한국통계학회논문집
2010, 17권, 1호, 1-8

데이터마이닝을 이용한 국민연금 부정수급 예측모형 개발 -손해배상금 불성실 신고를 대상으로-

차경엽[1,a]

[a]감사연구원

요 약

최근 사회복지분야에서 부정수급, 횡령 등이 빈번히 발생함에 따라 비리를 방지하기 위한 체계적인 관리 방안이 요구되고 있다. 데이터마이닝은 다수의 이해관계자와 많은 예산이 투입되는 사업을 관리하는데 효과적인 방법이다. 본 연구는 국민연금의 부정 수급자 관리방안으로 데이터마이닝을 이용한 예측모형을 개발하였다. 분석결과, 수급자의 급여, 연금 가입, 사고내역 정보가 부정수급의 특성 요인으로 나타났으며 이를 의사결정나무 모형, 로지스틱 회귀모형, 인공신경망 모형에 적용한 결과 의사결정나무 모형의 예측력이 가장 우수한 것으로 분석되었다.

주요용어: 국민연금, 부정수급, 데이터마이닝.

1. 서론

국민연금은 국민의 노령, 장애 또는 사망에 대하여 연금급여를 실시함으로써 국민의 생활안정과 복지증진에 이바지하는 것을 목적으로 하고 있다. 국민연금은 시행과정에서 저부담·고급여 구조와 인구의 고령화 및 생산가능 인구 감소 추세로 인해 향후 재정적자를 초래할 가능성이 높은 것으로 예측되고 있다. 또한 최근에는 이러한 구조적인 측면뿐만 아니라 국민연금공단의 수급자 관리가 미흡한 것으로 나타나 효율적인 관리방안 마련이 요구되고 있다. 2009년 국정감사에 따르면 지난 10년간 국민연금 부정수급 사례는 21만 5,524건이 적발됐으며 부정수급액은 1,211억 원인 것으로 파악되고 있다. 이는 국민연금공단이 적발한 사례이고 전체 부정수급 규모는 이 보다 클 것으로 예상된다. 부정수급의 문제는 수급자의 도덕적인 자질이 일차적인 책임이긴 하나 궁극적으로는 부정을 방지하기 위한 관리방안이 미흡한데 원인을 찾을 수 있다. 국민연금공단은 부정수급 담당자 87명이 253만 명의 수급자를 관리하고 있어 부정수급을 방지하는데 한계가 있다고 하나 고령화 사회로 인해 향후 수급자 규모가 증가될 것으로 예상되는 상황에서 인력 충원만이 대안일 수는 없다. 즉, 과학적인 부정수급 탐지 기법을 통해 부정수급 발생 가능성이 높은 유형의 수급자를 식별하여 이를 집중적으로 관리하는 방안 마련이 시급한 실정이다. 여기서 부정 탐지기법은 관련 기관 간 내부정보시스템을 서로 공유하여 부정을 적발하는 방법인 데이터 공유(data sharing)와 정보간의 상호 연관성을 파악하고 이를 모형화하여 부정을 탐지하는 데이터 마이닝(data mining) 기법 등이 있다. 이중 데이터 마이닝 기법은 자료간의 상호 연관성을 기초로 부정 가능성이 높은 집단을 추출하는 것으로 국민연금처럼 이해관계자가 많고 예산규모가 큰 사업을 관리하는데 효과적이다. 미국, 영국 등 선진국에서는 민간에서 뿐만 아니라 공공부문에서 데이터마이닝을 적극 활용하고 있다. 미국은 51개 부처 또는 공공기관에서 데이터 마이닝을 사용하고 있으며 활용 분야는 서비스 및 성과개선, 부정·방지·남용 탐지인 것으로 조사되었다 (GAO, 2003). 또한 영

[1] (110-793) 서울시 종로구 계동 현대빌딩 감사원 감사연구원, 연구관. E-mail: kycha@korea.kr

국은 연기금 부정수급자 탐지에, 호주 국세청은 세금탈세 탐지에, 뉴질랜드는 기초사회보장제도의 부정수급자 탐지에 데이터마이닝을 활용하고 있다 (오윤섭 등, 2009). 이에 비해 우리나라는 국민건강보험 등 일부 공공부문에서만 활용되고 있는 수준이다.

