

노인 장기요양보험의 등급판정을 위한 의사결정나무 연구

한상태^a, 강현철^a, 최보승^b, 이성건^{1,c}

^a호서대학교 정보통계학과, ^b대구대학교 전산통계학과, ^c성신여자대학교 통계학과

요약

노인장기요양보험은 고령이나 노인성질병 등으로 인하여 혼자서 일상 생활을 수행하기 어려운 노인 등에게 신체활동 또는 가사지원 등의 장기요양급여를 사회적 연대원리에 의해 제공하는 사회보험 제도로서, 우리나라에서는 2008년 7월부터 시행하고 있다. 신뢰성있는 등급판정을 위하여, 안정적인 수급자의 요양인정점수를 산출하는 것은 노인장기요양보험제도의 시행에 있어 매우 중요한 요소라고 할 수 있다. 본 연구에서는 요양인정점수 산출과 등급판정에 의사결정나무 모형을 사용하였고 안정적인 모형을 위해 원자료의 변환 및 절사, 다양한 분리기준(splitting criterion), 정지규칙(stopping rule)을 적용하였다. 본 연구에서 생성한 모형이 기존의 모형보다 안정적임을 확인하였다.

주요용어: 의사결정나무, 등급판정도구, 노인장기요양보험.

1. 서론

간병수발이 필요한 고령자가 인구고령화의 심화에 따라 늘어나고 있으며, 이들을 가족적 차원에서 보호해줄 수 있는 환경은 열악해지고 있어 간병수발의 문제가 사회적 리스크(social risk)화 되어가고 있다. 이에 대비하고자 정부는 간병수발의 사회적 리스크화될 것을 고려하여 그동안 장기요양보호정책을 준비해 왔다. 2007년 4월에는 정부가 그동안 추진해 온 노인 장기요양 보험 법안이 국회에서 통과되어 2008년 7월부터는 본격적으로 노인 장기요양 보험제도가 시행되고 있다. 본 시행에 앞서 정부는 시범사업을 2회 실시하였다(1차 시범사업: 2005년 7월 ~ 2006년 4월, 2차 시범사업: 2006년 4월 ~ 2007년 4월). 1차 시범사업평가는 제도시안으로 제시한 각종 도구의 현실적합성과 예상되는 문제점을 표면화하는 과정이었으며, 2차 시범사업평가는 1차 시범사업에서 나타난 문제점을 수정하여 제도시안을 보다 정교하게 만든 과정이라고 할 수 있다. 또한 1차 시범사업의 6개 시군구지역에서 2차 시범사업은 8개 시군구지역으로 확장됨과 동시에 대상자범위도 65세 이상의 모든 노인으로 확대 실시되었다(보건복지가족부, 2007).

요양인정점수의 산출 및 등급판정 도구는 노인 장기요양 보험제도의 시행에 있어 매우 중요한 요소이다. 1차와 2차 시범사업평가를 통해 개발 및 보완된 요양인정점수 산출 도구는 의사결정나무 분석(decision tree analysis)의 결과를 이용하여 개발된 것이다. 의사결정나무분석의 결과는 분리기준(splitting criterion), 정지규칙(stopping rule), 타당성 검증(validation)을 통한 모형의 수정 등 여러 가지 사항에 따라 달라질 수 있고 또한 설명변수 및 목표변수에 대한 변환과 수정에 의해서도 의사결정나무분석의 결과가 달라질 수 있는 것이다(Breiman 등, 1984). 본 연구의 목적은 그동안의 분석과정 및 결과를 전반적으로 검토하고 추가적인 분석을 통해 요양인정점수의 산출 도구를 개선 및 보완하려는 것이다. 일차적으로는 의사결정나무분석을 통해 다양한 분리기준, 정지규칙, 타당성 검증 방법 등을

¹ 교신저자: (136-742) 서울시 성북구 동선동 3가 249-1, 성신여자대학교 통계학과, 조교수.
E-mail: sklee@sungshin.ac.kr

표 1: 종속변수 8개 (8개 서비스군 분류표)

서비스군	서비스 내용
(1) 청결요양시간	세수/위생, 착의탈의, 목욕
(2) 배설요양시간	배설
(3) 식사요양시간	식사
(4) 기능증진보조시간	체위변경시간, 환자이동, 훈련보조
(5) 치매대응요양시간	문제행동대처
(6) 간접지원요양시간	의사소통 및 기타신체수발(청소, 일지작성 등)
(7) 간호처치시간	측정 및 검사시간, 순회시간, 약주사시간, 치료처치, 기타간호처치
(8) 재활훈련시간	평가 및 기능동작, 일상동작, 물리치료, 기타기능훈련

