

퍼지 ID3를 이용한 건설협력업체 평가 룰의 획득에 관한 연구

A Study on Acquisition of Evaluation Rules for Subcontractors of Construction Companies Using Fuzzy ID3

양상열* · 김성은** · 황승국*** · 원유동**** · Isao Hayashi*****

Sang-Yul Yang, Sung-Eun Kim, Seung-Gook Hwang, You-Dong Won, Isao Hayashi

* 경남대학교 경영학과 박사과정

** 경남대학교 산업공학과 강사

*** 경남대학교 산업공학과 교수

**** 경남대학교 경영학과 교수

***** Professor, Faculty of Informatics, Kansai University, Japan

Department of Industrial Engineering, Kyungnam University, Korea

요 약

건설업에서 기업경쟁력의 향상을 위해서는 우수한 협력업체의 확보가 매우 중요하며, 이들 협력업체는 종합적인 평가를 통하여 선정된다. 또한 이들 평가의 결과는 협력업체의 관리의 자료로도 사용된다. 그러므로 건설협력업체를 평가하는 룰을 사전에 알 수 있다면 업체를 관리하는데 효과적으로 사용될 수 있다. 그래서 주어진 데이터로부터 if-then 퍼지룰을 획득하는 방법인 퍼지 ID3을 이용하여 건설협력업체의 평가에 대한 룰을 구하고 그 유용성을 보이고자 한다.

Abstract

In construction industry, it is very important to secure the eminent subcontractors of construction companies in order to increase the competitive power, these subcontractors are selected through the total evaluation. Also, the results of evaluation is used management data the subcontractors. Therefore, if we know the evaluation rules of subcontractors previously, it can be used to manage the subcontractors effectively. Thus, using fuzzy ID3 which is obtain the if-then fuzzy rule from the given data, we acquire the rules for evaluation of subcontractors and show its usefulness.

Key words : Fuzzy ID3, Subcontractors of Construction Companies, If-then Rule

1. 서 론

건설업체들의 우수한 건설협력업체를 확보하는 것이 기업 경쟁력을 향상시키는 것과 직결되기 때문에 이들 건설협력업체에 대한 평가는 상당한 의의를 갖는다고 할 수 있다.

건설협력업체의 평가에 대한 객관적이고 합리적인 평가모델개발 및 건설협력업체의 평가등급에 대한 패턴을 분석하는 연구가 있다[1-2]. 이 연구는 일반적인 건설협력업체의 평가의 핵심역량모델로 간주하고, 건설협력업체를 평가할 수 있는 전문가들에 대한 설문조사를 통해 10개 부문 69개 평가항목을 선정하였다. 또한, 90개의 건설협력업체에 대한 평가와 이들 기업의 등급을 이용해서 퍼지 TAM 네트워크에 의한 패턴분석을 실시하여 퍼지 TAM 네트워크의 입력 및

출력데이터간의 if-then 퍼지룰을 생성하고 있다. 획득한 if-then 퍼지룰은 말로서 쉽게 설명하는 데는 한계가 있다. 이러한 한계점을 해결하기 위해서는 건설협력업체 평가를 위한 입출력데이터를 이용하여 기업평가와 관련하여 누구나 쉽게 이해할 수 있는 도구가 필요한 셈이다.

여기에 적합한 도구로서 데이터에서 if-then 룰을 획득하여 트리구조로 나타낼 수 있는 방법으로 ID3[3]가 있고 여기에 퍼지집합을 이용한 퍼지ID3[4-5]가 있다.

따라서, 본 논문에서는 퍼지ID3를 이용하여 기계설비전문인 건설협력업체의 평가에 대한 if-then 퍼지룰의 획득에 관한 연구를 하고자 한다.

2. 건설협력업체의 평가

본 논문에서 사용하는 건설협력업체의 핵심역량모델은 기존연구[1,2]에서 건설협력업체 평가 전문가의 설문을 통하여 선정한 10개 부문으로 표 1과 같다. 각 부문별 문항개수

접수일자 : 2007년 6월 29일

완료일자 : 2007년 8월 17일

감사의 글 : 본 연구는 2007학년도 경남대학교 학술연구장려금 지원으로 이루어졌음

의 합은 69개이다.

표 1에서 경영부문, 재무부문, 기술부문, 업체기여, 시공능력, 품질관리, 안전관리, 환경관리, 시공관리, 현장관리는 현장평가에 해당된다.