따라서 본 연구는 국민연금 운영의 효율적인 관리방안으로 데이터마이닝을 활용한 부정수급 예측모형을 개발하였다. 그동안 수급자의 특성을 분석한 연구는 있었으나 부정수급에 대한 특성을 분석한 연구는 없어 연구결과가 국민연금에서의 부정탐지를 위한 관리방안으로 활용될 수 있으리라 판단된다.

2. 국민연금의 부정수급에 대한 이해

부정수급이란 수급요건을 충족하지 못함에도 불구하고 급여를 받거나 자신이 받아야 할 급여수준 이상으로 급여를 받는 경우를 말한다. 부정수급집단의 존재는 제도의 건전성을 해치고 장기적으로 제도에 대한 사회적 지지를 약화시키는 부정적 효과를 초래하게 된다. 연기금의 운영상의 부정수급 현황을 살펴보면, 미국 요부양가족부조(Aid to Families with Dependent Children; AFDC)의 경우 부정수급률은 공식적으로 3~5%로 조사되었으나 전문가들은 30~50%에 이를 것으로 추정하고 있다 (Hutton, 1979). 이에 미국 정부는 품질관리 개념을 도입하여 허용오류수준을 초과하는 부정수급율이 발생하는 주정부 및 지방정부에 대해 이를 시정하도록 요구하고 있다. 여기서 미국 정부가 제시한 AFDC의 허용오류수준은 3%이다. 그리고 '04년 미국 노동부의 실업보험급여의 부정수급율은 5.07%, '03년 미국 주택도시개발부의 공동주택기금의 부정수급율은 5.1%인 것으로 조사되었다. 또한 영국 사회보장성 보고서에 따르면 2007/2008년도 수입지원 및 구직수당 프로그램(Income Support and Jobseek's Allowance)의 부정수급률은 4.9%, 연금급여(Pension Credit)에서는 5.1%로 추정하고 있다. 외국의 사례를 볼 때, 연기금 운영에서의 부정수급율은 5% 내외인 것으로 판단된다. 이와 관련하여 우리나라는 공식적인 통계는 없으나 이와 유사한 부정수급율이 나타날 것으로 예상되고 있다.

국민연금에서의 부정수급 정의는 국민연금법에 제시되어 있다. 국민연금법 제57조에 따르면 부정수급자를 거짓이나 그 밖의 부정한 방법으로 급여를 받거나 수급권이 소멸 또는 정지된 급여 그 밖의 과오납된 급여를 받는 자로 규정하고 있다. 국민연금공단은 외부자료의 활용 및 수급자 관리를 통해 부당이득이 확인된 경우 환수조치를 하고 있다. 즉, 외부자료의 활용은 행정안전부의 사망자료, 국세청의 소득자료, 건강보험공단의 건강보험 자료 등 타 기관으로부터 정기적으로 자료를 제출 받아 부정수급자를 추출하고 있다. 그리고 수급자 관리는 수급권자를 대상으로 수급관련 변동 여부 등을 지속적으로 관리하는 것으로 공단 차원에서 정기적으로 이루어지기 보다는 지점 및 관리자별로 이루어지고 있어 관리가 체계적이지 못하다.

2008년도 국민연금 수급자 중 부당이득을 환수한 사례를 살펴보면, 환수금액은 총 206억 원이며 환수사례는 총 27,834건으로 나타났다. 부당이득 환수사례를 보면 부양가족연금미해당(13,238건)이 가장 많으며 자격징수 내역변경(5,727건), 소득정지(4,341건) 순이었으며, 금액 순으로는 소득정지가 88억 원으로 가장 많은 것으로 조사되었다 (표 1).

부당이득 환수사례를 살펴보면, 대부분 기존 자료의 점검이나 외부자료를 활용한 데이터 매칭방법(data matching)으로 확인이 가능하다. 예를 들어, 수급자의 사망 등 소멸 여부는 행정안전부의 전 국민 주민전산 자료를 통해, 산재급여 수령여부는 근로복지공단의 산재보험 성립신고 자료 및 고용보험 사업장 자료를 통해 부당이득 여부를 확인할 수 있다. 그러나 환수사례 중 손해배상금 불성실 신고의 경우 민간 보험사를 통해 손해배상금을 수령하였는지, 수령한 금액을 정확히 신고하였는지를 확인하는 것이 필요하나 개인정보보호로 인해 자료 확보가 어려워 데이터 매칭방법을 통해 부당이득을 확인하기 어렵다. 따라서 본 연구는 국민연금 부정수급 사례 중 손해배상금 불성실 신고 여부를 데이터마이닝 기법을 이용하여 분석하고자 한다.

