

<http://dx.doi.org/10.17703/JCCT.2021.7.4.815>

JCCT 2021-11-100

빅데이터 분석기법을 활용한 숙박업체 운영 개선 방안에 대한 연구

A Study on Improvement of Pension Operation and Management using Big Data Analysis Techniques

윤선희*

Sunhee Yoon*

요약 빅데이터의 장점은 인터넷상의 대량의 데이터를 수집하여 가치 있는 데이터를 정제하여 사용하는 것이다. 즉, 비정형 데이터를 사용자가 필요한 관점에서 분석하여 활용할 수 있도록 가공하는 것이다. 본 논문은 실생활에 밀접하게 적용되어 마케팅에 활용할 수 있는 비정형 데이터를 기반으로 하며 실험 대상은 서울에서 한 시간 거리의 수도권에 있는 숙박업체를 모델로 하여 빅데이터를 사용자가 필요한 관점에서 분석하여 매출 증대, 비용 감소 및 수익률 증가 등의 효과를 나타낸 실험으로 소셜네트워크 등의 빅데이터를 분석하는 과정에서 입력되는 데이터가 숙박 정보로써 활용할 수 있는 데이터인지를 판별하여 필터링하는 시스템을 제안하여 숙박물의 향상 및 공실률을 감소시킬 수 있는 마케팅 전략을 구축하고자 한다.

주요어 : 빅데이터, 데이터마이닝, 비정형데이터, 필터링, 공실률

Abstract The advantage of big data is to collect a large amount of data on the Internet and refine and use valuable data. That is, the unstructured data is processed so that the user can analyze and utilize it from a necessary point of view. This paper is a relatively small project and is based on unstructured data that can be closely applied to real life and used for marketing. The subjects of the experiment were modeled on lodging companies in the Seoul metropolitan area an hour away from Seoul, and analyzed for the increase in lodging rates before and after marketing using big data. As an experiment that shows the effects of increasing sales, reducing costs, and increasing returns by users, we propose a system to determine and filter whether data input in the process of analyzing big data such as social networks can be used as accommodation-related information.

Key words : Big Data, Data Mining, Unstructured Data, Filtering, Vacant Ratio

1. 서론

정보통신의 발전으로 최근 실세계에서 스마트, 모바일, 소셜 네트워크는 밀접하게 상관되어졌다. 있다. 스마트한 정보기기들을 사용함에 따라 의료, 교육, 과학,

금융, 교통, 제조, 엔터테인먼트 등 사회 전 분야에서 다양한 유형의 대량의 데이터가 끊임없이 생성되어 사용되고 있다. 또한, 다양한 분야의 스마트한 정보기기들은 모바일 기술이 적용되어 언제 어디서, 무엇을 어떻게 인터넷에 연결되어 사용될 수 있는 유비쿼터스 환경

*정회원, 송의여자대학교 IT비즈니스과 부교수 (제1저자)
접수일: 2021년 10월 20일, 수정완료일: 2021년 10월 30일
게재확정일: 2021년 11월 4일

Received: October 20, 2021 / Revised: October 30, 2021
Accepted: November 4, 2021

*Corresponding Author: shyoon@sewc.ac.kr
Dept. of IT Business, SoongEui Women Univ, Korea

으로 발전되었으며 개인 또는 집단 간 상호의존적인 관계에 따라 소셜 네트워크라는 새로운 사회관계 구조가 형성되었다. 이러한 환경에서 사용되는 다양하고 거대한 정형 또는 비정형 데이터를 빅데이터라 하며 부피가 크고, 변화의 속도가 신속하며, 다양한 속성의 데이터로써, 규모(Volume), 변화속도(Velocity), 다양성(Variety), 즉 3V의 특성을 가진다. 최근 3V 이외에 불확실성(Veracity)를 포함하여 4V라고도 한다.

2018년 국내 빅데이터 시장은 전년 대비 50% 이상 성장한 것으로 나타났다[1]. 지난해 국내 빅데이터 시장 규모는 5843억원으로 전년 대비 28.5% 성장한 결과로 2015년 26343억원과 비교하면 50% 이상 증가한 추세다(그림 1).