표 2: 독립변수 57개(52개 기능평가항목 + 5개 영역별 총점)

영역	항목
신체기능영역(12항목)	(1) 옷 벗고 입기, (2) 세수하기, (3) 양치질하기, (4) 목욕하기, (5) 식사하기, (6) 체위변경하기, (7) 일어나 앉기, (8) 옮겨 타기(앉기), (9) 방밖으로 나오기, (10) 화장실 사용하기, (11) 대변 조절하기, (12) 소변 조절하기
인지기능영역(7항목)	(1) 며칠 전에 들었던 이야기나 일을 잊는다, (2) 오늘이 몇 월 며칠인지 모른다, (3) 자신이 있는 장소를 알지 못한다, (5) 자신의 나이나 생년월일을 모른다, (6) 지시를 이해하지 못한다, (8) 주어진 상황에 대한 판단력이 떨어져 있다, (9) 의사소통이나 전달에 장애가 있다
문제행동영역(14항목)	(1) 절도상해 위협(망상), (2) 환각이나 환청, (3) 슬픈상태 울기도함, (4) 불규칙수면 주야혼돈, (5) 도움에 저항, (6) 서성거림, 안절부절못함, (7) 길을 잃음, (8) 폭언, 위협행동, (10) 밖으로나가려고함, (11) 물건망가트리기, (13) 의미없거나 부적절한 행동을 자주보임, (18) 돈 물건 감추기, (19) 부적절한 옷입기, (22) 대소변불결행위
재활육구영역(10항목)	(1) 우측상지 마비, (2) 좌측상지 마비, (3) 우측하지 마비, (4) 좌측하지 마비, (5) 어깨관절, (6) 팔꿈치관절, (7) 손목 및 수지관절, (8) 고관절, (9) 무릎관절, (10) 발목관절
간호처치육구(9항목)	(1) 기관지결개관 간호, (2) 흡인, (3) 산소요법, (4) 욕창간호, (5) 경관영양, (6) 통증간호, 암성통증간호, (7) 도뇨관리, (8) 장루간호, (10) 복막투석, 투석간호

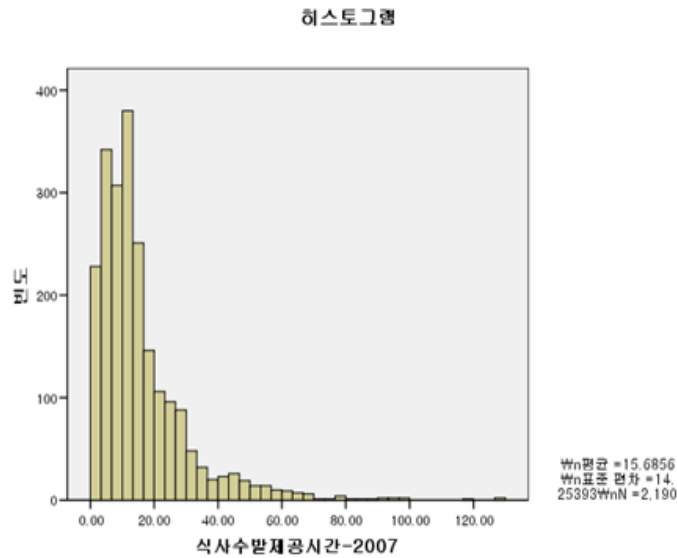
고려하여 통계적 분석을 수행하고, 이러한 과정을 통해 요양인정점수 산출의 정확성, 타당성, 유용성을 높일 수 있는 도구를 개발하고자 한다.

의사결정나무의 대안적인 방법으로 회귀분석 등 다른 형태의 통계기법을 적용하여 새로운 요양인정점수 평가표를 개발하여, 이를 통해 요양인정점수 산출 도구의 정확성 및 안정성을 비교하고 향후 새로운 대안으로의 사용가능성도 검토할 수 있을 것이다. 또한 2차 판정 도구로써, 52개 기능 평가항목에 대해 통계적 분석을 실시하여 항목별 가중치가 고려된 영역별 점수를 산출할 수도 있다. 이는 같은 영역에서 같은 기능 상태 점수를 받은 사람들은 기능상태가 비슷하며 필요한 서비스도 유사하다고 추측할 수 있으므로, 의사결정나무분석으로 산출된 요양 인정시간을 보완할 수 있는 자료로 사용할 수 있다. 그러나 본 연구에서는 의사결정나무 모형만을 고려하기로 한다.

