

기계학습 알고리즘을 사용한 스포츠 경기장 방문객 마케팅 적용 방안

박소현¹ · 임선영² · 박영호^{1*}

¹숙명여자대학교 공과대학 IT공학과

²숙명여자대학교 빅데이터활용 연구센터

A Study on Application of Machine Learning Algorithms to Visitor Marketing in Sports Stadium

So-Hyun Park¹ · Sun-Young Ihm² · Young-Ho Park^{1*}

¹Department of IT Engineering, Sookmyung Women's University, Seoul 04310, Korea

²Department of Big Data Research Center, Sookmyung Women's University Seoul 04310, Korea

[요 약]

본 연구에서는 마케팅 분야 중 스포츠 경기장을 찾는 관람객의 빅 데이터를 분석하여 소비자에게 맞춤형 마케팅 서비스를 제공하는 연구를 진행한다. 이를 위해 본 연구에서는 K-평균 군집화 방법을 사용하여 유사 관람객 그룹을 도출하고자 하며, K-근접 이웃 방법을 사용하여 새로운 방문객의 관심 매장을 예측하고자 한다. 실험 결과를 통해 상기 두 가지 알고리즘을 사용하는 것은 유사 관람객 그룹을 도출하며 신규 관람객 입장 시 신규 관람객의 특성에 맞는 적합한 마케팅 서비스를 제공 할 수 있게 하였다.

[Abstract]

In this study, we analyze the big data of visitors who are looking for a sports stadium in marketing field and conduct research to provide customized marketing service to consumers. For this purpose, we intend to derive a similar visitor group by using the K-means clustering method. Also, we will use the K-nearest neighbors method to predict the store of interest for new visitors. As a result of the experiment, it was possible to provide a marketing service suitable for each group attribute by deriving a group of similar visitors through the above two algorithms, and it was possible to recommend products and events for new visitors

색인어 : 데이터 분석, K-평균 군집화, K-근접 이웃, 기계학습 알고리즘, 스포츠마케팅

Key word : Big Data Analytics, K-Means Clustering, K-NN, Machine Learning Algorithm, Sport Marketing

<http://dx.doi.org/10.9728/dcs.2018.19.1.27>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 20 December 2017 ; **Revised** 23 January 2018

Accepted 29 January 2018

***Corresponding Author; Young-Ho Park**

Tel: +82-02-2275-4435

E-mail: dcs@naver.com

I. 서론

최근, 다양한 분야에 빅 데이터 분석 연구가 응용되고 있다. 특히 고객 정보 빅 데이터를 분석하여 마케팅에 활용하고자 하는 연구들이 활발히 진행되고 있다. 예를 들어, 고객 신용평가를 위해 기계학습 알고리즘으로 고객을 분류하거나[1], 사용자 맞춤형 아이템 추천하기 위한 기계학습 기반 알고리즘 연구가 진행되었다[2, 3].

본 연구에서는 스포츠 경기장을 찾는 관람객의 빅 데이터를 분석하여 소비자에게 맞춤형 마케팅 서비스를 제공하는 연구를 진행하고자 한다. 많은 연구들에서는 고객정보를 분석하여 이를 마케팅에 활용하였는데[1-4], 아직 스포츠 경기장에서의 마케팅을 위해 분석 기술이 적용된 연구는 거의 이루어지지 않았다. 따라서 본 논문에서는 스포츠 경기장을 찾는 관람객을 대상으로 관련 빅 데이터를 분석하여 소비자에게 맞춤형 마케팅 서비스를 제공할 수 있도록 스포츠 경기장 방문객 군집화 방법과 예측 방법을 제안하고자 한다.

먼저, 스포츠 경기장 방문객 군집화 방법은 실시간 대용량 데이터로 관람객의 선호도와 매장 이용 패턴을 파악하기 어렵다는 한계점을 해결하기 위한 방법으로, 관람객의 방문 패턴 및 방문 구역 내에 머문 시간 등의 이동 정보를 기준으로 유사한 경향을 보이는 관람객 그룹을 K-평균 군집화 알고리즘을 통해 군집화하여 패턴을 추출한다[3]. 신규 관람객을 대상으로 군집화한 결과를 기반 K-군집 이웃을 사용하여 새로운 관람객에게 적합한 마케팅 콘텐츠를 예측 및 추천한다[5].

