

분류회귀나무를 이용한 의료서비스 적기처리 예측모형

A Prediction Model of Timely Processing on Medical Service using Classification and Regression Tree

이 종 찬*, 정 승 우**, 이 원 영**

Jong-Chan Lee*, Seung-Woo Jeong**, Won-Young Lee**

Abstract

Turnaround time (called, TAT) for imaging test, which is necessary for making a medical diagnosis, is directly related to the patient's waiting time and it is one of the important performance criteria for medical services. In this paper, we measured the TAT from major imaging tests to see it met the reference point set by the medical institutions. Prediction results from the algorithm of classification regression tree (called, CART) showed "clinics", "diagnosis", "modality", "test month" were identified as main factors for timely processing. This study had a contribution in providing means of prevention of the delay on medical services in advance.

요 약

의학적 진단을 내리기 위해 시행되는 검사의 소요시간(turnaround time, TAT)은 환자대기시간과 직결되며 중요한 의료서비스 평가항목 중 하나이다. 본 연구에서는 주요 영상의학검사를 대상으로 TAT를 측정하고, 그 결과가 의료기관이 설정한 기준치를 달성하는지 여부를 분석하였다. 분류회귀나무 알고리즘을 이용한 예측 결과, "진료과", "상병", "검사종류", "실시월"이 적기처리 달성에 가장 큰 영향을 주는 요인으로 확인되었다. 본 연구는 의료서비스의 적기처리를 예측하는 모형을 통하여 의료서비스 지연을 사전에 조치할 수 있는 수단을 제공하였다는 데에 큰 의미가 있다.

Key words : quality of medical service, turnaround time, CART, imaging test, timely processing

1. 서론

의료서비스의 품질 향상은 다양한 측정 지표의 개발을 통하여 실현되고 있다. 그림 1에 나타내는 바와 같이 국내 의료기관평가인증원에서도 '기본가치체계'(Basic Value System), '환자진료체계'(Patient Care System), '행정관리체계'(Administrative Management System), '성과

관리체계'(Performance Management System)로 나누어진 4개의 평가영역을 통하여 환자의 안전 보장과 지속적인 의료 질 향상을 도모하고 있다 [1]. 이처럼 의료서비스에 대한 다양한 관점과 객관적인 평가를 통하여 점차 의료서비스에 대한 정량적 품질 측정이 가능해지고 있다. 그러나 궁극적인 평가의 목적이 서비스 질을 개선하기 위함이라면 평가항목을 점수화하는 것에만 그치

* Dept. of Industrial and Information Systems, Seoul National University of Science and Technology.

** The Catholic University of Korea. Uijeongbu St.Mary Hospital.

★ Dept. of Industrial and Information Systems Engineering, Seoul National University of Science and Technology (wonylee@seoultech.ac.kr, +82-2-970-6478)

※ Acknowledgment : This study was supported by the Research Program funded by the Seoul National University of Science and Technology.

Manuscript received Dec 10, 2015; revised Jan 26, 2016; accepted Jan 27, 2016

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.



Fig. 1. criteria for certification of medical services
그림 1. 의료기관 인증평가 기준항목

지 않고 각각의 평가 결과에 대한 원인을 분석하는 것이 중요하다. 특히, 의료행위는 고수준의 지식 통합적 노동을 요구하는 작업으로 단순 반복적인 작업 유형과 다른 특성을 보인다. 예를 들어 의학적 진단을 내리기 위한 검사결과 보고서는 최대한 빨리 작성되는 것이 환자에게 유리하다. 하지만, 검사결과의 정확성과 신속성 사이에서 트레이드오프(trade off)가 발생하기 때문에 검사결과 보고서의 제출시한을 검사 종류별로 지정하는 방식은 합리적인 문제해결이라고 볼 수 없다. 이러한 이유로 서비스 개선을 위해 측정지표에 영향을 주는 요인에 대한 분석을 실시하는 것은 지극히 자연스러운 사고 과정이다. 예를 들면, 작업시간의 정량화 척도인 반환시간(TAT)을 이용한 방사선 검사 시스템의 작업흐름 변경을 통한 보고시간 개선 연구[2]와 방사선 전문의의 성과지급을 위한 공정가중치에 대한 연구[3]에서는 다양한 입력 값에 따른 측정지표의 개선효과를 통계적 분석을 통해 설명하였다. 더 나아가, 병리검사의 적시 결과보고를 위한 공정 단계별 행동에 관한 연구[4]는 측정지표에 영향을 미치는 요인이 매우 복잡할 수 있음을 보여주었다. 따라서 지표 값의 변동을 설명 할 수 있는 잠재적 영향요인을 어떻게 도출 할 것인가에 대한 기

본적인 물음이 있어야 한다. 한편, 데이터마이닝(data mining)을 이용한 품질지표 도출에 대한 연구[5]는 이전 연구가 가지고 있는 탐색범위의 협소함을 극복하는 한 가지 방안을 제시하였고, 본 연구에도 다량의 데이터 속에 숨겨진 인과관계를 설명하기 위해 되도록 인간이 이해하기 쉬운 알고리즘과 데이터 마이닝 기법을 이용한 예측모형을 구현하고자 하였다.

