

# 제주국제자유도시 공개 데이터를 이용한 국가별 소비패턴 유사도 분석

## (Similarity Analysis for the Per-Country Consumption Patterns Based on JDC Open Data)

이 정 훈 <sup>\*</sup>      박 영 진 <sup>\*\*</sup>      이 윤 지 <sup>\*\*</sup>  
(Junghoon Lee)    (Youngjin Park)    (Yunji Lee)

**요 약** 본 논문에서는 제주국제자유도시 개발센터가 개방형 데이터 사이트에 공개한 월별 매출 데이터를 이용하여 각 국가별 매출 동향을 파악하고 국가별 유사성을 분석한다. 다운로드받은 원시데이터를 Python 프로그램을 통해 일련의 SQL 문장들로 변환한 후 MySQL 플랫폼상에서 수행하여 데이터베이스에 저장하고, RMySQL 라이브러리를 통해 R 통계분석 패키지의 작업공간으로 읽어온 다음 시계열 데이터 분석을 수행한다. 이 과정에서 각 국가별 월별 소비액 변화에 대해 동적 time warping 기법으로 플로우간 거리를 구하며 계층적 클러스터링에 의해 국가간 유사도를 구한다. 분석 결과 매출액 기준으로는 4개, 매출분포로는 3개의 그룹을 발견할 수 있었으며 각각의 패턴에 맞추어 그룹별로 판매촉진 전략 및 제품준비 계획을 수립할 수 있음은 물론 아직은 미진한 데이터 공개를 촉진할 수 있을 것이다.

**키워드:** 제주면세점, 공개데이터분석, 국가간 유사도, 동적 time warping, 계층적 클러스터링

**Abstract** The purpose of this paper is to investigate the sales pattern and the inter-country similarity applying the dataset made open by Jeju Free International City Development Center. A Python program converts the raw data downloaded from the open data site into a series of SQL insert commands to store the data archive in our database. Then, the developed R script reads the data through the RMySQL library and conducts a time series analysis for the monthly sales flow for visitors from different countries. The dynamic time warping procedure measures the distance between the sales series of each country and the hierarchical clustering procedure yields the inter-country similarity. Four and three groups are found for the whole sales revenue and for the sales distribution, respectively. By applying the findings, the management team can establish group-specific sales strategies and accelerate the publication of diverse dataset.

**Keywords:** Jeju DFS, open data analysis, inter-country similarity, dynamic time warping, hierarchical clustering

· 이 논문은 2019학년도 제주대학교 교원성파지원사업에 의하여 연구되었음  
· 이 논문은 2019 한국컴퓨터종합학술대회에서 '제주국제자유도시 공개 데이터를 이용한 국가별 소비패턴 유사도 분석'의 제목으로 발표된 논문을 확장한 것임

<sup>\*</sup> 정 회 원 : 제주대학교 전산통계학과 교수(Jeju Nat'l Univ.)  
jhlee@jejunu.ac.kr  
(Corresponding author임)

<sup>\*\*</sup> 비 회 원 : 제주대학교 전산통계학과 학생  
powersa53@naver.com  
tmxjel6279@naver.com

논문접수 : 2019년 8월 23일

(Received 23 August 2019)

논문수정 : 2019년 12월 12일

(Revised 12 December 2019)

심사완료 : 2019년 12월 19일

(Accepted 19 December 2019)

Copyright©2020 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.  
정보과학회 컴퓨팅의 실제 논문지 제26권 제2호(2020. 2)

## 1. 서론

공공데이터 포털은 2011년부터 서비스를 시작하여 정부 공공기관이 보유하고 있는 데이터들을 통합관리하면서 국민들이 이를 편리하게 이용하여 데이터의 활용을 높여 다양한 부가가치를 창출할 수 있도록 한다[1]. 각 지역자치단체들도 이와 같은 정부의 노력에 부응하여 보유한 데이터들을 적극적으로 공개하고 있다. 제주도는 관광과 전기자동차에 관련된 데이터 및 약 500여건의 데이터들을 이 서비스를 통해 공개하고 있으며 최근에는 제주데이터 허브도 개설하여 공개를 확장하고 있다[2]. 이중에서 국토교통부 산하 특수법인 공공기관인 제주국제자유도시개발센터(이하 JDC)도 아직은 절대적으로 미비한 점이 많지만 현재 운영중인 국내면세점, 항공 우주박물관, 영어도시, 지구촌 축제, 옥외광고 등에 관련된 정보를 개방하고 있으며 지속적으로 공개 데이터의 종류와 양을 확대해가고 있다.