본 논문의 공헌은 다음과 같다.

1) 본 논문에서는 스포츠 경기장에서 방문객 분석을 통해 마케팅에 활용할 수 있는 기계학습 알고리즘 기반의 방법을 제안한다. 전자상거래나 백화점 등의 구매고객 분석 위주로 이루어졌지만, 스포츠 마케팅 분야에서의 분석은 부족한 실정이다. 따라서 본 논문에서는 먼저, 스포츠 경기장 관람객의 유사한 방문객 그룹을 대상으로 개인화된 서비스를 제공하기 위한 방안으로 관람객의 선호패턴을 분석하고 선호패턴에 따라 군집화 한다. 이후에, 관람객의 선호패턴을 기반으로 군집화 된 주요 관람객 데이터를 대상으로 새로운 방문객 예측 방법을 제안한다.

2) 데이터 분석 결과에 따라 스포츠 마케팅 활용 방안을 대해 모색한다. 제안하는 방법을 통해 기존의 방문객들을 군집화 하면 유사 특성을 발견할 수 있으므로 군집 단위로 마케팅이 가능하다. 또한, 새로운 방문객이 유입되었을 때, 방문객의 군집을 예측하면 맞춤형 마케팅을 할 수 있을 것이다.

본 논문의 구성은 다음과 같다. 2장에서는 기계학습방법을 활용하여 빅 데이터 분석을 진행한 연구들에 대해 소개한다. 3장에서는 스포츠 경기장 방문객 군집화 방법과 방문객 이 데이터를 활용한 방문객 예측 방법을 제안한다. 4장에서는 3장에서 제안한 방법들의 실험 결과를 분석하고 활용방안을 제시한다. 마지막으로 5장에서는 결론을 내리고 향후연구에 대해 소개한다.

II. 관련 연구

본 장에서는 고객 분석 및 마케팅에 활용하기 위해 기계학습 알고리즘을 기반으로 분석한 관련 연구들을 소개하고자 한다. 먼저, 고객 정보를 분석하고 이를 마케팅 등에 활용하고자 하는 연구를 소개하고자 한다. 참고문헌 [1]에서는 신용평가를 위한 고객 분류 문제에 있어서 이원분류 모형을 설계하고 전통적인 분류 방법인 로지스틱 회귀분석과 서포트 벡터 기계를 이용하여 분류 성능을 비교하였다. 또한, 유비쿼터스 컴퓨팅 환경에서 고객에게 빠르고 정확하게 아이템을 추천하기 위해 선호도 기반의 군집 방법을 이용한 개인화 추천 방법이 연구되었다[2]. 제안된 방법은 고객점수 가중치 기반의 K-평균 군집화 기법과 베이저안 네트워크를 이용해 고객 선호도 계산을 할 수 있는 새로운 군집방법이다. Cho et al[3]는 실시간 유비쿼터스 컴퓨팅 환경이 요구되는 U-커머스 분야에서 사용자 맞춤형 추천 시스템을 구현하기 위해서 RFM(Recency, Frequency, Monetary)에 기반한 K-평균 군집화 알고리즘 방법을 제안하였다. Natchair et al[4]은 고객들의 만족을 고취시키기 위해 고객관계관리에 데이터마이닝 기법을 적용하였는데, 기존 고객관계관리 데이터에 노이즈가 많고 연밸런스해서 생기는 고객관계관리 분류 성능저하의 문제점을 해결하기 위해 새로운 특징추출기법을 제안하였다. 또한, 백화점 화장품 고객을 대상으로 마케팅 전략을 수립하기 위하여 고객 세분화 및 고객가치 점수화의 실증적 분석 연구도 진행되었다[6]. Na et al[7]은 K-평균 군집화 알고리즘의 단점인 각 데이터들의 거리와 모든 클러스터들의 중심점을 매번 계산해야한다는 점을 개선하기 위해 매번 계산시마다 다음 계산에서 사용될 정보들을 데이터구조에 저장하여 계산량을 줄였다. Ibrahim et al[8]은 M-커머스 분야에서 구매자의 신뢰도에 영향을 끼치는 요소들을 찾아내기 위하여 선호도 기반 K-평균 군집화 알고리즘 등의 데이터마이닝 기법을 사용하였다. 이외에도 K-평균 군집 알고리즘과 K-군집 이웃 알고리즘은 교통상황예측분야[4], 스포츠 분야[9, 10], 미디어 분야 [11], 인지재활분야[12] 등 다양한 분야에서 연구되었다.