그림 1. 국내 빅데이터 추세
Figure 1. Trend of Domestic Big Data

2017년과 2018년을 비교하면, 서비스와 소프트웨어의 비중은 증가했지만, 서버와 스토리지, 네트워크 등의 하드웨어는 감소한 것으로 나타났다. 서비스와 소프트웨어는 각각 3.0P와 2.4P 증가했으며 하드웨어는 5.4P 감소한 것으로 나타났다. 이는 빅데이터 관련 투자가 과거에는 하드웨어 중심으로 이루어졌던 것에서 빅데이터를 활용한 서비스, 이를 실행하기 위한 소프트웨어에 대한 투자로 변화가 이루어진 것을 의미한다(그림 2).

기업들이 빅데이터를 도입하는 핵심 이유 중 하나는 효과적인 의사결정으로 이를 통해 매출 증대, 비용 감소, 수익률 증가 등의 효과를 얻기 위해서이다. 본 논문은 실생활에 밀접하게 적용되어 마케팅에 활용할 수 있는 비정형 데이터를 기반으로 한다. 실험 대상은 서울에서 한 시간 거리의 수도권에 있는 숙박업체를 모델로 설정하여 빅데이터를 사용자가 필요한 관점에서 분석

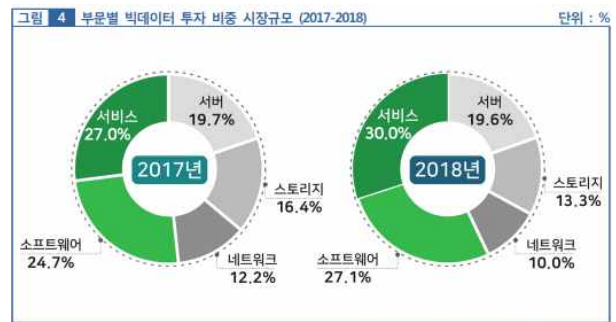


그림 2. 빅데이터 시장 현황
Figure 2. Market Condition of Big Data

하여 매출 증대, 비용 감소 및 수익률 증가 등의 효과를 나타낸 실험이다, 소셜네트워크 등의 빅데이터를 분석하는 과정에서 입력되는 데이터가 숙박 정보로써 활용할 수 있는 데이터인지를 판별하여 필터링하는 시스템을 제안하여 이를 기반으로 숙박률의 향상 및 공실률을 감소시킬 수 있는 마케팅 전략을 구축하고자 한다.

II. 관련 연구

1. 빅데이터 분류를 위한 딥러닝

빅데이터 분석 방법 중의 하나인 의미를 발견하기 위한 데이터처리 및 분석기술, 즉 의미 분석기술과 데이터마이닝기술 등 관련기술은 알고리즘이 필요하다[3]. 데이터마이닝(data mining)은 통계 및 수학적 기술뿐만 아니라 패턴인식 기술들을 이용하여 데이터 저장장소에 저장된 대용량의 데이터를 수집 및 분석하여 의미 있는 새로운 상관관계, 패턴, 추세 등을 발견하는 과정으로 본 논문에는 빅데이터를 활용한 데이터마이닝 기법을 적용하여 숙박정보를 필터링한 후에, 이를 기반으로 공실률 감소를 통해 향상된 마케팅 전략을 구축하고자 한다.

