<사회보장정보 Brief 제21호> 2018년10월19일

복지사각지대 대상자 발굴시스템 변수 구성에 따른 실험 결과 분석

이우식 부연구위원
(☎: 02-6360-4623)

1. 들어가며

- 복지사각지대 복지 대상자 발굴시스템은 데이터 수집부터 데이터 전처리와 모델의 처리 부분 등 다양한 요소가 결합된 시스템
- 통합 발굴 대상자 데이터 변수구성은 최종적으로 추출되는 대상자에 많은 영향을 미치며, 통계모형 기반 추론을 하는데 있어 상의한 결과를 가져옴.
- 통계 모형에서 유형이 비슷하거나 불필요한 변수 제거는 다중공선성(Multicollinearity)을 제거하여 정확도를 높일 수 있음. 반면 중요한 변수의 제거는 부정확한 결과를 가져올 수 있음.

※ 다중공선성: 독립변수들 간에 강한 상관관계가 나타나는 문제

2. 통계 모형의 변수 구성

- 독립 변수 (Independent Variable)
  - ○ 복지사각지대 대상자 발굴시스템에서 독립 변수는 다른 변수에 영향을 주는 변수로써 원인이 되는 변수임. 예를 들어, A라고 하는 변수가 최종 발굴대상자에 영향을 주는 변수가 되면 A는 통계모형에서 독립변수가 됨.
- 종속 변수 (Dependent Variable)
  - ○ 복지사각지대 대상자 발굴시스템에서 종속 변수는 다른 변수로부터 영향을 받는 변수로 결과물이나 효과를 나타내는 변수임. 예를 들어, 추론을 위한 A라고 하는 독립 변수를 넣어서 나온 결과가 B변수이면, B변수는 종속 변수가 됨.

3. 대상자 발굴을 위해 사용되는 변수 현황

- 복지사각지대에서 정보 연계 현황
  - ○ 복지사각지대 정보 연계를 위해 14개 기관 27종의 정보를 수집하여 위험 가구를 추론하여 지방자치단체를 통해 조사 및 상담을 하여 공공부조/민간서비스를 제공하고 있음.
- 대상자 발굴을 위해 사용되는 변수 현황 (27종)
  - ○ 단전, 단수, 단가스, 전기료체납, 국민연금체납, 건강보험료체납, 화재피해, 피부양의무자장기요양, 전세금액기준이하가구, 월세금액기준이하, 고용보험개별연장급여대상, 고용보험실직사유, 고용보험비대상, 방문건강집중관리, 기저귀조제분유지원, 신생아난청, 자살예방관리, 자살시도, 위기학생, 범죄피해, 시설입퇴소, 기초생활긴급지원수급자, 공공임대주택, 산재요양종결후근로단절, 재난피해자, 금융연체자, 의료비용과다지출가구

4. 변수 구성에 따른 실험 결과 분석

- 본 실험을 하기 위한 데이터는 복지사각지대 변수 27종을 가지고 하였으며, 머신러닝 기법으로 실험을 진행하였음.
- 불필요한 변수 제거에 따른 실험 결과 분석
  - ○ 불필요한 변수를 제거하기 전과 비교하였을 때 불 필요한 변수 5개를 제거하면 정확도(Accuracy)와 민감도(Sensitivity)가 증가되는 결과가 나옴.

<표 1> 불필요한 변수 제거에 따른 실험 결과

	As-is	To-be
Accuracy	74.8%	75.8%
Sensitivity	36.23%	48.46%

- 중요한 변수 제거에 따른 실험 결과 분석
  - ○ 중요한 변수 제거에 따른 실험결과는 전체적으로 정확도와 민감도 모두 떨어지는 결과를 보여주며, 불필요한 변수를 제거 하는 것에 비해 중요한 변수 제거가 더욱 더 많은 영향을 미친다는 것을 알 수 있음.

<표 2> 중요한 변수 제거에 따른 실험 결과

	As-is	To-be
Accuracy	74.8%	64.41%
Sensitivity	36.23%	12.69%

※ Accuracy = (TP/TN)/(TP_FP+FN+TN)
※ Sensitivity = TP/(FN+TP)
※ TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative

5. 맺음말

- 신규 및 파생변수 추가를 통한 대상자 확대 방안
  - ○ 실험 결과 불필요한 변수를 제거하면 정확도와 민감도가 증가되어 실제로 대상자가 뽑힐 확률이 증가하는 반면, 중요한 변수를 제거하게 되면 정확도와 민감도가 떨어진다는 결과를 얻었음. 이런 결과를 통해 중요한 변수의 추가와 불필요한 변수의 제거는 더욱 더 정교한 통계모형을 만들 수 있는 가능성을 보여줌.
- 비정형 데이터 확대를 통한 복지 수급 대상자 정확도 제고
  - ○ 음성, 문서, 상담자료 등 복지사각지대에 대한 비정형 데이터가 많이 증가하고 있지만 아직까지 활용도는 낮은 상황임. 만약 비정형 데이터 분석을 통해 활용도를 높여 중요한 변수를 증가 시킨다면 통계모형 정확도 향상에 도움이 될 것으로 예상 됨.

※ 본 실험 데이터는 데이터 생성기로 생성된 데이터로써 실 데이터로 실험 시 결과가 달라질 수 있음.