본 연구의 2절에서는 장기요양인정 서비스 자료에 대해서 설명하고, 3절에서는 대상 자료에 대한 기초통계분석을 실시하여 자료의 변환을 검토하고, 4절에서는 의사결정나무를 이용한 등급판정모형을 구축하고 이를 기존 모형과 비교한다. 마지막 5절에서는 결론 및 향후 연구에 대해서 논의한다.

2. 장기요양인정 서비스 자료

의사결정나무, 회귀분석, 대응분석 등에 사용된 장기요양인정서비스 자료는 2007년 5월 이전까지



백분위수

	백분위수									
	5	10	25	50	75	90	95	97.5	98	99
가중평균(정의 1) 식사수발제공시간-2007	1,7049	3,1947	6,4135	11,6996	19,7983	32,2200	44,8235	55,9011	58,6678	67,6198
Tukey의 Hinges 식사수발제공시간-2007			6,4405	11,6996	19,7923					

그림 1: 식사수발 제공시간의 분포 및 백분위수 (99 백분위수이상변환)

의 기존 타임스터디 자료이다. 타임스터디는 각 서비스별로 대상자의 신체특성 및 기능평가 항목에 따라 소요되는 서비스시간을 측정하는 것이다. 자료는 57개의 독립변수와 8개의 종속변수로 구성되어 있으며 분석대상 개체수는 2,200건이다. 데이터의 변수 내용은 표 1과 2에 나타나 있다.

3. 기초통계분석

본 절에서는 의사결정나무분석을 수행하기에 앞서, 설명변수와 목표변수들에 대한 탐색적 기초통계분석 및 가공을 수행하였다. 평균, 표준편차, 왜도 등 기초통계량을 중심으로 제변수들에 대한 분포를 파악하고, 기초통계분석의 결과를 통해 데이터의 변환 및 절사 등 모형의 안정성, 정확도, 효율성을 높이기 위한 데이터 가공방법을 도출하고 특이한 패턴을 보이는 관측개체의 존재를 파악하여 데이터의 안정성을 도모하였다. 또한 기존의 의사결정나무에서 현실 문제와는 다른 역전현상이 발생함으로써 모형의 신뢰도를 떨어뜨리는 결과를 가져왔으므로 이에 대한 이유를 살펴보았다.

3.1. 분포 탐색 및 데이터 변환

본격적인 분석에 들어가기 앞서 분석에 사용할 8개의 종속변수에 대한 분포를 살펴봄으로써 모형의 안정성 및 일반화에 영향을 줄 수 있는 특이값(outlier)을 탐색하고자 한다. 또한 통계적 모형은 자료에 대한 정규분포(normal distribution) 가정에 기반한 추론이 이루어짐으로 통계적 변환법을 적용하여 정규분포로의 변환을 시도하는 등 안정적인 모형이 도출될 수 있도록 하였다. 많은 통계모형 및 통계치들은 특이값에 민감하다. 의사결정나무분석의 경우에도 대상자에 대한 요양인정 점수는 의사결정

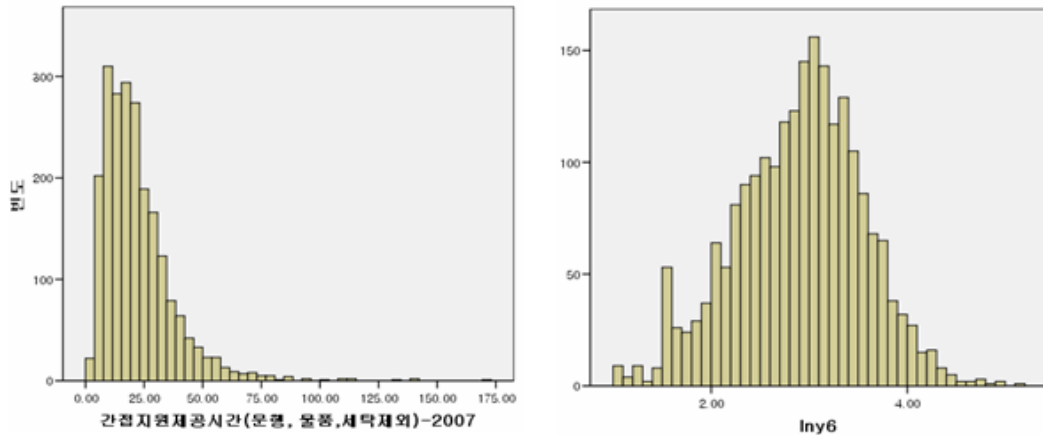


그림 2: 간접지원 제공시간의 로그변환 전과 후의 분포

나무의 잎마디에 속해 있는 개체들의 평균(mean)값으로 한다. 이러한 평균값은 특이치에 민감한 대표적인 통계치이므로 원데이터에서 이러한 특이치를 제거한 후 분석하는 것이 타당할 것이다 (강현철 등, 1999).