JDC는 내외국인 면세점을 운영하면서 지속적으로 많은 수익을 올리고 있고 운영 데이터의 일부를 공개하여 영업 성과를 알리는 한편 관련된 데이터 분석을 가능하게 하고 있다. 비록 아직은 공개되는 데이터의 종류가 다양하지 않고 그 활용도가 높지 않지만 일반 민간 영역에서도 이 공개되는 데이터를 이용한 분석을 수행하고 다양한 영업 전략을 수립할 수 있다면 점차적으로 심도있는 데이터를 공개하게 될 것이고 그 데이터의 활용도도 크게 향상될 것으로 전망된다. 이때 Fuzzy, 연관사상 등 다양한 인공지능 기법들이 도입되고 있으며 상품진열, 재고관리, 생산계획 등을 효율화시킬 수 있다[3].

JDC 면세점에서는 항공권과 여권 등의 신분증이 있어야 면세품 구매가 가능하기 때문에 국적별 데이터 분류가 가능하다. 제주도와 JDC에서는 각 관광객 국가별 데이터를 파악하여 제공하고 있어서 각 국가별 유사도와 행태 분석을 수행할 수 있다. 따라서 본 논문에서는 공개된 데이터를 효율적으로 처리하는 데이터 처리 구조를 제시하고 이 구조상에서 다양한 도구들을 결합하여 각 국가별 면세점 구매액 패턴을 구해 유사한 국가간 공통된 매출 전략을 세울 수 있도록 한다[4]. 더욱이 공항에서 항공에 관련된 매출과 더불어 다양한 상점, 음식점 등 공항에서 비행객 매출의 중요성이 점차 증가되는 흐름을 고려할 때 비단 JDC와 같은 면세점뿐만 아니라 공항내 음식점 혹은 기타 소매점의 영업 전략을 수립하여 관광수익을 증가시키는 기반이 될 수 있다.

본 논문의 구성은 다음과 같다. 1장에서 논문의 주제를 소개한 후 2장에서는 관련 연구를 리뷰한다. 3장에서는 논문에서 구축한 데이터처리 구조를 설명하고 4장에서는 국가간 시계열 유사도 분석 결과를 보인다. 마지막으로 5장에서는 논문을 요약하고 결론과 추후과제를 도출한다.

## 2. 관련연구

제품판매에 관련된 시계열 데이터를 클러스터링하는 연구로서 [5]는 과거 데이터를 기반으로 판매 패턴이 유사한 그룹을 발견함으로써 명시적인 이상치 제거 과정 없이도 이상치들의 영향을 최소화하고 이에 기반한 수요예측으로 상품의 준비와 재고관리의 효율성을 향상시킨다. 시계열 흐름간 유사도를 발견하는 과정에서 단기 예측에 중점을 두어 트렌드와 주기성을 찾아 이의 영향을 최소화하여 상품 고유의 흐름이 반영되도록 한다. 유사도를 판단하는데는 PLS(Partial Least Square) 회귀 방식을 사용하며 클러스터링을 위해서는 로컬 서치에 기반한 휴리스틱을 사용하는데 이 과정에서 초기 클러스터를 임의로 구성한 후 비용함수를 최소화하도록 항목들을 이동하는 과정을 반복한다.

공항내에서 승객의 국적에 따르는 분류 및 국적별 행동 분석에 기반한 응용의 확장이 시도되고 있는데 그 예로 [6]은 판매 기록상의 부족한 정보로부터 사전에 정의된 규칙에 기반하여 승객의 국적을 파악한다. JDC의 경우에는 여권을 파악할 수 있으므로 구매자의 국적을 추가적으로 분석할 필요는 없다. 국적별 분석에 있어서 [7]의 연구에서는 공항내에서 전반적인 상품 판매의 패턴을 분석하여 중요한 영향 요소들을 파악한 바 있는데 기본적으로 상품의 매력 요소를 포함하여 저가항공사 이용자, 출장자 등의 여행자 타입과 탑승대기 시간, 공항이용객 수, 공항 크기, 상점의 분포 등의 영향을 분석하였다. 이 분석에 의하면 상점에 들어설 때 구매의사는 없었으나 구매하게 된 경우를 충동구매라고 규정할 때 공항에서 70% 정도는 충동구매에 해당하며 심리적인 요소가 많이 작용한다는 결과를 도출하였다.