표 1: 2008년도 국민연금 부당이득 환수 현황 (단위: 백만원)

구분		합계	노령연금	장애연금	유족연금	반환일시금	사망일시금
합계	건수	27,834	19,588	3,017	4,364	738	127
	금액	20,641	14,348	2,594	2,428	1,178	91
소멸	건수	3,438	2,054	376	1,007	1	0
	금액	1,967	790	395	782	0	0
소득정지	건수	4,341	4,161	1	179	0	0
	금액	8,806	8,427	0.1	379	0	0
손해배상금수령	건수	156		38	118		
	금액	175		104	71		
산재급여수령	건수	136		19	117		
	금액	345		157	187		
구직급여수급	건수	7	6		1		
	금액	4	4		0.1		
부양가족연금 미해당	건수	13,238	10,893	807	1,538		
	금액	1,043	773	64	204		
수급권취소	건수	791	360	64	69	263	35
	금액	3,614	1,904	538	263	830	77
자격징수 내역변경 등	건수	5,727	2,114	1,712	1,335	474	92
	금액	4,683	2,448	1,333	540	348	13

자료: 국민연금공단 내부자료 (2009)

3. 손해배상금 불성실 신고의 특성분석

손해배상금 불성실 신고란 제3자의 가해로 인해 장애연금 및 유족연금 수급권자가 가해자로부터 손해배상금을 수령하였는데도 동 사실을 신고하지 않거나 수령금액 또는 합의금액을 허위로 감액 신고하여 지급 정지될 연금을 지급받는 경우를 말한다. 즉, 제3자의 가해로 인해 장애 및 유족연금을 지급받는 경우 총 연금액은 연금 지급액과 지급기간에 따라 결정되며 연금 지급액은 가입경력, 장애정도, 타급여 수령여부 등에 의해, 지급기간은 일실이익(손해배상금 – 의료비 – 장재비 등), 지급정지요건 등을 고려하여 산정한다. 여기서 지급정지기간은 손해배상금에서 의료비 등을 제외한 비용을 평균소득으로 나누어 산출하므로 손해배상금을 적게 신고할수록 지급기간은 늘어나며 수급자의 총 연금액은 원래보다 많아진다. 따라서 손해배상금 불성실 신고에 대한 예측모형을 개발하기 위해 전체자료를 제3자 가해로 인해 장애 및 유족연금을 수급한 자를 대상으로 하였다. 2003년부터 2009년 4월까지 장애연금 및 유족연금 수급자는 총 304,481명이며, 연금별로는 장애연금 89,449명, 유족연금 215,032명으로 구성되었다. 그리고 유족연금은 장애연금 및 노령연금 수급권자의 사망으로 인해 부양가족이 급여를 대신하여 수령하는 것으로 노령연금으로부터의 유족연금 수급자는 분석에서 제외하였다. 또한 수급권자가 손해배상금 불성실 신고자는 아니나 소득 정지 등 다른 부당이득 사유로 부정수급권자로 판정된 경우도 제외하였다. 왜냐하면 손해배상금 불성실 신고의 특성을 분석하는 것이 목적이므로 다른 부정수급 특성은 배제하는 것이 적정하다고 판단했기 때문이다. 그리고 자료의 중복 및 이상치를 파악한 결과, 자료의 중복은 없는 것으로 나타났으나 일실이익은 있으나 손해배상금이 없는 이상치 자료가 24건 존재하여 분석에서 제외하였다. 따라서 제3자 가해로 인한 연금수급자는 장애연금이 총 21,489명, 유족연금이 15,464명으로 총 36,953명이다. 그리고 공단으로부터 부당이득 환수조치가 이루진 수급자를 부정수급자로, 그렇지 않은 수급자를 정상수급자로 구분할 때 장애연금의 부정수급자 비율은 0.59%, 유족연금은 2.69%로 나타났다.