III. 기계학습 기반 스포츠 경기장 방문객 분석

본 장에서는 스포츠 경기장 방문객에 대해 마케팅을 적용하기 위해 기계학습 알고리즘을 기반으로 방문객을 분석하는 방법에 대해 소개하고자 한다. 스포츠 경기장에 방문하는 관람객들에게 맞춤형 마케팅을 제공하기 위해서는 먼저 유사 방문객들을 군집화 하고, 정해진 군집을 바탕으로 새로운 방문객의 성향을 예측해야 한다. 3.1절에서는 먼저 방문객들을 군집화하는 방법을 설명하고, 3.2절에서는 군집을 바탕으로 예측하는 방법을 설명하고자 한다.

3-1 스포츠 경기장 방문객 군집화 방법

본 절에서는 스포츠 관람객 빅 데이터를 의미 있는 정보를 추출하고, 자동적으로 데이터를 분석하기 위한 군집화 방법에 대해 설명한다. 대용량의 스포츠 관람객 빅 데이터로부터 정보

를 추출하기 위해 K-평균 군집화 방법을 사용하고자 한다. K-평균 군집화 방법은 데이터 분석 등을 위하여 방대한 데이터를 몇 개의 그룹으로 묶는 군집화 방법 중 하나이다.

구현하는 알고리즘의 입력 값은 방문매장유형과 선호도가 담긴 데이터와 군집화 하고자 하는 개수이다. 결과 값은 K 개의 군집으로 나뉜 군집 집합이다. 이를 위해 진행될 단계는 다음과 같다. 첫째, 최초의 중심 값을 랜덤으로 선택한다. 둘째, K개의 중심 값과 각 개별 데이터간의 거리를 측정하여 가장 가까운 군집에 각 개별 데이터를 할당한다. 셋째, 각 클러스터마다 새로운 중심 값을 계산한다. 넷째, 새로 선택된 중심 값의 변화가 없다면 멈추고, 변화가 있으면 첫 번째 단계부터 변화가 없을 때까지 반복한다.

그림 1. 스포츠 경기장 방문객 군집화 방법의 슈도코드
Fig. 1. Pseudocode for Sports Stadium Visitor Clustering Method

```

Algorithm : K-means Algorithm for Analyzing Preference Patterns
of Spectators in Sports Stadium

Input : Store type and preference data set , K (The number of clusters)
Output : A set of K clusters
Method :
Repeat
1. Arbitrarily select centroid of objects
2. Assign each object to the cluster with the closest cluster.
   (This is based on the average value of the objects in the cluster)
3. Update the cluster average
Until The newly selected center value stops if there is no change.
Repeat from 1 if there is a change
    
```

3-2 방문객 예측 방법

본 절에서는 스포츠 경기장에 새로운 방문객이 등장했을 때, 이 방문객이 어떤 매장에 관심을 가질 것인지를 예측해보고자 한다. 이를 위해 본 연구에서는 K-근접 이웃 방법을 사용하여 새로운 방문객의 관심 매장을 예측하고자 한다. K-근접 이웃은 새로운 데이터가 출현했을 때, 근접한 K개의 이웃 데이터를 이용하여 예측하는 방법이다.