딥러닝은 컴퓨터가 큰 규모의 데이터를 이용해 스스로 학습할 수 있도록 인공 신경망을 기반으로 구축한 기계 학습 기술이다. 본 연구에서 숙박정보 빅데이터를 분류하기 위해 적용되는 딥러닝 모델은 SONN모델을 기반으로 하고 있는 FE-SONN모델이다. SONN모델은 베즈덱(James Bezdek)의 퍼지 c-means 알고리즘의 퍼지 멤버십 등식을 신경망과 융합한 자율적인 자기조직화 신경망 모델이다[4,5]로써 외부의 교사를 필요로 하지 않는다. 이 모델은 주어진 입력에 대한 클러스터의 수나 클러스터의 중심에 대한 사전 지식 없이 자율적으

로 클러스터에 관한 정보를 제공하여 유사 패턴분류와 패턴인식 등에 적합하며 높은 신뢰도 결과를 보여주고 있다. SONN 모델에 쓰이는 SONN 알고리즘은 빅데이터를 분류할 때 자율적인 학습을 하며 유사한 숙박정보 입력 데이터 값을 분류 처리할 수 있는 알고리즘으로써 코호넨의 자기조직화 모델과 유사한 점이 있다[6].

본 논문에서 빅데이터의 분류를 위해 적용한 FE-SONN 모델은 <그림 3>에서 보는 바와 같이 입력벡터가 입력층으로 들어오고 거리층과 멤버십층에서 피드백 하면서 클러스터들의 정보를 제공한다. 숙박정보 빅데이터를 자기학습을 통해 Taxonomy를 생성하고 자기조직화 기능을 적용하며 이것은 입력데이터의 유사 클러스터 중심점 등에 관한 어떤 사전 정보도 없이 들어온 입력 숙박정보에 대해 클러스터와 멤버십에 관한 정보를 제공한다.

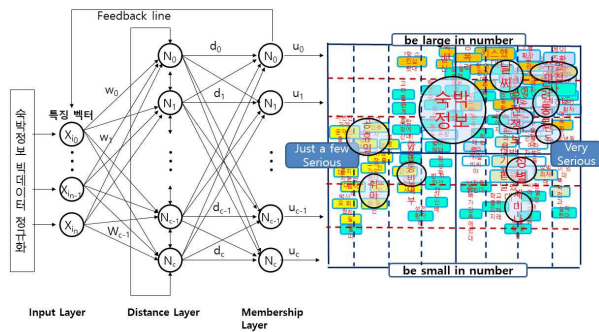


그림 3. FE_SONN 모델
 Figure 3. The Model of FE-SONN

2. 빅데이터기반 숙박정보 필터링 시스템 설계

본 논문에서 제안한 빅데이터 숙박정보 필터 시스템은 1단계 빅데이터 처리과정과 2단계 숙박정보 필터링 처리과정 등 2단계 과정으로 나뉘어 진다<그림 4, 그림 5>. 1단계 빅데이터 처리과정은 외부 데이터인 Social Medea(SNS, Blog, Community)와 Mass Medea(News, BBS, RSS)를 통해 데이터를 수집하고 수집된 데이터를 저장한다. 어휘 구분 분석단계에서는 문장분리→단어분리→구문분석→어휘분석을 거쳐 의미 분류를 시행한다. 의미 분류단계에서는 온톨로지를 생성하고, 의미 정보분류와 최종 어휘매칭분류를 수행한다. 2단계는 숙박정보 필터링과정으로 1단계의 어휘 매칭에서 분류된 자료와 숙박정보를 비교하여 일치하면 필터링하는 방식으로 숙박정보를 필터링하기 전단계에서 신뢰도 검사를 실시하여 정확도를 높였다.

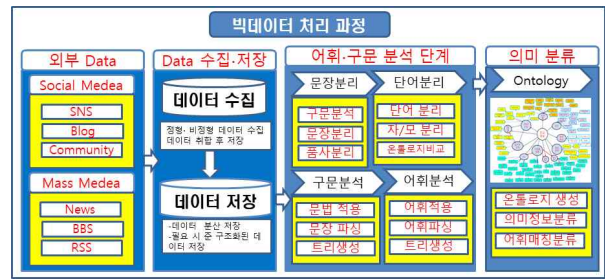


그림 4. 제안시스템 1단계 빅데이터 처리과정
 Figure 4. Process of Big data for the first step of Processed System