박능후와 송미영 (2008)은 국민기초생활보장제도에서 정상 집단과 부정수급 집단을 구별하는 변수

표 2: 분석대상 변수 내역

특성	변수명	설명
신상	연령대	20대, 30대, 40대, 50대, 60대 이상
	부양가족수	없음, 1명, 2명, 3명, 4명, 5명, 6명, 7명 이상
소득	수급전년도평균소득월액	
	최종5년간평균소득평균액	수급전 5개년의 소득 평균액
	소득변동액	수급전년도평균소득월액 – 최종5년간평균소득평균액
	평균가계소득대비개인소득	가구원별 가구소득 – 수급자 소득
연금가입	가입형태변경	사업장→사업장, 사업장→지역, 지역→사업장, 지역→지역
	지역가입경력	있음, 없음
	사업장가입경력	있음, 없음
	임의가입경력	있음, 없음
	가입기간연금지급율	
	가입기간	
	평균 보험료 등급	
급여	산재등타급여가입	미가입, 산재법상급여, 근로기준법상 급여, 선원법상 급여, 어선원 및 어선재해보상법상 급여
	타급여대상여부	대상, 비대상
사고내역	장애등급	1등급, 2등급, 3등급, 4등급, 해당사항 없음
	사고비용	손해배상금 – 일실이익
	일실이익	없음, 3천만원이하, 6천만원이하, 9천만원이하, 1억2천만원이하, 1억5천만원이하, 1억5천만원 이상
	손해배상수령여부	합의종결, 손해배상일부수령, 합의후 미수령, 구상권소멸시효경과, 국외사고합의불가, 미합의, 기타제외사유

주: 일실이익은 원래 금액기준인 연속형 자료이나 산포가 커서 분포를 고려하여 범주형으로 설정하였음

로 18세 이하 아동수, 장애여부, 민성질환유무, 주거형태, 주거유형, 경제활동 참여상태, 국민기초생활보장제도 인지여부, 미래의 주관적 경제생활, 부양의무자의 부양능력 등을 선정하였다. 또한 정우진 등 (2003)은 국민건강보험 지역보험료 체납에 영향을 미치는 요인으로 성별, 연령별, 가족수, 월소득액, 소득유형, 자가 및 자동차 소유여부 등을 선정하였다. 국민연금 부정수급과 직접적인 연관성은 적으나 앞에서 제시한 유사사례 및 체납영향요인을 살펴본 결과, 정상집단과 그렇지 않은 집단을 구별하는 요인은 수급자의 신상정보, 소득정보, 장애정보 등이 유의한 것으로 나타났다. 따라서 손해배상금 불성실 신고의 특성요인으로 신상, 소득, 연금가입, 급여, 사고 정보를 수집하였다. 공단의 「국민연금 DB 테이블 및 컬럼 현황」을 보면, DB테이블은 총 1,395개이며 세부적으로 자격정보가 362개, 징수정보가 475개, 급여 및 중증장애정보가 410개, 고객정보가 148개로 구성되어 있다. 이 중 특성요인과 연관성이 높은 DB테이블은 총 10개이며 관련 변수는 총 19개인 것으로 나타났다. 즉, 신상 정보는 연령, 부양가족수 등이, 소득 정보는 수급전년도평균소득월액, 최종 5년간 평균소득평균액 등이, 연금가입 정보는 가입기간, 가입형태변경 등이, 급여 정보는 산재 등 타급여가입, 타 급여대상여부 등이, 사고내역은 장애등급, 일실이익 등이다 (표 2).

부정수급과 연관성이 높은 변수를 선택하기 위해 변수 유형별로 구분하여 유의성을 검증하였다. 즉, 범주형 변수는 χ^2검정을 이용하였으며, 연속형 변수는 T검정 및 로짓모형의 일치도(concordant)를 통해 유의성을 검정하였다. 범주형 변수의 유의성 검정결과, 장애연금의 경우 유의수준 1%하에서 연령대, 임의가입경력을 제외하고 모든 변수가 유의한 것으로 분석되었으며, 유족연금은 가입형태 변경, 산재 등 타급여가입, 손해배상금수령여부가 유의변수로 나타났다. 또한 연속형 변수의 유의성 검정결과, 장애연금은 유의수준 1%하에서 수급전년도평균소득월액, 평균보험료 등급을, 유족연금은 가입기간, 평균보험료 등급을 제외하고 모든 변수가 유의한 것으로 분석되었다 (표 3).