임의의 프로세스가 개시되고 종료되기까지의 시간간격을 반환시간(turnaround time, TAT)으로 정의하였다. 예측 모형에서 분류회귀나무 알고리즘을 이용하여 TAT가 정상 범위 값을 갖는 데이터와 그 반대의 데이터를 학습하도록 하였다. 학습된 알고리즘은 일종의 TAT 분류기(TAT classifier)로서, 서비스 개선을 목적으로 다양하게 응용될 수 있으며, 특히 전문가시스템(expert system) 또는 의사결정지원시스템(decision support system, DSS) 분야에서 그 활용 가치가 높다[6],[7].

II. 관련연구

본 절에서는 결정트리 분석 기법에 대해 알아보고, 다양한 결정트리 알고리즘 중에서 CART 알고리즘에 대해 소개하였다.

2.1 결정트리 (Decision Tree) 분석

결정트리 분석은 다변량 분석의 도구로 다양한 분야에서 폭넓게 활용되고 있다. 특히, 알고리즘의 명칭에서 확인되는 ‘나무’의 구조적 특징 때문에 어떠한 규칙을 원인과 결과로 해석하는데 큰 장점을 갖는다. 결정트리는 일반적으로 순환적인 분할(recursive partitioning) 작업을 수행한다. 루트 노드(root node)에서부터 가장 순도 높은 데이터 그룹이 찾아질 때까지 나무의 가지분기가 반복된다. 이때 가지분기의 원리를 표 1에서 나타내는 바와 같이 [분할기준]과 [가지분기전략]으로 설명할 수 있다[8],[9]. 가지분기전략이 다양하게 존재하지만 공통의 목적은 최초 주어진 데이터에

원리의 C4.5가 대안으로 제시될 수 있다. CART의 경우, 비교적 대규모 데이터에서도 뛰어난 성능을 발휘하며 별도의 매개변수 없이도 분류 작업을 수행하는 반면, C4.5는 데이터 크기에 따라 트리의 크기가 민감하게 변화하기 때문에 비교적 소규모 데이터에 적합한 것으로 알려져 있다[11]. C4.5는 CART와 다르게 다지분리(multiway split)가 가능하며 범주형(categorical) 입력변수에 대해서는 범주의 수만큼 분리가 일어나는 특징이 있다. CHAID는 알고리즘의 복잡한 설정에 따라 최적 트리의 크기가 크게 영향을 받기 때문에 알고리즘 선정 시 예비 후보에서 제외하였다.

Table 1. major characteristics of decision trees
표 1. 결정트리별 주요 특성

algorithm	split criteria	data type	missing value	split strategy	outlier
ID3	information entropy	categorical	n/a	n/a	sensitive
C4.5	gain ratio	categorical/continuous	possible	error rate	sensitive
CART	gini index(node impurity)	categorical/continuous	possible	cost complexity	possible
CHAID	chi square-test	categorical/continuous	n/a	Bonferroni testing	n/a

대한 최적의 분할에 있다. 트리의 분할은 트리의 크기에 영향을 미친다. 트리의 크기가 클수록 트리의 복잡도가 증가하게 되는데 결정트리 알고리즘에서 주어진 데이터를 설명하기 위해 최적 트리의 크기를 결정하는 작업은 매우 중요하다. 따라서 분석에 앞서 다양한 결정트리 알고리즘에 대한 고유한 특성에 대해 이해한 후 데이터 분석을 수행하는 것이 중요하다.

2.2 분류회귀나무 (Classification and Regression Tree, CART)

본 연구에서는 데이터의 형태와 연구의 특성에 가장 부합하는 CART 모형을 선택하였다. CART 모형은 비모수적 검증을 사용하며 관측변수의 분포에 대한 가정이 필요하지 않다. CART의 장점 중 하나는 반복적인 샘플링 시 발생하는 결측치에 대한 처리가 우수하다는 점이다. 그리고 다양한 형태의 데이터에 대한 처리가 가능하다. 더욱이 비교적 대규모 데이터(over 1000 observations)에서도 CART는 매우 잘 작동한다[10]. 알고리즘 선택 과정에서 CART와 유사한

III. 의료서비스 적기처리 예측 모형

본 절에서는 예측 모형에 사용된 CART 알고리즘의 작동원리에 대해 알아보았다. 이어서, 기계학습에 사용되는 변수 선택과 본 연구의 절차에 대해 논의하였다.

3.1 CART 분류 모형

본 연구에서 제시하는 의료서비스 적기처리 예측 모형의 설계를 위해 CART 알고리즘의 작동 방식을 이해 할 필요가 있다. 먼저 그림 2와 같은 데이터 분포를 가정할 수 있다. 사각형 내부에 흩어져 있는 원이 본 연구에서 관심을 가지고 있는 서비스의 적기처리를 나타내는 종속변수이다. 내부가 하얀 원은 서비스가 적기처리됨을 의미하며 내부가 검은 원은 서비스가 적기처리에 실패함을 의미한다. 최초 Y축의 변수 $V1[y2]$ 을 기준으로 분할을 하게 되면 서비스의 적기처리 건($<y2$)과 적기처리 실패 건($>=y2$)의 정보획득