예를 들어 그림 1의 종속변수 ‘식사 수발 제공시간’에 대한 분포를 살펴보면 값의 범위는 0~130사이임을 알 수 있다. 그러나 대부분의 값은 70이하에 분포하고 있으며 70이상에는 몇 개의 값밖에 나타나지 않는다. 이러한 값들은 다른 일반적인 패턴과 다르다고 할 수 있으며 이 값들로 인하여 전체 평균이 상향되는 결과를 가져오게 된다. 따라서 이 변수의 99 백분위수인 67.6보다 큰 값들은 특이값으로 간주하여 99 백분위수 값으로 변환하게 되면 그 영향은 현저히 줄어들게 된다. 여기서 특이값으로 간주되는 개체들을 제거하는 방법도 있으나 해당 개체의 정보를 최대한 활용하기 위하여 값을 변환 후 사용하였다. 또한 특이값에 대한 기준으로 98 백분위수, 97 백분위수 등도 고려하였으며 가장 안정적인 의사결정나무를 보이는 값으로 선택하였다. 다른 종속변수에 대해서도 동일한 방법을 적용하였다.

정규분포를 위한 변환으로는 로그 변환을 고려하였다. 그림 2의 ‘간접지원제공시간’에서처럼 원래 분포는 오른쪽으로 긴 꼬리를 가진 모습을 가지고 있었으나 로그변환(log transformation)후에는 대칭의 모습으로 정규분포에 가까운 모습을 보이고 있다. 그러나 원점수로의 환원 단계에서 편의(bias) 발생이 우려되어 본 의사결정나무분석에서는 사용되지 않았으나 향후 고려해볼 수 있을 것이다.

기존의 의사결정나무는 다양한 통계적 절차 없이 소프트웨어(SPSS/Answer Tree)의 디폴트(default) 설정으로 수행된 결과였다. 이에 연구 목적과는 배치되는 규칙이 의사결정나무에 나타나 모형의 안정성과 일반화에 좋지 않은 영향을 주게 되어 이에 대한 개선이 필요하였다. 기존의 의사결정나무에서 역전현상이 발생하는 원인을 기초통계분석을 통하여 탐색하여 보았다.

역전현상이란 몸이 더 불편함에도 서비스 인정시간이 더 짧아지는 현상으로 일반적인 상식과 배치되는 현상이다. 각 독립변수(완전자립, 부분도움, 완전도움)에 대한 종속변수 평균의 역전현상 발생여부를 살펴보았는데 지면상 목록에 대한 간접지원제공시간만을 일례로 살펴보고자 한다.

목록이 완전자립일 경우의 수발시간의 평균은 18.9, 부분도움일 경우는 25.6, 완전도움일 경우는 23.2로 부분도움과 완전도움에서 역전이 발생함을 알 수 있다. 부분도움이 필요한 사람에서 수발시간이 완전도움에 필요한 사람의 수발시간보다 크게 나타난 것이다. 즉, 원 데이터에서 역전이 발생하고 있음으로 인하여 의사결정나무에도 영향을 미쳐 의사결정나무 내에서도 역전현상이 발생하고 있음을 알 수 있었다. 그러나 모든 역전현상이 문제가 되는 것은 아니다. 재활훈련 제공 시간의 경우에 회복

가능성이 있는 환자에 대해서만 재활치료를 하게 되므로 회복가능성이 거의 없는 환자의 경우에는 도리어 서비스 시간이 짧아지게 되는 것이다. 본 연구에서는 모형의 일반화를 위해 대체규칙을 이용하여 가능하면 역전현상이 의사결정나무에 나타나지 않도록 하였으며 문제가 되지 않는 역전현상에 대해서는 대체규칙을 사용하지 않았다. 여기서 대체규칙은 의사결정나무 구축시 사용되는 분리변수가 역전현상을 발생시킬 경우 예측 성능이 거의 동등한 다른 변수로 대체하여 역전이 없는 의사결정나무를 구축할 때 사용된다.

