

앞으로 어디가 핫할까?

# 2020 SEOUL 핫플레이스 예측

선형대수학팀

김수인, 강현주, 이수진, 오정민, 홍세정





1. 지난주

2. 추가전처리

3. 지도학습

4. 비지도학습

5. 의의 및 한계

## 문제점

**1**

2020년의 데이터가 없음 → 시계열이 아니고서는 예측이 불가

**2**

문화 데이터 변수 개수 부족 → 문화 지표를 대표할 수 없음

**3**

적절한 종속 변수 (Y) 의 부재 → 무엇을 예측할 것인가..?



1. 지난주

2. 추가전처리

3. 지도학습

4. 비지도학습

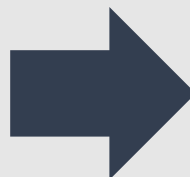
5. 의의 및 한계

출처: 지방 행정데이터 개방 > 인허가 데이터

## 문화 데이터 보강

### <인허가 데이터>

6110000\_서울특별시\_10\_33\_01\_P\_무도장업  
6110000\_서울특별시\_10\_33\_02\_P\_무도학원업  
6110000\_서울특별시\_10\_34\_01\_P\_빙상장업  
6110000\_서울특별시\_10\_35\_01\_P\_수영장업  
6110000\_서울특별시\_10\_36\_01\_P\_스키장  
6110000\_서울특별시\_10\_37\_01\_P\_종합체육시설업  
6110000\_서울특별시\_10\_38\_01\_P\_승마장업  
6110000\_서울특별시\_10\_39\_01\_P\_썰매장업  
6110000\_서울특별시\_10\_40\_01\_P\_요트장업  
6110000\_서울특별시\_10\_41\_01\_P\_체육도장업  
6110000\_서울특별시\_10\_42\_01\_P\_체력단련장업  
6110000\_서울특별시\_11\_43\_01\_P\_담배도매업  
6110000\_서울특별시\_11\_43\_02\_P\_담배소매업  
6110000\_서울특별시\_11\_43\_03\_P\_담배수입판매업체  
6110000\_서울특별시\_11\_44\_01\_P\_목욕장업  
6110000\_서울특별시\_11\_45\_01\_P\_물류주선업(국제)  
6110000\_서울특별시\_11\_45\_02\_P\_물류창고업체  
6110000\_서울특별시\_11\_46\_01\_P\_민방위급수시설  
6110000\_서울특별시\_11\_46\_02\_P\_민방위대피시설  
6110000\_서울특별시\_11\_47\_01\_P\_상조업  
6110000\_서울특별시\_11\_48\_01\_P\_승강기유지관리업체



[업종]

단란 주점

유흥 주점

주점 데이터 추가!



1. 지난주

2. 추가전처리

3. 지도학습

4. 비지도학습

5. 의의 및 한계

## Y 값을 정해보자

월별 (2012년 7월 ~ 2019년 12월), 법정동별 (467개동)

### 1) 경제적 지표

- ① 거주 인구수(남, 녀, 총)
- ② 부동산 거래가격 및 거래량
- ③ 대수선
- ④ 음식점 창 폐업 수, 영업기간

### 2) 문화적 지표

- ⑤ 공연장 수
- ⑥ 노래방 창업 수, 폐업 수

### 3) 트렌드 지표

- ⑦ 인스타그램 크롤링
- ⑧ 국민일보 크롤링



변수 간의 중요도를 파악해서 **가중치**를 부여하자!