또 판매 데이터를 분석하는 연구로서 [8]은 최근 온라인 쇼핑에 있어서 구매 패턴을 분석하였는데 사용자 클릭 스트림 패턴, 구매까지의 세션길이, 세션에서의 클릭 수 등과 구매연결 여부간의 예측모델을 구성하였다. 이 방식은 재귀분할 회귀트리 방식에 기반한 휴리스틱을 사용하였다. 이러한 행동양식 분석에서는 이상치를 제거하여야 하는데 [9]의 연구에서는 각 제품의 트렌드와 주기성들이 혼재된 상황에서 한 제품씩 주기성을 순차적으로 제거한다. 이 방식에서는 전체 데이터의 주기성을 발견하고 요소 분해 방식으로 트렌드를 제거한다. 이상치를 탐지하는 데는 GEST(Generalized Extreme Studentized Deviate) 방식을 사용한다.

이외에도 [10]은 단일 상품 판매 시리즈에서 복수의 영향요소들간의 상호관계에 대한 연구를 수행하였으며 아직 파악되지 않은 요소들을 영향을 고려한 예측모델을 구성하였다. 또, [11]에서는 미국경제위기 등 급격한

변동요소를 포함한 구간에서의 트렌드의 효과를 제거하고 인공신경망에 의해 제품별 월간 판매량 모델을 구축하고 분석하였다. 이와 같이 판매 데이터 혹은 그 시계열 데이터에 대한 분석은 다양하게 시도되고 있으며 각 응용과 목적과 가용한 정보에 따라 다른 대상과 특성을 가지고 있다. 특히 가용한 정보에 가장 큰 영향을 받는데 JDC 개방형 데이터는 현재까지는 월별 국적별 매출액을 공개하고 있다.

### 3. 데이터처리 구조

JDC는 국가공공데이터 포털에 연도별 월별 국가별 매출액을 공시하였다. 내부적인 사정과 아직은 낮은 데이터 활용도 등의 이유로 최신의 자료는 공시되어 있지 않고 2010년부터 2015년까지의 데이터만 CSV 파일의 형태로 다운로드가 가능하다. 또, 이와 별개로 JDC와 국적분류가 정확하게 일치하지는 않지만 제주도청도 관광객통계를 공개하고 있는데 두 데이터셋을 연계한다. 본 논문에서 구성한 데이터 처리구조는 그림 1과 같으며 우선적으로 위 사이트에서 각 연도별 파일을 개별적으로 다운로드하여 특정 폴더에 저장한다. 엑셀 혹은 CSV 파일을 전처리하는데 있어서 Python은 프로그래머가 사용하기 쉬운 라이브러리를 제공하고 있으며 프로그램에서는 이 라이브러리를 import하여 관련된 함수를 호출하면 된다. 개발된 Python 프로그램은 각 파일을 하나씩 열고 각 셀의 내용을 순차적으로 읽어들이어 적절한 포맷으로 변경할 수 있는데 우선적으로 텍스트 파일의 형태로 월별 국가별 매출액을 저장한다.

이와 아울러 데이터의 확장과 다양한 질의를 위해 MySQL 데이터베이스에 입력할 수 있으며 (국가, 연도-월, 매출액) 등으로 이루어진 테이블을 정의한 후 이 테이블에 데이터를 넣는 다수의 insert 문으로 CSV 파일들을 변환할 수 있다. 이후 이 SQL 문장들을 MySQL command line에서 배치 모드로 수행시켜 자료들이 데

이터베이스에 입력되도록 한다. MySQL 공간에서는 관광객 현황, 기상 등 다양한 정보와 효율적으로 결합할 수 있다. 현재 가용한 데이터의 양은 데이터베이스에서 충분히 처리할 수 있어서 하둡과 같은 빅데이터 분석도 구는 필요하지 않다. 다음 단계에서는 최근 널리 사용되고 있는 통계 패키지인 R에서 텍스트 파일 입력을 수행하는 read.table 혹은 MySQL 인터페이스인 RMySQL 라이브러리를 통해 JDC 매출액 데이터를 그 워크스페이스로 읽어온다. R은 다양한 인공지능, 데이터마이닝 및 시각화 라이브러리를 제공하는데 이를 활용하여 효율적인 분석을 수행할 수 있다[12].

