

헬스 케어 시스템에서 질병 관리를 위한 질병 규칙 생성 기법 설계+

이병관* · 정이나** · 정은희***

*,**관동대학교 컴퓨터학과, ***강원대학교 지역경제학과

A Design of Disease Rule Creation Scheme for Disease Management in Healthcare System

Byung-Kwan Lee* · INa Jung** · Eun-Hee Jeong***

*,**Kwandong University, ***Kangwon National University

E-mail : bklee@kd.ac.kr, lupinus07@nate.com, jeongeh@kangwon.ac.kr

요 약

본 논문에서는 헬스케어시스템에서 효율적인 질병 관리를 위한 질병 규칙을 생성하는 질병규칙생성 기법(Disease Rule Creation Scheme)을 제안한다. DRCS는 기본적으로 Rough Set Theory의 개념을 이용하며, 각 속성과 결정 속성간의 지지율을 계산한다. 그리고 결정 속성을 기준으로 지지율이 낮은 속성은 제거한 후에, 질병을 판단할 수 있는 질병 규칙을 생성한다. 그 결과, DRCS는 C4.5알고리즘 보다 적은 질병규칙의 수를 생성하였으며, 질병 규칙의 정확성은 향상되었다.

ABSTRACT

The paper proposed the DRCS(Disease Rule Creation Scheme) which generates the disease rules for efficient disease management in Healthcare system. The DRCS uses basically Rough Set Theory and computes support between each attributes and decision attributes. It creates the disease rules that judges disease after it removes the attribute which is the lowest support. Therefore, it reduces the number of disease rules and improves the exactness, compared with C4.5 algorithm.

키워드

DRCS(Disease Rule Creation Scheme), Rough Set Theory, Support

I. 서 론

헬스 케어 시스템, 병원에서는 매일 환자에 관한 수많은 양의 데이터를 수집하고 있으므로, 이러한 헬스 케어 빅 데이터 속에서 유용한 정보와 환자 질병에 대한 규칙을 추출하기 위한 기법이 필요하다.

본 논문에서는 데이터 마이닝 기법의 하나인

Rough Set Theory를 이용하여 효율적인 헬스케어 시스템의 관리를 위한 질병 규칙을 생성하는 질병 규칙 생성 기법인 DRCS(Disease Rule Creation Scheme)를 제안한다.

II. 관련연구

RST(Rough Set Theory)는 1980년대 초에 Pawlak[1]에 의해 소개 되었으며, 어떤 집합에서 확실하게 분류되는 하한 근사 공간(Lower Approximation)과 불확실하게 분류되는 상한 근사 공간(Upper Approximation)을 집합이론을 통해 나타낸다[2].

-하한근사집합: $\underline{B}X = \{x \in U | [x]_B \subseteq X\}$

+ 이 논문은 2013년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (NRF-2012R1A1A4A01012039)

*관동대학교 컴퓨터학과 교수, 제 1저자

**관동대학교 컴퓨터학과 박사과정, 제 2저자

***강원대학교 지역경제학과 부교수, 교신저자

-상한근사집합: $\overline{B}X = \{x \in U \mid [x]_B \cap X \neq \emptyset\}$
 -경계영역(Boundary Region): $BR(X) = \overline{B}X - \underline{B}X$

RST를 적용한 의사결정시스템은 $S = (U, C, D)$ 로 정의한다. 여기서 U는 개체의 전체집합 $U = \{x_1, x_2, \dots, x_n\}$, C는 조건 속성들의 집합 $C = \{c_1, c_2, \dots, c_n\}$, 그리고 D는 결정속성집합 $D = \{d\}$ 을 의미한다.

개체들의 속성 값 집합이 V_C 일 때, $A = C \cup D$, $C \cap D = \emptyset$ 이라 하자. 두 개의 개체 x, y 가 $P \subseteq A$ 인 속성 집합 P에 대해 같은 속성 값을 갖게 되면 x, y 는 속성집합 P에 대해 식별 불구분 관계 또는 동치관계이라 하고 $IND(P) = \{(x, y) \in U \times U : f(x, c) = f(y, c), \forall c \in P\}$ 로 표현한다[3].

