

생명보험 계리사의 새로운 도전

: 예측모형과 초개인화 계리적 가정의 시대

서 경석

리스크관리위원회 위원
윌리스타워스왓슨 이사

Disclaimer: 본고의 내용은 저자 개인의 의견이며 회사의 의견과는 무관합니다.

인공지능과 머신러닝의 시대를 맞아 보험계리사의 역할도 진화하고 있습니다. 특히 손해보험(P&C) 분야에서 활발히 활용되던 예측모형(Predictive Modeling)이 이제 생명보험(Life) 영역으로 확장되는 추세입니다. 이는 단순한 신기술의 도입을 넘어, 생명보험 계리사들이 데이터 기반의 의사결정과 정교한 계약자 행동 예측을 통해 보험산업의 미래를 재설계하는 혁신의 과정이라 할 수 있습니다.

이미 손해보험에서는 언더라이팅, 청구된 보험금 처리의 우선순위 분류 및 자동화, 보험사기 탐지, 소송 가능성 평가, 보고서 생성 자동화 등 다양한 영역에서 머신러닝 기반 모델이 실질적인 성과를 내고 있습니다. 대표적으로 자동차 보험에서는 연령, 성별, 차량 종류, 사고 이력 등 기준에 활용했던 다양한 요인을 GLM(Generalized Linear Model)으로 분석하여 요율을 산정하는 리스크 기반 가격 책정의 계산이 이루어지고 있으며, 이는 머신러닝 기법을 활용한 리스크 기반 가격 책정의 대표적인 사례입니다.

이러한 접근은 생명보험 분야에서도 충분히 응용 가능하며, 특히 계리적 가정 설정에 있어 새로운 가능성을 열어줍니다. 예를 들어, 사망률이나 질병 발생률 같은 생명보험의 핵심 가정은 과거 통계에 기반해 설정되어 왔지만, 최근에는 스마트워치나 웨어러블 기기를 통해 수집되는 생체정보와 결합하거나, 의료기록 데이터를 기준의 과거 통계 데이터에 붙여서 활용함으로써 보다 개인화되고 Dynamic한 가정 설정의 가능성을 열어주고 있습니다.

또한 계약자 행동 중 가장 중요한 변수 중 하나인 해지율은 개인의 자산 수준, 금융 이해도, 거시경제 지표(금리, 주가지수, 실업률 등)와 같은 보험 외 요인을 함께 고려함으로써 기존보다 훨씬 정교한 예측을 가능하게 합니다. 이러한 시도는 단순히 계리적 가정의 정밀도를 높이는 것을 넘어, 보험사의 리스크 관리, 상품 설계, 자산/부채 관리에 실질적인 영향을 줄 수 있으며, 생명보험 계리사의 역할을 데이터 기반의 전략적 의사결정자로 확장시키는 계기가 될 수 있습니다.

GLM은 복잡하고 Black-Box 형태인 한 인공 신경망 기반 모델에 비해 설명 가능성(Explainability)이 높다는 점에서 보험업에 특히 적합한 도구입니다. 모델 구조와 변수의 영향력을 명확히 해석할 수 있어, 계리적 가정 설정 시 방법론의 투명성과 이해 가능성성이 중요한 보험업계에서 큰 장점으로 작용합니다. 더 나아가 GLM은 다양한 위험 요인을 동시에 반영하고, 변수 간 상호작용과 상관관계를 고려할 수 있어 기존의 단순 모델보다 훨씬 유연합니다.

그러나 좋은 GLM모델이 구축되기 위해서는 양적으로 충분한 규모의 데이터, 정제된 데이터의 질, 데이터 모델링에 대한 전문역량 및 Domain에 대한 깊은 지식이 필요합니다. 많은 다른 머신러닝 기법들이 그러하듯, 설명변수를 무분별하게 증가시켜 불필요한 복잡성을 초래하여 모델 해석이 어려워지고 과적합이 될 수 있는 점 또한 유의해야 합니다.

다만 최근에는 Factor Importance 분석, PCA(Principal Component Analysis, 주성분 분석), 변수 변환 및 Interaction 적용 등을 통해 많은 변수를 효과적으로 다루는 방법론이 정립되어 있으며, 이미 해외 보험업계에서도 활발히 활용되고 있습니다. 따라서 데이터 수집이 충분히 가능하다는 전제 하에, 설명변수 수가 많아진다는 점은 분석의 앞단에서 수행해야 할 추가적인 작업이 더 존재하는 것으로 이해할 수 있습니다.