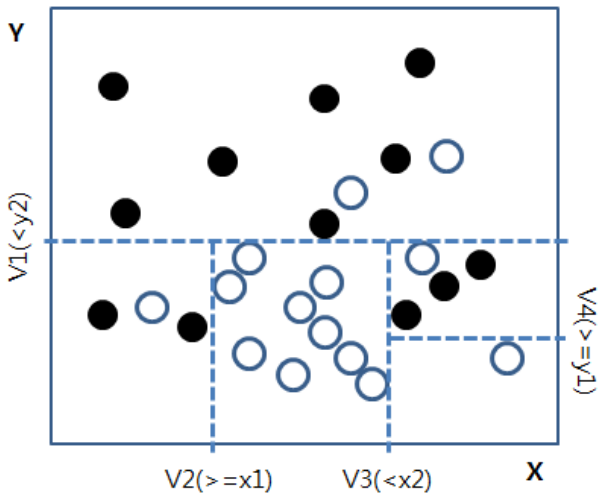


Fig. 2. classification principles of CART
그림 2. CART 분류 원리

량(information gain)이 가장 많은 것을 알 수 있다. 이어 X축의 변수 V2[x1]를 기준으로 분할하고 계속 V3[x2], V4[y1]의 순서로 영역을 분할해 나아갈 수 있다. 관측변수“V1~V4”를 사용하여 서비스 적기처리 결과를 표시하는 원을 서로 다른 영역으로 분할하는 과정은 더 이상 같은 영역 내에 서로 다른 부류의 원이 발견되지 않을 때까지 작업을 지속할 수 있다.

이러한 개념을 이진트리(binary tree)로 묘사한 결과는 그림 3에 나타나게 되는데 바로 이 구조

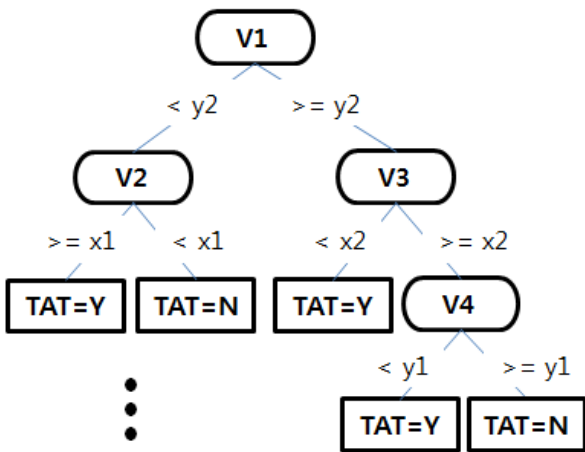


Fig. 3. classification model structure of CART
그림 3. CART 분류 모형 구조

가 본 연구의 CART 분류모형이 된다. 전체 트리 구조에서 원은 분할이 가능한 노드를 의미하고 사각형은 분할이 일어나지 않는 말단 노드를 의미한다. CART 모형은 비선형적인 입력과 출

력 간 연관관계를 해석하기가 매우 용이하다는 장점이 있다.

3.2 독립변수 선택

본 연구에서 사용된 관측변수는 표 2와 같다.

Table 2. data layout
표 2. 데이터 레이아웃

variable	remarks	type
TAT	report date - test date	number
test-mon	test month	categorical
holyday	number of non-business day(digit)	number
pat-type	patient visit(out/in/er)	categorical
severe	serious case(Yes=1/No=0)	categorical
age	patient age(digit)	number
sex	patient sex(male=M/female=F)	categorical
nat-ins	national insurance(Yes=1/No=0)	categorical
re-visit	continuous case(Yes=1,No=0)	categorical
special	selective treatment(Yes=1,No=0)	categorical
clinic	medical department	categorical
inhos	is inhospital?(Yes=1,No=0)	categorical
diagnosis	diagnosis(ICD10 standard)	categorical
test-day	test-week(MON,TUE,WED...)	categorical
modality	CTscan/MRIscan/Sonography	categorical

관측변수의 종류는 크게 독립변수와 종속변수를 포함하고 있다. 서비스 적기처리 예측을 위해 사용된 변수[실시월~모달리티]를 독립변수로 정하였고, 독립변수의 영향으로 나타난 서비스 적기처리 결과[TAT]를 종속변수로 정하였다. 독립변수의 데이터 타입(data type)은 숫자형과 범주형을 포함하고 있으며, 종속변수의 데이터 타입은 “달성[Y]” 또는 “미달성[N]”을 나타내는 범주형 값이다. “TAT”는 (검사결과보고일자) - (검사실시일자)에 의해 구해진 값으로 숫자형 데이터이다. “exec-mon”은 시간적 요인에 의해 서비스 적기처리에 영향을 받는지 확인하기 위한 변수이다. “holyday”는 “TAT” 내부에 존재하는 비근로일(non-business day)이다. 휴일일수가 서비스 적기처리에 미치는 영향을 확인하기 위한 변수이다. “pat-type”은 환자의 진료형태로 주로 외래, 입원, 응급으로 나뉜다. “severe”은 국가가 관리하는 중증 등록 관리 여부를 나타내므로 환자의 중증 여부가 서비스 적기처리에 영향을 미치는지 확인하기 위한 변수이다. “nat-ins”은 국가의료보험 여부를 나타낸다. 이 변수는 국가의료보험 외

비보험, 산업재해, 자동차보험 환자와 다른 영향을 미치는지 확인하기 위한 변수이다. "revisit"은 진료과 또는 질병에 대한 재방문을 의미하며 초진과 재진이 서비스 적기처리에 영향을 미치는지 확인하기 위한 변수이다. "special"은 숙련된 의사에게 받는 서비스를 의미한다. 진료의사의 숙련도에 따라 서비스 적기처리에 영향을 미치는지 확인하기 위한 변수이다. "clinic"은 병원 진료 시 흔히 볼 수 있는 진료과목에 해당한다. "in hos"은 환자의 병실 입원 상태를 의미한다. "diagnosis"은 의사의 진단을 나타내며, "test-week"은 검사가 실시된 날짜에 해당하는 요일을 의미한다. "modality"는 검사의 장비 구분으로 해당 정보로 어떤 종류의 영상의학검사를 실시하였는지 구분할 수 있다. 그 밖에 "age", "sex" 변수가 선택되었다.