본 연구에서는 K-근접 이웃을 활용한 예측을 위해 오픈 소스 분석 도구인 R을 사용하였다. R은 윈도우즈, MAC, 유닉스, 리눅스와 같이 다양한 운영체제하에서 사용되며, 최근 데이터 분석을 위해 많이 사용되고 있다. 본 연구에서는 스포츠 경기장의 새로운 방문객에 대한 예측을 R 스크립트를 통해 프로그램을 작성하였고, K-근접 이웃은 class 패키지를 사용하여 구현하였다. 코드 (1)은 K-근접 이웃 구현을 위해 필요한 class 패키지를 사용하기 위한 문장이다.

```
library(class) (1)
```

코드 (2)는 주어진 데이터로부터 예측을 위한 표본을 추출하는 과정이다. 실험을 위해 주어진 데이터를 학습데이터와 테스트데이터로 나누는데, 테스트데이터를 랜덤하게 추출하기 위하여 다음과 같은 코드를 작성하였다. 여기에서 테스트 데이터는 스포츠 경기장에서의 새로운 방문객을 의미한다.

```

sam <- sample(2, nrow(person_data), replace=T,
prob=c(0.98, 0.02)) (2)

person_train <- person_data[sam==1, ]
person_test <- person_data[sam==2, ]
    
```

마지막으로 코드 (3)은 K-근접 이웃 알고리즘을 실행하는 문장이다. 새로운 방문객에 대해 예측하기 위하여 코드 (3)에서는 3명의 근접 이웃을 확인한다.

```

person_knn <- knn(train=person_train,
test=person_test, cl=person_train_label, k=3) (3)
    
```

IV. 실험 결과 분석 및 활용

본 장에서는 3장에서 제안한 기계학습 기반의 방문객 분석 방법들을 실험한 결과와 이를 활용하는 방안에 대해 소개하고자 한다. 4.1절에서는 스포츠 경기장 방문객들을 군집화 하는 실험을 분석하고, 4.2절에서는 새로운 방문객을 예측하는 실험을 분석하고 각각 분석된 결과를 활용하는 방안을 소개한다.

4-1 스포츠 경기장 방문객 군집화 실험 결과 분석 및 활용

본 절에서는 스포츠 경기장 방문객을 군집화 한 결과를 분석하고, 이를 활용하는 방안을 소개한다. 실시간 대용량 데이터인 스포츠 관람객 관련 데이터는 관람객의 매장 선호도와 매장 이용 패턴이나 방문객 연령과 매장방문횟수에 대한 관련성을 파악하기 어렵다는 단점을 갖고 있다. 이를 해결하기 위해, 본 논문에서는 관람객의 방문 패턴 및 선호도에 대한 정보나 방문객 연령과 매장방문횟수를 기준으로 유사한 경향을 보이는 관람객 그룹을 K-평균 군집화 알고리즘으로 군집화한 후 패턴을 추출한다. 이를 위하여, 관람객의 매장 선호도와 매장 이용 패턴, 방문객 연령과 매장방문횟수에 대해 랜덤으로 생성한 데이터를 사용하였다.

그림 2는 스포츠 경기장 방문객 군집화 실험 결과로써 120개의 방문매장유형 및 선호도 데이터를 사용하였다. 군집의 개수는 5개로 설정되었다. 그림 2의 x축은 방문매장유형을 의미하고 y축은 선호도를 의미한다. 표 1과 같이 매장유형은 10가지로 구성되고, 선호도의 범위는 1부터 8까지이다.

Group A에서는 햄버거, 피자, 편의점에 대해 선호도가 상인 방문객들로 군집화 되었다. Group B에서는 동일한 매장 유형에 대해 선호도가 중인 방문객들로 군집화 되었고, Group C에서도 같은 유형에 대해 선호도가 하인 방문객들로 군집화 되었다. 또한, Group D는 카페, 아이스크림에 대해 선호도가 상인 방문객들로 군집화 되었다. Group E는 같은 매장 유형에 대해 선호도가 중-하인 방문객들로 군집화 된 것을 확인 할 수 있다. 군집화하기 이전에는 단순히 120명의 방문객에 대해 매장의 선호도 정보만 있었고, 방문객들을 적절하게 군집화 할 기준이 존

제하지 않았으나, 제안하는 방법을 통해 방문객들이 자동적으로 적절하게 군집화 된 것을 볼 수 있다.

스포츠 경기장에 신규로 입장한 방문객에 대한 선호방문매장 정보가 없는 상태에서 기존에 이와 관련된 군집화 실험 결과 분석 결과가 있는 경우, 신규 관람객과 유사한 관람객 그룹을 도출함으로써 각 그룹 속성에 맞는 마케팅 서비스를 제공할 수 있다.