이와 같은 점들을 고려하여, 계리적 가정 수립에 대한 GLM 방법론의 장단점을 정리해 보면 다음과 같습니다:

◆ 장점

- 설명가능성 높음: 모델 구조와 변수 영향 해석이 용이해 보험업에 적합
- 다양한 변수 반영: 여러 위험 요인을 동시에 고려해 정교한 예측 가능
- 상호작용 및 상관관계 처리: 변수 간 상호작용 관계를 반영해 예측 정확도 향상
- 적합성 평가 용이: AIC, p-value 등 통계적 지표로 가정의 적합성 평가 가능

◆ 단점

- 데이터 요구량 큼: 충분한 데이터 없으면 신뢰성 낮음
- 비선형 관계 제한: 기본적으로 선형 구조이므로 복잡한 패턴은 추가 기법 필요
- 과적합 위험: 과도한 설명변수 포함시 모델이 학습 데이터에만 지나치게 적합됨
- 또한 이러한 GLM을 의미 있게 사용하기 위해 준비되어야 할 데이터를 크게 4가지 영역으로 나누어 생각해 보았습니다:

◆ 보험사 내부 데이터

- 계약정보: 상품유형, 보험기간, 납입기간, 납입주기, 가입금액, 보험료, 선택특약, 약관대출, 추가납입/중도인출, 연금개시, 해지공제액 등
- 고객특성: 성별, 연령, 가입시점, 가입 후 경과년수 등
- 과거 경험 데이터: 질병, 장해 발생 등 과거 보험금 청구 이력, 과거 보험상품 해지 이력, 해지 시점의 경과 차월, 적립금 상태 등
- 심사서류, 콜센터 로그, 마케팅 반응 이력 (캠페인 참여 여부) 등

◆ 보험사 외부 공개 데이터

- 거시경제 지표: GDP, 실업률, 기준금리, 장단기 국고채 금리, 주가지수, 환율, 물가상승률, 금융시장 변동성, 주택가격 지수, 경기선행지수 등
- 보험시장 데이터: 경쟁사 보험료 수준, 경쟁사 상품 급부 및 특성 등, 업계 평균 해지율, 산업통계 기준 사망률, 질병발생률, 사고율 등
- 공공 통계: 인구 통계, 출생 사망률, 질병 발생률 및 유병률, 지역별 건강 지표 관련 통계 등

◆ 기타 개인 정보

- 라이프 스타일: 여행 빈도 및 목적지, 온라인 쇼핑 기록(건강 관련 제품구매 등), 외식 및 문화생활 지출, 차량보유 여부 등
- 위치 및 거주 정보: 거주 지역(의료접근성 참고), 이동 패턴(출퇴근 거리, 교통수단), 주거 형태, 위치기반 정보
- 사회적 특성: 학력 수준, 직업군 및 업종, 가족 구성(결혼여부, 부양가족 수) 등
- 심리 특성: 위험 선호도(투자 성향), 스트레스 지표 등

이러한 데이터 영역 중 보험사 내부 데이터는 일부 비정형 데이터를 제외하면 대부분 즉각적으로 이용 가능한 형태로 정제되어 있습니다. 보험사 외부 공개 데이터도 API 등을 통해 접근 가능하며 잘 정제되어 있는 경우가 많습니다. 재무, 의료 등 주요 개인정보는 대표적인 민감 정보로, 분명히 존재하는 데이터이지만 보험사 입장에서는 계약자 개인의 동의 없이는 활용할 수 없습니다.

기타 개인 정보 등도 마찬가지이며 이런 정보들은 계약자 개인의 동의가 있다 하여도 이를 효율적으로 기존 데이터에 결합시키려면 데이터 정제, 식별자 매칭, 동의 관리 체계 구축 등을 포함한 추가적인 기술적 작업이 필요할 수 있습니다.

결국 예측모형의 성패는 모델 자체보다는 데이터의 품질과 결합 방식에 달려 있습니다. 생명보험에서 GLM을 활용해 계리적가정을 제대로 설정하려면, 계약자의 의료 정보나 금융 정보 같은 민감 데이터를 보험사의 계약 데이터와 정교하게 연결하는 작업이 핵심입니다.

기술적으로 이러한 데이터 결합은 점점 더 쉬워지고 있지만, 개인정보 보호와 프라이버시를 침해하지 않으면서 분석 목적으로 활용할 수 있는 체계를 마련하는 것이 업계의 가장 큰 과제입니다. 특히 생명보험업계에서는 계약자 개개인의 특성과 행동 패턴을 반영한 “초개인화된 계리적 가정 설정”이 점차 현실화되고 있으며, 이는 기존의 획일적인 통계 기반 접근을 넘어, 보험 상품의 설계와 리스크 관리에 있어 새로운 전략적 가능성을 열어 줍니다. 데이터 활용과 보호의 균형을 어떻게 설계하느냐가 앞으로 보험계리 업무의 혁신을 좌우할 것입니다.

제 개인적으로는, 이러한 데이터 결합 방식과 계리적 가정 설정의 정교화에 대해 좀 더 연구를 진행하여, 내년 계리사회 뉴스레터에 후속 기고를 통해 구체적인 사례와 프레임워크를 공유할 수 있기를 기대하고 있습니다.

