

Travel Route Recommendation Utilizing Social Big Data

Yang Woo Yu*, Seong Hyuck Kim**, Hyeon Gyu Kim***

*Associate Professor, Dept. of Digital Contents Design, Ulsan College, Ulsan, Korea

**Senior Research Engineer, Dept. of Convergence Science, Sahmyook University, Seoul, Korea

***Associate Professor, Div. of Computer Science and Engineering, Sahmyook University, Seoul, Korea

[Abstract]

Recently, as users' interest for travel increases, research on a travel route recommendation service that replaces the cumbersome task of planning a travel itinerary with automatic scheduling has been actively conducted. The most important and common goal of the itinerary recommendations is to provide the shortest route including popular tour spots near the travel destination. A number of existing studies focused on providing personalized travel schedules, where there was a problem that a survey was required when there were no travel route histories or SNS reviews of users. In addition, implementation issues that need to be considered when calculating the shortest path were not clearly pointed out. Regarding this, this paper presents a quantified method to find out popular tourist destinations using social big data, and discusses problems that may occur when applying the shortest path algorithm and a heuristic algorithm to solve it. To verify the proposed method, 63,000 places information was collected from the Gyeongnam province and big data analysis was performed for the places, and it was confirmed through experiments that the proposed heuristic scheduling algorithm can provide a timely response over the real data.

▶ **Key words:** Social big data, Travel route recommendation, Shortest path, Travelling Salesman Problem, Hierarchical search

[요 약]

최근 여행에 대한 관심이 높아지면서, 번거로운 여행 일정을 대신 수립해주는 여행 일정 추천 서비스에 대한 연구가 활발히 진행되고 있다. 여행 일정 추천에 있어 가장 중요하면서도 공통적으로 제시되는 목표는 여행 목적지 근처의 인기 관광지를 포함한 최단 거리 여행 경로를 제공하는 것이다. 다수의 기존 연구에서는 개인 맞춤형 스케줄 제공에 초점을 맞추었으며, 사용자의 여행 이동 경로 이력이나 SNS 리뷰가 존재하지 않을 경우 설문 조사가 필요한 문제점이 있었다. 또한 최단 거리를 계산할 때 발생할 수 있는 현실적인 문제점도 명확히 지적되지 않았다. 이와 관련하여, 본 논문에서는 소셜 빅데이터를 활용하여 인기 관광지를 알아내기 위한 정량화된 방법을 소개하고, 최단 거리 알고리즘 적용시 발생할 수 있는 문제점과 이를 해결하기 위한 휴리스틱 알고리즘을 함께 제시한다. 제안 방법을 검증하기 위해, 경상남도를 대상으로 63,000여 개의 플레이스 정보를 수집하고 빅데이터 분석을 수행했으며, 실험을 통해 제안한 휴리스틱 스케줄링 알고리즘이 실제 데이터 상에서 실시간 처리가 가능함을 확인하였다.

▶ **주제어:** 소셜 빅데이터, 여행 경로 추천, 최단 경로, TSP, 계층적 탐색

-
- First Author: Yang Woo Yu, Corresponding Author: Hyeon Gyu Kim
 - *Yang Woo Yu (soft@uc.ac.kr), Dept. of Digital Contents Design, Ulsan College
 - **Seong Hyuck Kim (stellarlightttt@gmail.com), Dept. of Convergence Science, Sahmyook University
 - ***Hyeon Gyu Kim (hgkim@syu.ac.kr), Div. of Computer Science and Engineering, Sahmyook University
 - Received: 2022. 03. 24, Revised: 2022. 04. 29, Accepted: 2022. 05. 10.

I. Introduction

IT 기술의 발전으로 인해 단체 관광에서 개별 관광으로 패러다임 전환이 이루어지고 있다[1]. 개별 관광에 있어 사용자들은 관광 관련 정보 검색을 위해 모바일을 가장 보편적으로 이용하며, 콘텐츠 검색을 위한 채널로 SNS(61%), 블로그(34%) 등을 주로 이용하는 것으로 확인되었다[2]. 이에 반해, 수집된 정보를 바탕으로 여행 스케줄을 직접 작성하는 일은 많은 노력을 필요로 한다. 이러한 문제를 해결하기 위해, 하나투어의 트리플[3]이나 한국관광공사의 대한민국구석구석[4]과 같은 여행 스케줄 자동 생성 서비스들이 제공되고 있으며, 여행 일정 추천을 위한 다양한 연구가 함께 진행되었다.

여행 스케줄 생성에 있어 일반적으로 고려되는 사항은 아래와 같다.

- 여행 목적지에서 가장 인기 있는 관광지(Tour spots)는 어디인가?
- 관광지 근처의 인기 맛집이나 숙소는 어디인가?
- 관광지 간의 이동 거리는 얼마인가? 어떤 이동 경로가 최단거리인가?
- 이동 경로에 따른 여행 비용은 어느 정도인가?

위 내용을 알아내려면 먼저 여행 목적지 근처에 어떤 관광지나 맛집, 숙소가 있는지 파악할 수 있어야 한다. 그리고 SNS와 블로그에 남겨진 사용자들의 리뷰를 통해 어떤 플레이스가 인기 있는지 알아내야 한다. 이는 여행 스케줄 생성에 있어서도 소셜 빅데이터 분석이 필요함을 의미한다.

이와 관련하여, [5-10]을 포함한 다수의 연구에서 인기 관광지 정보를 얻기 위해 소셜 빅데이터를 활용하였으며, 이들은 주로 개인 맞춤형 여행 스케줄 제공 측면에서 차별화된 방법을 제시하고자 하였다. 개인 맞춤형 스케줄을 제공하려면 사용자의 관심 분야 파악이 필요하며, 이를 위해 사용자의 이동 경로 이력이나 사용자가 작성한 SNS 리뷰를 필요로 한다. 문제는 이들 자료가 존재하지 않을 경우 알고리즘의 정확도가 현저히 저하될 수 있다는 점이다. 사전 설문 조사를 통해 어느 정도 문제 해결이 가능하나, 이 역시 사용자가 다수의 문항을 직접 입력해야 한다는 점에서 한계가 있다.