RST에서는 불필요한 속성집합을 제거할 수 있는데 이것을 리덕트(Reduct)라고 하는데, RST에서 규칙을 생성할 때 사용할 수 있다. 즉, 불필요한 속성집합을 제거한 최소속성집합을 C' 라고 할 때, $C' \subset A$ 이고 $IND(C') = IND(C)$ 일 경우 C' 를 C 의 reduct라고 한다. 그리고 속성에 대한 reduct들의 교집합을 C 의 코어라고 한다. 따라서 코어의 속성은 한 개, 여러 개, 또는 0개 일 수도 있다.

본 논문에서는 이 코어를 이용하여 질병 규칙을 생성하는데, 각 속성에 대한 지지율(support)을 계산하여 지지율이 높은 속성순으로 규칙을 생성하도록 설계한다.

III. DRCS 설계

본 논문에서 제안하는 질병 규칙 생성 기법인 DRCS(Disease Rule Creation Scheme)은 Decision 속성을 기준으로 각 속성에 대한 지지율을 계산하여 불필요한 속성을 제거한 후, Rough Set Theory[1]의 식별 불구분 IND, BR, 속성에 대한 reduct, 그리고 최종적으로 Core를 구한다.

그림 1은 본 논문에서 제안하는 DRCS의 전체적인 흐름도를 설명한 것이며, DRCS가 질병 규칙을 생성하는 단계별 절차는 다음과 같다.

[1 단계] decision 속성을 기준으로 질병에 대한 속성들에 대한 각각의 지지율(attr.sup)을 계산한다.

[2 단계] 각 속성들 중에서 낮은 지지율을 갖는 속성을 질병규칙 생성 절차에서 제외시킨다.

[3 단계] BR(Boundary Region)=상한근사집합 - 하한근사집합인 데이터를 찾아 제외시킨다. 즉, 모든 속성은 같은데 결정 속성만 다른 데이터를 찾아 질병 규칙 생성 데이터에서 제외시킨다.

[4 단계] 결정속성을 기준으로 계산한 지지율이 낮은 속성이 제외하고, BR이 제거된 질병 규칙 생성 데이터의 각 속성에 대한 IND(attr)을 찾는다.

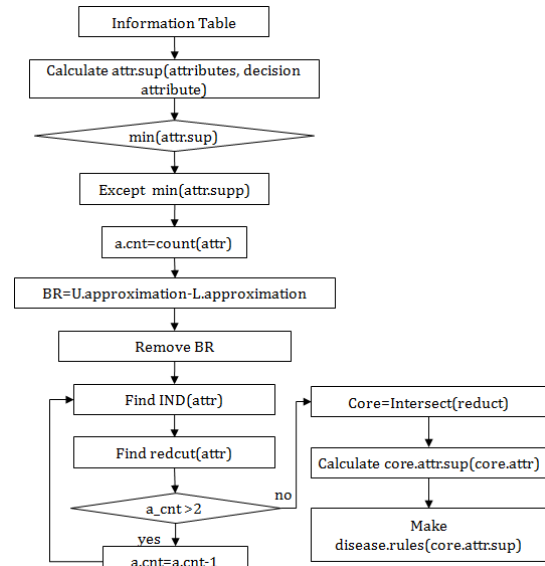


그림 1. DRCS 전체 흐름도

[5 단계] 4단계 결과 값을 이용하여 reduct를 구한다.

[6 단계] 4단계와 5단계를 반복하여 데이터의 각 속성에 대한 reduct를 모두 찾는다.

[7 단계] 6단계의 reduct들의 교집합이 Core가 된다. 이 Core의 속성들(core.attr)에 대한 지지율(core.attr.sup)을 계산하여, 지지율에 따라 질병 규칙(disease.rules)를 생성한다.

IV. DRCS 분석

본 논문에서 제안하는 DRCS는 20명의 환자 정보를 이용하여 맹기열을 판단하는 규칙을 생성하여, C4.5 알고리즘[4]과 비교하였다. 표1은 맹기열 실험 데이터[5]를 설명한 것이다. 그리고 표2는 DRCS의 7단계 후에 생성된 Core와 Core 속성의 지지율을 계산한 결과이다.

