하는 역할만 하지만 어떤 커버는 전자펜을 잘 관리할 수 있는 기능이 있다. 따라서 사용자는 액세서리 구매에 있어 다양한 선택의 폭을 가지고 있을 뿐 아니라 신중한 선택을 필요로 한다. 즉 고관여도(high involved) 상품일 수 있다. 본 연구의 목적은 디지털기기 액세서리의 구매과정을 통제하는 심리적 기제를 DSW의 관점에서 규명하는 것이다. 디지털기기 액세서

리 구매과정에서 사용자들은 DSW를 어떻게 인식하고 어떻게 관리하고 또 어떻게 발생하는지를 살펴 본다. 이를 위해 근거이론방법론(grounded theory methodology)에 의한 질적 접근법에 의해 디지털기기 액세서리 구매과정을 분석하고 그 결과를 기반으로 DSW와의 이론적 연관성을 도출한다.

II. 선행 연구

1. 디지털기기

정보기술 사용 분야에서 디지털기기에 대한 연구는 주로 기기의 수용이나 구매 또는 사용을 중심으로 비교적 활발히 이루어져 있다. 전통적인 기술수용모델(technology acceptance model)(Davis et al., 1989)을 기본 틀로 하여 PC나 스마트폰 구매 또는 채택을 조망하고 있다(예; Tzou and Lu, 2007; 이웅규, 박진훈, 2021; Yoo et. al., 2021). 하지만 기존 연구들의 초점은 디지털기기 자체에 두고 있는데 비해 디지털기기 액세서리에 관심을 둔 연구는 찾아보기 쉽지 않다.

2. 디지털그림자노동

디지털그림자노동(DSW)의 이론적 배경인 그림자노동(shadow work)은 보이지도 않고 인식되지 않는 노동을 의미한다(이웅규, 2021; Illich, 1981). 원래 자본주의의 모순을 비판하기 위해 제안된 개념이지만 디지털환경에서 그림자노동은 자본주의 비판의 관점이 아니더라도 항상 존재할 수 밖에 없다(이웅규, 조아라, 2022). 오늘날 DSW는 ‘디지털 기술을 이용한 사람들이 당연히 해야 하는 숨겨진

노력' 으로 '디지털 서비스에 수반되는 그림자노동' 을 의미한다(이용규, 조아라, 2022; 박상철, 이용규, 2019).

정보시스템 이론의 관점에서 DSW에 관한 연구는 양적으로는 부족하지만 의미있는 시도가 있었다(예; 이용규, 조아라, 2022; 이용규, 2021; 박상철, 류성열, 2020; 박상철 등 2020; 박상철, 2019; 박상철, 이용규, 2019). 이들 연구의 대부분은 디지털서비스에서 발생하는 DSW 역할과 영향 등을 다루고 있다. 하지만 디지털기기 액세서리는 물론이고 하드웨어인 디지털기기의 관점에서 DSW를 조망한 연구는 아직까지 없었다.

III. 연구방법론

근거이론 방법론(grounded theory methodology)에 의한 질적접근법을 채택하였다. 근거이론방법론은 데이터와 관찰에 의한 객관적인 실증주의(positivism)와 직관과 해석에 의한 주관적인 해석주의(interpretivism)를 동시에 추구하는 질적 연구방법론이다(이용규, 2020).

근거이론방법론은 자료 검토를 통해 현상을 발견하고 이를 범주화하는 개방코딩(open coding)과 개방코딩에 발견한 범주 간의 관계를 규명하는 축코딩(axial coding), 그리고 개방코딩과 축코딩 결과로부터 핵심이 되는 범주들을 정교화하는 선택코딩(selective coding) 그리고 선택코딩을 통해 얻어진 개념을 비교분석을 위해 타 이론과의 통합을 시도

하는 이론코딩(theoretical coding)과정 등으로 이루어진다(이용규, 2020; Strauss and Corbin, 1998).

본 연구의 대상은 아이패드 사용자의 액세서리 구매과정이다. 이를 위해 아이패드 사용자와의 인터뷰를 통해 얻어진 자료를 개방코딩과 축코딩에 의해 정리하고 구매과정을 모형화한다. 그리고 그 결과를 토대로 DSW의 관점에서 핵심범주를 찾아내고 핵심범주의 의미를 DSW 이론에 의해 해석한다.