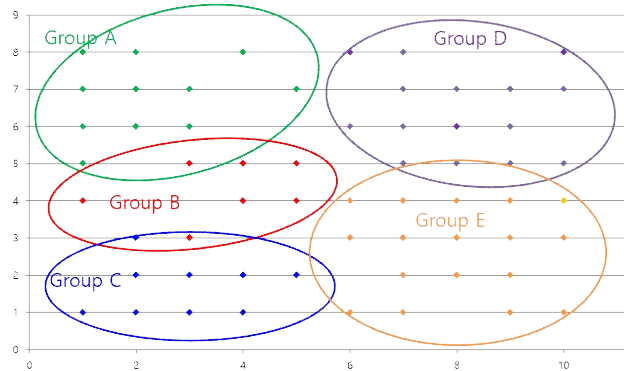


그림 2. 방문매장유형 및 선호도 데이터를 사용한 스포츠 경기장 방문객 군집화 실험 결과

Fig. 2. The experimental results clustering of sports stadium visitors using store type and visitor preference data

표 1. 매장유형

Table. 1. Store type

No.	Store type
1	Burgerking
2	Mcdonalds
3	Domino's Pizza
4	Pizza Hut
5	7-Eleven
6	CU
7	Starbucks
8	Angelinus
9	Baskinrobbins
10	Natuur

그림 3은 100개의 관람객 연령 및 매장방문횟수 데이터를 사용한 스포츠경기장 방문객 군집화 실험 결과이다. 군집의 개수는 5개로 설정되었다. 그림3의 x축은 관람객 연령을 의미하고, y축은 매장방문횟수를 의미한다. 관람객 연령의 범위는 5세 이상 및 60세 이하이고, 매장방문횟수 범위는 0회 이상 및 5회 이하이다.

연령이 10대 이상 및 60대 이하이면서 매장방문횟수가 상인 방문객들은 Group A로 군집화 되었다. 연령이 5세 이상 30대

이하이고 매장방문횟수가 중인 방문객들은 Group B로, 연령이 5세 이상 및 30대 이하이고 매장방문횟수가 하인 방문객들은 Group C로 군집화 되었다. 그리고 연령이 30대 이상이면서 60대 미만이고 매장방문횟수가 중인 방문객들은 Group D로, 관람객 연령이 30대 이상 60대 미만이고 매장방문횟수가 하인 방문객들은 Group E로 군집화 되었다. 만일 스포츠 경기장 방문객들의 매장 방문횟수를 높여 매상을 증가시키고자 한다면 방문횟수가 적은 Group C와 E에 집중적으로 마케팅을 할 필요가 있다고 볼 수 있다.

또한, 나이 정보는 존재하고 선호방문매장 정보가 없는 고객이 경기장에 입장했을 경우, 기존 실험을 통해 군집화 된 나이와 선호매장 관련 분석 결과를 활용하여, 신규 관람객과 유사한 관람객 그룹을 도출함으로써 각 그룹 속성에 맞는 마케팅 서비스를 제공할 수 있다.

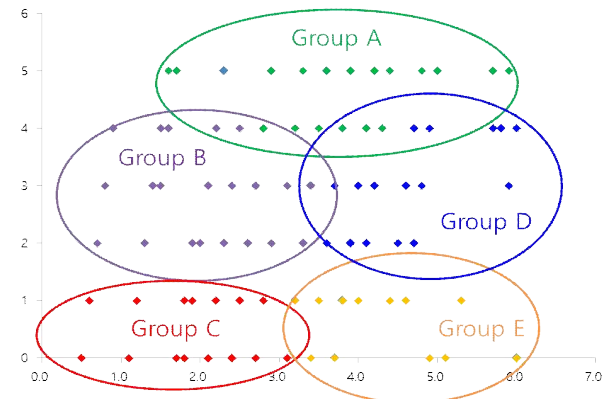


그림 3. 관람객 연령 및 매장방문횟수 데이터를 사용한 스포츠 경기장 방문객 군집화 실험 결과

Fig. 3. The experimental results clustering of sports stadium visitors using visitor age and number of visits

