

인스타그램에서 나타나는 인간 활동 패턴의 연간 변화 분석

Analysis of Annual Changes in Human Activity Patterns on Instagram

정종구

Jonggu Jung

한국과학기술원

KAIST

whdrn9722@kaist.ac.kr

한수민

Sumin Han

한국과학기술원

KAIST

suminhan@kaist.ac.kr

이동만

Dongman Lee

한국과학기술원

KAIST

dlee@kaist.ac.kr

요약문

이 연구는 각 달과 계절에 따른 사람들의 활동 패턴을 객관적이고 일반적으로 알아보는 데 목적이 있다. 이를 위해 인스타그램 게시물을 이용하여 사용자들의 시간에 따른 활동 패턴을 통계적으로 분석하였다. 연구 결과 사람들은 공통적으로 1월 12월을 연회의 달, 3월을 시작을 달 이라고 생각을 했다. 또한, 계절이 사람들의 활동 패턴에 미치는 영향력은 매우 높았다. 이 연구는 각 달이 가지는 상징성과, 계절이 사람들의 활동 패턴에 미치는 정보를 통해 다양한 분야에 활용할 수 있다는 점에 의의가 있다.

주제어 : 인간 활동 패턴, 시계열 분석

용도 적으며 보다 객관적으로 사람들의 행동 패턴을 파악할 수 있다.

이로부터 분석된 인간의 활동 분류를 이용한 마케팅, 여행 추천 등 다양한 분야에 적용하여 사용할 수 있다. 또한 이용자에게 맞는 광고를 함으로서 효과적인 게시물을 추천해 줄 수 있다. 본 연구에서는 수집된 인간행동분류 데이터를 활용하여 인간 활동 패턴의 연간 변화를 분석해보고자 한다. 이를 통해 사람들이 필요로 하는 정보를 보다 객관적이고 일반적으로 제시하고자 한다.

1. 서론

1.1 연구의 배경

사람들은 본인의 일상과 경험을 다른 사람들과 공유하는 것이 일상이 되었다. 특히 인스타그램(Instagram)은 한국에서 최근 몇년간 가장 활발하게 활용되는 SNS플랫폼이다. 인스타그램은 포스트를 업로드할 때 캡션을 통해 사진과 관련된 단어나 게시자의 감정을 함께 입력할 수 있다. 위치정보가 있는 인스타그램의 포스트의 경우 사용자가 실제로 해당 위치에서 본인들이 경험을 바탕으로 글을 쓴 경우가 많다. 그리고 캡션과 사진을 통해 특정 활동의 목적을 알 수 있다. 따라서 이 데이터를 활용하면 실제 현실에서 사람들의 활동 패턴을 파악하기위한 자료로 사용할 수 있다.

1.2 연구의 필요성

우리나라는 사계절이 뚜렷하며 날씨에 따라 사람들이 즐겨하는 활동의 패턴에도 변화가 생긴다 [2]. 하지만 이를 설문조사 이외의 방법을 통해 객관적인 데이터를 바탕으로 분석한 연구는 드물다. 설문조사를 통한 연구에 비해 방대한 SNS를 활용한 연구는 정리도 쉽고, 비

2. 데이터 셋

2.1 데이터 수집

본 연구에서는 대표적인 도심 지역에서의 사람들의 활동에 대한 데이터인 서울 지하철 2호선의 주요 25개 역 주변에서 업로드된 인스타그램 포스트를 크롤링하여 진행하였다.

2.2 데이터 라벨링

모아진 데이터 중 27명의 실험 참가자 3명씩 총 9팀으로 나누어 ATUS(American Time Use Survey)[3]의 카테고리에 알맞게 분류하는 작업을 거쳤다. 각 팀에 속하는 사람들은 독립적이지만 같은 포스트에 대해 태깅을 하여 3명의 라벨링 결과를 비교하여 두 명 이상이 동의한 카테고리를 해당 포스트의 정답 카테고리로 고려하였다.