본 논문에서는 소셜 빅데이터 분석을 통해 관광지의 대중적인 인기 정도를 정량화하고, 이를 여행 스케줄 추천에 적용하기 위한 방법에 대해 소개한다. 제안 방법은 한국관광공사에서 제공하는 Tour API[11]를 통해 지역별 관광지 정보를 수집한다. 지역별 맛집이나 숙소 정보는 소상공인진흥공단[12]에서 제공하는 공공데이터를 이용한다. 지역

별 관광지, 맛집, 숙소 등의 플레이스 정보가 준비되면, 각각의 플레이스 별로 소셜 빅데이터 분석을 수행한다. 이를 위해 [13, 14]에서 소개한 방법을 이용하며, 분석을 통해 정제된 리뷰 수를 바탕으로 플레이스별로 정량화된 인기 점수를 산정해 낸다.

플레이스별 정량화된 점수를 통해 대중적으로 인기 있는 플레이스 정보를 파악할 수 있으며, 이들로 구성된 여행 경로를 제공함으로써 사용자에게 대한 사전 정보가 없어도 만족도가 높은 여행 스케줄을 제공할 수 있다. 제안 방법에서는 TSP(Travelling Salesman Problems) 알고리즘[15]을 이용하여 플레이스 간의 최단거리 여행 경로를 찾아낸다. TSP를 이용할 경우, 모든 가능한 조합을 체크해야 하므로 처리 시간이 기하급수적으로 증가할 수 있다. 이와 관련하여 제안 방법에서는 휴리스틱 알고리즘의 일종인 계층적 탐색(Hierarchical search) 방법[16]을 적용하여 처리 시간을 줄이고자 하였다.

본 논문의 구성은 다음과 같다. 2장에서는 여행 경로 추천과 관련한 기존 연구들을 소개한다. 3장에서는 플레이스 및 소셜 리뷰 수집 방법과 플레이스 별 인기 점수 산정 방법, 그리고 인기 점수를 활용한 최단 경로 생성 알고리즘 등에 대해 설명한다. 4장에서는 온라인에서 실제 수집된 데이터를 바탕으로 제안 알고리즘의 성능을 검증하기 위한 실험 결과를 제시한다. 5장에서는 결론 및 추후 연구 방향 제시로 마무리한다.

II. Related Work

여행 스케줄 추천 알고리즘은 공통적으로 최대한 많은 인기 관광지를 포함하는 동시에 이동 경로를 최소화하는 것을 목표로 하며[5], 알고리즘에 입력되는 파라미터나 데이터 종류, 개인 맞춤형 스케줄 지원 여부 등의 측면에서 차이가 있다.

[17]에서는 여행의 특성을 최대한 반영하기 위해, 여행지 외에 숙소, 공항, 기차역, 교통편, 여행경비, 여행지별 선호도, 방문가능시간 등을 추천에 고려하였으며, 이들 파라미터는 TripAdvisor로부터 얻을 수 있다고 가정하였다. 최단거리 여행 경로를 찾기 위해 여행지, 숙소, 공항, 기차역을 대상으로 TSP 알고리즘을 적용하였다. [18]은 입력 파라미터로 출발지/도착지, 교통편, 관광명소 및 쇼핑 등의 여행 목적, 여행지의 범위를 입력받아 스케줄링을 수행하며, 최단거리 여행 경로를 찾기 위해 관광지를 대상으로 A* 및 다익스트라(Dijkstra) 알고리즘을 이용하였다.

여행 스케줄을 실시간에 동적으로 생성하거나 수정하는 방법에 대한 연구도 진행되었다. [19]는 밀집도, 날씨, 교통 정보 등 예상할 수 없는 문제점을 개선하고자 IoT 기반의 실시간 스케줄링 방법을 제안하였으며, 경로 생성을 위해 Greedy 알고리즘을 이용하였다. [6]은 핫플레이스 검색 모듈을 제공하여, 사용자가 관광지 인근의 맛집이나 카페, 호텔 등을 검색하고 여행 스케줄에 추가할 수 있도록 제공하였으나, 사용자의 수동적인 입력이 필요하다는 점이 한계로 지적되었다.

개인 맞춤형 여행 스케줄을 지원하고자 하는 시도는 다수의 연구에서 진행되었으며, 공통적으로 사용자들의 선호도 파악을 위해 여행 이동 경로 이력이나 SNS 리뷰 글들이 존재한다는 가정을 바탕으로 연구가 진행되었다. [20]은 사용자의 GPS 정보를 수집해 이동 패턴을 추출하고, 패턴이 유사한 사용자의 여행 일정을 추천하였다. [7]은 위키피디아 태그 정보를 이용해 관광지별로 시맨틱 벡터를 생성하고, Flickr[21]를 통해 수집된 사용자의 선호도와 비교하여 유사도가 가장 큰 여행 일정을 추천하였다. [5]는 [7]에서 제안한 내용을 바탕으로 정확도를 높이기 위해 장소별 체류 시간을 일정 추천에 고려하였다.

[22]은 설문 형식의 데이터 수집과 사용자의 상품 페이지 방문 패턴 분석을 통해 개인화된 여행 스케줄 생성 방법을 제안하였다. [8]은 페이스북, 트위터, 인스타그램 등의 SNS 플랫폼으로부터 사용자가 특정 지역에서 체크인한 정보와 해당 지역의 평가나 조언 등의 의견을 남긴 피드 글들을 분석하고 제공 가능한 서비스에 대해 기술하였으며, [9]에서는 해당 서비스 중 SNS 피드 정보를 활용한 개인 맞춤형 여행 스케줄링 추천 서비스에 대해 소개하였다. [10]은 한글로 작성된 SNS 피드 글을 분석하여 키워드를 추출한 후 이를 개인 맞춤형 여행 일정 추천에 활용하였다. 해당 방법에서는 SNS 수집을 위해 Twitter Streaming API[23]와 관광지 정보 수집을 위해 한국관광공사의 Tour API를 이용하였다.