### 4. 데이터 분석

#### 4.1 판매량 패턴 클러스터링

각 국가별 월별 매출액은 2010년도부터 2015년도까지 시계열 데이터를 형성하게 된다. 공개된 데이터에는 한국, 일본 등 9개 국가의 매출이 분류되어 있으며 일차적으로 단순히 매출액을 비교해본 결과 그림 2와 같이 내국인의 소비가 외국인의 소비를 압도하고 있다. 2010년 초반에는 데이터가 없는 상태이며 전반적으로 점차 증가하는 추세를 보이고 있는데 계절, 기상, 기타 이벤트 등에 영향을 많이 받고 있다. 또 그림에는 나타나지 않았지만 교포로 분류된 항목이 있는데 주로 재일교포인 것으로 보이며 매출에서 큰 비중을 차지하고 있지는 않아서 생략하였다.

그림 3은 내국인을 제외한 외국인들만의 매출액을 보여주고 있다. 중국 관광객들의 구매량이 가장 많은데, 전반적으로 연말에 급격히 감소하면서 등락을 반복하고 있다. 각 국가별로 등락의 패턴이 조금씩 다른 것을 확

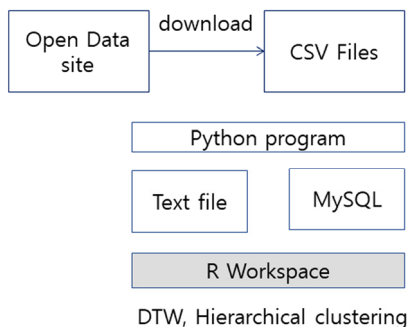


그림 1 데이터처리 구조

Fig. 1 Data processing architecture

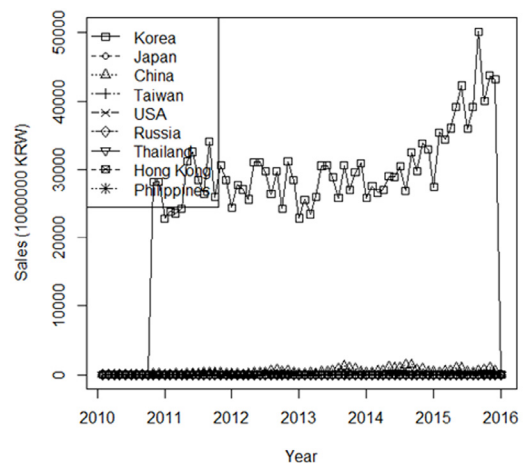


그림 2 국가별 월별 매출액 추이

Fig. 2 Per-country monthly sales pattern

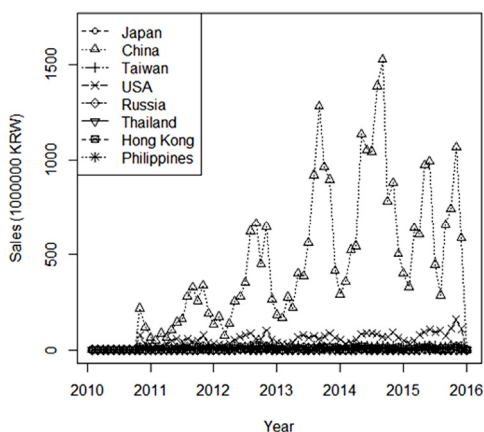


그림 3 외국관광객들만의 구매액 추이

Fig. 3 Per-country monthly sales pattern for foreign countries

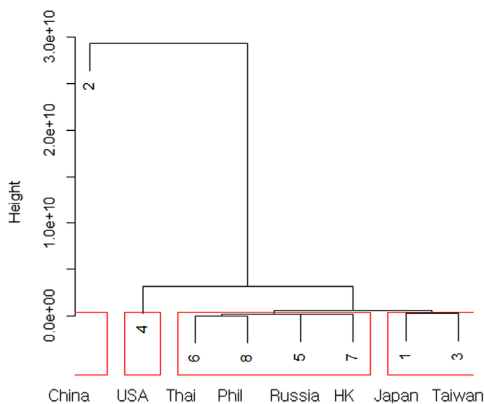
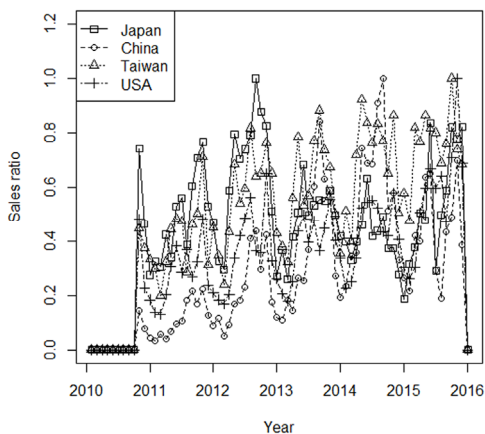