4. 의사결정나무 구축

의사결정나무분석은 의사결정규칙(decision rule)을 도표화하여 관심대상이 되는 집단을 몇 개의 소집단으로 분류(classification)하거나 예측(prediction)을 수행하는 분석방법이다. 분석의 과정 및 결과가 나무구조에 의해서 표현되기 때문에, 분류 또는 예측을 목적으로 하는 다른 방법들(예를 들면, 신경망(neural networks), 판별분석(discriminant analysis), 회귀분석(regression analysis) 등)에 비해 연구자가 분석과정을 쉽게 이해하고 설명할 수 있다는 장점을 가지고 있다.

4.1. 서비스 군별 의사결정나무분석 결과

본 연구에서의 의사결정나무는 기존의 의사결정나무에서 설명할 수 없는 역전현상 제거를 통한 모형의 안정화와 훈련용 자료와 검증용 자료를 이용한 모형의 일반화에 중점을 두어 개발되었다. 기존의 의사결정나무는 전체 데이터를 모두 의사결정나무 생성에 사용하여 모형의 타당성 검증은 할 수가 없었던 점을 개선한 것이다. 특이치가 제거된 목표변수에 대한 서비스군별 의사결정나무분석결과를 살펴본 아래의 결과는 통계적인 모형의 안정화를 위하여 여러 번의 의사결정나무 구축을 시도한 것 중 가장 좋은 모형을 선택한 것이다. 모형구축을 위한 설정은 아래와 같으며 종속변수의 하나인 청결서비스에 대한 의사결정나무 결과는 그림 3과 같다.

▶ 모형 구축을 위한 설정

- 목표변수: 청결수발 서비스 시간 (특이치 변환)
- 설명변수: 52개 기능평가 항목 및 5개 영역별 100점 득점
- 분리 방법: CART
- 불순도의 최소 변화: 0.0001
- 모형평가 방법: 훈련데이터와 검증데이터를 분리하여 모형 구축 및 평가 실시
- 최대 분할 수준(maximum tree depth): 6
- 분할될 부모마디(parent node)의 최소 크기: 100
- 분할된 자식마디(child node)의 최소 크기: 40
- 사용자 분리 및 가지치기: 설명할 수 없는 역전현상과 일반화에 문제가 있는 분리를 유사한 경쟁 규칙(competing rule)을 이용하여 분리함. 이 과정에서 장기요양보험 담당자와 협의 하여 설명 가능한 규칙인지 아닌지를 확인 후 의사결정나무를 구축하였음.