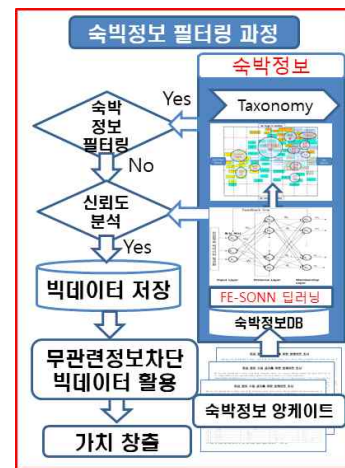


그림 5. 제안시스템 2단계 무관련정보 차단과정
 Figure 5. blocking Process of non-related data for the second step of Processed System

본 논문에서 사용된 데이터는 숙박업체 자체에서 보유중인 데이터, 여행 커뮤니티, 카페, 블로그 등 SNS에 올라오는 데이터와 숙박 포털 사이트에서 제공되는 데이터를 수집하여 분석되었다. 수도권에 위치한 숙박업체의 경우, 주말의 공실률은 매우 낮으나 주중의 공실률이 상대적으로 높아 주중의 공실률을 낮추는 것이 숙박업의 중요한 요소가 된다. 따라서 주중의 공실률에 따른 유지 관리 비용의 최소화를 위해 빅데이터의 분석이 필요하다. 본 논문에서 활용된 숙박업체의 경우, 수도권에 위치한 장점 때문에 팀, 또는 부서 단위의 브레인스토밍 회의를 겸한 세미나, 워크숍 또는 집단 주거지역에 거주하는 도시인들이 전원주택에서 휴식을 즐기기 위한 가족 모임이 주요고객이다. 현재 주중과 주말의 공실률은 주말의 경우, 공실률이 거의 없으나 60% 이상의 주중의 공실률을 빅데이터를 활용하여 가능한 공실률을 낮추는 것이 목적이다. 숙박업체의 마케팅은 펜션 홈페이지, 펜션 포털사이트, 키워드광고, 블로그 마케팅, 예약 대행 및 소셜커머스

등을 통해 이루어졌다.

III. 연구 내용

본 연구는 숙박업체가 보유한 데이터, 숙박포털사이트에서 제공한 데이터, SNS 데이터를 활용한 빅데이터와 빅데이터 분류를 위한 딥러닝 기술을 활용하여 연관상품과 관련된 분석 결과를 도출하였다. 도출된 시사점과 현재 보유한 데이터, 수집 가능한 외부 데이터를 고려하여 다음과 같이 5개의 빅데이터 주요 분석 서비스 시나리오를 선정하였다[7].

- 주요 이슈별 시즌 및 이벤트를 정의하고, 매출 데이터와 시즌 정의 데이터를 연관 분석하여 시즌 이벤트별 숙박상품 고객군 추천지수를 산출
- 주말과 주중 매출을 요일별 매출량과의 상관관계를 분석하여 요일별 숙박상품 고객군 추천지수를 산출
- 숙박 고객 중 직업 특성을 선정한 후, 고객군별 전체 평균 매출과 직업 특성에 따른 매출을 비교하여 직업 특성에 따른 숙박상품 고객군 단위의 추천지수를 산출
- 숙박 고객 중 지역 특성을 산정한 후 고객군별 전체 평균 매출과 지역 특성에 따른 매출을 비교하여 지역 특성에 따른 숙박상품 고객군 단위의 추천지수를 산출
- SNS 데이터를 수집하여 검색 대상 키워드에 대한 연관 키워드를 필터링하고 해당 키워드 중 여행 및 숙박정보와 관련된 데이터를 빅데이터 분류를 위한 딥러닝 기법에 따라 정보를 필터링한 후 연관 키워드를 추출

위의 시나리오를 기반으로 숙박정보를 필터링하기 위해 숙박업체 자체에서 보유 중인 데이터, 여행 커뮤니티, 카페, 블로그 등 SNS에 올라오는 데이터와 숙박포털 사이트에서 제공되는 데이터를 기반으로 단어 및 문장으로 구성된 온톨로지를 생성하게 된다.