그림 4 매출액에 따르는 클러스터링 결과

Fig. 4 Clustering result for sales volume



인할 수 있으며 국가별 휴일의 분포와 시기별 방문자의 취향에 따라 JDC 면세점에서의 구매액이 다르게 된다. 필리핀의 경우에는 최근에 판매량이 나타나기 시작했으며 2014년 이후 동남아 관광객들의 구매가 늘어나고 있다.

각 국가별 구매액에 해당하는 시계열 데이터에 대해 그림 4는 R에서 제공하는 계층적 클러스터링 라이브러리에 의해 분류한 계통수(dendrogram)이다. 여기서 x축은 각 원소 혹은 시계열들의 번호이고 y축은 원소들간의 거리 혹은 비유사도를 나타낸다. 우선적으로 DTW (Dynamic Time Warping)를 이용하여 각 시계열간의 거리를 구하였으며[13] 이를 2차원 배열로 저장한 후 이 행렬을 인자로 넘겨주어 클러스터링을 수행하였다[14]. 계통수를 기준으로 그룹을 어떻게 나눌지 결정하게 되는데 계통수의 높이를 볼 때 4개의 그룹으로 나누는 것이 적절하다. 그 결과 큰 영향력을 갖는 중국과 개인별 소비가 큰 미국은 독립적인 그룹에 분류되었으며 일본, 대만이 한 그룹, 동남아 국가들과 러시아가 또 하나의 그룹을 형성하였다. 이 분류에서는 시간적인 흐름과 아울러 매출액도 그룹을 나누는데 영향을 주고 있다. 따라서 이들 그룹별로 판매와 홍보 전략을 수립한다면 JDC 면세점의 운영이 좀더 효율화할 것이다.

#### 4.2 매출액 패턴 형태

각 국가 내에서의 매출액 크기 요소를 최소화하고 오직 패턴만으로서의 추이를 분석하기 위해서 각 시계열 데이터를 0에서 1사이의 구간으로 매핑하였다. 각 시계열에서 심각한 이상값을 갖는 outlier가 발견되지 않았기 때문에 시계열 데이터에서의 최대값을 구하고 각 값들을 이 최대값으로 나누어 정규화시킨다. 그림 5는 8개 국가에 대해 매출비율로 정규화된 시계열의 흐름을 보이고 있는데 전반적으로 보면 겨울에 매출액이 줄고 봄

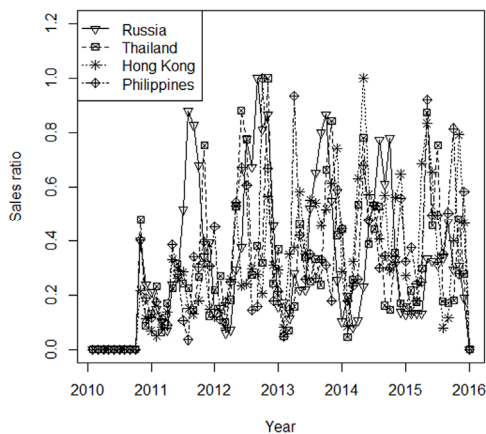


그림 5 국가별 매출 분포의 추이

Fig. 5 Per-country sales distribution patterns

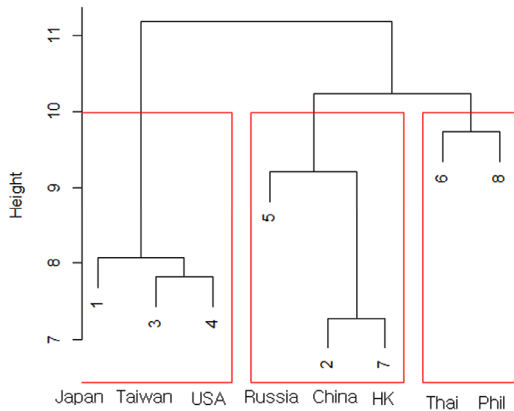


그림 6 매출 분포 흐름에 의한 클러스터링 결과  
Fig. 6 Clustering results of the sales distribution flow