[24]는 협업 필터링을 이용하여 개인별 맞춤형 여행 스케줄을 제공하고자 하였다. 협업 필터링은 초기 사용자가 없을 경우 추천 신뢰도가 떨어지는 문제점이 있으므로, 이를 보완하기 위해 한국관광공사의 Tour API를 이용하여 TF-IDF 방식을 통해 여행지별로 키워드를 추출한 후, 회원 가입 시 설문을 통해 조사된 사용자 선호도와 비교하여 유사도가 높은 여행지를 추천하였다. [25, 26]은 사용자의 만족도를 높이기 위해 장소별 체류시간과 장소 간 이동시간을 함께 고려하였으며, 유전 알고리즘을 적용하여 개인 맞춤형 여행 스케줄을 생성하였다.

앞서 언급한 바와 같이, 개인 맞춤형 여행 스케줄 관련 한 기존 연구에서는 사용자의 관심 분야 파악에 필요한 이동 경로 이력이나 리뷰 등의 사전 정보가 없을 경우, 알고리즘의 정확도가 현저히 저하될 수 있다는 문제점이 있다. 아래에서는 소셜 빅데이터 분석을 통해 대중적으로 인기 있는 플레이스 정보를 추출하고, 이를 기반으로 여행 경로를 생성하기 위한 방법에 대해 소개한다.

III. Proposed Method

본 논문에서는 소셜 빅데이터를 활용한 여행 경로 추천 방법을 제안한다. 제안하는 방법은 플레이스별 인기 점수 산정 모듈과 여행 스케줄 생성 모듈로 구성된다. 전자는 소셜 빅데이터를 활용하여 지역별 유명 관광지와 맛집, 숙소에 대한 정량화된 인기 점수를 산정하며, 후자는 사용자가 입력한 여행 목적지의 인기 플레이스들을 추출하고 이들을 연결한 최단거리 여행 경로를 생성하는 역할을 수행한다.

1. Popularity Score Estimation

플레이스별 인기 점수 산정 모듈에서는 먼저 스케줄링의 대상이 되는 플레이스 정보를 수집한다. 앞서 언급한 바와 같이, 관광지 정보는 한국관광공사에서 제공하는 Tour API[11]를 이용하며, 맛집이나 숙소 정보는 소상공인진흥공단[12]에서 제공하는 공공데이터를 이용한다.

지역별로 관광지, 맛집, 숙소 등의 플레이스 정보가 준비되면, 각각의 플레이스에 대해 소셜 빅데이터 분석을 수행한다. 제안 방법에서는 [13, 14]에 소개된 소셜 빅데이터 분석 방법을 이용하며, 플레이스 이름과 지역명을 검색어로 입력하면, 리뷰 수집, 형태소 분석, 노이즈 리뷰 필터링 단계를 거쳐 주어진 검색어와 연관성이 높은 리뷰만을 추출해 낸다.

보다 자세히, 리뷰 수집 단계에서는 네이버, 다음, 구글, 페이스북, 인스타그램 등의 온라인 사이트로부터, 주어진 검색어를 만족하는 SNS 피드 글이나 블로그 리뷰 등을 수집한다. 리뷰 수집에는 웹 크롤러나 온라인 사이트에서 제공되는 검색 API가 이용될 수 있으며, 수집된 결과에는 주어진 검색어가 포함된 모든 리뷰가 포함된다. 문제는 수집된 리뷰 중 연관성이 떨어지는 리뷰가 다수 존재할 수 있다는 점이다. 예를 들어, "스타벅스 강남"을 검색어로 설정할 경우, 수집된 리뷰에는 "강남 스타벅스 2층의 미용실"이나 "강남 스타벅스 건물 왼쪽의 김밥집" 등 검색어는 포함하나 실제 다른 플레이스를 지칭하는 리뷰들이 다수 포함

될 수 있다. 이들을 걸러내기 위한 목적으로 형태소 분석 및 노이즈 리뷰 필터링 단계가 수행된다.

형태소 분석 단계에서는 수집된 리뷰의 단어를 표준화하는 역할을 수행한다. 해당 과정에서는 형용사나 동사의 활용형을 원형으로 변환한다. 예를 들어, "간다", "갔다", "갔었다" 등의 활용형은 "가다"라는 원형으로 변환한다. 그리고 분석에 영향을 미치지 않는 부사나 조사는 제거될 수 있다. 특히 소셜 리뷰는 신조어나 고유 명사를 포함하는 경우가 많으며, 한나눔[27] 등의 사전에 기반한 기존 형태소 분석기로는 정확한 형태소 분석 결과를 얻기 어려울 수 있다. (사전에는 신조어나 고유 명사가 포함되지 않기 때문이다.) 따라서 제안 방법에서는 사전을 이용하지 않고 리뷰에 나타난 단어들의 빈도수 등을 활용한 통계 데이터 기반의 형태소 분석 기법[28, 29]을 활용한다.

노이즈 필터링 단계에서는 수집된 각각의 리뷰를 단어 벡터의 형태로 변환한다. 그리고 해당 벡터를 기반으로 베이저언 네트워크(Bayesian Network)나 LSTM(Long Short-Term Memory) 등의 답러닝 기법을 이용하여 주어진 검색어와 연관이 없는, 소위 "노이즈" 리뷰를 걸러내는 필터링 작업을 수행한다. 높은 정확도가 요구되지 않을 경우, TF-IDF 등의 간단한 방법을 이용해 필터링에 소요되는 시간을 줄일 수도 있다.

