

2-b. 지도기반 추천 및 패키징 방식 상세 설명

Service Introduction

[그림] : 60 대 어르신 | GPS | 문화체험

두 번째로 제안할 서비스는 60 대 이상 어르신들을 위한 GPS 기반의 문화 프로그램 패키징 서비스입니다. 앞서 언급 했듯이, 60 대 이상 어르신들이 문화시설을 이용하지 못하는 주된 이유는 주변 가맹점에 대한 정보 부족과 장거리 이동에 대한 부담 때문입니다. 또한, 어르신들은 문화 프로그램을 찾아보고, 구매를 결정할 때까지 다른 연령층 보다 많은 노력과 시간이 소요되기 때문에 결정 과정의 장벽이 높고 다소 비합리적인 결정을 내리는 특징이 있습니다. 따라서 저희는 이러한 문제점을 해결하고자, 사용자의 접근성을 고려하여 전체적인 탐색 및 결정 시간을 단축하고, 보다 다양한 문화시설 체험의 기회를 제공할 수 있는 서비스를 제안하고자 합니다.

Data



한국문화예술위원회

문화예술 개인/단체 문화소비 단체 CSV JSON 무료

차상위계층 카드 가맹점 정보

한국문화예술위원회 >

유형 CSV,JSON 가격 무료 데이터 갱신주기 Quarterly

2021.05.26 업데이트 👁 630 ⬇ 44

평점 ★★★★★ 0.0 평가하기 5 평가하기

데이터 이름 : '차상위계층 카드 가맹점 정보'

데이터 포맷 : CSV, 무료 공개, 분기별 수집

데이터 출처 : 한국문화예술위원회

데이터 링크 : [문화 빅데이터 - 포털- 전체 상품 \(bigdata-culture.kr\)](http://bigdata-culture.kr)

데이터 요약 : 부산지역 취약계층 가맹점 데이터,

기초생활 수급자/차상위 계층이 활용할 수 있는 문화누리카드 가맹점 데이터

데이터 활용 : 제공 가맹점 대상지역 차상위 계층의 카드 사용 활성화,

지원 카드 사용률과 제공 분야 개선

Data Load

```
1 # Data Load
2 card_shop_df = pd.read_csv('data/ak_lwicc_card_mrhst_info_202103.csv')
3 card_shop_df.columns=return_col('차상위계층 카드 가맹점 정보_컬럼정의서.xls')
4 print('card_shop_df.shape :', card_shop_df.shape)
5 card_shop_df.head(2)
```

card_shop_df.shape : (5027, 12)

	일련번호	가맹점명	가맹점위도	가맹점경도	가맹점시도코드	가맹점시도명	가맹점시군구코드	가맹점시군구명	가맹점행정동코드	가맹점행정동명	가맹점구분코드	가맹점구분코드명
0	14	GS25 R부산동구2점	35.136637	129.065045	21	부산광역시	21030	동구	2103071	범일1동	B	급식카드가맹점
1	15	GS25 R부산동구3점	35.136637	129.065045	21	부산광역시	21030	동구	2103071	범일1동	B	급식카드가맹점

데이터 선정의 이유와 목표 :

저희가 본 서비스 구현에 사용한 데이터는 '한국문화예술위원회'의 '차상위계층 카드 가맹점 정보' 데이터 입니다. GPS 기반의 추천 서비스를 제공하기 위해서는 문화누리카드 가맹점의 위치정보가 필요했습니다, 따라서 부산 지역 내 문화누리카드 가맹점의 위치정보를 담은 해당 데이터를 사용하게 되었고, 추가적으로 필요한 데이터와 특성에 대해서는 Service 의 Demo Version 구현을 위해 임의로 직접 만들어 보충하였습니다. 데이터는 csv 형태로 '문화 빅데이터 플랫폼'에 무료로 공개 되어 있으며, 로컬 환경으로 다운받아 개발을 진행했습니다.

Service Description

● Abstract

GPS 기반의 문화 프로그램 패키징 서비스는 사용자의 위치를 기준으로 설계한 결정규칙에 따라, 부합하는 문화시설을 패키지로 추천해주는 서비스입니다. 먼저 사용자의 단말기로부터 현재 GPS 정보와 등록된 문화누리카드의 잔금을 불러오게 되면, 사용자의 위치와 주변 접근성을 고려하여 잔금 내에서 최대한 다양한 카테고리의 문화 프로그램을 즐길 수 있도록 프로그램 리스트를 추천해줍니다.