1. 지난주

2. 추가전처리

3. 지도학습

4. 비지도학습

5. 의의 및 한계

## 가중치 HOW?

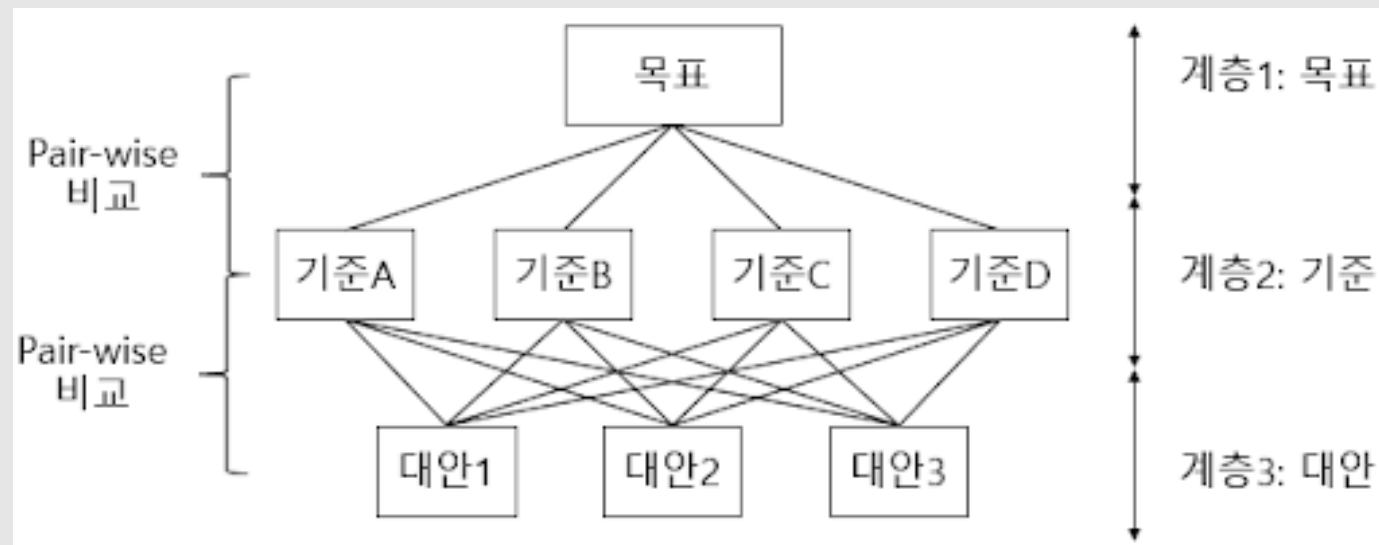
### AHP 분석

aka 계층화분석법

A: Analytic

H: Hierarchy

P: Process



'의사결정의 목표 또는 평가기준이 다수이며 복합적인 경우,  
이를 계층(Hierarchy)화해 주요 요인과 그 주요 요인을 이루는 세부 요인들로 분해하고,  
이러한 요인들을 쌍대 비교(Pairwise Comparison)를 통해 중요도를 산출하는 분석 방법이다.'



1. 지난주

2. 추가전처리

3. 지도학습

4. 비지도학습

5. 의의 및 한계

## AHP 결과

변수

인구 변화율  
부동산 거래금액 변화율  
편의점 수 변화율  
대수선  
음식점 창 폐업 수  
공연장 수  
노래방 창 폐업 수  
주점 수

젠트리피케이션  
위험도 수치화 ✕  
( -2 ~ 2 )

<구한 가중치 값>

0.1434  
0.2366  
0.1456  
0.2706  
0.2038  
0.2764  
0.2626  
0.4610

젠트리피케이션  
지수



문화 지수



핫플 지수



1. 지난주

2. 추가전처리

3. 지도학습

4. 비지도학습

5. 의의 및 한계

## Y변수 생성의 한계

|    | 구    | dong | year | month | Y      |
|----|------|------|------|-------|--------|
| 1  | 강남구  | 논현동  | 2015 | 12    | 0.8210 |
| 2  | 은평구  | 갈현동  | 2015 | 01    | 0.7108 |
| 3  | 강남구  | 삼성동  | 2015 | 01    | 0.6973 |
| 4  | 강남구  | 삼성동  | 2015 | 01    | 0.6857 |
| 5  | 관악구  | 신림동  | 2015 | 10    | 0.6828 |
| 6  | 영등포구 | 여의도동 | 2019 | 04    | 0.6621 |
| 7  | 마포구  | 서교동  | 2014 | 07    | 0.6450 |
| 8  | 구로구  | 구로동  | 2017 | 01    | 0.6290 |
| 9  | 관악구  | 신림동  | 2012 | 08    | 0.6203 |
| 10 | 마포구  | 서교동  | 2013 | 10    | 0.5968 |
| 11 | 강남구  | 삼성동  | 2019 | 09    | 0.5838 |
| 12 | 강서구  | 화곡동  | 2016 | 04    | 0.5490 |
| 13 | 관악구  | 신림동  | 2017 | 01    | 0.5100 |
| 14 | 은평구  | 갈현동  | 2016 | 06    | 0.5200 |
| 15 | 마포구  | 서교동  | 2014 | 03    | 0.5047 |
| 16 | 서초구  | 서초동  | 2016 | 06    | 0.5034 |

강남구

논현동

여의도

홍대

서초동



최종 Y값 내림차순 정렬을 해보니,  
문화적 특색보다는 상업적으로 활발한 지역



1. 지난주

2. 추가전처리

3. 지도학습

4. 비지도학습

5. 의의 및 한계

Y 변수 생성의 한계

## 핫플레이스 정의

특색있는 문화 + SNS 인기로 사람들이 몰려 상권의 변화가 일어나는 곳



### 1) 경제적 지표

젠트리피케이션 현상이  
일어나는 곳



### 2) 문화적 지표

특색이 풍부한 곳



### 3) 트렌드 지표

SNS 유행이 번지는 곳

SNS트렌드 지표를 포함해서 **클러스터링**을 해보자!