본 연구에서는 아이패드 사용자들을 대상으로 총 3차에 걸친 인터뷰를 실시하였다. 1차 인터뷰에서는 5명의 사용자들에게 아이패드 액세서리 전반에 걸친 사용 행태를 얻었다. 2차 인터뷰에서는 1차 인터뷰를 기반으로 10명의 사용자를 대상으로 아이패드 액세서리 구매의 일반적인 행태를 분석하였다. 최종적으로 3차 인터뷰에서는 연구의 대상을 아이패드 커버에 중점을 두고 9명의 사용자들을 대상으로 인터뷰하였다.

IV. 연구결과

1. 개방코딩

개방코딩은 1차, 2차 인터뷰를 참조하여 3차 인터뷰를 토대로 대규모의 코드를 정리하는 방식으로 실시하였다. <표 1>에서 보는 바와 같이 총 20개의 코드가 도출되었고 각 코드를 검토한 결과 7개의 범주로 정리되었다.

〈표 1〉 개방코딩 결과

범주	코드	인터뷰대상자 답변예	대상자
기기에 대한 관심	기기 가격	일단 이게 제가 거의 200만원 가까이 주고 샀기 때문에 파손이 될까 봐서 무섭기도 하고	3-3
	파손 우려	아이패드 자체가 좀 잘 깨질 것 같다는 그런 인식이 있어서...	3-2
	기기의 외적 미관	그냥 아이패드 순정 자체가 되게 예쁘단 말이에요...	3-5
기기관리의 어려움	안전관리	밖에서는 조금 그래도 위험하니까 떨어뜨릴 수도 있고 그렇잖아요..	3-1
	오염 방지	오래 쓰고 싶고 기스 남기는 게 싫어가지고 ..	3-8
	고가의 수리비	이게 어쨌든 떨어뜨리면 수리비가 굉장히 많이 나오고, ... 파손되면은 원가에 반절 정도를 수리비로 내야 된다..	3-3
기기 기능 확대	활용성 확대	... 세워서 보고 싶을 때는 커버가 있어야지 좀 세워지거든요.	3-2
	타 액세서리 관리	... 케이스 옆에 애플 펜슬을 꽂을 수 있는 그런 이제 공간이 필요하다고 생각을 하고..	3-5
구매시 고려사항	가격	학교 입학할 때 돈이 너무 많이 들어가지고 아무래도 디자인이나 회사보다는 가격이 저렴한데	3-8
	내구성	커버를 살 때 가장 중요하게 생각하는 거는 저는 튼튼함이라고 생각해요.	3-5
	디자인	디자인 같은 경우에는 제가 약간 색깔 정도하고, 뒤에 이제 애플 로고가 보이는 거...	3-7
	기능	저는 그 위 앞면 말고, 옆에 끼우는 게 없었으면 좋겠어 가지고 폴리오 케이스를 구매했어요.	3-6
액세서리 구매	온라인 중심 구매	..공식 홈페이지에 들어가서 어떤 제품들이 있는지를 보다가.... 그걸 필터링을 해두고,...제품 하나하나 들어가서 리뷰를 살펴보고, 또 리뷰가 괜찮은 것들 중에 ... 골랐어요.	3-3
	크로스오버 쇼핑(온오프라인 병행 구매)	.. 실용적인 것들을 오프라인에서는 직접 더 많이 느껴보려고 하는 거였고, ...온라인에서는 디자인적인 걸 더 많이 봤던 것 같아요.	3-5
구매경험 결과	구매과정의 즐거움	그냥 나름 재미있었어요.	3-3
	제품선택의 피로도	근데 아무래도 태블릿 PC는 그렇게 비싼 거는 아니기 때문에 ... 점점 지치는 건 있는 것 같아요.	3-2
	전문성 확보	이 제품 사용법에 대해서 이제 제가 잘 알고, 잘 아니까 그렇다고 생각해요.	3-6
	타인에게 구매 조언	내가 여기저기서 찾아봤는데, 나는 다 비슷한데 나는 여기에 색깔이랑 디자인이랑 이런 디테일로 디테일이랑 가격 이런 걸로 이거를 골랐다, 이렇게 했어요.	3-9
구매결과	구매 만족	그냥 색깔도 그냥 무난하게 잘 산 것 같고, 별로 안 질리고요. 그리고 어디 잠깐 챙겨 나갈 때 용도로는 좀 튼튼한 것 같아요.	3-1
	재구매	근데 요즘에는 조금 제가 시간이 있어서 그냥 시간 나는 대로 생각나는 대로 틈틈이 찾아보고 있어요.	3-8

1) 기기에 대한 관심

대부분의 인터뷰 대상자들은 상대적으로 고가로 지각될 수 밖에 없는 디지털기기의 가격과 노출이 많으면서도 늘 휴대해야 하는 속성 때문에 파손에 대한 우려를 가지고 있다. 더욱이 아이패드의 경우 외적 미관이 빼어나기 때문에 더 많은 관리의 필요성을 느끼고 있다.