● Decision Rule

	가맹점명	위도	경도	비용	카테고리
0	삼천리자전거동래점	35.202679	129.083916	30000	여행
1	행복한스튜디오	35.152427	129.054766	13000	생활체육
2	네이쳐엔트리	35.061973	128.984193	10000	여행
3	신평태권도장	35.093446	128.973558	33000	여행
4	송무인 신금 태권도	35.252920	129.013612	18000	생활체육

패키징을 추천해주는 결정규칙은 다음과 같습니다. 먼저 WGS84 좌표계 (위도 경도 좌표계)를 기준으로 사용자로부터 인접 가맹점들과의 거리를 계산합니다. 그리고 가맹점을 사용자와의 거리 순으로 오름차순 정렬한 뒤, 최종적으로 위와 같은 데이터셋을 구축한 후 본격적인 패키지 탐색을 시작합니다.

```
# 탐색
remain_cat = ['생활체육', '문화공연관람', '여행'] # 아직 채택되지 않은 카테고리
accepted = [] # 채택된 문화시설
idx = 0 # 체크할 가맹점의 Index
neg = 0 # 탐색한 문화시설이 연속해서 비채택된 횟수
n = 10 # 탐색을 얼마나 진행할지 결정하는 Parameter
remain_num = 30 # 추가적으로 표시할 인접 문화시설의 수
```

탐색 과정에서 주의깊게 살펴 볼 요소는 크게 세 가지입니다.

첫째로, 가능한 여러 카테고리의 문화시설을 체험할 수 있는 패키지를 추천하는가 입니다. 아무리 인접하고 저렴한 문화시설이라고 해도, 한정된 금액 안에서 동일한 카테고리의 문화체험을 중복해 경험한다면 사용자 또한 지루할 뿐만 아니라 문화누리카드의 본래 출시 취지도 달성할 수 없을 것입니다. 따라서 탐색 과정 중 채택되지 않은 카테고리에 대해 우선적으로 가맹점을 탐색하도록 만들어 패키지의 다양성을 확보하였습니다.

둘째로, 사용자의 잔금이 해당 문화시설의 프로그램을 이용할 수 있을 만큼 남아있는지 입니다. 본 추천 알고리즘은 사용자와 거리가 가까운 문화시설부터 Greedy한 방식으로 탐색을 진행하여 패키지를 구성합니다. 따라서 채택할 때마다 해당 프로그램에 상응하는 비용을 잔금에서 차감하게 되는데, 만약 잔금이 충분히 남아있지 않다면 탐색을 멈추고 해당 시점까지 패키지로 구성한 리스트를 반환합니다.

셋째로, 현재까지 구성한 패키지가 합리적인가 입니다. 앞서 정의한 결정규칙대로 가맹점을 탐색하다가 규칙에 부합하지 않는 가맹점이 탐색되었다고 해서 그 시점에서 탐색을 바로 중지하는 것은 합리적이지 못합니다. 거리상 손해를 보더라도, 몇 미터 차이 나지 않는 곳에 더 좋은 문화시설과 체험의 기회가 있을지도 모르는 일이기 때문입니다. 따라서 저희는 일종의 Tolerance Parameter (용인 인자)를 설정하여 처음 채택이 거부되더라도 n 번 까지는 더 탐색을 진행하도록 설정하였습니다.

이렇게 채택된 문화시설은 데이터프레임 형태로 반환되며, 여러 부가 정보를 포함하여 최종적으로 사용자에게 제공되게 됩니다.

● Dashboard Mapping : Folium Map

앞서 추천 서비스 구축을 위한, 패키지 결정규칙에 대해 소개하였습니다. 저희는 이렇게 반환된 정보를 사용자가 보다 직관적인 형태로 받아들일 수 있도록, 출력된 정보를 지도 대쉬보드 상에 나타냈습니다. Python 환경에서 Folium Library를 Import 하는 형태로 개발을 진행하였으며, 완성된 대쉬보드는 HTML 형태로 Export하여 배포하였습니다.

Dashboard에 매핑한 정보로는 우선, 사용자의 위치를 기준으로 채택된 문화시설과 채택되지 않은 인근 문화시설을 다른 색으로 마킹하여 나타내었고, 채택된 문화시설 중 가장 먼 거리에 있는 가맹점을 기준으로 패키지의 거리상 수렴 범위를 나타내었습니다. 해당 수렴 범위를 근거로, 본 패키지를 이용했을 때 예상되는 최대 이동 거리를 사용자에게 제공하여 패키지 이용여부를 결정할 때 접근성 면에서 참고할 근거로 제공하고자 대쉬보드에 나타내었습니다.

또한, 위치정보 외에도, 특정 마커에 마우스를 오버랩할 경우, 해당 마커에 위치한 문화시설의 세부 정보를 표시하도록 하였습니다. 이를 통해, 사용자는 관심 문화시설에 대한 정보를 쉽고 빠르게 얻을 수 있으며, 패키지 이용 결정 여부에도 도움이 될 뿐만 아니라, 인근 시설에 대한 추가 정보도 얻을 수 있게 됩니다.