3.3 예측절차

본 연구의 진행 방식은 그림 4와 같다. 병원의 영상의학검사 데이터베이스(DBMS)로부터 1차적으로 원시데이터를 추출한다. 이 원시 데이터(source files)는 본 연구에서 사용한 데이터 분석 도구인 R(R i386 3.1.2)에 의해 직접 처리될 수

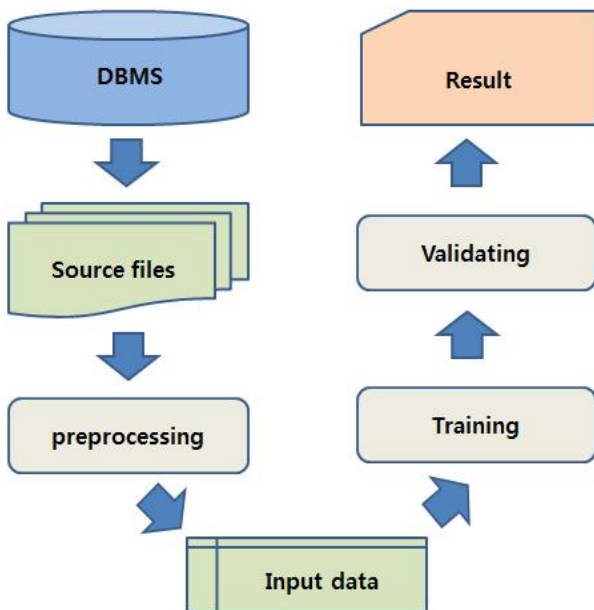


Fig. 4. procedural flow of study
그림 4. 연구절차

있는 형태가 아니므로, 데이터 분석 도구 R에서 데이터를 적재하기 위한 데이터 전처리

(preprocessing) 작업이 요구된다. 이렇게 전처리 작업이 끝난 결과물(input data)은 분석 알고리즘이 구현되어 있는 데이터 분석 도구에 적재 후 바로 분석이 가능하다. "Training" 단계에서는 기계학습(machine learning)을 수행하게 되는데, 이 과정에서 학습과 검증을 통하여 알고리즘의 성능을 최적화하게 된다. 학습의 주된 내용은 균질의 구성원을 갖는 말단 노드를 만들 때까지 나뉘어지를 계속 분기(recursive partitioning)하는 것이다. 이 때 CART는 나뉘어지를 분기하는 전략으로 지니 계수(Gini index)를 사용하여 현재 노드의 불순도(impurity)를 측정한다.

획일적인 학습 데이터는 아직 확인되지 않은 데이터에 대한 과대적합을 초래할 수 있다. 이를 예방하기 위해 학습 데이터와 검증 데이터 간 교차 검증(cross validation)이 될 수 있도록 분석 데이터를 분할하여 사용하는 것이 바람직하다. 본 연구에서는 그림 5에서 나타나는 바와 같이 전체 데이터를 2개의 그룹으로 균등분할 후 각각의 그룹을 7:3의 비율로 나누어 각각 학습 데이터와 검증 데이터로 사용하는 소단위 검증 단계를 진행 후, 다시 전체 규모의 데이터를 7:3의 비율로 나누어 각각 학습 데이터와 검증 데이터로 사용하는 대단위 검증 단계를 거침으로써 교차검증을 수행하였다.