4-2 스포츠 경기장 방문객 예측 실험 결과 분석 및 활용

본 절에서는 새로운 스포츠 경기장 방문객에 대해 예측하는 실험 결과에 대해 분석하고, 이를 활용하는 방안에 대해 소개한다. 방문 경험이 없는 신규 관람객에게 적합한 마케팅 콘텐츠를 추천하기 위해서는 3.1절에서 소개한 것과 같이 유사한 특징을 보이는 관람객들끼리 묶는 군집화 단계가 필요하며, 그 다음으로는 군집화한 결과를 참고하여 새로운 관람객에게 적합한 마케팅 콘텐츠를 추천해주는 과정이 필요하다. 이를 위하여, 3.1절과 4.1절에서 소개하였던 관람객의 방문 패턴 및 선호도에 대한 정보나 방문객 연령과 매장방문횟수에 대한 스포츠 관람객 군집화 결과 값을 K-군집 이웃을 사용하여 신규 관람객에게 적합한 마케팅 콘텐츠를 예측한다.

그림 3에서 사용한 데이터는 4.1절의 방문매장유형 및 선호도 데이터를 사용한 스포츠 경기장 방문객 군집화 실험 결과 값인 5개의 군집 집합이다. 그림 3과 같이 3개의 신규 관람객 데이터를 실험 데이터로 사용하였다. 표 2에서 확인할 수 있는 것

처럼 3개의 실험 데이터 모두 정확한 군집으로 예측되었다. 만일 마케팅 비용을 줄여 매출을 늘리고 싶다면, Group C에게 선호도가 높은 매장 0-2에 대해 중점적으로 마케팅을 늘리고 선호도가 낮은 매장 5-9에 대한 마케팅을 줄이는 것이 필요하다고 볼 수 있다.

방문 경험이 없는 신규 관람객을 대상으로 맞춤형 마케팅 서비스 제공하기 위해서는 K-근접 이웃 알고리즘을 통해 신규 관람객의 매장 선호도와 매장 이용 패턴을 고려하여, 기존에 생성한 유사 관람객 그룹 중 가장 근접한 군집으로 신규 관람객을 예측할 수 있다.

표 2. 방문매장유형 및 선호도 데이터를 사용한 스포츠 경기장 방문객 예측 실험 결과

Table. 2 The experimental results of prediction of sports stadium visitors using store type and visitor preference data

No.	Test data	Predicted Cluster	Correct Answer
1	Store type : 2 Preference : 8	Group A	Group A
2	Store type : 7 Preference : 6	Group D	Group D
3	Store type : 7 Preference : 5	Group D	Group D

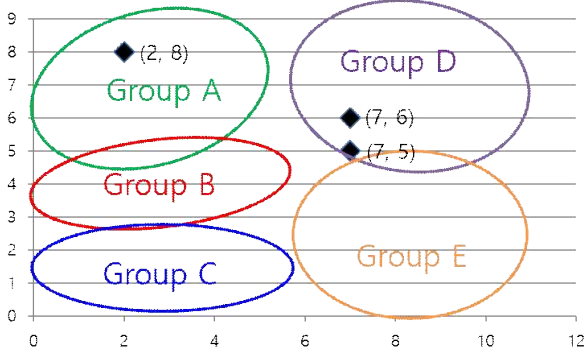


그림 3. 방문매장유형 및 선호도 데이터를 사용한 스포츠 경기장 방문객 예측 실험 결과

Fig. 3. The experimental results of prediction of sports stadium visitors using store type and visitor preference data

그림 4에서 사용된 데이터는 4.1절의 방문객연령 및 매장방문횟수 데이터를 사용한 스포츠 경기장 방문객 군집화 결과 값인 군집 집합이다. 그림 4와 같이 3개의 신규 관람객 데이터를 실험 데이터로 사용하였다. 표3을 보면 3개의 실험 데이터 모두 정확한 그룹으로 예측된 것을 확인할 수 있다. 만일, 기존에 방문횟수가 많은 집단에 대하여 추가적으로 매출을 올리고 싶다면, 방문횟수가 많은 Group A, B, D를 중점적으로 마케팅을 늘리고 방문횟수가 적은 Group C, E는 마케팅을 줄이는 것이 필요하다고 볼 수 있다.