2.3 데이터 분석

그 중 가장 빈번하게 발생한 상위 7개 카테고리(자기관리-1114개, 식사-5013개, 사교 활동-2609개, 휴식-926개, 예술 및 엔터테인먼트-2721개, 스포츠-465개, 여행-1651개)에 대해 분석을 진행하였다. 분석 이후 비슷한 경향을 띄는 항목을 제거하여 4개의 항목을 추렸다.

3. 장소성 시각화

3.1 데이터 전처리

데이터 전처리를 위해, 각 카테고리에서 발생한 소셜미디어의 시간당 빈도수를 기준으로 하였으며, 1년 주기인 365일에 대한 각각의 24시간에 대하여 해당 카테고리의 포스트가 발생한 횟수를 데이터로 삼았다. (예: 1월1일 23시 - 3회) 데이터 셋은 특정 지역에서 업로드된 인스타그램의 게시물 개수이다. 기간은 2012년 7월부터 2019년 인테 2월 29일이 포함된 년도는 2016년이 유일하다. 2016년 2월 29일에 해당 지역에서 업로드된 게시물의 개수는 0개였다. 따라서 윤년을 제외하고 1년을 365일로 두고 본 연구는 진행되었다.

3.2 데이터 형식

본 연구에 사용되는 data는 다음과 같이 정리가 된다. 세로축의 항목은 시간을 나타내고 가로축은 카테고리의 수이다. 세로축의 시간은 날짜와 시간을 동시에 의미한다. d번째/365일의 h시/24일 경우 $(d-1) \times 365 + h$ 의 row에 해당된다. 가로축은 '2.데이터 셋'에서 정리한 7개의 카테고리를 의미한다. 즉, $(d-1) \times 365 + h$ 번째 row의 n번째 column은 h시~h+1시에 업로드된 게시물 중 n category에 속하는 게시물의 개수를 의미한다. 따라서 데이터의 row는 (일×시)의 총 개수이므로 $365(\text{일}) \times 24(\text{시}) = 8760$ 이고 column는 7개로 '8760 rows × 7 columns'인 데이터 셋이다

3.3 분석 방법

본 연구의 목적은 1년이라는 기간을 기준으로 데이터들의 경향성과 주기성을 파악하는 것이 가장 중요하다. 그래서 시간의 흐름에 따라 데이터를 분석하는 방법인 '시계열 분석법'을 사용하였다. 본 연구에서는 Python을 이용하여 시계열 데이터를 Decompose하여 경향성과 주기성을 나타내는 그래프를 도출할 것이다.

경향성과 주기성을 파악하기 위해 가장 많이 사용되는 방법은 STL Decomposition (Seasonal Trend Decomposition using Loses)이다.

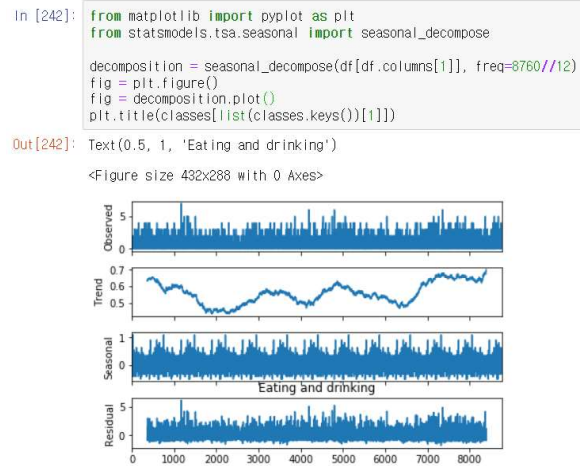


Figure 1. Decomposition code

Figure1 은 7개의 카테고리 중 가장 많은 식사 데이터를 제시된 Decomposition을 통해 데이터를 분리하고 python의 matplotlib의 pyplot을 이용해 나타낸 그래프다.