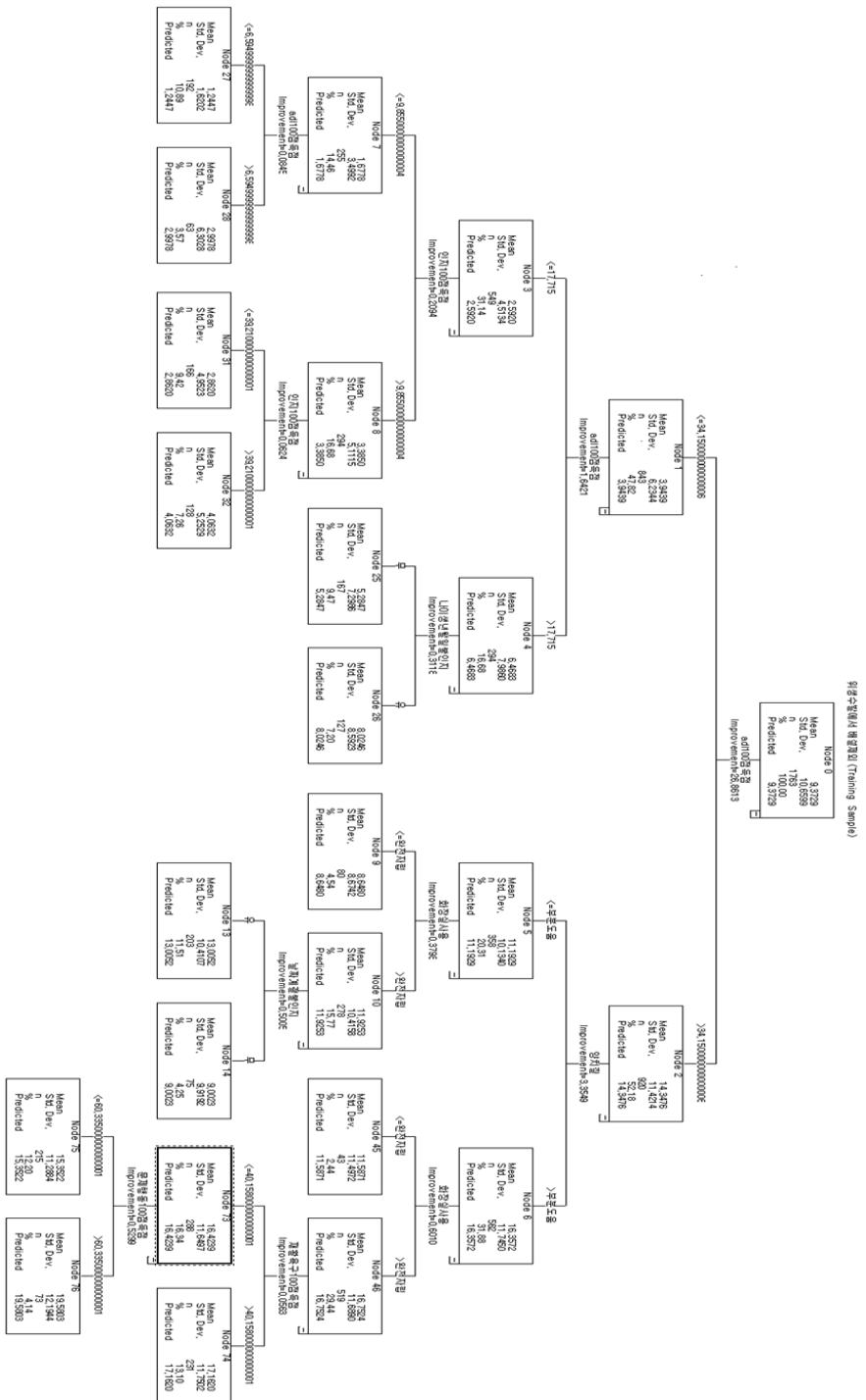


그림 3: 청결서비스군의 의사결정나무

표 3: 기존인정시간과 개선된 인정시간 평균

	청결	배설	식사	기능보조	문제행동	간접지원	간호처치	재활훈련
서비스시간	9.55	5.77	15.69	7.45	1.02	22.22	9.79	4.34
개선인정시간	9.37	5.59	15.26	7.17	0.76	12.97	9.31	4.26
기존인정시간	9.15	5.52	15.18	7.44	0.96	22.12	9.56	4.36

표 4: 기존인정시간과 개선된 인정시간의 MAE

	청결	배설	식사	기능보조	문제행동	간호처치	재활훈련
개선인정시간	6.62	4.13	6.75	5.18	1.22	5.98	4.11
기존인정시간	6.37	3.92	6.47	5.21	1.32	6.09	4.14

표 5: 기존인정시간과 개선된 인정시간의 RMSE

	청결	배설	식사	기능보조	문제행동	간호처치	재활훈련
개선인정시간	10.07	7.06	10.44	8.56	3.83	12.35	6.23
기존인정시간	9.72	6.61	9.76	8.55	3.86	12.01	6.23

4.2. 의사결정나무 타당성 검증

표 3은 모형 구축에 사용된 2,200건을 이용한 분석 결과 가운데에서 8개 서비스군에 따른 원 서비스 시간과 기존의 의사결정나무분석과 개선된 의사결정나무 분석에 의하여 산정된 인정 시간의 평균을 정리한 표이다. 각 서비스 영역에서 모형구축 과정에서 반응변수로 사용된 서비스시간과 의사결정나무분석 모형에 의하여 산출된 인정시간 간에는 큰 차이를 보이지 않는다. 다만 간접지원 시간의 경우 개선된 인정시간에서 기존인정시간과 서비스 시간 사이에 큰 차이를 보이고 있다. 이는 안정적이고 효율적인 최종 인정 등급을 산출하기 위하여 간접지원에 대하여 변경된 서비스 시간을 모형구축에서의 반응변수로 이용하였고, 개선된 인정시간은 이 새로운 간접지원 시간을 기준으로하여 모형을 구축하였으며 이에 따라 생성된 인정시간은 기존의 인정시간과 서비스 시간과 일정 정도 차이를 보이게 된다. 따라서 이후 모형의 평가에서는 간접지원 경우 정확한 평가치를 계산할 수 없으므로 생략하기로 한다.

모형의 평가 방법으로는 다양한 통계량이 사용될 수 있다. 본 연구에서는 가장 보편적으로 사용되는 평균절대오차(mean absolute error; MAE)와 평균제곱오차제곱근(root mean squared error; RMSE)를 모형 평가를 위한 통계량으로 사용하였다. 각 통계량은 다음과 같다.

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n},$$

$$RMSE = \frac{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}}{n - k}$$

여기서 y_i 는 서비스 시간, \hat{y}_i 는 의사결정나무분석에 의해 산출된 인정시간, n 은 전체 개체수, k 는 모형 구축에 사용된 전체 마디의 수를 의미한다.