표 1. 평가부문
Table 1. Departments of Evaluation

평가 부문	경영 부문 A	재무 부문 B	기술 부문 C	업체 기여 D	시공 능력 E	품질 관리 F	안전 관리 G	환경 관리 H	시공 관리 I	현장 관리 J
문항 개수	10	4	3	4	3	10	10	8	10	7

3. 퍼지 ID3

데이터에서 if-then rule을 획득하는 방법인 ID3[3]에서는 입력데이터와 출력패턴과의 데이터집합이 주어졌으면 상호정보량이 최대가 되는 속성을 선택하고, 그 속성에 의하여 데이터집합을 분할하고, 트리구조의 if-then rule을 구성한다. 이 트리구조를 의사결정트리라고 한다.

ID3에 퍼지집합을 사용한 퍼지 ID3[4]는 멤버십값을 이용한 상호정보량을 정의하고, 데이터집합과 퍼지집합이 주어진 경우에 상호정보량이 최대가 되는 속성을 선택하고, 퍼지룰을 표현하는 의사결정트리를 구성한다. 퍼지집합을 이용하기 때문에 수치데이터를 취급하는 것이 가능해서 비교적 소수의 룰을 구성할 수 있다. 또 각 데이터를 멤버십값의 정도를 가지는 패턴의 클래스로 분할 하는 것이 가능하다. 그러나 멤버십값에서 데이터집합을 분할하여 의사결정트리를 작성하기 때문에 퍼지룰의 구성이 멤버십함수의 형상이나 연산자에 크게 의존한다.

퍼지 ID3에 파라메타가 있는 t-norm 연산자를 이용하여 퍼지룰의 AND 연산자를 정의하고, 황금분할법으로부터 상호정보량을 최대가 되도록 하기 위하여 AND연산자를 조정하는 퍼지 ID3[5]가 제안되어 있다.

본 논문에서 사용한 퍼지 ID3의 알고리즘은 다음과 같다.

지금 속성 $x_j, j=1,2,\dots,r$ 에 대해서 N 개의 데이터 $x_i = (x_{i1}, x_{i2}, \dots, x_{ie}), i=1,2,\dots,N$ 이 있다면 각 x_j 는 r 개의 클래스 $C_k, k=1,2,\dots,r$ 로 분류된다. 단, 클래스 C_k 는 보통 집합이다. 또, 각 x_j 에 대해서 $m(j)$ 개의 퍼지집합 $F_{tj}, t=1,2,\dots,m(j)$ 를 정의한다.

의사결정트리는 이하의 단계를 반복하여 구축한다. 이 반복회수를 기호 $L, L=0,1,2,\dots$ 로 나타낸다.

[단계 1] 의사결정트리의 제 L 번째 클래스 C_k 의 출현정도 G_k^L 을 식(1)에 의해 구한다.

$$G_k^L = \left(\frac{\sum_{i \in C_k} \mu^L(i)}{\sum_{i=1}^N \mu^L(i)} \right), k=1,2,\dots,r \quad (1)$$

단, $\mu^L(i)$ 는 제 L 번째에 있어서 데이터 x_i 의 적합정도이고, $\mu^0(i) = 1$ 로 한다.

[단계 2] 제 L 번째에 있어서 데이터 D 의 엔트로피 $H(D)$ 를 식(2)에 의해 구한다.

$$H(D) = - \sum_{k=1}^r G_k^L \log_2 G_k^L \quad (2)$$

[단계 3] 제 $L+1$ 번째의 속성 x_j^* 와 AND 연산자를 이하의 단계로부터 결정한다.

[단계 3-1] 속성 x_j 에 관한 조건부 엔트로피 $H(D|j,p)$ 를 식(3)에 의하여 구한다.

$$H(D|j,p) = \sum_{t=1}^{m(j)} (g_{tj} \times (- \sum_{k=1}^r G_k^{L+1} \log_2 G_k^{L+1})), j=1,2,\dots,n \quad (3)$$

단,

$$g_{tj} = \left(\frac{\sum_{i=1}^N \mu F_{tj}(x_{ij})}{\sum_{t=1}^{m(j)} \sum_{i=1}^N \mu F_{tj}(x_{ij})} \right) \quad (4)$$

$$\mu^{L+1}(i) = \mu^L(i) \textcircled{1} \mu F_{tj}(x_{ij}) \text{ for } \forall t \quad (5)$$

여기서, $\textcircled{1}$ 는 파라메타 p 를 가지는 t-norm 연산자를 나타낸다.