표 3: 변수의 유의성 검정 결과

범주형 변수	변수명	유의확률	
		장애연금	유족연금
신상	연령대	0.1800	0.1600
	부양가족수	0.0001	0.0570
연금가입	가입형태변경	0.0001	0.0001
	지역가입경력	0.0053	0.0327
	사업장가입경력	0.0001	0.7820
	임의가입경력	0.6539	0.1798
급여	산재등타급여가입	0.0001	0.0001
	타급여대상여부	0.0001	0.8260
사고내역	장애등급	0.0001	-
	일실이익	0.0001	0.0001
	손해배상금수령여부	0.0001	0.0001

연속형 변수	변수명	장애연금		유족연금	
		유의확률	일치도	유의확률	일치도
소득	수급전년도평균소득월액	0.5693	29.0	0.0005	44.6
	최종5년간평균소득평균액	0.0001	47.5	0.0002	49.6
	소득변동액	0.0004	31.1	0.0001	29.5
	평균가계소득대비개인소득	0.0001	39.1	0.0001	28.3
	가입기간연금지급율	0.0001	48.3	0.0001	22.3
연금가입	가입기간	0.0001	48.0	0.0213	43.0
	평균보험료등급	0.0884	28.6	0.5499	16.5
사고내역	사고비용	0.0001	72.7	0.0001	52.7

4. 손해배상금 불성실 신고에 의한 부정수급 예측모형

손해배상금 불성실 신고에 의한 부정수급 예측 모형은 부정수급 여부를 목표변수로 하는 지도학습 모형인 로지스틱 회귀모형(logistic regression model), 의사결정나무 모형(decision tree model), 인공신경망 모형(neural network model)을 사용하였다. 여기서 모형구축은 불성실 신고자의 특성분석에서 유의확률이 1% 이하인 변수를 이용하였다.

그리고 모형 선택을 위해 분석용 자료는 부정수급자와 정상수급자를 동일하게 추출하였으며 이를 위해 정상 수급자를 임의추출방식으로 부정수급자 수만큼 추출하였다. 이는 기존 자료의 부정수급 비율이 너무 낮아 분석의 신뢰성이 낮을 수 있다는 판단에 따라 분석용 자료를 재구성하였다. 또한 검증용 자료도 앞에서의 사례에서 언급하였듯이 외국의 연기금 부정수급율이 5% 내외로 발생하고 있어 현실 적용 가능한 모형을 수립하기 위해 임의추출방식으로 부정수급자의 비율을 5%로 맞추어서 표본을 구성하였다. 그리고 모형의 타당성은 일반적으로 사용하고 있는 평균자승오차(average square error), 오분류율(misclassification rate), 리프트 차트(lift chart)를 이용하였다. 장애연금을 분석한 결과, 평균자승오차, 오분류율에서 의사결정나무가 다른 모형에 비해 우수한 것으로 나타났다. 또한 리프트 차트에서도 의사결정나무 모형이 상위등급에서의 리프트가 크고 하위등급의 리프트가 급격히 감소하여 로지스틱 회귀모형, 인공신경망 모형에 비해 예측력이 우수하였다. 그리고 검증용 자료의 경우 정상수급자를 정상수급자로 예측하는 비율이 83.8%, 부정수급자를 부정수급자로 예측하는 비율이 86.5%로 나타나 예측력을 우수한 것으로 나타났다. 그리고 유족연금의 경우도 평균자승오차, 오분류율, 리프트 차트에서 의사결정나무가 다른 모형에 비해 우수한 것으로 나타났다 (표 4, 5, 그림 1).