또한, 방문 경험이 없는 신규 관람객을 대상으로 맞춤형 마케팅 서비스를 제공하기 위해서는 K-근접 이웃 알고리즘을 통

해 신규 관람객 나이를 고려하여, 기존에 생성한 유사 관람객 그룹 중 가장 근접한 군집으로 신규 관람객을 예측할 수 있다. 결론적으로, 기존 데이터를 토대로 신규 관람객을 위한 상품 및 이벤트 추천 알고리즘 적용이 가능하다.

표 3. 관람객 연령 및 매장방문횟수 데이터를 사용한 스포츠 경기장 방문객 예측 실험 결과 표

Table. 3. The experimental results of prediction of sports stadium visitors using visitor age and number of visits

No.	Test data	Predicted Cluster	Correct Answer
1	Visitor Age : 16 Number of visits : 4	Group B	Group B
2	Visitor Age : 28 Number of visits : 1	Group C	Group C
3	Visitor Age : 44 Number of visits : 5	Group A	Group A

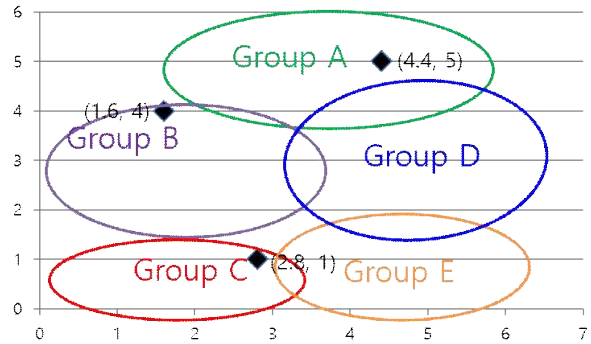


그림 4. 관람객 연령 및 매장방문횟수 데이터를 사용한 스포츠 경기장 방문객 예측 실험 결과

Fig. 4. The experimental results of predicting of sports stadium visitors using visitor age and number of visits

V. 결 론

본 논문에서는 스포츠 경기장의 방문객 대상으로 한 마케팅을 위해 기계학습 알고리즘을 기반으로 방문객데이터 분석하는 방법을 제안하였다. 이를 위해 먼저 스포츠 경기장을 방문한 방문객들을 K-평균 군집화 알고리즘을 바탕으로 군집화 시키고, 새로운 방문객이 유입되었을 때, 방문객의 특성을 K-근접 이웃 알고리즘을 바탕으로 예측하였다. 실험 결과를 통해 상기 두 가지 알고리즘을 사용하는 것은 유사 관람객 그룹을 도출하며 신규 관람객 입장 시 신규 관람객의 특성에 맞는 적합한 마케팅 서비스를 제공 할 수 있게 하였다. 뿐만 아니라, 실험 결과를 분석한 결과, 두 가지 알고리즘은 신규 관람객을 위한 상품 및 이벤트 추천에 다양하게 활용될 수 있음을 보였다.

향후연구에서는 실시간으로 유입되는 스포츠 경기장 방문객 데이터 처리 방안에 대한 연구를 진행하고자 한다.

감사의 글

This Research was supported by Sookmyung Women's University Research Grants(과제번호 1-1503-0079)