숙박정보에 대한 온톨로지는 실세계에서 느끼고 사고하는 것에 대하여 상호 의견의 내용을 개념적이고 컴퓨터에서 다룰 수 있는 형태로 표현한 모델로써 개념의 타입이나 사용상의 제약조건들을 명시적으로 정의한 기술이다. 즉 언어로 표현된 개념 간 연관 관계 지식이

드러나는 망으로 표현되며 온톨로지를 통한 개념 간의 분석은 정보시스템의 대상이 되는 자원의 개념을 명확하게 정의하고 상세하게 기술하여 정확한 정보를 찾을 수 있도록 하며 용어와 용어 사이의 관계를 컴퓨터가 이해할 수 있는 형태로 정의한 것으로써 지식을 공유하고 새로운 지식이 추가되었을 경우 기존 지식과의 연계를 유연하게 할 수 있다.

데이터 생성 및 분류를 위한 온톨로지는 숙박정보에 대한 빅데이터 분석을 위해 온톨로지 분석을 통해 검색룰(rule)과 결합하여 연역적 논리 추론이 가능하다. 최근에는 대규모 소셜 미디어 및 동적 소셜 네트워크에서 분석을 수행하는 방법으로 시멘틱 기술을 사용하기도 한다. 이 방법의 특징은 복잡하고 이질적인 그래프 구조를 트리플로 구성된 RDF(Resource Description Framework) 그래프를 표현하는데, RDF는 웹에 있는 자원에 관한 메타 정보를 표현하기 위한 언어로 상호 이질적 정보들을 단일적 표현체계로 통합 표현 및 연계가 가능하고, 각 노드의 특성 정보를 통합 표현이 가능하다[6].

그림 3의 우측 부분과 같이 온톨로지 생성 및 분류 정보를 기반으로 Taxonomy가 생성된다. 이때 생성된 Taxonomy는 다층구조 형태의 신경망을 기반으로 하는 머신러닝으로써 빅 데이터의 자동 분류 방법인 딥러닝 기법을 적용하는 방법으로 FE-SONN모형을 제안했다, 딥러닝 모델은 SONN모형을 기반으로 하한 FE-SONN 모델이며 SONN모형은 베즈덱(James Bezdek)의 퍼지 c-means 알고리즘의 퍼지 멤버십 등식을 신경망과 융합한 자율적인 자기조직화 신경망 모델로 입력벡터가 입력층으로 들어오고 거리층과 멤버십층에서 피드백하면서 클러스터들의 정보를 스스로 제공해주는 특징이 있다[6,7]. 따라서 숙박정보 빅데이터를 자기학습을 통해 Taxonomy를 생성하는 역할을 한다.

본 연구는 빅데이터를 분석할 때 입력되는 데이터가 숙박정보인지를 필터링하고자 하는 부분에 있어서, 딥러닝의 부분에 FE-SONN 모델이 제안되었고, 신뢰도 분석 및 성능평가를 통해 본 연구의 타당성을 입증하고 최종적으로 필터링된 데이터를 기반으로 산출된 지수를 통해 실험 전후의 향상된 공실률의 변화를 보여주고자 한다.

1. 딥러닝의 신뢰도 분석

본 연구에서 제시된 신뢰도분석은 동일한 측정대상에 대해 같거나 유사한 측정도구로 반복적으로 측정할 경우에는 동일하거나 유사한 결과를 얻고자 하며, 신뢰도 분석은 일관성, 안정성, 정확성 및 예측가능성을 의미한다 [8].