노이즈 리뷰 필터링 단계가 종료되면 주어진 검색어와 연관성이 높은 "진성" 리뷰만 남게 된다. 일반적으로 진성 리뷰의 수는 네티즌의 관심도나 인기 정도와 비례한다. 따라서 플레이스 별로 진성 리뷰 수를 이용하여 해당 플레이스에 대한 인기를 정량적으로 측정할 수 있다. 제안 방법에서는 플레이스의 인기 점수 산정을 위해 아래 식을 이용한다.

$$f = \sum_{i=1}^N \{n_i w_i - m_i u_i\} \quad (1)$$

위 식은 월별로 수집된 리뷰 수에 가중치를 곱한 값을 합산하고 있다. 보다 자세하게, n_i 와 w_i 는 i 번째 달의 긍정(positive) 리뷰 수와 가중치에 해당하며, 가중치는 가장 최근 달부터 순차적으로 값이 줄어들도록 정의된다. 그리고 m_i 와 u_i 는 i 번째 달의 부정(negative) 리뷰 수와 가중치에 해당하며, 일반적으로 u_i 는 w_i 에 비해 큰 값으로 정의된다. 아래 표는 최근 1년($N = 12$)동안 수집된 리뷰 수를 바탕으로 인기지수를 산정한 예를 보여준다.

Table 1. Estimation of the popularity score based on the number of reviews collected for one year ($N = 12$)

	12	11	10	9	8	7	6	5	4	3	2	1	Sum
n_i	27	25	23	17	12	15	13	16	19	22	21	29	
w_i	1.5	1.4	1.3	1.2	1.1	1	1	0.9	0.8	0.7	0.6	0.5	
$n_i w_i$	40.5	35	29.9	20.4	13.2	15	13	14.4	15.2	15.4	12.6	14.5	239.1
m_i	3	2	2	0	1	0	1	0	1	1	2	2	
u_i	3	2.8	2.6	2.4	2.2	2	2	1.8	1.6	1.4	1.2	1	
$m_i u_i$	9	5.6	5.2	0	2.2	0	2	0	1.6	1.4	2.4	2	31.4
Popularity score													207.7

2. Scheduling Algorithm

여행 스케줄 생성 모듈에서는 플레이스별로 산정된 인기 점수를 바탕으로 먼저 스케줄링 대상을 선정한다. 대상 선정은 여행 목적지에 따라 아래와 같은 방식으로 선정될 수 있다.

- 여행 목적지가 사용자의 현재 위치와 다른 지역인 경우: 지역 내 모든 플레이스를 스케줄링 대상으로 선정. 예를 들어 목적지가 경상남도 통영시일 경우, 통영에 포함된 모든 관광지 및 맛집, 숙소를 대상으로 함
- 여행 목적지가 사용자의 현재 위치 근처인 경우: 사용자 위치를 기준으로 반경 R km (일반적으로 $R \leq 10$) 이내의 플레이스를 스케줄링 대상으로 선정

스케줄링 대상 선정이 완료되면, 인기순으로 플레이스를 정렬한 후 여행 스케줄을 생성한다. 스케줄의 시작점은 여행 목적지에서 가장 인기 있는 관광지로 설정한다. (가장 인기 있는 관광지의 경우 사용자들이 먼저 방문하고자 할 확률이 높기 때문이다.) 아래는 경상남도 거제시 관광지를 인기 점수 순으로 나열한 예이다. 거제시의 경우, 여행 스케줄의 시작점은 "외도보타니아"가 된다.

Table 2. Tour spots and their popularity scores in Geoje city

ID	Tour Spot	Score	ID	Tour Spot	Score
1	외도보타니아	1024	8	거제자연휴양림	238
2	해금강	597	9	구조라해수욕장	227
3	공곶이	448	10	거제조선해양문화관	202
4	포로수용소 유적공원	396	11	와현해수욕장	201
5	씨월드	389	12	해금강테마 박물관	140
6	거제관광 모노레일	252	13	신선대전망대	104
7	명사해수욕장	248			

제안 방법에서는 주어진 모든 관광지를 방문하는 최단 거리 여행 경로를 찾기 위해 TSP 알고리즘을 이용한다. 논의의 편의를 위해 모든 관광지가 서로 연결되어 있고, 임의의 두 위치 간의 거리는 km 단위의 직선거리로 가정한다. 이러한 가정을 바탕으로, 관광지들의 GPS 정보를 이용해 노드 간의 가중치가 주어진 완전 그래프(Weighted complete graph)를 생성할 수 있으며, 최단 거리 여행 경로를 찾는 문제는 해당 그래프에서 가장 작은 가중치를 가지는 해밀턴 순환(Hamiltonian path)[15]을 구하는 문제가 된다. 단, 시작점과 종료점이 같은 해밀턴 순환과는 달리 제안 방법에서는 시작점으로 반드시 돌아와야 한다는 제약 조건을 강제하지 않으며, 이 경우에도 계산 복잡도는 변하지 않는다.

완전 그래프의 형태로 문제가 주어질 경우, 모든 가능한 경로를 찾는 문제는 모든 순열을 찾는 문제로 단순화될 수 있다. 위 예제의 경우, 13개의 관광지를 모두 순회하는 순열을 찾아야 하며, 경우의 수는 13!이 된다. 모든 순열이 구해지면, 각 순열에 포함된 관광지들의 거리 합을 계산하여, 최소값을 가지는 순열을 해로 반환한다. 아래는 주어진 모든 관광지를 방문하는 최단거리 여행 경로를 찾기 위한 알고리즘을 JavaScript로 구현한 코드이다.

```

1 function getShortestPath(graph, permut) {
2   var min = 9999, spath = null;
3   for (var i=0; i<permut.length; i++) {
4     var permut = permut[i];
5     var cost = 0;
6     for (var j=0; j<permut.length-1; j++) {
7       var idx1 = permut[j] - 1;
8       var idx2 = permut[j+1] - 1;
9       cost += graph[idx1][idx2];
10    }
11    if (cost < min) {
12      min = cost;
13      spath = permut;
14    }
15  }
16  return spath;
17 }

