21. 11. 17

Dart 크롤링 <- 동헌

재무지표 크롤링 <- 도형

형태소 분석기 : Khaiii

User dictionary 만들기 < - 우리 모두

주가변동 데이터(시계열) 에 쓸 알고리즘

실루엣, 시각화

Kmeans <- 실루엣만

Kmedoid

DBSCAN

계층적 군집화 (HAC)

(Fuzzy 군집화)

Spectral clustering 도형

MeanShift 도형

최대 2개만 해오기 계속해서 카톡으로 특정 알고리즘 가능여부나, 실행여부 공유!

(기업개요, 재무지표는 중간발표 때까지는 Kmeans로 하고, 추후 다른 알고리즘 실험해보기)

도형

3번으로 진행하되, 계층적 클러스터링을 진행해서 그 결과를 2차 클러스터링에서 반영한다.

1번 대안은 클러스터링을 진행하고 결과를 해석하는 데에 너무 해석에 근거가 부족한 느낌.

은수

3번으로 진행하되, 1번 클러스터링의 결과 ( 클러스터 간 거리 등) 를 이용해,

2차 클러스터링에서 거리계산 측면에서 더 합리적인 클러스터링이 되도록 한다.

정규화 진행 후에, [ 8716 + 276 + 재무지표의 개수 ] 로 하나의 벡터로 만든다.

동헌

클러스터의 중심좌표 간의 거리로 나타내고, ( , , ) 정규화를 통해 거리를 맞춰준다. 그 후에 이걸 metric 으로 사용해서 다시 클러스터링한다.

각 기업이 3개의 vector 를 가지는데, [[8716] ,[276] [재무지표의 개수]]

0.001 …… 거리값 = >? 500 centroid 가 5개다 . 500 400 200 300 Z

0.01 거리값 = >? 20 Z

**3번 대안 :**

**“ 각각의 data에 대해 클러스터링 진행 후, 그 결과를 바탕으로 2차 클러스터링”**

다음 회의 : 토요일 13시!!!! -> 일요일 (14시) 까지 수합해서 정리해서 확정