표 1. 맹기열 실험 데이터

Patient	Conditional Attributes			Decision Attribute
	blotched_red_skin	muscular_pain_articulations	temperature	dengue
1	no	no	normal	no
2	no	no	high	no
3	no	no	very high	yes
4	no	yes	high	yes
5	no	yes	very high	yes
6	yes	yes	high	yes
7	yes	yes	very high	yes
8	no	no	high	no
9	yes	no	very high	yes
10	yes	no	high	no
11	yes	no	very high	no
12	no	yes	normal	no
13	no	yes	high	yes
14	no	yes	normal	no
15	yes	yes	normal	no

16	yes	no	normal	no
17	yes	no	high	no
18	yes	yes	very high	yes
19	yes	no	normal	no
20	no	yes	normal	no

표 2. Core와 Core 속성의 지지율

Patient	Conditional Attributes			Decision Attribute
	blotched_red_skin	muscular_pain_articulations	temperature	dengue
1	no	no	normal	no
3	no	no	very high	yes
4	no	yes	high	yes
support (yes)	yes : 2/2	yes : 1/2		

표 3은 DRCS에 의해 생성된 최종 결과물인 질병 규칙의 결과를 설명한 것이다.

표 3. DRCS과 C4.5 알고리즘의 규칙 수 비교

C4.5 규칙	Rule 1	if (blotched_red_skin=no) and (temperature=normal) then (dengue=no)(100%)
	Rule 2	if (blotched_red_skin=no) and (temperature= high) and (muscular_pain_articulation=yes)and then (dengue = yes)(100%)
	Rule 3	if (blotched_red_skin=no) and (temperature= high) and (muscular_pain_articulation=no)and then (dengue=no) (100%)
	Rule 4	if (blotched_red_skin=no) and (temperature= very high) then (dengue=yes) (83%)
DRCS 규칙	Rule 1	if (blotched_red_skin=no) and (muscular_pain_articulation=no) and (temperature=normal) then (dengue=no)(100%)
	Rule 2	if (blotched_red_skin=no) and (muscular_pain_articulation=no)and (temperature=very high) then (dengue = yes)(100%)
	Rule 3	if (blotched_red_skin=no) and (muscular_pain_articulation=yes) and (temperature=high) then (dengue = yes)(100%)

그 결과, DRCS는 C4.5 알고리즘에 의해 생성된 규칙 수보다 1개 줄였으며, 각 규칙의 정확도는 DRCS가 높은 것으로 나타났다.

V. 결론

본 논문에서는 효율적인 헬스케어시스템의 관리를 위한 질병 규칙을 생성하는 질병규칙생성 기법(Disease Rule Creation Scheme)을 제안하였다.

본 논문에서 제안한 DRCS는 각 속성과 결정 속성간의 지지율을 계산하여, 지지율이 낮은 속성은 제거한 후에, 질병을 판단할 수 있는 질병 규칙을 생성하도록 설계하였다.

따라서 불필요한 데이터를 제외시킴으로써 좀 더 정확한 질병 규칙을 생성하였으며, 기존의 C4.5 알고리즘에 비해 규칙 수를 줄이고, 정확성

도 향상되었다.

그 결과, DRCS가 생성한 질병규칙을 헬스 케어 시스템에서 활용함으로써 응급상황에 신속하게 대비할 수 있으므로 불의의 사고를 미연에 방지할 수 있다.

참고문헌

- [1] Z. Pawlak, "Rough Sets," International Journal of Information and Computer Sciences, Vol.11 No.5, pp.341-356, 1982.
- [2] 서완석, 김재련, "러프집합과 SOM을 이용한 연속형 속성의 이산화," 한국산업경영시스템학회 2003 추계학술대회 논문집, pp.36-40, 2003.
- [3] 한상욱, 김재련, "러프셋 이론과 개체 관계 비료를 통한 의사결정나무 구성," Journal of the Korean Institute of Industrial Engineers, Vol.33, No.2, pp.183-190, June 2007.
- [4] Quinlan, J. R., "C4.5: Programs for Machine Learning," Morgan Kaufmann, Los Altos, 1993
- [5] Silvai Rissino, Germano Lambert-Torres, "Rough Set Theory - Fundamental Concepts, Principals, Data Extraction, and Applications," Data Mining and Knowledge Discovery in Real Life Applications, Book edited by: Julio Ponce and Adem Karahoca, ISBN 978-3-902613-53-0, pp. 438, February 2009, I-Tech, Vienna, Austria