```

Fig. 1. Algorithm for finding the shorted path from all possible permutations combined by tour spot IDs

getShortestPath() 함수는 graph와 permut를 파라미터로 입력받는다. permut는 관광지 ID로 구성될 수 있는 모든 가능한 순열을 배열 형태로 포함하며, graph는 관광지 간의 거리 정보를 지닌 2차원 배열에 해당한다. 예를 들어, ID가 1, 2, 3인 세 개의 관광지가 있고, 각각의 GPS 정보로부터 1과 2간의 거리는 5, 2와 3간의 거리는 7, 1과 3간의 거리는 9로 계산되었다고 가정할 경우, graph는 아래의 2차원 배열로 표현된다.

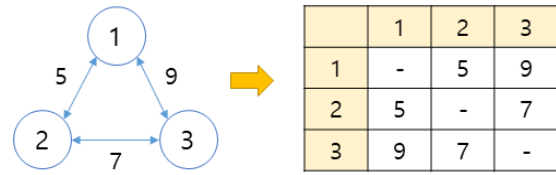


Fig. 2. Algorithm for finding the shorted path from all possible permutations combined by tour spot IDs

min과 spath는 각각 최단 거리 값과 최단 경로를 나타내며, permut에 포함된 각각의 순열에 대해 5~10행을 통해 거리 값을 계산한다. 계산된 거리 값이 기존의 min 값보다 적을 경우, 계산 값을 min 값으로 수정하고 해당 순열을 spath에 저장한다. permut의 모든 순열에 대해 위 과정을 반복한 후 얻어진 spath를 결과 값으로 반환한다.

한편, 위 알고리즘은 모든 가능한 조합을 체크해야 하므로, 관광지 수에 따라 처리 시간이 기하급수적으로 늘어날 수 있다. 이러한 문제를 해결하기 위해, 제안 방법은 계층적 탐색(Hierarchical search)[16]을 이용해 고려해야 하는 경우의 수를 줄이고자 하였다. 계층적 탐색에서는 그룹 간 탐색과 그룹 내 노드 간의 탐색을 분리하여 수행한다. 예를 들어, 관광지간의 거리가 N km 이내의 관광지들을 그룹으로 묶을 수 있다. 아래 표는 N을 1로 설정한 후 관광지를 그룹으로 묶은 예를 보여준다. 아래에서 관광지 그룹의 인기 점수는 그룹에 포함된 관광지들의 점수를 모두 합산한 값이 된다.

Table 3. Tour spot groups and their popularity scores in Geoje city

ID	Tour Spot Group	Score
1	외도보타니아	1024
2	씨월드, 거제조선해양문화관, 와현해수욕장	793
3	해금강, 신선대전망대	701
4	포로수용소유적공원, 거제관광모노레일	648
5	공곶이	448
6	명사해수욕장	248
7	거제자연휴양림	238
8	구조라해수욕장	227

위 예제의 경우, 13개의 관광지가 8개의 그룹으로 축소되었으며, ID가 2, 3, 4인 관광지 그룹의 경우 각각 3, 2, 2개의 관광지를 멤버로 포함한다. 그 결과 계산해야 하는 경우의 수가 13!, 약 62억 개에서 (8! + 3! + 2! + 2!) ≈ 8!, 약 4만 개로 크게 줄어든다.

결과적으로, 휴리스틱을 적용한 제안 방법에서는 관광지 그룹을 대상으로 TSP 알고리즘을 적용하여 최단 여행 경로를 추출한다. 제안 방법에서는 N을 1로 설정한다. 그

이유는 1 km 이내일 경우 도보로 이동 가능한 거리이기 때문이다. 일반적으로 N이 커져서 그룹 수가 줄면 알고리즘의 성능이 향상된다. 극단적인 경우, N이 너무 커서 모든 관광지가 1개의 그룹으로 병합되면 휴리스틱을 적용하기 전과 성능이 동일해진다.

제안 알고리즘의 아이디어는 단순하나, TSP 적용에 있어 계산해야 할 경우의 수를 크게 줄일 수 있다는 점에서 의미가 있다. 그리고 이는 제안 방법을 실제계의 스케줄링 문제에 실제 적용 가능한지 여부를 판가름한다. 아래 표는 관광지 수에 따라 계산에 필요한 순열의 수와 메모리 크기를 보여준다. 표에서 볼 수 있듯이, 관광지의 수가 12개만 되더라도 순열 계산을 위해 약 5.7GB의 메모리가 필요하며, 13개가 되면 80GB 이상의 메모리가 필요해 일반 컴퓨터로는 계산이 불가능한 수준임을 알 수 있다.

Table 4. Required memory size according to the number of places

Number of places	Number of permutations	Size of input array (bytes)	Memory size (bytes)
10	3,628,800	10	36,288,000
11	39,916,800	11	439,084,800
12	479,001,600	12	5,748,019,200
13	6,227,020,800	13	80,951,270,400

제안 방법을 이용할 경우, 관광지 그룹 수와 그룹 내에 포함되는 관광지 수가 계산 가능한 수준이 되도록 N 값을 조정할 수 있다. 예를 들어 그룹 수와 그룹 내에 포함되는 관광지 수를 10개로 설정한다면, 스케줄링에 포함되는 관광지 수는 100개가 된다. 실제계에서는 100개 이상의 관광지를 대상으로 스케줄을 계산하는 경우가 거의 없다. 따라서 제안 방법은 N 값의 조정을 통해 실제계의 관광 스케줄링 문제에 효과적으로 적용될 수 있다. 이와 관련하여 4장에서는 온라인상에서 수집된 실제 데이터를 바탕으로 진행된 실험을 통해, N을 1로 설정하더라도 실시간 여행 스케줄 계산 및 추천이 가능함을 보여준다.