따라서 <그림 5>에서 제시된 것은 빅데이터를 분석할 때 입력되는 데이터가 숙박정보인지를 판별하여 필터링하고자 하는 부분에 있어서, 딥러닝의 부분에 FE-SONN 모델을 제안했다. 딥러닝의 구조는 기존의 다층 인식기를 이용해서 인식 처리했을 때 입력층의 노드 수는 전역적 특성을 가진 36개 특징 벡터와 지역적 특성(오프라인상에서 추출) 12개의 특징 벡터 정보 총 48개가 된다. 또한, 기존 신경망 알고리즘의 경우 목표값은 비트의 조합에 의해 출력한다. 그러나 목표값을 비트 조합으로 정의할 때 오인식 처리할 경우가 발생한다. 즉, 테스트 데이터의 결과가 임계치를 0.75로 두었을 때, 신경망이 학습된 목표값에서 어느 하나의 유니트 결과가 0.75를 초과하게 되면, 예측이 불가능한 인식 결과를 출력하게 된다. 이러한 단점을 해결하기 위해 본 논문에서는 출력 유니트의 개수를 입력된 원본 서명 개수와 동일하게 함으로서 인식된 결과가 학습 목표값과 일치하는지를 검색하고, 이로 인해 오인식 요소를 줄이며, 이를 기반으로 검증할 수 있다. 그러나 시험대상인 48개의 특징 추출 값 가운데 가장 큰 값은 0.9328이고 두 번째 큰 값은 0.1007이 된다. 이렇게 인식된 결과에 대한 신뢰도 측정을 하게 된다. 신뢰도 수식은 <식 1>과 같이 정의했다.

$$RF = \frac{MAX(bit_array(i)) - 2nd_MAX(bit_array(i))}{MAX(bit_array(i))} = 1 - \frac{2nd_MAX(bit_array(i))}{MAX(bit_array(i))} \dots\dots\dots(1)$$

따라서 FE-SONN망에 의해 처리된 목표값을 나타내는 빅데이터 분석 시 숙박정보로 인식되는 값의 범주의 신뢰도에서 최대값 $MAX(bit(i))$ 은 0.8914이고 $2nd_MAX(bit(i))$ 는 0.2204이므로 신뢰도 지수를 구하면 $1 - (0.2204/0.8914) = 0.752748$ 이 된다. 본 논문은 <표 1>과 같이 RF의 값이 0.75이상이면 인식 결과의 신뢰성을 갖는 것으로 간주하여 인식 결과를 출력한다. 표 1에서

신뢰도 지수를 0.75을 신뢰도 경계값으로 지정한 이유는 숙박정보 빅데이터로 인식 결과의 비트 배열에서 가장 큰 $MAX(bit(i))$ 와 두 번째로 큰 $2nd_MAX(bit(i))$ 차에 대한 비율 값으로 신뢰도 지수로 평가했을 때 0.75 이상이면 정확도 부분에 만족하는 것으로 인정한다.

표 1 신뢰도 지수 비교표
 Table 1. Compare table of Reliability Factor

| Max(bit(i)) | 2nd_Max(bit(i)) | Reliability Factor |
|-------------|-----------------|--------------------|
| 0.9328 | 0.1007 | 0.892045 |
| 0.9264 | 0.1411 | 0.847689 |
| 0.9102 | 0.1816 | 0.800483 |
| 0.8914 | 0.2204 | 0.752748 |
| 0.8798 | 0.2628 | 0.701295 |
| 0.8654 | 0.3029 | 0.649988 |
| 0.8518 | 0.3433 | 0.596971 |
| 0.8354 | 0.3849 | 0.539262 |
| 0.8239 | 0.4254 | 0.483675 |
| 0.8109 | 0.4661 | 0.425206 |

2. 실험 및 결과

본 논문은 빅데이터 수집 및 분석할 때 숙박정보를 필터링하기 위해 크게 두 개 단계로 나누어 진다. 빅데이터 처리 단계와 숙박정보를 필터링하는 단계로 나누어 수행 및 실험했다. <표 2>는 빅데이터 수행 과정에서 데이터 수집단계와 저장단계를 수행했다. 어휘구문