2) 기기 관리의 어려움

기기에 대한 관심도가 높을수록 기기관리에 대한 어려움을 인식하고 있다. 무엇보다도 기기 파손에 대한 안전관리에 신경을 쓸 수 밖에 없고 기기에 흠이 가지 않고 깨끗이 쓰는데 신경이 쓰일 수 밖에 없기 때문이다. 특히 안전관리나 오염방지가 제대로 되지 않았을

때 기기 가격에 맞먹는 수리비에 대해서도 걱정을 하는 사용자들이 많았다.

3) 기기 기능 확대

기기에 대한 관심과는 별도로 아이패드가 가지고 있지 않은 기능들에 대해 관심을 가지고 있다. 가령, 세워서 볼 수 있는 기능이나 애플 펜슬을 보관할 수 있는 기능의 필요성을 지각하고 있다.

4) 액세서리 구매

액세서리 구매의 대부분은 온라인을 중심으로 이루어지고 있었다. 주로 액세서리 홈페이지 검색, 온라인상에서 상품리뷰 검토, 좋은 리뷰 상품 중심으로 상품 선택 등의 과정으로 구매가 이루어지고 있다.

사용자에 따라서는 크로스오버 쇼핑 즉, 오프라인을 동시에 이용하여 쇼핑을 하는 경우도 있었다. 이 때 오프라인에서는 주로 기능적인 관점을 주로 보았고 온라인에서는 디자인적인 관점을 주로 보고 있다.

5) 구매시 고려사항

구매시 고려사항은 가격, 내구성, 디자인, 기능 등이 있었다. 대부분의 사용자들은 온라인 검색 또는 오프라인에서의 경험을 통해 구매시 고려사항을 상대적으로 매우 상세히 검토하여 구매하고 있었다.

6) 구매경험 결과

구매경험은 어떤 사용자에게 있어서는 즐거움으로 지각되었다. 상품 자체를 검색하고 선택하

는 것이 즐거움으로 지각되고 있었다.

반면 상품 선택에 있어 피로도를 느끼는 경우도 있었다. 별로 비싸지도 않은데 여러 종류의 상품 가운데 선택해야 하는 것에 지치는 사용자들이 많았다. 하지만 이들도 상품 선택 이후에는 나름 즐거움을 느끼고 있었다.

구매과정에서 발생하는 상품 검색은 해당 액세서리에 대한 전문성을 높여 주었을 뿐 아니라 다른 사람들에게 구매를 조언해 주기도 하였다.

7) 구매 결과

대체적으로 심사숙고하여 구매하였기 때문에 자신이 구매한 상품에 만족하고 있었지만 경우에 따라서는 재구매 의도를 보여주고 있었다. 구매경험의 결과 얻어진 상품에 대한 지식이 재구매시 많은 도움을 줄 것으로 기대하고 있었다.

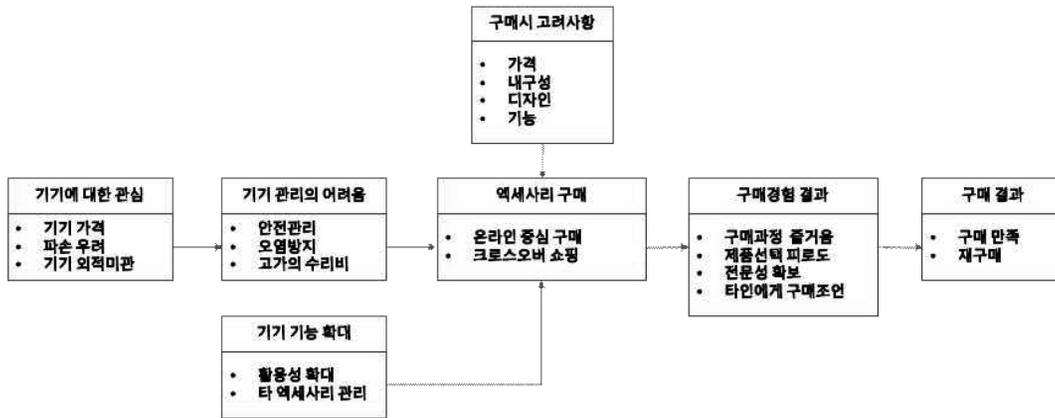
2. 축코딩

범주 간의 관계를 설정하는 축코딩을 위해 Strauss와 Corbin(1998)의 패러다임 모형을 본 연구에 맞도록 수정하였다. 축코딩의 결과는 <그림 1>에서 보는 바와 같다.

