

제주도 여행 루트 추천 시스템

이수경 이승우 이은서 주혜인 홍현경





태그 정제

서로 유사한 의미를 지니는 태그가 많을 경우, 개별 태그가 지닌 중요도가 떨어짐
따라서 태그의 분포를 알기 위해 태그 클러스터링 진행!

클러스터링을 위해서!

1

사전 학습된 Fasttext 모델을 불러온다

2

모델에 우리 데이터를 학습 시켜 fine-tuning한다!

새로운 데이터셋을 추가로 학습시켜
모델의 파라미터 업데이트 하는 것



키워드 추출

태그가 설명하지 못하는 관광지의 특성을
보충하기 위해 리뷰 데이터에서 키워드를 추출할 것!

VISIT JEJU에서 크롤링 한 관광지 태그 81개

EX) #맑음 #연인 #숲길 #액티비티

Preprocess

1 띄어쓰기, 오타 수정

2 문장 단위로 끊음

3 Tokenize

4 Stemming

5 Stopwords 제거



키워드 추출 모델 선정

1

2

3

4

5

TF-IDF

plc	result
김영갑갤러리	['김영갑', '사랑', '루게릭병', '유품', '전시관']
애월한담해안산책로	['유채꽃', '감상', '매력', '바위', '바다', '산책로']
오설록티뮤지엄	'녹차', '오설록', '케익', '공간', '전통', '의미', ...]

Word
Rank

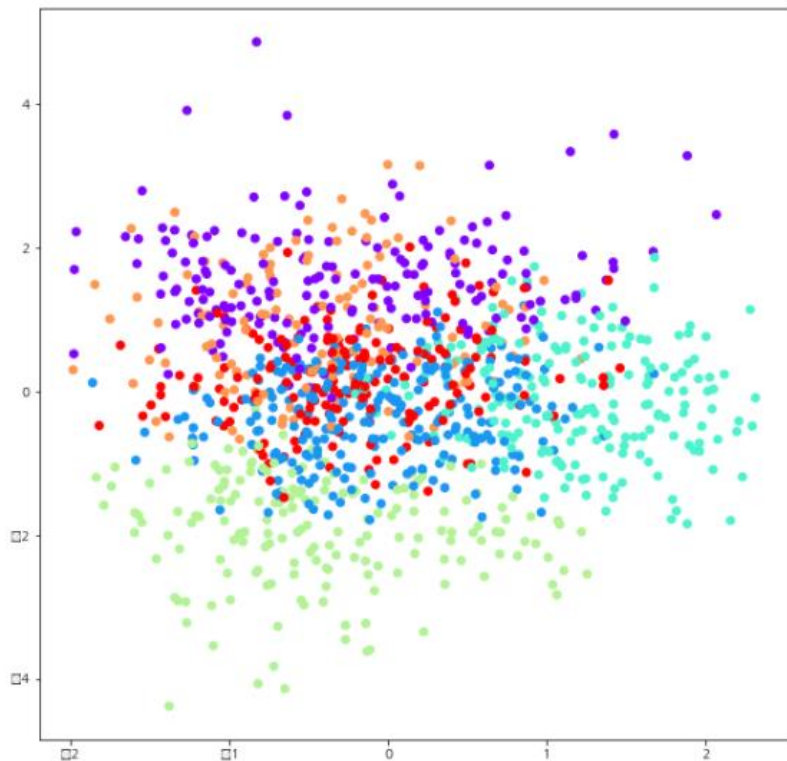
plc	result
김영갑갤러리	['있다', '김영갑', '전시', '열정', '루게', '병원']
애월한담해안산책로	['있다', '유채', '바다', '아름답다', '산책', '애월']
오설록티뮤지엄	녹차, '문화', '의미', '있다', '푸른', '오설', '공간' ...

비교 결과 TF-IDF가
더 의미 있는 키워드들을
추출해내는 것을 확인할 수 있음



키워드 정제

K = 6



?

산발적으로 분포하는 키워드

!

키워드 정제의 필요성 ↑

키워드가 속한 관광지가
3개 이하인 키워드 삭제

&

태그를 기준으로
병합 or 제거



Bidirectional Model

What is **Bidirectional**?