부정수급 예측을 위해 의사결정나무 모형을 적용한 결과, 수급자의 연금가입정보, 급여정보, 사고

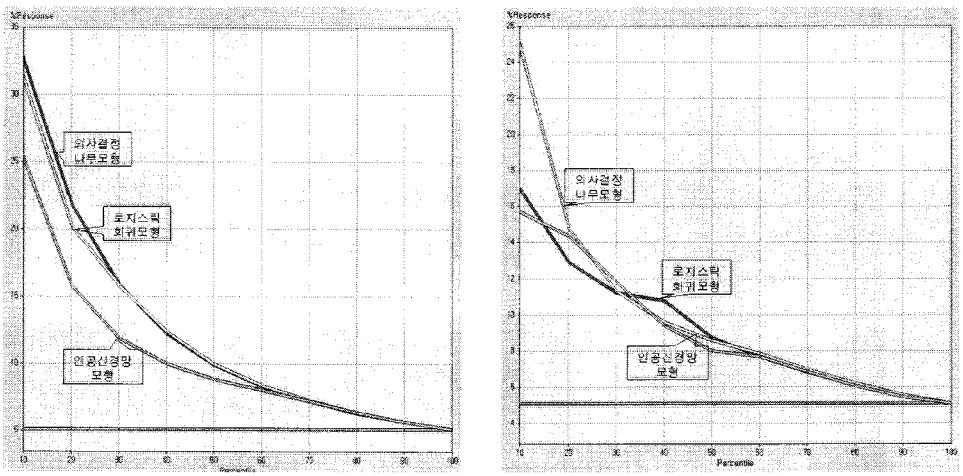
표 4: 연금별 부정수급 예측모형 결과

구분	장애연금				유족연금			
	평균자승오차		오분류율		평균자승오차		오분류율	
	분석용	검증용	분석용	검증용	분석용	검증용	분석용	검증용
의사결정나무	0.3246	0.3526	0.1507	0.1607	0.4141	0.4142	0.2464	0.2358
로지스틱회귀모형	0.3530	0.3697	0.1905	0.2052	0.4336	0.4583	0.2824	0.3357
인공신경망모형	0.4326	0.4414	0.3293	0.5682	0.4253	0.4183	0.2873	0.3503

표 5: 의사결정나무모형의 정오분류표

구분		장애연금		유족연금	
실제	예측	분석용	검증용	분석용	검증용
정상수급자	정상수급자	83.3% (105/126)	83.8% (2006/2394)	76.7% (319/416)	76.5% (6050/7904)
	부정수급자	17.7% (21/126)	16.2% (388/2394)	23.3% (97/416)	23.5% (1854/7904)
부정수급자	정상수급자	13.5% (17/126)	13.5% (17/126)	25.9% (108/416)	25.9% (108/416)
	부정수급자	86.5% (109/126)	86.5% (109/126)	74.1% (308/416)	74.1% (308/416)
오분류율		15.1% (38/252)	16.1% (405/2520)	24.6% (205/832)	23.6% (1962/8320)
정분류율		84.9% (214/252)	83.9% (2115/2520)	75.4% (627/832)	76.4% (6358/8320)

주: 괄호안은 분석 대상 표본수와 조건에 해당되는 표본수를 기술하였음(해당표본수/대상 표본수)



(a) 장애연금 (b) 유족연금

그림 1: *장애연금 및 유족연금의 리프트 차트*

내역정보가 유의한 것으로 나타났다. 세부적으로 살펴보면 장애연금의 경우, 연금가입기간, 산재 등 타급여가입, 일실이익, 장애등급이, 유족연금은 산재 등 타급여가입, 일실이익, 손해배상금수령여부가 유의한 변수로 선정되었다. 즉, 장애연금의 경우 연금가입기간이 100.5개월 미만인 수급자가 장애등급 4급 이하의 장애를 얻어 공단에 일실이익을 수령하였다고 신고한 경우의 부정수급율이 35.9%로 가

표 6: 의사결정나무 예측모형

종류	구분	부정수급율	표본수
장애연금	(일실이익 = 있음) & (장애등급 < 4등급) & (가입기간 < 100.5개월)	35.9%	198
	(일실이익 = 있음) & (장애등급 = 4등급) & (산재등타급여가입 = 있음)	23.3%	30
	(일실이익 = 있음) & (장애등급 < 4등급) & (가입기간 ≥ 100.5개월)	14.2%	134
	(일실이익 = 없음) & (산재등타급여가입 = 없음)	8.9%	135
	(일실이익 = 있음) & (장애등급 = 4등급) & (산재등타급여가입=없음)	5.2%	172
	(일실이익 = 없음) & (산재등타급여가입 = 있음)	0.4%	1851
유족연금	(일실이익 ≤ 천만원) & (산재등타급여가입 = 있음)	23.4%	501
	(일실이익 = 있음) & (산재등타급여가입 = 없음) & (손해배상금수령 = 미종결)	21.0%	376
	(일실이익 > 3천만원) & (산재등타급여가입 = 있음)	11.1%	226
	(일실이익 = 있음) & (산재등타급여가입 = 없음) & (손해배상금수령 = 종결)	4.4%	3854
	(일실이익 = 없음)	0.8%	3363