범주는 크게 자신이 가지고 있는 디지털기기에 인식(“기기에 대한 관심”, “기기관리의 어려움”, “기기 기능 확대”)과 구매과정에서 인식의 흐름(“구매시 고려사항”, “액세서리 구매”, “구매경험 결과”, “구매결과”)으로 나뉠 수 있다.

Strauss와 Corbin의 패러다임 모형에 의하면 축코딩은 인과적 조건-현상-전략-결과 그리고 현상에 영향을 미치는 맥락적 조건과 전략에 영향을 미치는 중재적 조건으로 나뉜다.

〈그림 1〉 축코딩 결과 - 수정된 패러다임 모형



기기에 대한 관심이 일종의 인과적 조건이다. 액세서리 구매의 가장 근본적인 원인은 기기 자체에 대한 관심 때문이다. 기기 자체의 관심은 기기 관리의 어려움이라는 현상을 낳게 되고 기기가 가지고 있는 부족한 기능을 확대하고 싶은 현상을 아울러 갖게 한다. 결국 사용자는 관리의 어려움과 기능 확대를 해결하기 위해 액세서리 구매라는 행동 즉, 전략을 취하게 된다. 이 때 구매시 고려사항은 액세서리 구매에 대한 중재조건의 역할을 한다. 액세서리는 구매경험을 통한 직접적인 경험과 구매 이후의 만족 또는 재구매를 유도한다.

불하면서 수리를 해야 하는 번거로움을 감수해야 하기 때문이다.

다시 말해 디지털기기 사용에서 발생할지도 모르는 여러 가지 DSW를 액세서리 구매를 통해 피하려는 것이다. 이응규(2021)의 분류에 따르면 ‘기기 관리의 어려움’은 ‘should’ 형이면서 관리지향인 코어형 DSW가 발생할 가능성이 높고 ‘기기 기능의 확대’는 ‘want’ 형이면서 운영지향인 루틴형 DSW를 발생시킬 수 있다. 따라서 액세서리의 구매는 사용자들이 시도하는 DSW의 아웃소싱(outsourcing)이라고 볼 수 있다.

3. 디지털기기 액세서리 구매와 DSW

1) DSW의 아웃소싱

본 연구에서 도출한 축코딩에 따르면 액세서리의 구매는 디지털기기에서 어려움을 겪을 것으로 예상되는 관리와 부족하다고 생각하는 기능을 액세서리 구매를 통해 대체하려는 것이다. 만일 액세서리가 없다면 사용자들은 파손이나 오염방지를 위해 지속적인 관리를 해야 하고 문제가 생긴다면 고가의 수리비를 지

2) 새로운 DSW의 생성

본 연구에서 제시한 바와 같이 대부분의 사용자는 디지털기기의 액세서리 상품의 구매 과정에 깊이 관여한다. 상품 검색은 물론이고 상품 리뷰를 검토한 후 자신에 맞는 상품을 고르는 것이 일반적인 과정이다. 그리고 이 과정은 많은 사용자들이 재미있어 한다. 무엇보다도 액세서리는 필수적이거나보다는 선택 사항이기 때문에 굳이 하지 않아도 되는 쇼핑 과정이다.

따라서 액세서리 구매 과정은 일종

의 ‘want’ 형이면서 운영지향인 퀘스트형 DSW로 볼 수 있다(이웅규, 2021). 특히 본 연구의 결과에 따르면 액세서리 구매과정은 본인의 의지와 상관없이 나름대로의 전문지식을 확보하는 경우가 많은 것으로 밝혀졌다. 전문성의 확보는 퀘스트형 DSW가 갖는 전형적인 특성이다(이웅규, 2021).

V. 결 론

본 논문의 목적은 휴대용 디지털기기의 액세서리 구매과정을 DSW 관점에서 분석하려는 데 있다. 이를 위해 근거이론방법론에 총 3차에 걸쳐 총 24명의 아이패드 사용자를 인터뷰하여 개방코딩과 축코딩에 의해 정리하였고 DSW의 관점에 의해 분석하였다. 그 결과 디지털기기 액세서리 구매과정을 수정된 패러다임 모형에 의해 모형화하였고 DSW의 아웃소싱과 새로운 DSW가 생성됨을 밝혔다.