균형 잡힌 정보를 인코딩 하기 위한 방법

과거의 정보

의미 파악이 필요한 단어

미래의 정보

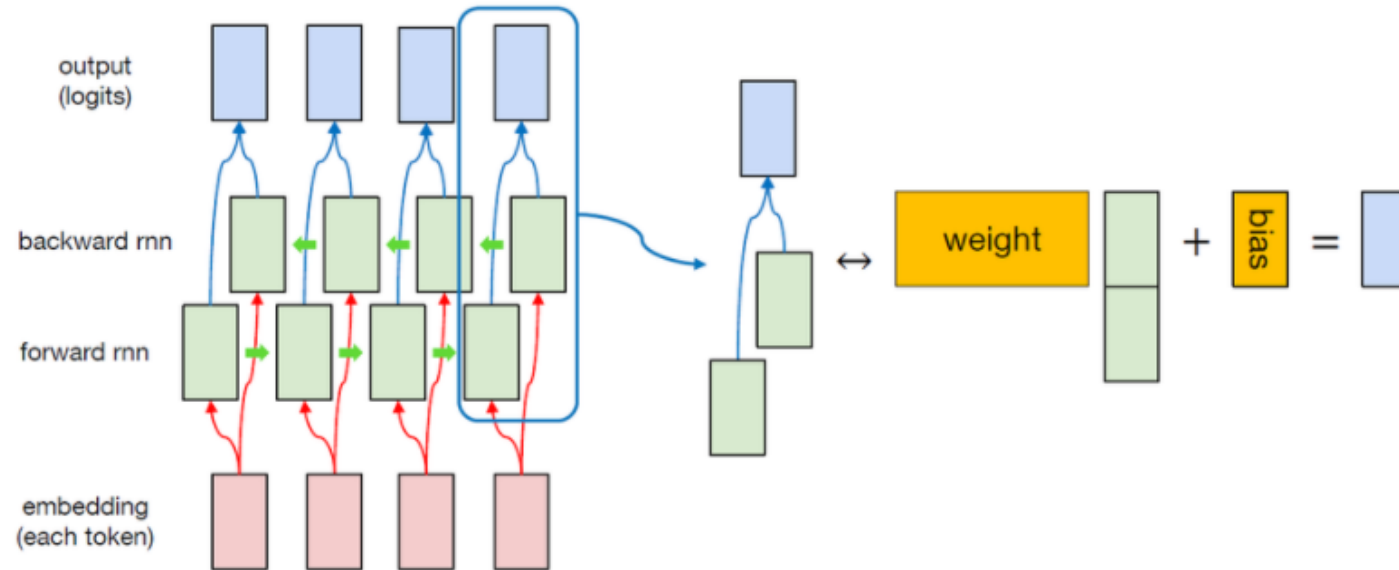
정방향뿐 아닌 역방향으로 처리하는 계층 추가

→ 각 단어에 대응하는 은닉 상태 벡터에 양쪽 방향의 정보 집약



Bidirectional RNN

Bidirectional RNN 구조



Forward RNN(이전 → 현재 시점)의 은닉층과 Backward RNN(이후 → 현재 시점)의 은닉층을 concatenate하는 방식으로 output 생성



최종모델 선정

각 모델에서 산출된 관광지 벡터를 이용한 관광지 클러스터링 결과를 바탕으로 선정

클러스터링

K means 클러스터링

K=7을 선택



관광지 데이터를 크롤링한 웹사이트의 관광지 분류기준이 7!

자연 문화관광 테마관광지 도보 레저/체험 섬 속의 섬 웰니스

시각화

t-sne를 통한 공간축소

t-sne란?

n 차원에 분포된 데이터를 차원을 축소하며
거리정보를 보존하되, 거리가 가까운 데이터의
정보를 우선하여 보존하기 위해 고안된 방법.



