

중간 점검 제출(고려대학교)

한국의 미래 ODA 시행 방향에 대한 제언

한국의 미래 ODA 시행 방향에 대한 제언

진행 상황 정리(11/17)

1. 아시아경제 K원조 추적기 내용 정리
2. 연관성 분석: NET ODA에 영향을 미치는 주요 변수 선정
3. Clustering: 우리나라와 지표 상의 비교로 정책적으로 참고할 만한 국가 선정

논리 전개 정리

1. 서론
 - 1-1. 문제 의식
 - 1-2. 문제 해결 계획
2. 본론
 - 2-1. 연관성 분석
 - 2-2. Clustering
3. 결론

질문 사항 정리

진행 상황 정리(11/17)

1. 아시아경제 K원조 추적기 내용 정리

- 유엔 ODA/GNI 목표치를 충족한 나라는 노르딕 주요 5개국으로, 이들을 ‘선진공여국’이라 할 수 있다.
- 한국은 노르딕 국가 및 일본과 비교했을 때, 낮은 무상원조 비율, 정책과 집행의 담당 기관 분리로 인한 원조 분절화, ‘예비검토제’에 기인한 느린 정책 추진 등의 문제를 가지고 있다.

2. 연관성 분석: NET ODA에 영향을 미치는 주요 변수 선정

- 60개의 데이터 수집
 - 단순 USD 규모가 아닌, per capita나 % of GDP의 형식으로 국가마다 규모의 차이를 완화하는 지표로 선정
 - Robust Scaling, Imputation 등 전처리
- Feature selection을 위한 10개 이상의 모델 실험

- 상관계수, 변수 중요도, Lidge, Lasso, Random Forest 등
- 모델의 공통된 주요 변수를 종합해 정리
- 향후 과제
 - 해당 변수들이 ODA 액수에 영향을 미친 이유에 대해 심화 탐구 예정

3. Clustering: 우리나라와 지표 상의 비교로 정책적으로 참고할 만한 국가 선정

- Hierarchical Clustering(Average Link) 선택
- 중간 피드백을 토대로 변수 간 Scaling 진행
- imputation을 통해 안정적인 Clustering 결과를 내도록 조치할 예정
- 현재 PCA 적용에 대한 우려가 있으며, 좋은 결과를 낼 수 있는지 실험할 예정

논리 전개 정리

1. 서론

1-1. 문제 의식

- ODA/GNI 비율은 DAC 평균 0.39%이나, 한국은 0.17% 수준
- CPA의 비중 감소와 효율성 의문 및 실질적인 증액 여부 불투명
- 낮은 무상원조 비율과 정책과 집행의 담당 기관 분리, '예비검토제'에 기인한 느린 정책 추진 등의 정책적 문제

1-2. 문제 해결 계획

- 연관성 분석: 데이터와 머신러닝 기법을 활용해 NET ODA와 여러 지표 사이의 연관성을 분석하고, 주요한 변수를 선정
- Clustering: 주요 변수를 포함하여 ODA 이행 성과에 따른 DAC 회원국을 군집화 하고 제도적으로 참고할 타 국가 선정

2. 본론

2-1. 연관성 분석

- t년도 변수를 활용해 t+1년도 NET ODA를 예측하는 Task를 정의하고, 모델의 성능이 좋은 경우에 주요 변수를 잘 선정한 것으로 가정
- 적용한 Method(Filter Method, Embedded Method)에서 공통적으로 선정된 상위 N개의 변수를 정리
- 아래는 대략적인 결과로, 각 Method가 공통된 변수를 선정하는 경향이 있음
- 변수에 Scaling(StandardScaler, MinMaxScaler, RobustScaler, Normalizer)을 적용한 결과도 도출
- Filter Method
 - Pearson, Spearman, Kendall Tau, Feature Importance(Random Forest), Permutation Importance, OLS Regression, F-regression (총 7가지)

*아래는 선정된 주요 변수

- Military spending
- Gini index
- Agriculture, forestry, and fishig, value added
- Trade
- R&D spending
- Medium and high-tech exports
- Average wages
- Exports of goods and services
- Population ages 0-14
- atc.
- Embedded Methods
 - Lasso, Ridge, Elastic Net (총 3가지)

*아래는 선정된 주요 변수

- Poverty rate
- Military spending
- R&D spending
- Population growth
- Life expectancy at 65

- Survival to 65_male
- Population ages 0-14
- Business extent of disclosure index
- Old-age dependency ratio
- Labor force
- atc.
- 향후 액션스텝 : 선정된 주요 변수에 대해 조사하여 ODA 액수에 영향을 미치는 이유에 대한 심층 분석