이나 가을에 증가하는 것을 알 수 있으며 그 패턴은 국가별로 상이하다. 이 국가별 패턴에 있어서 6년간 월별 매출액의 표준편차는 0.21에서 0.27 구간에 속해 있는데, 가장 표준 편차가 큰 국가와 적은 국가는 각각 대만과 미국이다. 표준편차가 큰 경우에는 매출액이 많거나 적은 구간을 집중적으로 분석할 필요가 있다.

그림 6은 이 매출 분포의 흐름에 대해 4.1에서와 같은 방식으로 클러스터링을 수행한 계통수이다. 높이에 따라 3 개의 그룹으로 나누는 것이 적절한데 3.1에서 독립 그룹에 속했던 미국이 일본, 대만과 패턴이 유사한 것으로 나타났다. 또 중국의 경우 홍콩과 러시아와 유사했으며 태국과 필리핀이 하나의 그룹으로 형성되었다. 매출비율 흐름에 의한 클러스터링은 각 국가의 정치 및 지리적인 특성이 JDC 면세점에서의 구매 패턴에 영향을 주는 것을 보이고 있다.

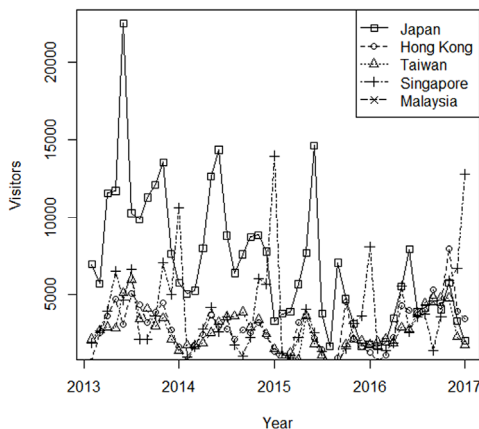


그림 8 제주도 관광객 수의 패턴(중국 제외)

Fig. 8 Pattern of the number of visitors to Jeju (Excluding China)

#### 4.3 관광객당 소비액의 패턴 비교

JDC의 매출액은 제주를 방문하는 관광객의 수와 밀접한 관계를 가지게 되는데 방문 관광객의 수는 JDC가 아닌 제주도청에서 집계하여 공개하고 있다. 따라서 집계되는 국가가 JDC 매출액의 경우와 차이를 보인다. 그림 7은 제주도청에서 공개한 관광객 수 데이터인데 싱가포르, 말레이시아, 인도네시아, 베트남 등이 추가적으로 보이는 반면 러시아와 필리핀은 포함되지 않았다. 그림에서 보는 바와 같이 2017년도까지 역시 중국 관광객의 수가 압도적으로 많다.

그림 8은 중국을 제외한 국가들에 대해서만 관광객 수를 플로팅한 결과이다. 2016년경에 감소하기는 했지만 일본 관광객 수가 가장 많으며 최근 동남아 국가들의 방문 수가 늘어나고 있는 추세이다.

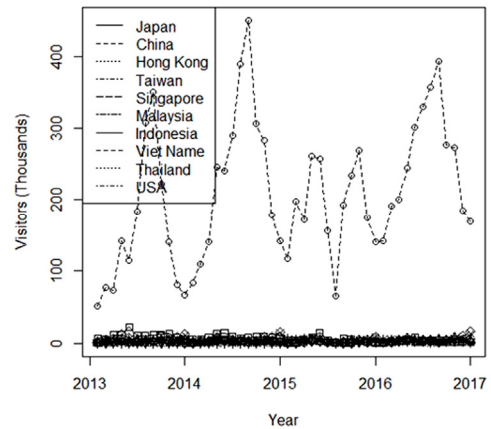
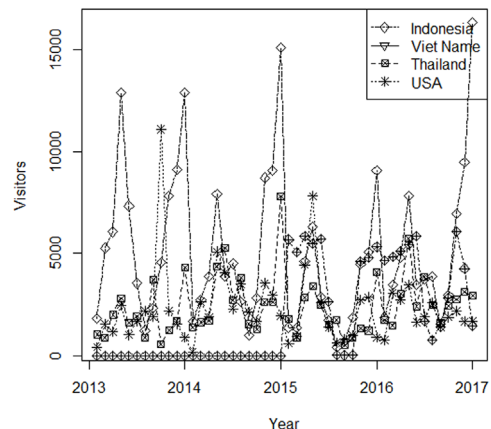