본 연구의 이론적 시사점은 다음과 두 가지로 요약할 수 있다. 첫째, DSW에 대한 새로운 이해를 할 수 있게 되었다. 기존의 DSW 연구는 대부분의 디지털 서비스에 초점을 맞추고 있었지만 본 연구에서는 하드웨어 사용에서의 DSW를 다루었다. 특히 액세서리 구매 과정에서 DSW의 아웃소싱과 새로운 DSW의 생성을 규명한 것은 이론적으로 매우 흥미로운 발견이다.

둘째, 디지털기기 액세서리에 대한 이론적 의미를 부각시켰다. 대부분의 기존 이론은 디지털기기 사용이나 수용 또는 구매과정을 다루고 있다. 하지만 액세서리에 대한 관심은 상대적으로 부족했다. 본 연구에서는 디지털기기 액세서리가 정보시스템 사용 이론의 관점에서 볼 때 충분히 의미를 가지고 있을 뿐 아니라 새로운 영역으로 부각될 수 있음을 보였다.

실무적으로는 사용자가 디지털기기에 대해 지각하는 DSW의 의미가 무엇인지 보여 주었다. 근본적으로 사용자의 인식 여부와 상관없이 코어형과 같은 관리지향 DSW가 발생하는 것은 공급자에게 책임이 있다. 공급자가 디지털기기에 대한 관리를 좀더 용이하게 만들 경우 DSW는 그 만큼 경감하게 되고 사용자의 만족도는 증가할 것이다. 본 연구의 결과는 앞으로 디지털기기 공급자들이 DSW를 어떻게 관리해야 하는지를 시사하고 있다.

본 연구는 디지털기기 가운데 특정 브랜드를 갖는 태블릿에 초점을 맞추었고 또 태블릿의 커버에 중점을 두었다. 물론 연구의 결과물을 일반화하기 위해 노력했지만 한계가 있을 수 있다는 점을 인정할 수 밖에 없다. 스마트폰의 커버나 태블릿의 전자펜슬 등은 본 연구의 연구대상과는 다른 결과를 제공할 가능성이 있고 디지털기기 액세서리 또는 주변기기에 대한 이해의 폭을 넓혀 줄 것이다.

참고 문헌

박상철, “그림자노동 차원에서 본 IT 사용자의 비밀번호 변경행동 연구”, 정보시스템연구, 제28권 제2호, 2019, pp. 93-107.

박상철, 류성열, “맥도날드제이션 관점에서 본 SST 사용자의 그림자 노동”, 인터넷전자상거래연구, 제20권 제1호, 2020, pp. 99-115.

박상철, 이용규, “문헌 고찰을 위한 근거이론방법의 활용: 디지털 환경에서의 그림자 노동 개념 도출”, 지식경영연구, 제20권 제2호, 2019, pp. 183-195.

박상철, 이용규, 고준, 류성열, “디지털 기술 환경에서의 그림자 노동 메커니즘 규명 연구”, 경영학연구, 제49권 제1호, 2020, pp. 31-50.

이용규, “우리나라 정보시스템 연구에서 근거이론방법론 현황의 분석과 적용 지침 제시”, 정보시스템연구, 제29권 제1호, 2020, pp. 181-199.

이용규, “디지털그림자노동의 분류와 동태성 및 연구 방향,” 정보시스템연구, 제30권 제1호, 2021, pp.105-121.

이용규, 박진훈, “브랜드 애착과 심미성이 새로운 스마트폰에 대한 구매의도에 미치는 효과”, 정보시스템연구, 제23권, 제4호, 2014, pp. 147-168.

이용규, 조아라, “메이크업형 디지털그림자노동 사용을 지배하는 기제에 관한 연구: 프로필 사진 관리 사례”, 정보시스템연구, 제31권, 제3호, 2022, 게재예정.

Tzou, R-C., and Lu, H-P., “Exploring the Emotional, Aesthetic, and Ergonomic Facets of Innovative Product on Fashion Technology Acceptance Model”, Behaviour & Information Technology, Vol. 28, No. 4, 2009, pp. 311-322.

Yoo, J., Choi, S., Hwang, Y., and Mun, Y. Y., “The Role of User Reistance and Social Influences on the Adoption of Smartphone: Moderating Effect of Age”, Journal of Organizational and End User Computing, Vol. 33, Issue 2., 2021, pp. 36-55.

Strauss A. and Corbin, J., Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory, 2nd ed.Sage., 1998.