[단계 3-2] 상호정보량 $I(j,p)$ 를 식(6)에 의하여 구한다.

$$I(j,p) = H(D) - H(D|j,p) \quad (6)$$

[단계 3-3] $I(j,p)$ 가 최대가 되는 파라메타 p^* 를 황금분할법을 이용하여 구한다. 여기서, 황금분할법을 이용하는 이유는 파라메타 p 에 관해서 상호정보량 $I(j,p)$ 가 단봉성을 만족할 필요가 있기 때문이다.

[단계 3-4] 식(7)을 만족하는 속성 x_j^* 를 구한다.

$$I(j^*, p^*) = \max_j I(j, p^*) \quad (7)$$

[단계 3-5] 속성 x_j^* 를 제 $L+1$ 번째의 속성으로 하고, p^* 의 t-norm 연산자를 AND 연산자로 한다.

[단계 4] 식(8)의 조건을 만족하는 경우에는 데이터집합 D 에 관한 퍼지룰이 획득된 것으로 하고, 이후의 층에서의 알고리즘의 단계를 정지한다.

$$\exists k; G_k^{L+1} > \alpha \quad \text{또는} \quad \sum_{i=1}^N \mu^{L+1}(i) < \beta \quad (8)$$

단, α 는 출현정도가 α 이상인 클래스가 존재하면 정지하기 위한 역치이고, β 는 데이터 x_i 의 적합정도의 합계가 β 이하이면 정지하기 위한 역치이다.

[단계 5] [단계 4]의 정지 규칙을 만족하지 않는 경우에는 $L=L+1$ 로 하여 [단계 1]로 돌아가고, 알고리즘을 계속 진행한다.

4. 사례 연구

퍼지 ID3을 이용한 건설협력업체 평가에 표 2와 같이 29개의 데이터를 사용한다. 표 3은 건설협력업체에 대한 평가 등급을 나타내고 있다. 건설협력업체의 평가시에 등급을 표 3과 같이 10등급으로 나누는데 이것이 너무 세밀하기 때문에 다시 미흡, 보통, 양호, 우량의 4 등급으로 나누어 퍼지 ID3에 적용하였다. 본 연구에서 사용한 자료에는 우량의 데이터를 가진 건설협력업체는 없었다.

표 2. 건설협력업체 평가 데이터
Table 3. Data of Subcontractor of Construction Company

No	A	B	C	D	E	F	G	H	I	J	Class
1	42	45	26	43	26	44	38	42	68	64	LC
2	48	65	38	45	32	50	44	44	56	56	LC
3	48	50	32	55	32	72	48	36	64	64	NC
4	56	70	44	55	64	62	60	58	64	60	NC
5	56	85	44	50	68	64	62	54	66	62	NC
6	64	35	40	45	42	60	70	48	58	60	NC
7	66	75	34	45	48	62	58	56	64	56	NC
8	64	65	32	48	26	64	64	66	72	66	NC
9	70	60	32	48	26	60	54	52	56	58	NC
10	54	55	32	55	26	58	44	46	70	62	NC
11	48	50	38	50	32	52	44	36	66	56	NC
12	58	60	32	40	54	62	52	42	54	60	NC
13	56	55	32	55	70	50	46	36	52	62	NC
14	66	75	34	45	48	62	58	56	64	56	NC
15	58	60	32	40	54	62	52	42	54	60	NC
16	56	85	44	50	68	64	62	54	66	62	NC
17	64	65	32	48	26	64	64	66	72	66	NC
18	76	65	38	48	36	70	76	78	74	72	NC
19	56	75	38	35	64	48	74	58	76	72	NC
20	68	60	46	65	52	64	66	60	76	76	NC
21	62	75	26	65	60	64	66	68	74	66	NC
22	60	80	32	65	42	64	70	58	74	68	NC
23	64	60	46	55	58	60	60	48	64	64	NC
24	60	80	32	65	42	64	70	58	74	68	NC
25	90	80	80	80	84	92	86	80	84	80	GC
26	66	70	52	80	70	82	84	48	88	86	GC
27	76	70	80	65	80	68	66	52	66	70	GC
28	94	95	80	80	84	94	94	70	90	80	GC
29	92	80	80	80	84	96	90	74	78	80	GC

표 3. 건설협력업체의 등급
Table 3. Grade of Subcontractor of Construction Company

		미흡 LC		보통 NC		양호 GC		우량 SC	
매우 불량	불량	미흡	보통 이하	보통	보통 이상	양호	우량	매우 우량	초 우량
1	2	3	4	5	6	7	8	9	10

표 4는 10개의 평가부문에 대한 퍼지집합이다.