Service Implementation

● Data Load & Preprocessing

```
# Example Coordination
busan_coord = [35.1796, 129.0756]
busan_station_coord = [35.1152, 129.0422]
busan_namgu_coord = [35.1366, 129.0844]
```

	가맹점명	위도	경도	비용	카테고리
0	삼천리자전거동래점	35.202679	129.083916	30000	여행
1	행복한스튜디오	35.152427	129.054766	13000	생활체육
2	네이쳐앤트리	35.061973	128.984193	10000	여행
3	신평태권도장	35.093446	128.973558	33000	여행
4	송무인 신금 태권도	35.252920	129.013612	18000	생활체육

부산광역시 내 문화누리카드 가맹점 데이터를 기준으로 Demo Version Service를 구현해보았습니다. 먼저, 사용자의 위치를 대신할 예제 좌표 세 개를 정의한 후, '차상위계층 카드 가맹점 정보' 데이터를 활용하여 DB를 정의하였습니다. 원 데이터를 전처리하는 과정은 크게 { Feature Selection -> Adding Cost Columns -> Adding Category Columns } 이상 세 과정으로 이루어져 있으며, 해당 과정을 거쳐 총 1,255개의 Instance로 이루어진 DataBase를 구축하였습니다.

● Recommendation Algorithm

```
# 추천 시스템
def Recommendation_Algorithm(df, coord, deposit):
    # 사용자의 거주지를 기준으로 인근 문화시설과의 거리 계산.
    df['거리'] = distance(df, coord)

    # 중복되는 카테고리 없이 예산 Deposit 안에서 이용할 수 있는 Top n개의 문화시설 추천.
    df_accepted, df_not = decision_rule(df, deposit)

    # 사용자 위치를 중심으로 추천 문화시설 및 주변시설 매핑
    m = Mapping_Node(df_accepted, df_not, coord)

    return df_accepted, m

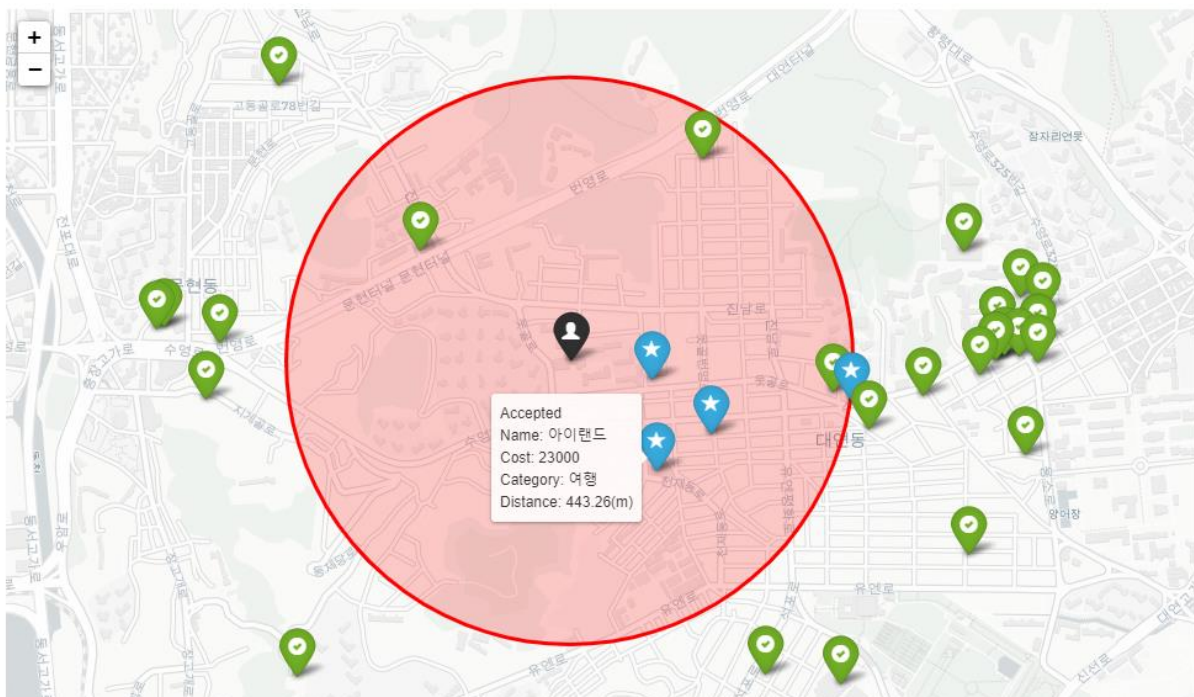
1 # Service Implementation
2 df, m = Recommendation_Algorithm(map_subset, busan_namgu_coord, 100000)
3 df
```