1. 지난주

2. 추가전처리

3. 지도학습

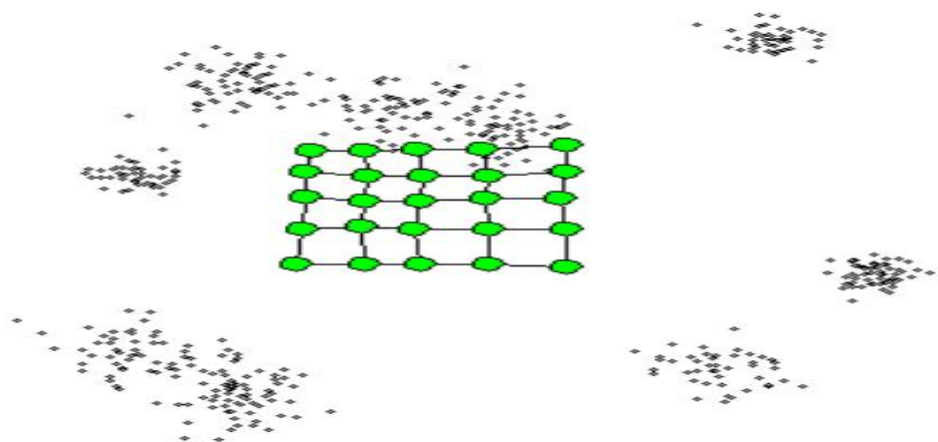
4. 비지도학습

5. 의의 및 한계

SOM

**S**elf **O**rganized **M**ap  
----- 자기조직화지도

격자벡터, 즉 클러스터의 중심점을 관측치와 가깝게  
업데이트해가는 DNN 기반의 클러스터링 방법



초기 랜덤한 값 (또는 PCA 등 방법을 이용)을 가진  
격자 벡터로 학습을 시작



1. 지난주

2. 추가전처리

3. 지도학습

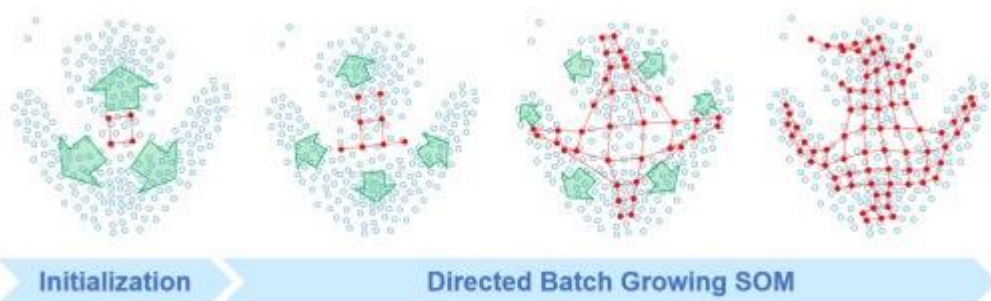
4. 비지도학습

5. 의의 및 한계

## Growing SOM

# GROWING SOM

SOM에서 적절한 지도 크기를 찾지 못하는 문제를  
극복하기 위해 개발된 알고리즘



최소의 노드 수로 시작하여  
**Heuristic**을 기반으로  
경계에서 새로운 노드를 **성장**시킨다.



1. 지난주

2. 추가전처리

3. 지도학습

4. 비지도학습

5. 의의 및 한계

## 클러스터링 결과

C  
L  
U  
S  
T  
E  
R  
3

| gu   | dong   | year | month |
|------|--------|------|-------|
| 강동구  | 암사동    | 2014 | 6     |
| 도봉구  | 쌍문동    | 2013 | 7     |
| 동대문구 | 신설동    | 2017 | 12    |
| 송파구  | 가락동    | 2013 | 9     |
| 송파구  | 석촌동    | 2018 | 8     |
| 영등포구 | 당산동3가  | 2013 | 5     |
| 영등포구 | 영등포동3가 | 2016 | 2     |
| 영등포구 | 영등포동3가 | 2018 | 8     |
| 종로구  | 예지동    | 2019 | 9     |
| 종로구  | 종로5가   | 2015 | 7     |
| 강서구  | 가양동    | 2019 | 3     |
| 용산구  | 후암동    | 2019 | 12    |
| 종로구  | 돈의동    | 2017 | 1     |
| 종로구  | 동숭동    | 2018 | 2     |



레트로한 분위기 / 유명역이 있는  
신설동



힙합의 정석 / 해방촌  
후암동



송리단길 / 석촌호수와 함께 카페가 있는  
석촌동

고즈넉한 카페가 많이 있는 지역  
우리가 원하는 핫플...?



1. 지난주

2. 추가전처리

3. 지도학습

4. 비지도학습

5. 의의 및 한계

## 클러스터링 결과

그럼에도 불구하고 예측해보자

클러스터 3에 속한 2019년 데이터를 통해 본 2019년의 핫플은?



후암동 (해방촌)



예지동 (시계골목, 청계천)

특색 있는 지역으로 묶이기는 한 것 같다



## 결론

1

지도학습을 통해 핫플 지수를 생성하여  
강남구, 서초동, 논현동 등 상업적으로 활발한 지역을 찾아낼 수 있었다.

2

여러 클러스터링 모델을 통해  
상업적 핫플, 문화적 핫플, 이외 주거지역을 분석할 수 있었다.

3

2019 핫플 예측은 후암동와 예지동!  
이후 2020 데이터를 통해 추후 예측이 가능하다.



1. 지난주

2. 추가전처리

3. 지도학습

4. 비지도학습

5. 의의 및 한계

## 한계

**1**

핫플레이스를 나타낼 수 있는 독특화된 변수에 대한 고려가 부족했다.

**2**

접근 가능한 데이터가 부족했다.

**3**

지도학습 및 비지도학습의 일관된 결과를 내지 못했다.