3. Consideration for Detail Schedules

제안 방법에서는 관광지 외에 식당과 숙박 등의 플레이스 정보도 함께 지니고 있으므로, 각 관광지 근처의 인기 식당과 숙박 장소를 함께 추천해줄 수 있다. 관광지를 중심으로 반경 M km 이내의 식당과 숙박 장소를 인기 순으로 정렬한 후, 가장 인기 있는 K개의 플레이스를 함께 추천할 수 있다. M 값은 걸어서 이동하는 경우와 차로 이동하는 경우에 따라 다르게 설정될 수 있으며, 전자의 경우는 1 km, 후자의 경우는 2~4km 정도로 설정될 수 있다.

Table 5. Tour spot groups and popular restaurants and lodge places near the groups in Geoje city

ID	Tour Spot Group	Restaurants	Lodge
1	외도보타니아	유경식당	-
2	씨월드, ...	어부빚집	홈포레스트리조트
3	해금강, ...	-	블루마우리조트
4	포로수용소 ...	하면옥, 두리원 ...	장승포비치호텔
5	공곶이	바룻, 어방가, ...	소노캄거제
6	명사해수욕장	-	라마다스위츠
7	거제자연휴양림	머그정, 노자산 ...	흑진주마리나리조트
8	구조라해수욕장	에이제게장백반	-

위 내용은 Table 3의 관광지 그룹 별로 추천된 인기 식당과 숙박 장소 리스트를 보여준다. 해당 내용을 바탕으로 여행 일자별 세부 일정을 구성할 수 있으며, 일정 수립을 위해서는 관광지별 체류 시간 등에 대한 보다 상세한 정보가 필요하다. 관련 알고리즘은 향후 연구에서 보다 구체화 시킬 예정이다.

IV. Experimental Results

1. Data Sets

제안 알고리즘의 실험을 위해 경상남도 지역의 플레이스 데이터를 수집하였다. 앞서 언급한 바와 같이, 관광지의 경우 한국관광공사의 Tour API를 이용하며, 맛집이나 숙소 정보는 소상공인진흥공단에서 제공하는 공공데이터를 이용하였다. 경상남도 전체를 대상으로 수집된 플레이스 수는 63,819개였다. 이들 플레이스에 대해 네이버 검색 API[30]를 이용해 플레이스 별로 블로그 글들을 수집하고 필터링하였다.

아래 표는 경상남도 내 18개 지역(시군구)을 대상으로 수집된 플레이스의 수와 진성 리뷰 수를 보여준다. 리뷰 수는 최근 3년 이내에 작성된 블로그 글을 대상으로 집계하였으며, 진성 리뷰의 총합은 475,871개로 조사되었다.

Table 6. Numbers of places and true reviews in 18 cities of the Gyeongnam province

City name	# of places	# of reviews	City name	# of places	# of reviews
거제	4,991	57,634	의령	537	3,232
거창	1,334	4,566	진주	6,443	41,878
고성	1,246	5,354	창녕	1,605	5,961
김해	7,404	53,544	창원	19,423	121,005
남해	1,498	27,881	통영	3,862	39,542
밀양	2,603	16,853	하동	1,310	12,982
사천	2,645	15,525	함안	1,161	6,416
산청	1,004	6,943	함양	937	5,307
양산	4,823	44,724	합천	993	6,524

위 데이터 중 관광지와 관련한 플레이스 수와 리뷰 수는 아래와 같다.

Table 7. Numbers of tour spots and their true reviews in 18 cities of the Gyeongnam province

City name	# of places	# of reviews	City name	# of places	# of reviews
거제	29	3,999	의령	9	1,104
거창	16	2,019	진주	9	2,536
고성	11	2,205	창녕	10	1,944
김해	25	4,099	창원	27	4,231
남해	29	4,889	통영	24	5,861
밀양	17	4,101	하동	17	2,606
사천	19	2,570	함안	13	2,046
산청	17	3,071	함양	21	3,330
양산	17	3,236	합천	15	2,733

2. Performance Test

제안 알고리즘의 성능 검증을 위해, 기존 TSP 알고리즘과 제안한 Heuristic TSP 알고리즘의 실행 시간을 비교하였다. Intel Core i5-8250U 1.8GHz CPU, 8GB 메모리 환경에서 실험이 진행되었다.

실험 데이터로는 Table 6의 관광지 정보 중, 인기 점수가 100점 이상인 관광지만을 선별하여 이용하였다. 이는 인기 관광지만을 대상으로 스케줄 추천이 이루어지는 경우가 많으며, 추천 대상에 모든 관광지가 포함될 경우 사용자의 만족도가 떨어질 수 있기 때문이다. 실험 데이터에 포함된 관광지의 수와 그룹 수는 아래와 같다.

Table 8. Numbers of tour spots with popularity score ≥ 100 and their groups in 18 cities of the Gyeongnam province

City name	# of places	# of groups	City name	# of places	# of groups
거제	14	8	의령	3	3
거창	4	4	진주	7	4
고성	6	6	창녕	2	2
김해	17	8	창원	17	8
남해	16	7	통영	15	9
밀양	10	8	하동	8	7
사천	8	6	함안	5	5
산청	6	5	함양	11	7
양산	7	7	합천	6	6

실험에는 관광지 수가 11개 이하인 도시 5개(밀양, 사천, 양산, 진주, 함양)의 데이터가 이용되었다. 그 이유는 Table 4에서도 언급한 바와 같이, 관광지 수가 12개일 경우 순열 계산을 위해 약 5.7GB의 메모리가 필요하며, 실제 알고리즘 실행 시에도 메모리 오버플로우 문제가 발생했기 때문이다.

아래 표는 기존 및 제안 알고리즘의 실행 시간을 보여준다. 10회 실험을 통해 얻어진 평균값을 측정하였다.