8개 서비스 영역에 따라 각 통계량이 계산되었으며 그 값이 작을 수록 모형의 설명력과 모형의 적합도가 더 높다고 할 수 있다. 표 4와 5는 간접지원을 제외한 나머지 7개 서비스 영역에서 산출된 MAE와 RMSE를 정리한 표이다. 전반적인 수치를 비교하면 MAE에서는 개선된 인정시간의 적합도와 기존 인정시간의 적합도에 큰 차이를 보이고 있지 않으며 RMSE에서도 큰 차이를 비고 있지 않다. 그러나 모형 구축에서 언급한 바와 같이 개선된 의사결정나무분석에서는 역전현상의 제거와 함께 보다 작은 수의 가지와 노드에 의하여 모형이 구축되었다. 보다 작은 수의 가지와 노드를 이용하여 모형이 구축되었기 때문에 구축된 모형의 확장에서 보다 안정적인 결과를 줄 수 있으며 이와 더불어 모형이 보

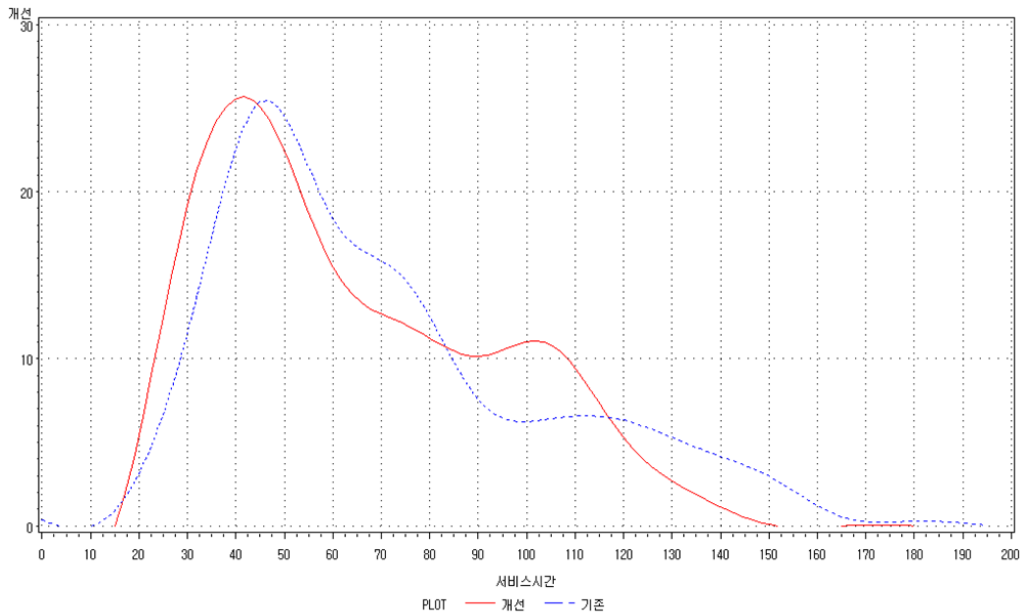


그림 4: 기존 인정시간과 개선된 인정시간의 분포

다 간소화 되었다 볼 수 있다. 이는 모형 구축에서의 간결성에서도 보다 바람직한 방향이라 할 수 있으며 일반적으로 모형이 간소화 되면 모형의 적합도는 일정 정도 떨어지는 것이 사실이다. 이를 감안 하였을 때 본 연구에서 개선한 모형이 보다 간소화 함에도 불구하고 기존의 의사결정나무분석에 비하여 MAE와 RMSE에서 큰 차이를 보이고 있지 않으므로 본 연구의 개선모형의 적합도가 보다 개선되었다고 할 수 있다.

모형의 타당성 및 안정성을 평가하는 또 다른 방법은 모형 구축과정에서 사용되지 않은 데이터를 이용하여 모형 적합을 수행하여 보는 것이다. 즉 기존의 의사결정나무방법과 개선된 의사결정나무방법에서 생성된 인정점수의 산출 규칙을 모형 구축에 이용되지 않은 새로운 데이터에 적용하여 이로부터 생성된 인정점수를 이용하여 평가를 수행하는 것이다. 실제적으로 모형의 안정성 및 예측능력을 평가하는데 있어서는 이와 같이 모형 구축에 사용되지 않은 별도의 검증용 자료를 이용하여 평가를 수행하는 것이 보다 타당하다 할 수 있다. 특히 모형의 예측능력 및 안정성을 평가하는데 있어서는 모형의 구축에 사용된 데이터를 가지고 다시 모형 평가에 사용하는 것은 실제 모형이 가지고 있는 예측능력을 과대평가하는 오류를 범할 수 있기 때문이다.