'df'는 Instagram_data.csv 를 python에서 불러온 값이다. $df[df.columns[1]]$ 에는 2번째 항목인 음식 항목의 값이 들어가 있다. seasonal_decompose에 사용할 frequency는 그래프에서 얻을 수 있는 여러 주기성 중 원하는 주기성을 설정하는 것이다. 본 연구의 주기성(seasonal)은 크게 달(month)을 기준으로 한다. 따라서 12개의 펄스파가 나오기 위해 $freq = 7860//12$ (7860을 12로 나눈 몫)으로 두었다.

Decomposition이후 총 4개의 그래프를 볼 수 있다. 4개의 그래프의 관계는 $Observed = Trend + Observed + Residual$ 이다. Observed는 데이터 셋에서 가로축의 변수에 따라 세로축에 값을 입력해 나열한 기본적인 그래프이다. Seasonal은 데이터의 주기성을 나타낸 값으로 freq에 따라 반복되는 파형의 개수가 결정된다. Trend는 데이터의 전체적인 경향성을 나타내는 그래프이다. 남은 값은 Residual 그래프에 표현되었다.

4. 분석 결과

4.1 자기관리 카테고리

자기관리 카테고리는 잠을 자거나 씻거나 옷을 입는 행위 등 자신의 개인적인 활동이다. 특히 자신의 모습을 찍는 행위가 인스타그램에 표현되는 대표적인 자기관리라고 할 수 있다.

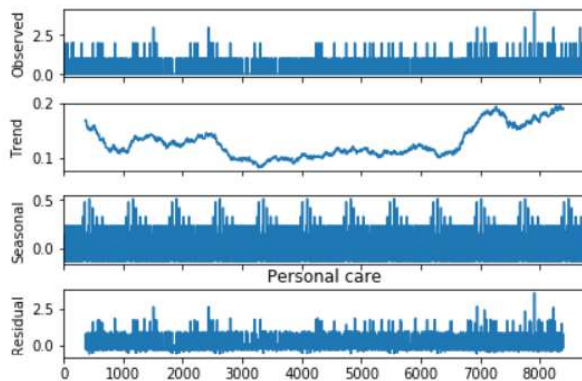


Figure 2. 자기관리 그래프

자기관리 카테고리의 특징은 Figure2에서 볼 수 있다. 경향성 그래프를 보면 4월 말부터 낮아져서 일정한 값을 이루고 10월 말에 최고 점을 찍고 12월 말에 다시 최고 점을 찍는 것을 알 수 있다. 인스타그램에 자기관리로 분류되는 대부분의 사진은 셀카와 예쁜 풍경에 자신의 전신을 찍은 사진이 업로드 되어있다.

1월과 12월에는 여러가지 행사가 있다. 행사에 참여하면 평소보다 더 꾸미게 되고, 예쁜 장소에 있는 평소보다 아름다운 자기 자신을 기록하고 싶어진다. 또한, 10월 말은 가을에서 겨울로 넘어가는 간절기이라서 입는 외투의 종류가 다양하다. 패딩, 블레이저, 코트, 가디건, 야상, 니트 등 모든 종류의 외투를 입을 수 있다. 즉, 자신의 패션을 보여주는 가장 적절한 시기이기 때문에 많은 사람들이 자신의 옷을 찍고 인스타그램에 업로드를 하는 것이다.

4.2 식사 카테고리

식사 카테고리는 먹거나 마시는 행위를 하는 모든 활동을 의미한다. 카페에서 음료를 마시거나, 식당이나 집에서 먹는 사진 또는 음식이 있는 사진을 이 카테고리로 분류했다. 식사는 모든 사람이 매일 하는 활동이다. 그래서 음식 사진은 더 특별해야한다. 이 게시물은 특별하게 만든 음식, 맛집, 분위기 좋은 식당에서 찍은 사진이다.

Figure3는 식사 카테고리에 관한 그래프이다. 경향성은 1월과 12월에 최고를 찍고, 5월과 7월 말에 게시물이 많고, 3월, 10월 초에 최소가 된다.