표 2. 빅데이터 처리과정단계 실험결과
 Table 2. Results of Big Data Process Step Experiment

| 단계 분야 | 수행 단계 | 실험 데이터 수 | 성공 데이터 수 | 부분 성공률 (%) | 단계별 최종 성공률 (%) |
|------------------|-----------|----------|----------|------------|----------------|
| 수집/ 저장 단계 | 데이터 수집 단계 | 12897 | 12885 | 99.91% | 99.89% |
| | 빅데이터 저장단계 | 12885 | 11998 | 93.12% | 91.18% |
| 어휘 구문 분석 단계 | 문장 분리단계 | 11998 | 11971 | 99.78% | 99.43% |
| | 단어 분리단계 | 11971 | 11936 | 99.71% | 97.63% |
| | 구문 분석단계 | 11936 | 11178 | 93.65% | 87.11% |
| | 어휘 분석단계 | 11178 | 11063 | 98.98% | 86.16% |
| 의미 분류 단계 | 의미 분류 단계 | 11063 | 11012 | 99.54% | 85.23% |
| | 온톨로지 생성 | 11012 | 10908 | 99.06% | 84.13% |
| 빅데이터 처리 단계 수행 결과 | | 12897 | 10908 | 84.60% | |

분석단계에는 문장분리단계, 단어분리단계, 구문분석단계, 어휘분석단계를 수행했고, 의미분류단계에서는 의미분류단계와 온톨로지를 생성하는 과정을 수행했다. 그 결과 실험에 수행된 빅데이터 수는 12897개 중 온톨로지 실험까지 한 결과 10912으로써 최종 성공률은 84.60%로 나타났다.

<표 3>에서는 숙박정보를 필터링하는 단계로써 숙박정보학습단계에서 데이터분석, 어휘분석 및 구문분석, FE-SONN 딥러닝 단계에서 Taxonomy 수행단계를 수행했을 때 실험 데이터 수 4180개 중 성공 데이터 수는 4104개로 98.20%의 성공률을 보였다. 성공률이 높은 이유는 이미지 데이터와 같이 모호한 데이터를 학습하는 것이 아니라, 문자 데이터를 패턴매칭하기 때문이다.

표 3 숙박정보 필터링 단계 실험결과
Table 3. Test results of Pension Information Filtering Process Step

| 단계 분야 | 수행 단계 | 실험 데이터 수 | 성공 데이터 수 | 부분 성공률 (%) | 최종 성공률 (%) |
|-------------------|-------------|----------|----------|------------|------------|
| 숙박정보 학습 단계 | 숙박정보 데이터분석 | 4180 | 4180 | 100.00% | 100.00% |
| | 어휘분석/ 구문 분석 | 4180 | 4163 | 99.45% | 99.38% |
| | FE-SONN 딥러닝 | 4163 | 4132 | 99.25% | 99.13% |
| | Taxonomy 수행 | 4132 | 4096 | 99.12% | 99.05% |
| 숙박정보 필터링 단계 수행 결과 | | 4180 | 4096 | 98.20% | |

필터링한 숙박정보를 기반으로 빅데이터 분석 전의 매출 데이터와 분석 후 평일 상품을 개발한 후 매출 데이터를 분석한 결과, 분석전의 주말 매출은 91.67%로 우수하였으나 평일매출은 58.25%로 공실률이 심각한 수준이었으나 분석 후, 주말 평균은 상승률이 높지 않

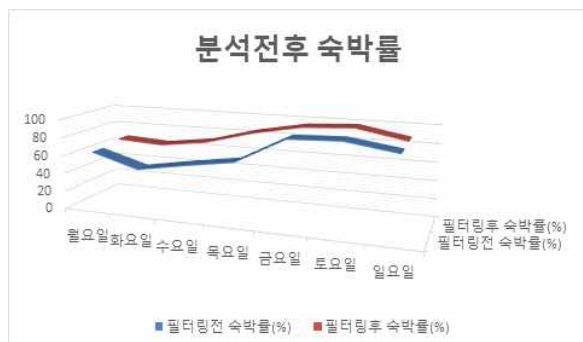


그림 6. 실험 전후 공실률 분석
Figure 6. Analysis of Before and After Exprement of Vacant Ratio

았으나 평일의 경우 13%이상 공실률이 감소하여 전체적으로 7.75% 감소한 것으로 나타났다<그림 6>, <표 4>.