장 높았다. 그리고 유족연금의 경우는 산재 등 타급여에 가입되어 있는 수급자가 일실이익이 3천만원 이하인 것으로 신고한 경우의 부정수급율이 23.4%로 나타났다 (표 6).

5. 결론

본 연구는 국민연금의 부정수급을 방지하기 위해 데이터마이닝을 이용하여 예측모형을 개발하였다. 국민연금 부정수급 유형 중 기관 간 정보 공유로서 해결되지 않는 손해배상금 불성실 신고를 대상으로 분석하였다. 분석결과, 부정수급에 영향을 미치는 요인은 수급자의 연금가입, 급여, 사고내역인 것으로 나타났다. 세부적으로 살펴보면, 장애연금의 경우, 연금가입기간, 산재 등 타급여가입, 일실이익, 장애등급이, 유족연금은 산재 등 타급여가입, 일실이익, 손해배상금수령여부가 정상수급과 부정수급을 구분하는 유의한 변수로 나타났다. 그리고 의사결정나무 모형, 로지스틱 회귀모형, 인공신경망 모형을 이용하여 모형의 예측력을 측정한 결과 의사결정나무 모형이 우수한 것으로 나타나 이를 이용하여 예측모형을 수립하였다. 이 예측모형은 국민연금공단의 연금심사에 활용할 수 있으리라 판단된다. 즉, 수급권자가 연금청구 시 예측모형을 이용하여 리스크를 측정하고 리스크가 높을 경우 심층심사를 통해 부정수급을 사전 예방할 수 있으리라 기대된다.

데이터마이닝 기법은 연기금 등 정책운영기관의 관리방안으로 활용될 뿐 아니라 감사원의 비리감사에도 활용될 수 있다. 외국 감사원의 경우 데이터마이닝을 이용하여 리스크가 높은 감사대상을 우선 선정하여 감사하고 있어 효율적인 감사방법론으로 활용되고 있다. 따라서 우리나라도 감사기법으로 데이터마이닝을 적극 활용하는 방안이 요구되고 있으며 이를 위한 향후연구로서 연기금 등 공공부문의 부정 방지를 위한 예측모형을 개발하고자 한다.

참고 문헌

박능후, 송미영 (2008). 국민기초생활보장제도 수급자격 적격성 연구 -사각지대와 부정수급집단간의 특성을 중심으로-, <사회복지연구> 2008년 겨울호.

오윤섭, 안창보, 이오, 김건수, 김찬수, 원성연, 이혜승 (2009). 사회복지 관련 비리에 대한 주요 선진국의 접근방법, <감사원 감사연구원>.

정우진, 이선미, 김원훈, 신승호, 조우현 (2003). 국민건강보험 지역보험료 체납 결정요인 및 체납확률 예측모형, <보건행정학회지>, **13**, 85–100.

GAO (2003). Results and Challenges for Government Program Audit and Investigations(GAO-03-591T)

Hutton, G. W. (1979). Welfare Fraud and the Police, The Police Chief. Nov. 1979

2009년 11월 접수; 2009년 12월 채택

An Application of Data-Mining Tool in Fraud Pension Payment Prediction

Kyung Yup Cha[1,a]

[a]Audit and Inspection Research Institute

Abstract

This study tested the applicability of a Data mining tool in the analysis of massive National Pension data for the purpose of developing fraud pension payment prediction model. This study is identified significant variables for fraud pension payment through the statistical analysis process and developed prediction models using data mining methodology.

Keywords: National pension, fraud, Data mining.

[1] Research Fellow, Audit and Inspection Research Institute, Board of Audit and Inspection, 7th Hyundai Bldg., Kye-Dong, Jongro-Gu, Seoul 110-793, Korea. E-mail: kycha@korea.kr