참고문헌

- [1] J. G. Eom, H. S. Seo, Y. B. Yoon, M. Yoon, "Comparison of Customer Classification Performance Using Machine Learning," *Journal of the Korean Data Analysis Society*, Vol.14, No.5, pp. 2441-2450, 2012.
- [2] H. B. Park, Y. S. Joe, H. H. Ko, "Clustering Method of Weighted Preference Using K-means Algorithm and Bayesian Network for Recommender System," *Journal of Information Technology Applications & Management*, Vol. 20, No.3, pp. 219-230, 2013.
- [3] Y.S. Cho, S.C. Moon, S.C. Noh, "Implementation of Personalized Recommendation System using k-means Clustering of Item Category based on RFM," *International Conference on Management of Innovation & Technology (ICMIT)*, 11-13 June, 2012, Sanur Bali, Indonesia.
- [4] S. U. Natchiar, S. Baulkani, "Customer Relationship Management Classification using Data Mining Techniques," *International Conference on Science Engineering and Management Research (ICSEMR)*, 27-29 Nov. 2014, Chennai, India
- [5] H. J. Kim, S. H. Park, G. T. Jang, "Short-term Traffic States Prediction Using k-Nearest Neighbor Algorithm : Focused on Urban Expressway in Seoul," *Journal of Korean Society of Transportation*, Vol.34, No.2, pp.158-167, 2016.
- [6] Y. J. Joe, J. Heo, "A Application Strategy of Customer Scoring Model : Focusing on Cosmetics part of Department Store," *Journal of the Korean Data Analysis Society*, Vol.8, No.1, pp. 335-348, 2006.
- [7] S. Na, L. Xumin, G. Yong, "Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm," *The International Symposium on Intelligent Information Technology and Security Informatics*, 2-4 April, 2010, Jingganshan, China.
- [8] O. Ibrahim, M. Nilashi, K. Bagherifard, N. Hashemi, N. Janahmadi, M. Barisami, "Application of AHP and K-Means Clustering for Ranking and Classifying Customer Trust in M-commerce," *Australian Journal of Basic and Applied Sciences*, Vol.5, No.12, pp. 1441-1457, 2011.
- [9] S. C. Han, S. Y. Jun, S. H. Jin, "Clustering Korean Professional Basketball Players by Using k-Medoids Clustering," *Journal of the Korean Data Analysis Society*, Vol.10, No.6, pp. 3423-3433, 2008.
- [10] W. I. Jang, S. Y. Ihm, A. Nasridinov, Y. H. Park, "A Study on Predicting Candidate Players for National Baseball Team using k-Nearest Neighbor," *International Journal of Applied Engineering Research*, Vol.9, No.23, pp. 18655-18662, 2014.
- [11] N. J. Kang, J. Y. Lee, H. M. Lee, "Classifying Media Repertoires Groups Using K-mean Cluster Analysis Method," Vol. 22, No. 3, pp. 7-46, 2008.
- [12] G. T. Choi, "Image Recognition and Clustering for Virtual Reality based on Cognitive Rehabilitation Contents," *Journal of Digital Contents Society*, Vol. 18, No. 7, pp. 1249-1257, 2017.



박 소 현 (So-Hyun Park)

2014년 : 숙명여자대학교 기약학부 (학사)
2014년~2016 : 숙명여자대학교 멀티미디어학과 (이학석사)
2016년~ 현재 : 숙명여자대학교 IT공학과 박사과정 (공학박사과정)

※ 관심분야 : 데이터 마이닝(Data Mining), 스마트 교육(Smart Education), 추천 시스템(Recommendation System), IT 융합(IT Convergence)



임 선 영 (Sun-Young Ihm)

2011년 : 숙명여자대학교 멀티미디어학과 (이학사)
2013년 : 숙명여자대학교 멀티미디어학과 (이학석사)
2017년 : 숙명여자대학교 IT공학과 (공학박사)
2017년~ 현재 : 숙명여자대학교 빅데이터활용 연구센터 책임연구원

※ 관심분야 : 빅데이터(Big Data), Top-k 질의 처리(Top-k Query Processing), 데이터베이스(Database), 기계학습(Machine Learning), 계층 인덱스(Layer-based Index), IT 융합(IT Convergence)



박 영 호 (Young-Ho Park)

1990년 : 동국대학교 컴퓨터공학과 (공학사)
1992년 : 동국대학교 컴퓨터공학과 (공학석사)
2005년 : 한국과학기술원 전산학과 (공학박사)

1993년~1999년: 한국전자통신연구원 교환전송연구단 선임연구원
1999년~2006년: COSMO 책임연구원
2002년~2005년: 한국기술대학교 겸임교수
2005년~2006년: 동국대학교 겸임교수
2005년~2006년: 한국전산기술원 AITrc 연구원
2006년~현 재: 숙명여자대학교 공과대학 IT공학과 교수

※ 관심분야 : 데이터 분석 (Data Analytics), 정보검색(Information Retrieval), 감성공학(Emotional Computing), 기계학습(Machine Learning), 데이터베이스 관리 시스템(DBMS), IT 융합(IT Convergence), XML