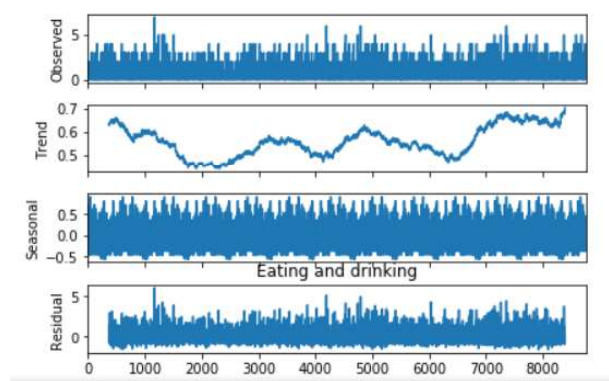


Figure 3. 식사 그래프

1월과 12월에 높은 이유는 행사에는 특별한 음식이 필요하기 때문이다. 그래서 연말의 음식 카테고리 사진의 캡션에는 ‘#크리스마스’, ‘#신년’이 있다. 5월에는 가족과 함께 하는 사진이 늘어난다. 어버이 날에는 부모님, 어린이 날은 자식과 함께 맛집을 가고 싶어 한다. 그래서 많은 사람들이 평소에 가지않는 좋은 곳에 가서 음식 자랑을 한다. 많은 사람들은 5월을 가정의 달로 인식한다.

7월의 음식 사진은 캡션에 ‘여행’이라는 단어가 많이 있었다. 7월은 많은 대학생들이 방학을 맞거나 직장인들이 휴가를 받아서 여행을 가기 적절한 시기이다. 여행을 가면 누구나 사진으로 남긴다.

4.3 예술 및 엔터테인먼트 카테고리

예술 및 엔터테인먼트 카테고리는 스포츠를 제외한 문화 활동을 포함하는 항목이다. 대표적인 활동으로는 영화보기, 박물관과 전시회 관람, 그리고 콘서트나 영화제 관람이다.

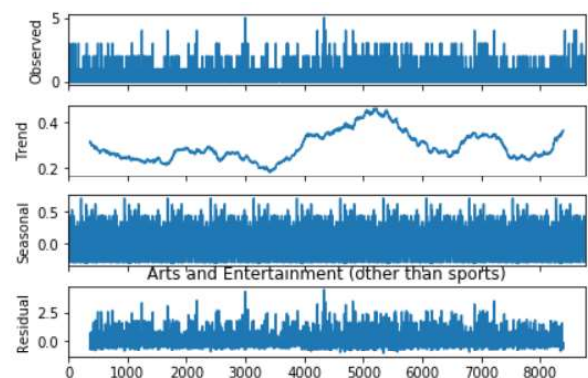


Figure 4. 예술 및 엔터테인먼트 그래프

이 카테고리는 다른 카테고리와는 다른 분포를 보여준다. Figure2, 3은 1월과 12월에 높았

지만 Figure4는 6월초에서 8월 중순까지 매우 높은 분포를 보여준다. 이 항목은 1년 중 언제든지 즐길 수 있는 활동인데 특정기간에만 높다. 이를 분석하기 위해 게시물의 해시 태그와 함께 분석했다.

6월 말부터 8월 중순 동안 예술 및 엔터테인먼트로 분류된 게시물에는 ‘더위’와 관련된 단어가 많이 포함되어 있었다. 이 시기는 여름이라서 밖에서 다른 활동을 하기가 힘들다. 하지만 영화 관람과 전시회 관람은 실내활동이면서 2시간 정도 진행 가능한 활동이기 때문에 더위를 피할 수 있는 활동이다. 많은 사람들의 예술 및 엔터테인먼트의 활동은 여름에 가장 큰 영향을 받는다.