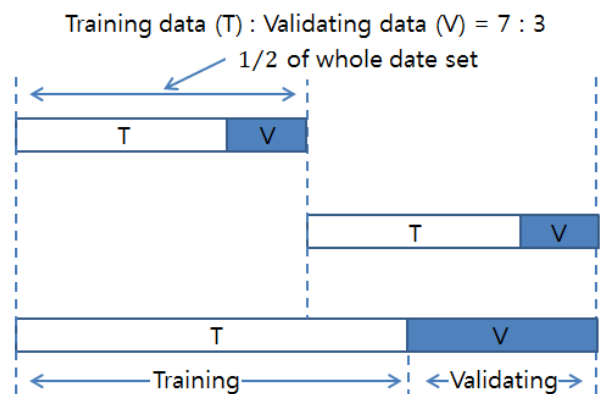


Fig. 5. cross validation for preventing overfitting
그림 5. 과대적합 방지를 위한 교차검증

IV. 사례 분석 : C의료원

본 절에서는 예측 모형의 적용 사례를 제시하였다. 먼저 수집된 데이터를 소개하고 이어서

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17
1	2	JAN	2	1	0	0	0	11	F	1	1	1	OS	0	D	SAT	CT
2	1	JAN	0	1	0	0	0	70	F	1	0	1	NS	0	I	MON	CT
3	8	JAN	8	1	0	0	0	20	M	1	1	0	MC	0	Z	TUE	CT
4	2	APR	2	1	0	0	0	38	M	1	1	1	MG	0	K	SAT	CT
5	0	JAN	0	1	0	0	0	22	F	1	0	1	NS	0	G	TUE	CT
6	0	JAN	0	1	0	0	0	74	M	0	0	0	MC	0	R	TUE	CT
7	4	JAN	2	1	0	0	0	41	M	1	0	1	NS	0	G	THU	CT
8	3	JAN	2	1	0	0	0	41	M	1	0	1	NS	0	S	THU	CT
9	6	MAR	2	1	0	0	0	72	F	1	1	1	NS	0	I	TUE	CT
10	1	JAN	0	1	0	0	0	47	M	1	0	1	NS	0	G	MON	CT
11	4	FEB	2	1	0	0	1	52	M	1	1	1	MG	0	C	FRI	CT
12	1	JAN	0	1	0	0	0	51	F	1	0	1	NS	0	G	MON	CT
13	0	JAN	0	1	0	0	0	47	M	1	0	1	NS	0	G	MON	CT
14	5	JAN	2	1	0	0	0	59	F	1	0	1	NS	0	G	FRI	CT
15	6	MAR	2	1	0	0	0	78	F	1	1	1	NS	0	I	TUE	CT
16	0	JAN	0	1	0	0	0	66	M	1	0	1	MP	0	R	THU	CT
17	4	JAN	2	1	0	0	1	48	M	1	1	1	MG	0	C	FRI	CT

Fig. 6. example of collected data for CART analysis
그림 6. CART 분석 데이터 예시

CART 알고리즘을 통해 구축된 분류나무를 보여 준다.

4.1 데이터

C 의료원은 국내 의료기관 인증을 취득한 7개의 부속 병원으로 구성되어 있으며, 그 중 1개 병원은 국제 의료기관 평가기관 인증(joint commission international: JCI)을 소유하고 있다. 2009년 이후 의료원 산하 5개 병원의 정보시스템을 통합 구축하여 운영하고 있으며, 서비스 수준 협약(service level agreement: SLA)을 토대로 IT서비스를 제공하고 있는 비교적 업무 프로세스가 선진화된 의료기관으로 볼 수 있다. 이러한 C 의료원의 부속 기관 중 700여 병상 규모의 병원을 한 곳 선정하여 영상의학검사의 TAT 분석을 실시하였다. 데이터는 2014년 1월~4월 사이 발생된 약 1만2천여 건의 요약 데이터를 사용하였으며, 관측변수는 영상의학검사의 TAT 달성여부(종속변수)와 TAT에 영향을 줄 것으로 예상되는 관측변수(독립변수)로 구성되어 있다. 그림 6에서 CART 알고리즘의 입력으로 사용할 수 있도록 전처리가 완료된 데이터의 모습을 보여주고 있다.

4.2 CART 분류 수행 결과

CART의 트리는 일반적으로 분할-성장종료-가지치기 단계로 구축된다. 분할 기준(criterion)은 그 하한선을 0.9, 0.95, 0.99와 같이 3가지 경우로

설정하였다. 최소 분할 수(number of split)로 10개, 30개, 50개, 100개와 같이 4가지 경우로 설정하였다. 나무 깊이(depth)는 0, 10, 5와 같이 3가지 경우를 고려하였다. 이와 같은 파라미터 세트를 행렬로 만들면 표 3에서 나타나는 바와 같이 최적의 파라미터 세트를 순위별로 비교 할 수 있다. 즉, CART 나무의 구축을 위해 가장 효율적인 파라미터 세트를 찾기 위함이다. 가장 첫 번째 리스트가 분할 기준, 분할 수, 깊이에 대해서 모두 최소의 값을 가지고 있으며, 그 외의 평가항목에서도 다른 조합에 비해 평가 결과가 우수하게 나타나고 있다. 따라서 본 연구에서는 가장 첫 번째 리스트를 최적의 파라미터 세트로 선정하였다. 하지만 나무의 복잡도(complexity)는 예측의 성능과 비례관계에 있다고 보기 어렵다. 때문에 최적의 파라미터 선정의 기준은 도메인 특성에 따라 얼마든지 다르게 적용 할 수 있을 것이다. 최적 파라미터 세트를 가지고 총 12,000개의 레코드(record)에 대해 CART 분석을 실시한 결과, 그림 7과 같이 노드 수가 37개이면서 깊이가 5인 트리가 구축되었다. 각 분기점 마다 99.9% 이상의 신뢰 수준으로 자료를 분할하였으며, 서비스의 적기처리로 간주하고 있는 TAT 달성과 실패에 대한 2가지 범주로 분류가 이루어졌다. 영상의학검사 후 6일 이내 판독완료를 의미하는 TAT의 달성이 80% 이상인 말단 노드를 TAT 적기처리로 보았고, TAT가 적기처리된 노