남은 금액 : 3000

	가맹점명	위도	경도	비용	카테고리	거리
0	지성도서	35.136089	129.087233	25000	문화공연관람	263.806919
1	아이랜드	35.133482	129.087437	23000	여행	443.255061
2	세븐브릭스	35.134532	129.089295	37000	생활체육	501.002211
3	팬레코드	35.135508	129.094187	12000	여행	898.216027

이제 앞서 정의한 'Recommendation_Algorithm(df, coord, deposit)'에 (조회할 데이터베이스, 사용자 좌표, 잔금)을 입력하여 사용자에게 적합한 문화 프로그램 패키지를 추천받습니다. 저희는 '부산 남구' 좌표를 사용자 좌표로 사용하였으며, 예산은 문화누리카드의 최초 지급액인 100,000 원을 기입하였습니다. 그 결과, 결정 규칙 조건에 맞는 총 4 개의 프로그램을 추천해주었으며, 해당 패키지를 이용한 후 잔금까지 출력된 것을 볼 수 있습니다.

● Dashboard Mapping : Folium Map



위에서 앞서 정의한 결정규칙에 근거하여 '부산 남구' 좌표를 기준으로 총 4 개의 프로그램이 추천된 알고리즘 수행 결과를 보았습니다. 위 figure 는 출력 결과를 Folium Map 위에 나타낸 그림입니다. 그림을 보면, 검정색으로 마킹된 사용자 위치를 기준으로 4 개의 파란색 별 마킹과 초록색 체크 마킹이 찍힌 것을 알 수 있습니다. 파란색 별 마킹은 추천 알고리즘이 채택한 시설이며, 초록색은 채택은 되지 않았지만, 인근 문화시설 정보를 나타내는 마커들입니다. 또한, 채택된 문화시설 중 가장 거리가 먼 '펜레코드' 가맹점 까지의 거리를 기준으로 빨간색 활동 범위가 그려진 모습입니다. 그리고 마지막으로 관심 마커에 마우스를 오버랩할 경우, 가맹점 이름, 비용, 카테고리 등 관련 부가 정보가 나타나는 것을 볼 수 있습니다.

추천 시스템 설계 과정 중 Code Level에서의 모든 Code 는 Objectification, Soft Coding, Modularity 하여 관리하였음으로 확장성과 호환성이 보장되도록 작성하였습니다. 따라서 어떤 Coord, 어떤 DataSet 에도 적용이 가능하며, 지속 가능한 개발이 가능하겠습니다.

Codes

● Github URL :

[깃허브 : URL]

● 핵심 코드

```

# 중복되는 카테고리 없이 예산 Deposit 안에서 Top n개의 문화시설 추천.
def decision_rule(df, deposit):
    # 거주지와 거리 차이 순으로 오름차순 정렬
    df = df.sort_values('거리').reset_index(drop=True)

    # 탐색
    remain_cat = ['생활체육', '문화공연관람', '여행'] # 아직 채택되지 않은 카테고리
    accepted = [] # 채택된 문화시설
    idx = 0 # 체크할 가맹점의 Index
    neg = 0 # 탐색한 문화시설이 연속해서 비채택된 횟수
    n = 10 # 탐색을 얼마나 진행할지 결정하는 Parameter
    remain_num = 30 # 추가적으로 표시할 인접 문화시설의 수

    while True:

        # 다음 문화시설 데이터 객체화
        next_shop = df.iloc[idx]
        next_cat = next_shop['카테고리']
        next_cost = next_shop['비용']

        # 채택(Accepted) : 아직 선택되지 않은 카테고리이며, 예산 안에 이용가능할 경우.
        if (next_cat in remain_cat) & (next_cost <= deposit):
            remain_cat.remove(next_cat)
            accepted.append(idx)
            deposit -= next_cost
            neg = 0
        else:
            neg += 1

        # 정지 규칙 : 탐색에 연속해서 n번 이상 실패할 경우 종료
        if neg >= n:
            break

        # 다음 루프를 위한 파라미터 조정
        idx += 1
        if not bool(remain_cat):
            remain_cat = ['생활체육', '문화공연관람', '여행']

    print('남은 금액 :', deposit)
    accepted_df = df.iloc[accepted].reset_index(drop=True)
    not_accepted_df = df.drop(accepted).reset_index(drop=True)

    return accepted_df, not_accepted_df[:remain_num]

```

Summary

이처럼 본 서비스를 통해 60 대 이상의 어르신들은 사용자의 거주지를 기준으로 주위에 인접한 문화누리시설에 대한 정보를 쉽고 빠르게 얻을 수 있으며, 더 이상 탐색과 결정 과정에 많은 시간을 사용하지 않아도 되겠습니다. 또한, 본래 문화누리카드의 출시 취지와 부합하게 지속 가능하고 다양한 문화체험의 기회를 제공할 수 있으며, 전 연령대의 문화체험 불균형을 해소할 수 있습니다.