그림 4는 기존 인정시간과 개선된 인정시간간의 분포를 비교한 그림이다. 그림에서 실선으로 표시된 것은 개선된 인정시간의 분포를 나타내고 점선으로 표시된 부분인 기존의 인정시간의 분포를 나타낸다. 기존의 인정시간에서는 시간의 폭이 개선된 인정시간보다 상당히 큰 것을 볼 수 있다. 개선된 인정시간은 30분을 중심으로 하여 비교적 안정적으로 분포하고 있는 것을 볼 수 있으나 기존의 인정시간에서는 40분을 중심으로 하여 0분부터 200분 까지 큰 폭으로 분포를 하고 있다. 개선된 인정시간이 보다 안정적으로 분포를 하고 있다고 할 수 있다.

5. 결론

본 연구는 그동안의 노인장기요양보험제도의 등급판정도구의 분석과정 및 결과를 전반적으로 검

또하고 추가적인 분석을 통해 요양인정점수의 산출 도구를 개선 및 보완하려는 것이다. 일차적으로는 의사결정나무분석에 있어 보다 다양한 분리기준, 정지규칙, 타당성 검증 방법들을 고려하여 분석을 수행하였고, 이러한 과정을 통해 요양인정점수 산출의 정확성, 타당성, 유용성을 높일 수 있는 도구를 개발하였다. 본 연구를 통하여 기존 의사결정나무분석 모형의 안정성 및 정확성을 통계적인 분석 과정을 보다 심도있게 함으로써 기존 노인요양인정점수 산출 도구의 신뢰성을 제고할 수 있었다. 또한 의사결정나무분석 모형의 경험적 및 이론적 타당성을 확보하여 노인장기요양보험제도 시행에 있어 노인요양인정점수 산출 도구의 적절성을 제고할 수 있었다. 어떤 분야에서든지 예측모형은 지속적으로 보완, 개선되어야 하므로 본 연구의 결과를 활용하여 정확성, 안정성, 타당성을 지속적으로 제고할 수 있는 정책적 시사점을 도출할 수 있을 것이다. 충분히 의미있는 결과들을 향후 지속적으로 유지하기 위해서는 다음의 두 가지 사항이 고려되어야 할 것이다. 첫째, 데이터의 시의성이다. 본 연구에서 개발된 모형은 2007년 중반까지의 데이터에 기반한 것이므로 미래의 데이터에는 편이가 발생할 수 있으므로 지속적인 조사 데이터 갱신을 통해 개발된 모형을 수정 보완해 나가야 할 것이다. 둘째, 데이터의 정확성이다. 분석 데이터가 오염되어 있다면 분석 결과도 오염되어 있을 수 밖에 없다. 따라서 데이터 획득 단계에서 신뢰성 있는 측정을 위한 도구 및 제도가 갖추어져야 할 것이다.

참고 문헌

- 강현철, 한상태, 최종후, 김차용, 김은석, 김미경 (1999). <SAS Enterprise Miner를 이용한 데이터 마이닝-방법론 및 활용>, 자유아카데미, 서울.
- 보건복지가족부 (2007). <노인장기요양제도 등급판정도구 요양인정점수 산출 통계기법 연구 최종보고서>.
- Breiman, L., Friedman J., Olshen, R. and Stone, C. (1984). *Classification and Regression Trees*, Pacific Grove.

A Study on the Judgement Rating for Level of Need for Long-term Care Insurance Using a Decision Tree

SangTae Han^a, HyunCheol Kang^a, BoSeung Choi^b, SeongKeon Lee^{1,c}

^aDepartment of Informational Statistics, Hoseo University

^bDepartment of Computer Science and Statistics, Daegu University

^cDepartment of Statistics, Sungshin Women's University

Abstract

Long-term care insurance is a social insurance system that provides benefits to the elderly who have difficulty taking care of themselves for a period of at least 6 months. This system was started in July, 2008 and it is very important to set proper judgement ratings for the approval process. We try to develop and improve the judgement rating system using decision tree models. Our tree model is found to be more stable and efficient than the previous one.

Keywords: Decision tree, long-term care insurance, judgement rating.

¹ Corresponding author: Assistant professor, Department of Statistics, Sungshin Women's University, Seoul 136-742, Korea. E-mail: sklee@sungshin.ac.kr