Table 3. top 20 ranks of optimal parameter sets

보고가 완료되어야 하는 제한사항이 기인한 것

표 3. 최적 파라미터세트 순위 20

criteria	split	depth	recall	precision	accuracy	F measure	AUROC	node
0.9	10	5	0.946531	0.928403	0.88827	0.937379	0.818798	19
0.9	30	5	0.946531	0.928403	0.88827	0.937379	0.818798	19
0.9	50	5	0.946531	0.928403	0.88827	0.937379	0.818798	19
0.9	100	5	0.946531	0.928403	0.88827	0.937379	0.818798	19
0.95	10	5	0.946531	0.928403	0.88827	0.937379	0.818798	19
0.95	30	5	0.946531	0.928403	0.88827	0.937379	0.818798	19
0.95	50	5	0.946531	0.928403	0.88827	0.937379	0.818798	19
0.95	100	5	0.946531	0.928403	0.88827	0.937379	0.818798	19
0.9	10	0	0.939622	0.93123	0.88535	0.935407	0.818172	27
0.9	10	10	0.939622	0.93123	0.88535	0.935407	0.818172	27
0.9	30	0	0.939622	0.93123	0.88535	0.935407	0.818172	27
0.9	30	10	0.939622	0.93123	0.88535	0.935407	0.818172	27
0.9	50	0	0.939622	0.93123	0.88535	0.935407	0.818172	27
0.9	50	10	0.939622	0.93123	0.88535	0.935407	0.818172	27
0.9	100	0	0.939622	0.93123	0.88535	0.935407	0.818172	27
0.9	100	10	0.939622	0.93123	0.88535	0.935407	0.818172	27
0.99	10	0	0.954941	0.925743	0.892516	0.940115	0.818112	16
0.99	10	10	0.954941	0.925743	0.892516	0.940115	0.818112	16
0.99	30	0	0.954941	0.925743	0.892516	0.940115	0.818112	16
0.99	30	10	0.954941	0.925743	0.892516	0.940115	0.818112	16

드는 8번, 13번, 19번, 24번, 25번, 26번, 31번, 33번, 34번, 36번, 37번으로 나타났다. TAT의 적기처리가 60% 미만으로 나타나는 노드는 TAT 미달로 정하였다. 즉, TAT 미달이 된 노드는 검사 후 7일이 경과한 후 판독 결과가 나오는 경우가 40%를 초과했다는 의미이다. TAT가 현저히 미달성된 노드는 5번, 9번, 16번, 17번, 20번, 30번으로 나타났다. 가장 첫 분할이 이루어진 1번 노드[V13:진료과]는 진료과를 기준으로 분할을 수행하였기 때문에 TAT의 적기처리를 예측하기 위해 첫 번째 데이터 분기가 발생한 변수는 “진료과목”으로 나타났다. 진료과목을 기준으로 자식노드를 통해 말단 노드를 형성할 때까지 반복적인 분기가 발생하게 되는데, 이 때 혼동하지 말아야 할 것은 부모 노드에서 사용된 변수가 자식 노드에서 발견 될 수 있다는 점이다. 두 번째 분할은 루트(root) 노드를 기준으로 좌측 분기된 2번 노드[V4:외래여부]에서 발생하였으며, 그 값이 “1”인 경우, 외래 내원 환자를 의미한다. 외래 내원 환자의 경우가 입원 또는 응급 환자보다 더 많은 건수가 TAT 적기 처리됨을 보여준다. 그 원인으로는 병원 서비스의 특성상 외래 방문 환자가 검사 후 다음 재방문 이전까지 검사결과

으로 해석할 수 있다. 루트(root) 노드를 기준으로 우측 분기된 21번 노드[V13:진료과]는 좌측 동일 위상의 2번 노드[V14:외래여부]보다 4배 이상 큰 레코드를 하위에 거느리고 있다. 이는 데이터 발생 빈도를 고려하였을 때 TAT 적기 처리율에 대한 인과관계를 설명할 때 좌측 분기와 비교하여 좀 더 큰 신뢰도를 가지고 해석할 수 있음을 나타낸다. 21번 노드[V13:진료과] 하위로 좌측 분기된 22번 노드[V17:모달리티]는 검사 분류에 따라 발생한 분기이다. 22번 노드[V17:모달리티] 하위의 분기를 관찰해 보면 초음파검사와 그 외의 검사인 CT와 MRI검사로 나누어지는 것을 확인 할 수 있다. 초음파 검사는 100% TAT가 적기 처리되고 있으며, 결국 CT와 MRI검사인 경우에서 TAT 적기 처리 여부가 갈리는 것으로 확인된다. 계속해서 22번 노드[V17:모달리티] 하위 노드인 23번 노드[V15:상병]가 최종 말단 노드를 거느리는데 상병 변수의 내용은 세계보건기구(WHO)에서 질병과 증상을 분류한 코드 체계이며, 국제통계분류 10차 개정판을 의미하는 ICD-10 코드로 명명한다. 나열된 “A”, “C”, “D”, “E”, “H” 등의 코드 구간은 각각 “전염병 또는 기생충병”, “종양”, “종양 또는 면역 체계와 관련