Table 9. Average execution times of the existing TSP and the heuristic TSP algorithms for 5 cities in Gyeongnam province (unit: milliseconds)

City Name	Existing TSP algorithm	Heuristic TSP algorithm
밀양	160.95	3.92
사천	3.72	0.21
양산	1.74	1.94
진주	1.29	0.04
함양	1701.47	3.90

알고리즘의 성능을 정량적으로 비교하기 위해, Table 9에서 시군구 별 5개의 측정값을 합산하여, 알고리즘 별로 평균값을 도출하였다. 아래 그림은 두 알고리즘의 평균 실행 시간을 보여준다. 기존 알고리즘에 비해 제안 알고리즘이 약 180배 빠른 성능을 제공하는 것으로 조사되었다.

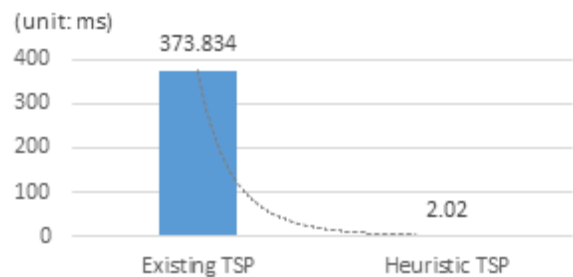


Fig. 3. Average execution times of the existing TSP and the heuristic algorithms

마지막으로 관광지 그룹 수에 따른 제안 알고리즘의 실행 시간 추이를 비교하였다. 관광지 그룹이 10개 이하일 경우, 실행 시간은 0.2초 미만으로 실시간 스케줄링이 가능한 수준이었다. 한편 그룹 수가 11개로 증가할 경우 실행 시간이 1.8초 정도로 다소 증가하나, 이 역시 사용자가 받아들일 수 있는 수준의 응답 시간인 것으로 조사되었다.

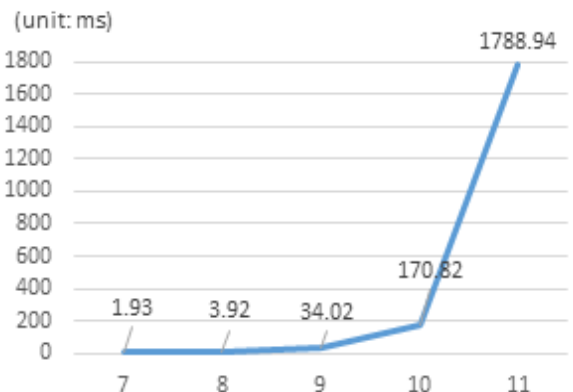


Fig. 4. Average execution time of the heuristic TSP algorithm according to the number of tour spot groups

V. Conclusion and Future Work

본 논문에서는 소셜 빅데이터를 활용한 여행 경로 추천 방법을 제안하였다. 제안 방법은 먼저 여행 목적지 근처의 인기 관광지 및 맛집, 숙소 정보를 파악하기 위해 소셜 빅데이터 분석을 수행하며, 플레이스 별 인기 점수는 수집된 소셜 리뷰 수를 기반으로 정량적으로 산정된다. 인기 점수 산정이 완료되면 점수가 높은 관광지들을 연결한 여행 경로 생성 작업이 수행된다. 제안 방법에서는 주어진 관광지를 모두 포함한 최단 경로를 얻기 위해 TSP 알고리즘을 이용한다. 단, TSP 알고리즘은 고려 대상의 수에 따라 계산에 필요한 메모리가 기하급수적으로 증가한다. 이러한 문제를 해결하기 위해, 제안 방법에서는 인근 관광지들을 그룹화하여 스케줄링 대상을 축소하는 계층적 탐색 기반의 휴리스틱 방식을 채택하였다. 제안 방법의 검증을 위해 경상남도 관광지, 맛집, 숙소 정보를 온라인으로부터 수집하여 소셜 빅데이터 분석을 수행하였으며, 실험 결과 휴리스틱 방식을 적용한 알고리즘이 그렇지 않은 알고리즘에 비해 약 180배 빠른 성능을 보였으며 실시간 처리 역시 가능한 수준임을 확인하였다.

현재, 제안 알고리즘은 국내 관광지를 대상으로 한 최단 경로 이외에 일정별 세부 가이드나 여행 경비 등에 대한 구체적인 여행 정보는 제공하지 못하고 있다. 또한 개인 맞춤형 여행 스케줄 역시 지원되지 않는다. 따라서 제안 알고리즘에서 위 내용을 지원하기 위한 연구를 지속할 예정이다.

ACKNOWLEDGEMENT

This research was supported by the Sahmyook University Research Fund in 2022.