2-2. Clustering

- ODA 실적에 대한 지표를 사용해 Clustering 하여 참고할 수 있는 국가를 선정
- 사용하는 변수는 연관성 분석에서 선정된 변수 포함 약 20개
- 현재 두 가지 방법론에 대해 구현 및 성능 측정 중이나, 1번 방법론에서 PCA 적용에 대한 우려 사항 존재
 - 1번 방법론 (PCA 적용에 대한 우려 사항 존재)
 - PCA를 통해 고유값으로 70%의 설명력을 갖는 주성분 개수를 파악해 m개 ($m < p$)의 주성분이 필요함을 확인 ($m = 2, 3$ 인 경우가 가능한지 확인 → 시각화의 용이성 확보 가능)
 - 축소된 주성분을 활용해 Clustering
 - (우려 사항) K-means 방법 등에서는 변수의 개수가 많아질 때 잘 작동하지 않는 경향이 있어 차원 축소 기법을 활용해 변수의 개수를 줄이는 Tandem 군집 분석이 활용됨. 하지만 이 역시 차원 축소 방법을 이용해서 찾아낸 축소 차원이 반드시 군집에 대한 구조를 잘 반영할 것이라는 보장이 없음. 특히 군집의 구조와는 상관없는 변수들의 분산 또는 공분산이 클 때, 주성분 분석을 통한 차원 축소는 오히려 군집의 구조를 가릴 수 있다는 문제.
 - 우려 사항 등의 이유로 Clustering의 결과가 좋지 않다면, PCA를 적용하지 않고, 적은 피쳐 개수로 Combination해서 Clustering을 여러 번 수행하는 방법을 선택해야 함
 - 2번 방법론
 - NET ODA와 ODA/GNI로 Clustering 실시
 - K원조 추적기에서 제시한 바와 유사한 결과 확인

- 북유럽국가 (제 1그룹), 한국 포함된 국가(제 2그룹), 미국/일본 등 국가(제 3그룹)
- 그룹 별 특징이 명확하고, 각종 ODA 저널에서 명시한 그룹 별 ODA 집행에서의 공통된 특징을 확인 가능
- 문제 제기한 바와 일맥상통하는 결과를 얻어 결론에서 방향을 제시할 때 근거로 채택 가능

3. 결론

- 1번 방법론을 적용한 Clustering 결과를 제시
 - 우려 사항 해소, 성능 검증 등의 작업 진행 예정이며, 한국과 지표상 유사한 국가 Cluster를 구별하는 것이 목적
- 2번 방법론을 적용한 Clustering 결과를 제시
 - 서론에서 제기한 문제에 상응하는 결과를 도출
 - 상업주의 모델, 인도주의 모델 등으로 나뉘며, 한국이 가지는 문제를 해소하기 위해 Cluster별 모델을 참고하는 것이 목적
- Clustering 결과로 도출한 여러 국가의 사례를 참고해야 한다고 제안
 - 가령, ODA 관장 일원화 기구가 있는 국가를 참고해 우리나라 방향 설정

질문 사항 정리

Q. 앞서 언급한 대로, 차원 축소를 실시했을 때의 통계학적 관점에서 우려가 있습니다. 하지만 여러 변수를 동시에 활용해 Clustering 할 수 있다는 점에서 차원 축소를 얻을 수 있는 이점이 큰 상황인데, 결과적으로 PCA 수행 결과 더 적은 차원(2 또는 3개의 주성분)으로 잘 적합할 수 있다면 그 결과를 활용해 Clustering을 해도 괜찮을까요?

Q. 차원 축소를 적용하는 방식이 문제가 있다면, 이를 적용하지 않고 적은 차원으로 nC_2 또는 nC_3 Clustering을 실시하고자 합니다. 여러 번의 Clustering을 진행하여 최종적으로 '한국과의 ODA 집행 유사성을 가진 국가'를 '한국과 같은 Cluster로 묶이는 횟수가 많은 국가'로 선정해도 괜찮을까요? 여러 번 Clustering을 실시하는 해당 방법에 대한 조언을 주실 수 있으신가요? (한국과 유사한 국가를 군집화 하는 더 똑똑한 방법이 있을지 궁금합니다.)

- 본 방식에 대해 예시를 들면, A, B, C, D라는 4개의 변수를 활용한다고 할 때 $4C_2 = 6$ 번의 Clustering을 수행할 수 있습니다. 이 때 6번 중 가장 많은 4번 정도 한국과 같은 Group으로 묶이는 국가를 '한국과의 ODA 집행 유사성을 가진 국가'로 선정합니다.