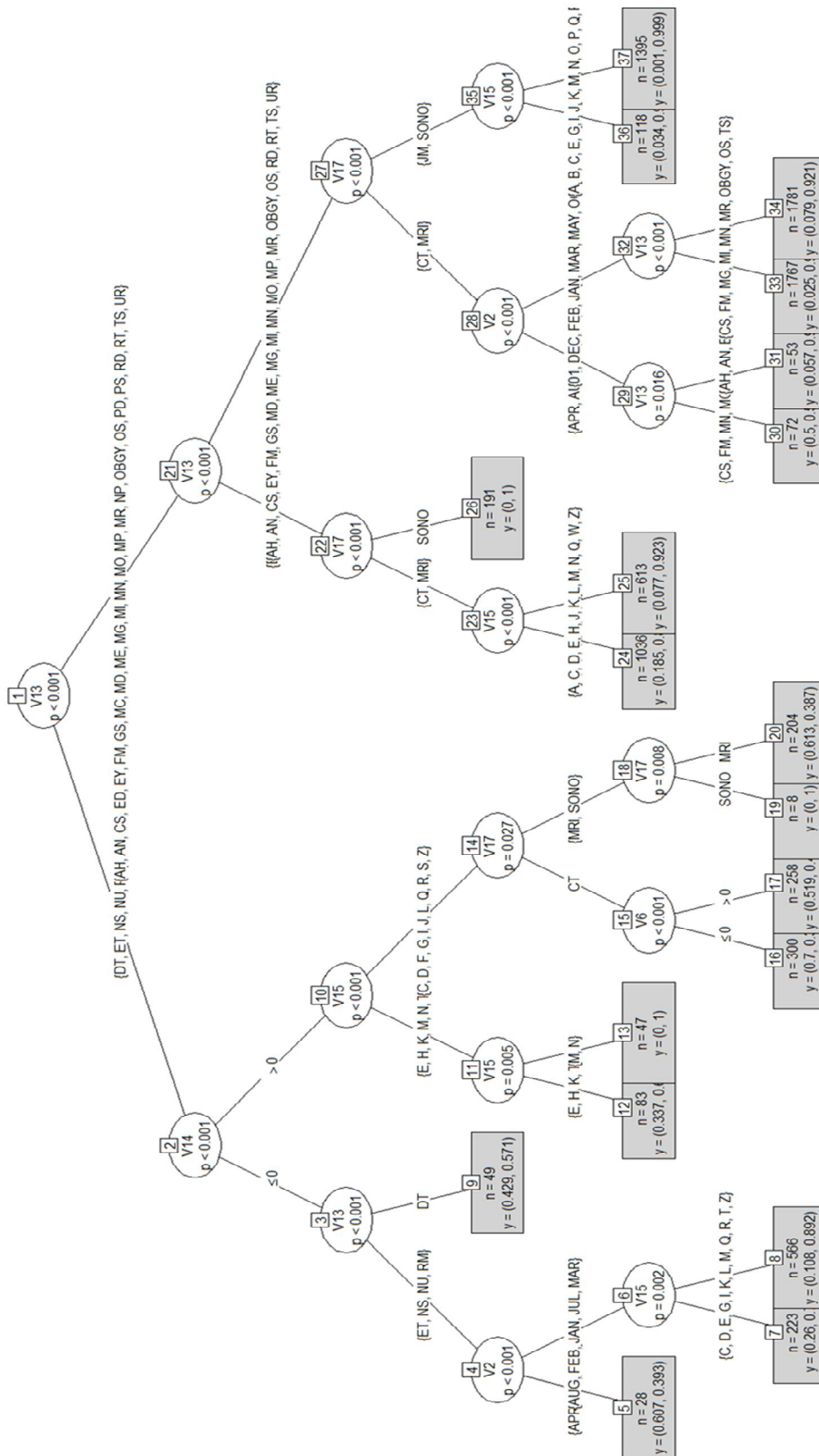


Fig. 7. completed tree
 그림 7. 완성된 의사결정나무

된 병”, “내분비선 영양 및 순환계”, “눈과 그에 딸린 기관의 질병” 등으로 ICD-10 코드 목록에서 확인 가능하다. 따라서 진료과목과 상병분류는 TAT 적기 처리에 많은 영향을 주는 요인으로 해석 될 수 있다. 나무를 거슬러 올라 21번 노드[V13:진료과] 하위로 우측 분기된 27번 노드[V17:모달리티]에 의해서도 초음파는 거의 모든 경우 TAT가 적기 처리됨을 확인 할 수 있다. 이 때 TAT 적기 처리가 확인되는 말단 노드에 이르기 직전 35번 노드[V15:상병]를 거치고 있으나, TAT 적기 처리에는 큰 영향력을 미치지 못 하고 있음을 확인 할 수 있다. 그러나 27번 노드[V17:모달리티] 하위로 좌측 분기된 28번 노드[V2:실시월]은 TAT 적기 처리에 계절적인 요인이 영향을 미칠 수 있음을 시사한다. 특히, 28번 노드[V2:실시월] 하위로 좌측 분기된 29번 노드

[V13:진료과]에 의해 “4월”이라는 계절 요인과 진료과목이 “CS”, “FM”, “MN”, “MC” 등에 해당하는 경우, TAT 적기 처리가 미달하는 경우가 발생되었다. 28번 노드[V2:실시월] 하위로 우측 분기된 32번 노드[V13:진료과]를 거치고 있으나, TAT 적기 처리에는 큰 영향을 미치지 못 하였다. C 의료원의 사례분석 결과, 영상의학검사의 TAT 적기처리는 “진료과”, “상병”, “모달리티”, “계절” 등의 요인에 의해 크게 영향을 받는다는 것을 확인하였다. 이와 같이 본 연구에서 개발한 예측 모형을 C 의료원에 적용한 결과, 기존 TAT 적기 처리율을 높일 수 있는 개선점을 도출할 수 있다.