길이가 300인 벡터들을 2차원에 시각화

1

2

3

4

5

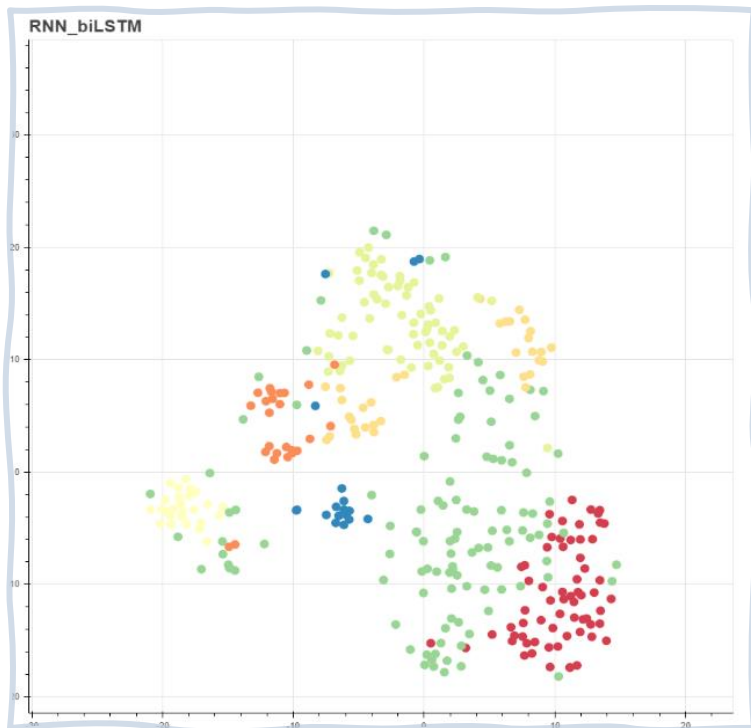


최종모델 선정

각 모델에서 산출된 관광지 벡터를 이용한 관광지 클러스터링 결과를 바탕으로 선정



Bidirectional RNN



최종 선정모델 : Bidirectional RNN

- 5개의 모델 중 가장 명확한 클러스터 구분
- 동적 시각화로 관광지명 확인 결과 가장 논리적인 관광지 클러스터링



사용자 벡터 생성 과정

1

각 태그 별 태그벡터 생성

plc	vector
부모	[0.098, 0.137, 0.157, ..., 0.094, 0.118]
4.3	[0.652, 0.267, 0.253, ..., 0.508, 0.336]
감귤	[0.0, 0.533, 0.0, ..., 0.356, 0.0]
...	...
건물	[0.054, 0.274, 0.094, ..., 0.108, 0.012]
문화유적지	[0.377, 0.251, 0.218, ..., 0.349, 0.397]
겨울	[0.155, 0.271, 0.737, ..., 0.481, 0.156]

태그의 unique 개수

최종적으로 길이가 300인 162개의 태그 벡터 생성



사용자 벡터 생성 과정

2

사용자로부터 태그를 입력받는다




세가지 항목 이상 선택해주세요!

여행의 목적

날씨

자연

☐ 바람
 ☒ 지질
 ☐ 오름
 ☐ 산

연령대

동반상태

계절

테마

방문하고 싶은 곳

시간대

하고싶은 것

특산물

기타

경로 추천 시작!



최종 태그 벡터

3

태그 벡터를 이용하여 최종 사용자 벡터를 구한다

plc	vector
✓ 부모	[0.098, 0.137, 0.157, ..., 0.094, 0.118]
✓ 4.3	[0.652, 0.267, 0.253, ..., 0.508, 0.336]
감굴	[0.0, 0.533, 0.0, ..., 0.356, 0.0]
...	...
건물	[0.054, 0.274, 0.094, ..., 0.108, 0.012]
✓ 문화유적지	[0.377, 0.251, 0.218, ..., 0.349, 0.397]
겨울	[0.155, 0.271, 0.737, ..., 0.481, 0.156]

[0.098, 0.137, 0.157, ..., 0.094, 0.118]

[0.652, 0.267, 0.253, ..., 0.508, 0.336]

[0.377, 0.251, 0.218, ..., 0.349, 0.397]

선택된 벡터들의 평균 구함!



최종 사용자 벡터

[0.375, 0.218, 0.209, ..., 0.317, 0.283]



유사도 행렬 생성

사용자 벡터	
[0.375, 0.218, 0.209, ..., 0.317, 0.283]	
관광지	vector
곶자왈	[0.0, 0.205, 0.067, ..., 0.225, 1.780, 0.296]
관덕정	[0.006, 0.104, 1.067, ..., 0.028, 0.632, 0.295]
화조원	[0.178, 0.0, 0.281, ..., 0.0, 0.464, 0.541]
...	...

cosine similarity

$$\frac{A \cdot B}{||A|| \times ||B||}$$

곶자왈	관덕정	화조원	...
0.353	0.898	0.548	...

사용자 벡터와 모든 관광지 벡터의 코사인 유사도 계산



거리 기반 추천

코사인 유사도로 내림차순 정렬된 관광지 목록

관광지	코사인 유사도
관덕정	0.898 ✓
너븐숭이 4.3기념관	0.895
제주 4.3 평화공원	0.884
...	...

1 사용자 벡터 - 관광지벡터 유사도가 높은 관광지들 추천

2 유사도가 가장 높은 여행지에서 부터 시작!



거리 기반 추천

출발지: 관덕정



관광지	출발지와의 거리
너븐송이 4.3기념관	24342 ✓
제주 4.3 평화공원	35753
다랑쉬굴	52363
...	...

가자!
너븐송이 4.3 기념관으로!

출발지를 기준으로 가장 **가까운** 관광지를 다음 경로로 추천