4.4 여행 카테고리

여행 카테고리는 이동과 관련된 활동을 포함한다. 여행을 가는 것 외에도 출퇴근길에 찍은 경치 사진이나, 산책을 나가서 찍은 야경 사진, 크리스마스 트리 구경을 여행 활동으로 간주한다.

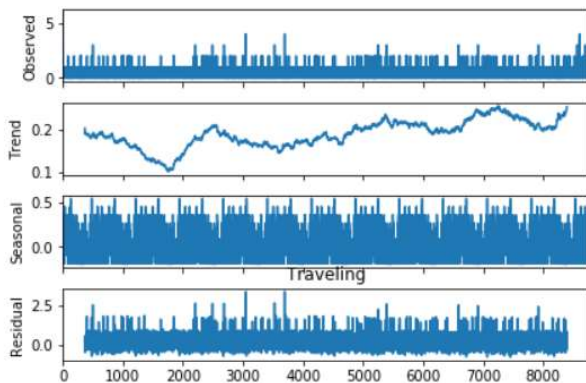


Figure 5. 여행 그래프

여행 카테고리는 전체적으로 일정한 경향성을 나타내지만 3월에는 특히 매우 낮다. 여행 카테고리의 대부분을 이루고 있는 것은 출퇴근길에 찍은 경치 사진이다. 그리고 경치 사진에는 ‘힘들다’와 관련된 캡션이 많이 포함되어 있다. 즉, 출퇴근길에 올린 아름다운 풍경 사진은 ‘지금은 힘들지만 풍경을 보면서 견디자.’라는 의미를 담고 있다. 즉, 여행 카테고리는 사용자의 심리의 안정도를 나타내어 준다.

5. 결론

본 연구의 목적은 서울 지하철 2호선 근처에서 업로드되는 인스타그램 데이터를 통해 사

람들의 활동 패턴이 시간에 따라 어떻게 변하는지를 확인하고, 활동 패턴에 대한 객관적인 자료를 얻는 데에 있다.

연구결과 우리는 각 달이 사람들에게 의미하는 상징성을 알 수 있었다. 1월과 12월은 한해의 시작과 끝을 의미하고 있어 다양한 행사와 활동이 일어난다. 대부분의 카테고리는 많은 사람들이 참여하는 행사와 관련 있어서 1월과 12월에 전체적인 게시 수가 증가했다. 또한 대한민국에서 3월이 의미하는 것은 ‘일의 시작’이라고 볼 수 있다. 그래서 사람들의 마음속의 여유를 대변해주는 여행 항목에서 3월에 급격하게 낮아지는 결과를 얻을 수 있었다. 게다가 사람들의 활동 패턴과 계절 사이에 큰 연관성이 있음을 밝혔다. 특히 예술 및 엔터테인먼트 카테고리에서 가장 큰 영향을 미친 부분은 여름이라는 계절이고, 자기관리 카테고리에서는 가을 풍경이 미치는 영향이 가장 크다는 것을 알아내었다.

이 연구는 각 달이 사람들이 가지는 상징성과, 계절이 활동 패턴에 미치는 영향을 확인했다는 것에 의의가 있다. 이렇게 객관성과 일반성을 증명 받은 정보를 통해 개개인에게 필요한 정보를 주고, 유익한 활동을 추천해 준다면 다양한 분야에 큰 영향을 끼치게 될 것이다.

사사의 글

본 연구는 자가 학습 기반 Autonomic IoT 엣지 컴퓨팅 핵심 요소 기술 연구 (2019-0-01126) 의 재원으로 진행되었음

참고 문헌

1. Klepeis, Neil E., et al. "The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants." *Journal of Exposure Science and Environmental Epidemiology* 11.3 (2001): 231.
2. 이용수. "계절별 날씨 변화가 신체활동 참여 수준에 미치는 영향." *한국웰니스학회지* 14.1 (2019): 449-459.
3. Shelley, Kristina J. "Developing the American time use survey activity classification system." *Monthly Lab. Rev.* 128 (2005).3.