그림 7 제주도 관광객 수의 패턴

Fig. 7 Pattern of the number of visitors to Jeju



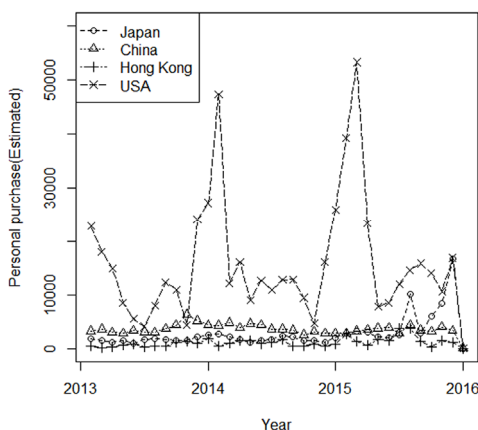


그림 9 개인별 구매액 추산

Fig. 9 Estimation of personal purchase

그림 9는 두 데이터 셋이 공통적으로 포함하고 있는 국가인 일본, 중국, 홍콩, 미국 등에 대해 역시 두 데이터 셋이 공통적으로 포함하고 있는 시간구간인 2013년부터 2015년까지의 구간에 포함된 기록들을 병합하여 개인별 구매액을 추산한 결과이다. 물론 제주도를 방문한 관광객이 모두 JDC 면세점을 이용한 것은 아니고 방문 목적과 방문 장소에 따라 면세점 이용에 주는 영향은 달라지겠지만 전반적인 개인별 구매액을 추산해볼 수 있다. 그림에서 보는 바와 같이 방문객 수에서는 많지 않은 미국국적 방문객들의 개인별 구매액이 타국가보다 월등하게 높는데 이는 구매잠재력과 아울러 일반 관광지에서 구매하기 보다는 면세점을 선호하는 경향때문인 것으로 파악된다. 달러당 환율은 2014년도 중반에 1,007원으로 최근 가장 낮았으며 2014년 초와 2015년 이후는 1,090원 이상에서 등락하였다. 따라서, 환율과 매출액의 연관성이 크다고 하긴 어렵다. 타국가들은 등락의 폭도 크지 않고 비교적 꾸준한 흐름을 보인다.

## 5. 결 론

본 논문에서는 제주국제자유도시개발공사가 공개한 매출 데이터를 시험적으로 분석하여 각 국가별 소비 유사도를 구하였다. 데이터 분석에 있어서는 Python, MySQL, R 등 공개소프트웨어에 기반한 분석 플랫폼을 구축하고 동적 time warping, 계층적 클러스터링 등 관련 라이브러리들을 결합하여 분석을 수행하였다. 매출액에 따른 그룹핑에서는 미국과 중국 관광객이 각각 고유한 패턴을 보이고 일본과 대만이 유사한 패턴을 보였다. 또 0과 1사이의 구간으로 매핑된 순수 패턴만을 분석한 결과는 미국, 일본, 대만이 같은 그룹, 그리고 홍콩, 중국, 러시아가 같은 그룹, 그리고 동남아 국가의 그룹 등을 발견

할 수 있었다. 이와 아울러 제주도청에서 공개한 제주도 관광객 수 데이터셋과 결합하여 일인당 소비액을 추산한 결과는 미국국적 관광객들이 JDC 면세점에서 가장 많이 소비하고 있는 것으로 나타났다.

분석된 결과는 특히 국가별로 판매촉진 계획을 수립하여 관광수익을 증가시키는데 기본 자료가 될 것으로 기대하며, 향후 연령대, 성별, 상품 목록별 데이터가 제공된다면 관광객 수, 기후 요소 등과 결합시켜 수요 패턴을 예측하고 이 예측을 기반으로 관련된 상품들을 준비하여 매출액을 향상시키는 물론 관광객들의 구매 편의성을 증진시킬 것이다. 더욱이, 제주도는 전기자동차, 스마트그리드 등 스마트 시티에 관련된 데이터들이 축적되고 있어서 이들에 대한 융합과 빅데이터 분석 모델 수요가 늘어나고 있다. 결국, 국가적으로 공개데이터의 중요성이 강조되어 각종 공공기관들도 다양한 데이터를 공개하고 있는데 아직은 공개되는 데이터의 다양성과 양이 미흡하지만 이러한 데이터 활용의 시도가 많아진다면 다양한 기관들이 데이터 공개에 적극적이 될 것으로 기대한다. 또, 데이터가 축적되면 딥러닝 등 다양한 인공지능 기법을 도입할 수 있어서 효과적인 제품 준비 및 제고관리가 가능할 것이다[15].

