

Week5-3

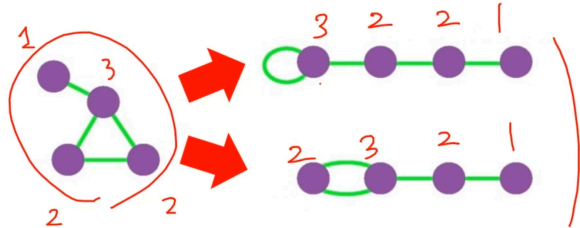
1. 학습정리
 - 배치모형

2.1 비교 대상: 배치 모형

성공적인 군집 탐색을 정의하기 위해 먼저 **배치 모형(Configuration Model)**을 소개합니다

주어진 그래프에 대한 배치 모형은,

- 1) 각 정점의 연결성(Degree)을 보존한 상태에서
- 2) 간선들을 무작위로 재배치하여 얻은 그래프를 의미합니다



배치 모형에서 임의의 두 정점 i 와 j 사이에 간선이 존재할 확률은 두 정점의 연결성에 비례합니다

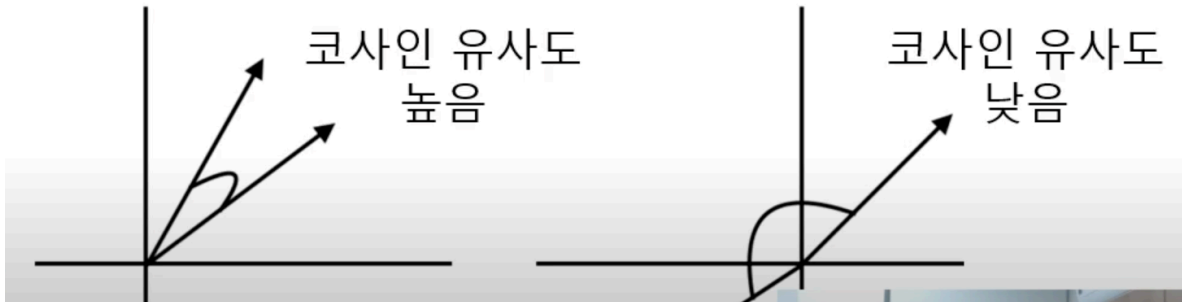
- 군집성
 - 배치 모형과 비교했을 때 군집 내 간선의 수가 많으면 좋음
 - 군집내 간선수 - 배치모형 간선수
 - $-1 \sim +1$
- 군집탐색 알고리즘
 - Girvan-Newman 알고리즘
 - 하향식
 - 그래프 탐색하며 군집간 다리 역할을 하는 간선 제거
 - 브릿지 찾는법 : 매개중심성
 - 매개중심성
 - 정점간의 최단 경로 위에 놓인 횟수
 - 군집성이 최대가 될 때 까지 반복
 - Louvain 알고리즘
 - 상향식
 - 군집성이 최대화 되도록 각 정점의 군집을 결정
 - 군집성이 증가하지 않을 때 까지 반복
 - 각 군집을 하나의 정점으로 하는 군집레벨 그래프로 변환
 - 정점이 한 개 남을 때 까지 반복
 - 중첩군집
 - 위 두 알고리즘은 중첩이 없다고 가정함
 - 실제로 있음
- 완화된 중첩 군집 모형
 - 각 정점이 각 군집에 속해있는 정도를 실수값으로 표현
- 내용기반 추천 시스템
 - 각 사용자가 구매했던 상품을 기반으로 추천
 - 각 상품을 원-핫 벡터로 표현
 - 사용자 프로필
 - 구매했던 상품들의 원핫벡터들의 가중평균
 - 상품 추천

- 사용자 프로필 벡터와 상품 프로필 벡터의 코사인 유사도 계산

사용자 프로필 벡터 \vec{u} 와 상품 프로필 벡터 \vec{v}

코사인 유사도 $\frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \|\vec{v}\|}$ 를 계산합니다

즉, 두 벡터의 사이각의 코사인 값을 계산합니다



- 장점
 - 새 상품도 추천 가능
 - 다른 사용자의 구매 기록 불필요
 - 추천 이유 제공
- 단점
 - 상품에 대한 부가 정보 필요
 - 구매 기록이 있어야 함
 - 과적합으로 인한 협소한 추천
- 협업 필터링 추천시스템
 - 사용자-사용자 협업 필터링
 - 추천 대상자와 유사한 취향의 사용자들 찾음
 - 그들이 선호한 상품을 추천
 - 취향 유사도 계산
 - 상관 계수

취향의 유사성은 상관 계수(Correlation Coefficient)를 통해 측정합니다

사용자 x 의 상품 s 에 대한 평점을 r_{xs} 라고 합시다

사용자 x 가 매긴 평균 평점을 \bar{r}_x 라고 합시다

사용자 x 와 y 가 공동 구매한 상품들을 S_{xy} 라고 합시다

사용자 x 와 y 의 취향의 유사도는 아래 수식으로 계산합니다

$$sim(x, y) = \frac{\sum_{s \in S_{xy}} (r_{xs} - \bar{r}_x)(r_{ys} - \bar{r}_y)}{\sqrt{\sum_{s \in S_{xy}} (r_{xs} - \bar{r}_x)^2} \sqrt{\sum_{s \in S_{xy}} (r_{ys} - \bar{r}_y)^2}}$$

즉, 통계에서의 상관 계수(Correlation Coefficient)를 사용해 취향의 유사도를 계산합니다

- 장점

- 상품의 부가정보 없어도 가능
- 단점
 - 충분한 데이터가 누적되어야 함
 - 새 상품, 새 사용자에게 대한 추천 불가능
- 추천 시스템의 평가

2. 피어세션

1. 강의리뷰
2. 피어 세션 질문 준비
3. 알고리즘 문제풀이
 1. <https://leetcode.com/problems/find-peak-element/>
 2. <https://leetcode.com/problems/all-paths-from-source-to-target/>
 - 3.