V. 결론

본 연구에서는 의료서비스의 적기처리를 예측하기 위한 모형 개발 절차를 제시하였다. 결정트리 분석의 강점은 인간이 이해하기 쉬운 형태로 데이터의 인과관계를 해석할 수 있게 해 준다는 점이다. 본 연구에서는 다양한 결정트리 분석 기법 중 분류회귀나무 알고리즘을 모형에 적합화한 후 데이터 마이닝 절차에 따라 C 의료원에 대한 사례 분석을 실시하였다. 그 결과, 노드 수가 37개이면서 깊이가 5인 트리가 만들어졌으며, “진료과”, “상병”, “검사구분”, “실시월”이 의료서비스 적기 처리에 영향을 미치는 요인으로 나타났다. 본 연

구의 결과는 산업분야 전 영역에서 서비스혁신을 목적으로 한 다양한 연구로 확장될 수 있다. 또한 전문가시스템 또는 의사결정지원시스템 분야에서 그 활용 가치가 더욱 높을 것으로 전망됨에 따라, 본 연구에서 개발된 예측 모형의 활용도를 높이기 위한 사례분석 및 시뮬레이션 방안에 대한 활발한 논의와 추가 연구가 요구된다.

References

- [1] Korea Institute for Healthcare Accreditation (KOIHA) (<https://www.koiha.or.kr/home/index.act>), “Accreditation Standards”, Jun. 2015.
- [2] J. P. Crabbe., C. L. Frank., and W. W. Nye., “Improving Report Turnaround Time: An Integrated Method Using Data from a Radiology Information System”, American Journal of Roentgenology (AJR), Vol. 163, No. 6, pp. 1503-1507, Jul. 1994.
- [3] Giles W. L. Boland., Elkan F. Halpern., and G. Scott Gazelle., “Radiologist Report Turnaround Time: Impact of Pay-for-Performance Measures”, American Journal of Roentgenology (AJR), Vol. 195, No. 3, pp. 707-711, Sep. 2010.
- [4] Joan H. Howanitz., and Peter J. Howanitz., “Timeliness as a Quality Attribute and Strategy”, American Journal of Clinical Pathologist (AJCP), Vol. 116, No. 3, pp. 311-315, 2001.
- [5] Young M. Chae, Hye S. Kim, Kwan C. Tark, Hyun J. Park and Seoung H. Ho, “Analysis of healthcare quality indicator using data mining and decision support system”, Expert System with Applications, Vol. 24, No. 2, pp. 167-172, 2003.
- [6] Yun Seok Ko, “The Construction Methodology of a Rule-based Expert System using CART-based Decision Tree Method”, Journal of The Korea Institute of Electronic Communication Science, Vol. 6, No. 6, pp. 849-854, Dec. 2011.
- [7] Jun Tae Park and Soo Beom Lee, “Effects Analysis of Traffic Safety Improvement Program Using Data Mining : Focusing on Urban Area”,

Journal of Transport Research, Vol. 18, No. 2, pp. 77-91, Jun. 2011.

[8] Sonia Singh, "COMPARATIVE STUDY ID3, CART AND C4.5 DECISION TREE ALGORITHM: A SURVEY", International Journal of Advanced Information Science and Technology (IJAIST), Vol. 27, No. 27, pp. 97-103, Jul. 2014.

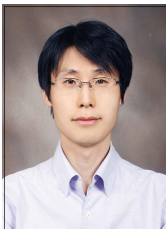
[9] Ture, Mevlut, Fusun Tokatli, and Imran Kurt. "Using Kaplan - Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4. 5 and ID3) in determining recurrence-free survival of breast cancer patients." Expert Systems with Applications Vol. 36, No. 2, pp. 2017-2026, 2009.

[10] Tzung-I Tang, Gang Zheng, Yalou Hualng, Guangfu Shu and Pengtao Wang, "A comparative study of medical data classification methods based on decision tree and system reconstruction analysis." Industrial Engineering and Management Systems, Vol. 4, No. 1, pp. 102-108, 2005.

[11] Antipov, Evgeny, and Elena Pokryshevskaya. "Applying CHAID for logistic regression diagnostics and classification accuracy improvement." Journal of Targeting, Measurement and Analysis for Marketing, Vol. 18, No. 2, pp. 109-117, 2010.

BIOGRAPHY

Jong-chan Lee (Member)



2005 : B.S. degree in Computer Engineering, Korea Polytechnic University, Gyeonggi, Korea.

2013 : M.S. degree in Information Systems

Engineering, Seoul National University of Science and Technology, Seoul, Korea.

2014 ~ Present : Doctoral Course in Information Systems Engineering, Seoul National University of Science and Technology, Seoul, Korea.

2013 ~ Present : Development Engineer, MIRACOM Inc Co., Ltd.

Areas of Interest : Data Mining, Business Analytics, ICT for Health

Seung-woo Jeong (Member)



1999 : B.S. degree in Administration, Kyonggi University, Gyeonggi, Korea.

2014 : M.S. degree in Information Systems Engineering, Seoul

National University of Science and Technology, Seoul, Korea.

2013 ~ Present : ICT Strategy Planner, The Catholic University of Korea, Uijeongbu St.Mary Hospital.

Areas of Interest : Management of Technology, Economic Feasibility Evaluation, ICT for Health

Won-young Lee (Member)



1978 : B.S. degree in Industrial Engineering, Seoul National University, Seoul, Korea.

1983 : M.S. degree in Industrial and Systems Engineering, Ohio State

University, Ohio, USA.

1990 : Ph.D. degree in Industrial Engineering, University of Louisville, Kentucky, USA.

1991 ~ Present : Professor, Dept. of Industrial and Information Systems Engineering, Seoul National University of Science and Technology, Seoul Korea.

Areas of Interest : Database Management System, Management Information System