## References

- [1] [Online]. Available: <https://www.data.go.kr/dataset/15020490/fileData.do>
- [2] [Online]. Available: <https://www.jejudatahub.net/main>
- [3] C. Ezhilarasan, S. Ramani, "Performance Prediction Using Modified Clustering Techniques with Fuzzy Association Rule Mining Approach for Retail," *International Conference on Intelligent Computing and Control*, 2017.
- [4] M. Gurnani, Y. Korke, P. Shah, S. Udmale, V. Sambhe, S. Bhirud, "Forecasting of Sales by Using Fusion of Machine Learning Techniques," *International Conference on Data Management, Analytics and Innovation*, 2017.
- [5] A. Jha, S. Ray, B. Seaman, I. Dhillon, "Clustering to Forecast Sparse Time-Series Data," *IEEE 31<sup>st</sup> International Conference on Data Engineering*, 2015.
- [6] A. Mottini, R. Acuna-Agost, "Relative Label Encoding for the Prediction of Airline Passenger Nationality," *IEEE 16<sup>th</sup> International Conference on Data Mining Workshops*, pp. 671-676, 2016.
- [7] N. Volkova, "Determinants of Retail Revenue for Today's Airports," *German Airport Performance*, 2009, available at <http://www.gapprojekt.de/web/papers.html>.
- [8] U. Parkhimenka, M. Tatur, A. Zhvakina, "Heuristic Approach to Online Purchase Prediction Based on Internet Store Visitors Classification Using Data Mining Methods," *International Conference on Information and Digital Technologies*, pp. 304-307, 2017.



- [9] K. Kono, Y. Yamamoto, "Discussion of Outlier Detection Methods of Purchasing Data," *4<sup>th</sup> International Conference in ICT and Knowledge Engineering*, pp. 12-18, 2016.
- [10] T. Chen, H. Yin, L. Wu, H. Wang, X. Zhou, X. Li, "TADA: Trend Alignment with Dual-Attention Multi-Task Recurrent Neural Networks for Sales Prediction," *IEEE International Conference on Data Mining*, pp.49-58, 2018.
- [11] G. Nunnari, V. Nunnari, "Forecasting Monthly Sales Retail Time Series: A Case Study," *IEEE 19<sup>th</sup> Conference on Business Informatics*, 2017.
- [12] Y. Zhao, *R and Data Mining: Examples and Case Studies*, 1<sup>st</sup> Ed., Academic Press, 2012.
- [13] M. Brill, T. Fluschnik, V. Froese, B. Jain, R. Niedermeier, and D. Schultz, "Exact Mean Computation in Dynamic Time Warping Spaces," *Data Mining and Knowledge Discovery*, Vol. 33, Issue 1, pp. 252-291, 2019.
- [14] P. Rodrigues, J. Gama, and J. Pedroso, "Hierarchical Clustering of Time-Series Data Streams," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 20, Issue 5, pp. 615-627, 2008.
- [15] H. Choi, J. Jang, and S. Lee, "Document Feature Composition using Event-based Neural Tensor Network," *KIISE Transactions on Computing Practices*, Vol. 25, No. 8, pp. 407-411, Aug. 2019. (in Korean)



이 정 훈

1988년, 1990년, 1996년 2월 서울대학교 컴퓨터공학과 학사, 석사, 박사. 1990년~1991년, 1996년 대우통신 종합연구소 선임 연구원. 1997년 2월~ 제주대학교 전산통계학과 교수. 2004년, 2009년 University of Texas at Austin 방문교수. 관심분야

는 Cyber Physical System



박 영 진

2020년 2월 제주대학교 전산통계학과 학사. 관심분야는 Cyber Physical System



이 윤 지

2019년 8월 제주대학교 전산통계학과 학사. 관심분야는 Cyber Physical System