REFERENCES

- [1] B. H. Joeng and K. H. Kim, "Analysis of Differences of Importance and Satisfaction of Smart Tourism IT Services Among Steps of Tourism Life Cycle," *The Journal of Internet Electronic Commerce Research*, Vol. 19, No. 1, pp. 289-306, 2019.
- [2] J. Y. Kim et al., "A Development of Optimal Travel Course Recommendation System Based on Altered TSP and Elastic Search Algorithm," *Journal of Korea Multimedia Society*, Vol. 22, No. 9, pp. 1108-1121, 2019.
- [3] Tripple, HanaTour, <https://play.google.com/store/apps/details?id=com.titicacacorp.triple&hl=ko>
- [4] VisitKorea, Korea Tourism Organization, <https://play.google.com/store/apps/details?id=com.visitkorea.kr&hl=ko>
- [5] S. H. Park and S. Park, "Personalized Itinerary Recommendation System based on Stay Time," *KIISE Transactions on Computing Practices*, Vol. 22, No. 1, pp. 38-43, Jan. 2016.
- [6] G. Kim, S. Park, S. Park, J. Yoon, and J. Ahn, "Design of Recommendation System for Self-directed Travel Planning," *Proceedings of KIIT Conference*, pp. 774-776, Nov. 2021.
- [7] I. Brillhante, J.A. Macedo, F.M. Nardi, R. Pergo, and C. Renso, "Where Shall We Go Today?: Planning Touristic Tours with Tripbuilder," *Proc. of the 22nd ACM international conference on Conference on information & knowledge management*, pp. 757-762, 2013.
- [8] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel, "Recommendations in Location-Based Social Networks: a Survey." *GeoInformatica*, Vol. 19, No. 3, pp.525-565, 2015.
- [9] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel, "Location-based and Preference-Aware Recommendation using Sparse Geo-Social Networking Data," *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, pp. 199-208, 2012.
- [10] J. Park, S. Yoo, and O Jeong, "Social Context-Aware based Travel Recommendation System," *Proceedings of KIISE Conference*, pp.1926-1928, June. 2016.
- [11] Tour API 3.0, Korea Tourism Organization, <https://kto.visitkorea.or.kr/kor/gov30/tourapi.kto>
- [12] Small Business Promotion Agency, <https://www.semas.or.kr>
- [13] H. G. Kim, "Developing a Big Data Analysis Platform for Small and Medium-Sized Enterprises," *Journal of the Korea Society of Computer and Information*, Vol. 25, No. 8, pp. 65-72, 2020.
- [14] H. G. Kim, "Social Big Data Analysis for Franchise Stores," *Journal of the Korea Society of Computer and Information*, Vol. 26, No. 8, pp. 39-46, 2021.
- [15] Traveling Salesman Problem, Wikipedia, https://en.wikipedia.org/wiki/Travelling_salesman_problem
- [16] F. Bacchus and Q. Yang, "The Expected Value of Hierarchical Problem-Solving," *Proc. of the 10th National Conference on Artificial Intelligence*, pp. 369-374, 1992.
- [17] I. Hwang and H. Kim, "New Mathematical Model for Travel Route Recommendation Service," *Journal of Korean Society of Industrial and System Engineering*, Vol. 40, No. 3, pp. 99-106, Sep. 2017.
- [18] J. H. Lee and J. H. Yang, "Efficient Travel Route Recommendation System Based on User Moving Route and Tourist Information," *Proceedings of the KIISE*, pp. 1616-1618, Dec. 2014.
- [19] C. Kim, E. Boo, S. Park, E. Lee, and Y. Ko, "Real-Time Travel

- Route Recommendation Service Based on IoT," Proceedings of Symposium of the Korean Institute of communications and Information Sciences, pp. 1084-1085, 2017.
- [20] H. Yoon, Y. Zheng, X. Xie, and W. Woo, "Smart Itinerary Recommendation Based on User-Generated GPS Trajectories," *Ubiquitous Intelligence and Computing*, Springer Berlin Heidelberg, Vol. 6406, pp. 19-34, Jan. 2010.
- [21] Flickr API, <https://www.flickr.com/services/developer/api/>
- [22] J. Park, H. Jeong, and H. Ru, "Personalized Travel Schedule Creation System Based on Recommendation System," *Proceedings of the KSCIC*, Vol. 25, No. 1, pp. 105-108, Jan. 2017.
- [23] Twitter Streaming API, <https://developer.twitter.com/en/docs/tutorials/consuming-streaming-data>
- [24] H. Park et al., "Travel Route Recommendation System Using Collaborative Filtering and Content-Based Recommendation," *Proceedings of the KIIT Conference*, pp. 452-454, Oct. 2020.
- [25] B. S. Wibowo and M. Handayani, "A Genetic Algorithm for Generating Travel Itinerary Recommendation with Restaurant Selection," *Proceedings of 2018 IEEE International Conference on Industrial Engineering and Engineering Management*, pp. 427-431, 2018.
- [26] S. H. Kwon and S. M. Park, "User Evaluation-based Travel Planning Modeling using Genetic Algorithm," *Journal of Korean Institute of Intelligent Systems*, Vol. 31, No. 4, pp. 311-317, Aug. 2021.
- [27] Hannanum, <http://semanticweb.kaist.ac.kr/hannanum/index.html>
- [28] H. G. Kim, "Efficient Keyword Extraction from Social Big Data Based on Cohesion Scoring," *Journal of the Korea Society of Computer and Information*, Vol. 25, No. 10, pp. 87-94, 2020.
- [29] Y. W. Yu and H. G. Kim, "Interactive Morphological Analysis to Improve Accuracy of Keyword Extraction Based on Cohesion Scoring," *Journal of the Korea Society of Computer and Information*, Vol. 25, No. 12, pp. 146-154, 2020.
- [30] Naver Search API, <https://developers.naver.com/docs/search/blog/>

Authors



Yang Woo Yu received the B.S. M.S. and Ph.D degrees in Computer Science from University of Ulsan, Korea, in 1995, 1997 and 2005, respectively. Dr. Yu joined the Department of the Digital Contents Design at

Ulsan College, Ulsan, Korea, in 2000. He is currently an Associate Professor in the Department of the Digital Contents Design at Ulsan College. He is interested in VR programming, web programming and big data processing.



Seong Hyuck Kim received the B.S. degree in Computer Science from Sahmyook University in 2017. He is currently a senior research engineer in OnGoodPlace Inc. He is interested in game programming, web

programming and big data processing.



Hyeon Gyu Kim received the B.S. and M.S. degrees in Computer Science from University of Ulsan, and Ph.D. degree in Computer Science from Korea Advanced Institute of Science and Technology, Korea, in 1997,

2000 and 2010, respectively. Dr. Kim joined the faculty of the Division of Computer Science and Engineering at Sahmyook University, Seoul, Korea, in 2012. He is currently an Associate Professor in the Division of Computer Science and Engineering, Sahmyook University. He is interested in big data processing, data stream processing, and mobile computing.