**Your Neighbors Affect Your Ratings: On Geographical Neighborhood Influence to Rating Prediction**

비즈니스에 대한 사용자의 등급은 비즈니스의 본질적 특성과 지리적 이웃의 외적 특성에 의해 결정된다고 가정.

기업이 위도/경도 좌표를 가진 특정 지리적 위치에 물리적으로 존재(노래나 영화와는 다름)

대부분의 비즈니스가 다른 비즈니스와 거리적(물리적)으로 독립적이지 않음(떨어져있지 않음)

결국, 사용자가 어떤 비즈니스를 방문할 때, 걸어갈만한 거리에 위치해있으면 다른 비즈니스를 지나갈 가능성이 높음

해당 지역의 전반적인 환경은 비즈니스에 대한 사용자의 관점에 영향을 미칠 수 있고, 그 후 비즈니스에 대한 사용자의 검토 및 등급에 영향을 미칠 수 있음.

예를 들어, 한 지역의 위생 표준은 해당 지역에 위치한 많은 레스토랑에 대한 사용자의 등급에 영향을 미칠 수 있음.

반면, 한 지역은 다른 지역과 구별되고 종종 방문객들에게 매력적이 됨. (그 지역에 몇 개의 좋은 비즈니스가 있음)

<기업의 등급과 지리적 이웃의 등급 사이에는 어떤 상관관계?>

사업의 범주와 관계 없이 기억의 등급과 주변 기업의 등급 사이에 약한 긍정적 상관관계가 존재한다고 봄.

매트릭스 요인화(Matrix-Factorization)에서 실현한 잠재 요인 모델을 기반으로 한 비즈니스 등급 예측 모델에 지리적 지역적 영향력을 통합.

사용자 리뷰, 비즈니스 범주, 비즈니스 인기 등 다른 요인의 영향과 함께, 제안된 모델이 다른 최신 기준보다 우수.

\* 데이터 분석을 실시하며, 기업의 범주에 관계없이 기업의 평균 등급과 주변 기업의 평균 등급 사이에 약한 긍정적인 상관관계가 있음을 관찰.

\* 사용자와 비즈니스 간의 지리적 거리는 중요하고 효과적인 요소임에도 불구하고, 비즈니스 등급 예측에 부정적인 영향을 미침

Memory-based :

유사성 측정을 사용하여 유사한 사용자/항목을 찾는 것.

일반적으로 사용되는 유사성 측정은 사용자-항목 등급, 애플리케이션에 따라 다른 사용자/항목별 특징을 기반으로 계산된 Pearson의 상관 관계 및 코사인 유사성

Neighborhood based CF라고 불리기도 함

Model-based :

기계 학습 기술을 사용하여 관찰된 사용자 항목 등급에서 모델을 만든 다음, 관찰되지 않은 등급을 예측.

Latent factor model이 CF중 가장 성공적인 모델 중 하나(사용자와 항목을 훨씬 낮은 차원의 공유 잠재 공간에 공동으로 매핑)

Latent factor model의 성공적 실현 - Matrix factorization(MF) -> Yahoo의 음악 등급 예측 등 다양한 추천자 시스템에 성공적으로 적용

SVD++는 MF 모델 중 가장 성공적인 모델 중 하나 :

사용자의 암묵적 피드백 정보를 항목에 통합. (사용자의 구매 내역 또는 검색 기록에 기반)

@ 지리적인 영향을 잠재 요인으로 쓴건 우리가 하려는 사용자 기반의 협업 필터링과는 좀 다른 듯. 모델 부분에서는 도움을 못 받을 것 같고, reference로 모델 언급은 해볼 수 있을 듯.

@ Rating Prediction 부분은 숫자라서 안 읽었는데, 어차피 모델을 구축해놓으면 이건 읽어보고 참고할 수 있으면 참고하는 것도 나쁘지 않을 것 같음