표 4 공실률 분석
Table 4. Anlysis of vacant ratio

| 구분 | 요일 | 분석전 숙박률 (%) | 분석후 숙박률 (%) | 공실률 증감률(%) |
|----|------|-------------|-------------|------------|
| 평일 | 월요일 | 63 | 67 | -4 |
| | 화요일 | 48 | 64 | -16 |
| | 수요일 | 57 | 71 | -14 |
| | 목요일 | 65 | 85 | -20 |
| 주말 | 금요일 | 93 | 95 | -2 |
| | 토요일 | 95 | 98 | -3 |
| | 일요일 | 87 | 88 | -1 |
| 평균 | 평일평균 | 58.25 | 71.75 | -13.5 |
| | 주말평균 | 91.67 | 93.67 | -2 |
| | 평균 | 74.96 | 82.71 | -7.75 |

IV. 결 론

본 연구는 숙박업체 자체에서 보유중인 데이터, 여행 커뮤니티, 카페, 블로그 등 SNS에 올라오는 데이터와 숙박 포털 사이트에서 제공되는 빅데이터를 분석할 때 딥러닝 기법으로 데이터가 숙박정보인지를 필터링하는 시스템을 제안했다. 제안된 시스템의 구조는 빅데이터 처리과정과 숙박정보 필터 처리과정 등 2단계 과정으로 나누어지며, 빅데이터 처리과정은 4단계로 빅데이터 수집에서 분석 및 적용하게 된다. 데이터수집/저장→어휘 분석 및 구문분석→의미분류→숙박정보 필터→신뢰도 분석을 통해서 비로소 숙박정보가 필터링된 빅데이터를 저장하여 업무에 적용되는 단단계별 형태로 설계되었다. 실험 결과, 필터링정보의 최종 성공률은 84.60%까지 나타났으며 필터링한 숙박정보를 기반으로 빅데이터 분석 전의 매출 데이터와 분석 후 지수 산출에 따라 평일 상품을 개발한 후 매출 데이터를 분석한 결과, 최종적으로 분석 필터링전의 주중 공실률은 58.25%로 심각한 수준이었으나 분석 후, 주말 평균은 상승률이 높지 않았으나 평일의 경우 13%이상 공실률이 감소하여 전체적으로 7.75% 감소한 것으로 나타났다.

References

[1] NIA, K-ICT, "2018 Annual Big Data Market Trend Analysis", NIA, K-ICT Big Data Center,

- 2018.
- [2] National Information Society Agency(NIA), "IT & Future Strategy," No. 3. April, 2012.
 - [3] S. Koo, W. Choi, M.Jang, M, Lee, S. Heo, "Big Data Platform Strategy", EtNews, 2013.
 - [4] J. S. Choi, Personal Information Protect How to Use & Protect, Personal Information Protect Practice guide Vol. 6., ecurity news, June, 2016.
 - [5] S. Yoon, "A Study on Employment Strategy Based On Employment Information Filtering", The Journal of the Convergence on Culture Technology(JCCT), Vol. 5 No. 4, pp. 251-258, November, 2019.
 - [6] G. S. Koo, "Edge Feature Extract CBIRS for Car Retrieval : CBIRS/EFI", Journal of The Korea Society of Computer and Information, Vol. 15 No. 11, pp. 75-82, November, 2010.
 - [7] S. Yoon, "An Efficient Vacant Ratio Reducing Strategy", Journal of The Korean Association for Science Education, Vol. 31, No. 1, pp. 15-22, November, 2016
 - [8] J. park, E. Choi, M. Kang, Y, Jung, "Dropout Generic Algorithm Analysis for Deep Learning Generalization Error Minimization", Information Journal of Advanced Culture Technology(JACT), Vol. 5 No. 2, pp. 82-89, April, 2017.
 - [9] K, Jang, "Big Data Technology", Communication Books, 2013.

| |
|--|
| ※ 본 논문은 2021년도 송의여자대학교 학술 연구비 지원에 의한 것임. |
|--|