표 4. 부문별 퍼지집합
Table 4. Fuzzy Sets by Departments

구 분		1점		2점		3점		4점	
		포인 트	멤버 집합 수	포인 트	멤버 집합 수	포인 트	멤버 집합 수	포인 트	멤버 집합 수
A	Big	70	0	75	0.5	80	1	100	1
	Medium	45	0	55	1	65	1	75	0
	Small	0	1	0	1	30	1	55	0
B	Big	70	0	77.5	0.5	85	1	100	1
	Medium	50	0	60	1	70	1	80	0
	Small	0	1	0	1	40	1	55	0
C	Big	55	0	65	0.5	75	1	100	1
	Medium	20	0	40	1	50	1	70	0
	Small	0	1	0	1	20	1	40	0
D	Big	60	0	67.5	0.5	75	1	100	1
	Medium	35	0	50	1	60	1	75	0
	Small	0	1	0	1	30	1	45	0
E	Big	80	0	85	0.5	90	1	100	1
	M.B	60	0	70	1	80	1	90	0
	Medium	45	0	55	1	65	1	75	0
	S.M	30	0	40	1	50	1	60	0
F	Small	0	1	0	1	30	1	40	0
	Big	90	0	95	0.5	100	1	120	1
	M.B	65	0	75	1	85	1	95	0
	Medium	45	0	55	1	65	1	75	0
G	S.M	25	0	35	1	45	1	55	0
	Small	0	1	0	1	20	1	30	0
	Big	90	0	95	0.5	100	1	120	1
	M.B	65	0	75	1	85	1	95	0
H	Medium	45	0	55	1	65	1	75	0
	S.M	25	0	35	1	45	1	55	0
	Small	0	1	0	1	20	1	30	0
	Big	70	0	75	0.5	80	1	100	1
I	Medium	45	0	55	1	65	1	75	0
	Small	0	1	0	1	40	1	50	0
	Big	70	0	75	0.5	80	1	100	1
J	Medium	45	0	55	1	65	1	75	0
	Small	0	1	0	1	40	1	50	0

표 4에서는 10개의 부문별 퍼지수를 표현하는 4개의 점에 대하여 그 실수 값과 멤버십함수로 표현하였다.

표 2와 표 4를 이용하여 퍼지 ID3의 알고리즘을 적용한 결과는 다음과 같다. 여기서, $\alpha=0.8$, $\beta=1.0$ 으로 하였다.

1. (Management,Big [p:99.972])
-> GC (LC:0.00, NC:0.14, GC:0.86)
2. (Management,Medium [p:99.972])
-> NC (LC:0.02, NC:0.94, GC:0.05)
3. (Management,Small [p:99.972]),
(Quality,M.B [p:0.128])
-> NC (LC:0.00, NC:1.00, GC:0.00)
4. (Management,Small [p:99.972]),
(Quality,Medium [p:0.128]),
(Finance,Medium [p:99.972]),
(Technology,Medium [p:0.128]),
(Construction_Ability,S.M [p:99.972])
-> LC (LC:1.00, NC:0.00, GC:0.00)
5. (Management,Small [p:99.972]),
(Quality,Medium [p:0.128]),
(Finance,Medium [p:99.972]),
(Technology,Medium [p:0.128]),
(Construction_Ability,Small [p:99.972]),
(Environment,Medium [p:99.972])
-> NC (LC:0.00, NC:1.00, GC:0.00)
6. (Management,Small [p:99.972]),
(Quality,Medium [p:0.128]),
(Finance,Medium [p:99.972]),
(Technology,Medium [p:0.128]),
(Construction_Ability,Small [p:99.972]),
(Environment,Small [p:99.972]),
(Contribution,Medium [p:99.972]),
(Safety,S.M [p:99.972]),
(Construction_Control,Medium [p:99.972]),
(Field,Medium [p:99.972])
-> w LC NC (LC:0.61, NC:0.39, GC:0.00)
7. (Management,Small [p:99.972]),
(Quality,Medium [p:0.128]),
(Finance,Medium [p:99.972]),
(Technology,Small [p:0.128])
-> NC (LC:0.12, NC:0.88, GC:0.00)
8. (Management,Small [p:99.972]),
(Quality,Medium [p:0.128]),
(Finance,Small [p:99.972])
-> NC (LC:0.00, NC:1.00, GC:0.00)
9. (Management,Small [p:99.972]),
(Quality,S.M [p:0.128])
-> LC (LC:0.97, NC:0.03, GC:0.00)

이 결과를 의사결정트리로 표현하면 다음의 그림 1과 같다. 이 의사결정트리를 보면 건설협력업체를 평가한 물을 쉽게 이해할 수 있다.

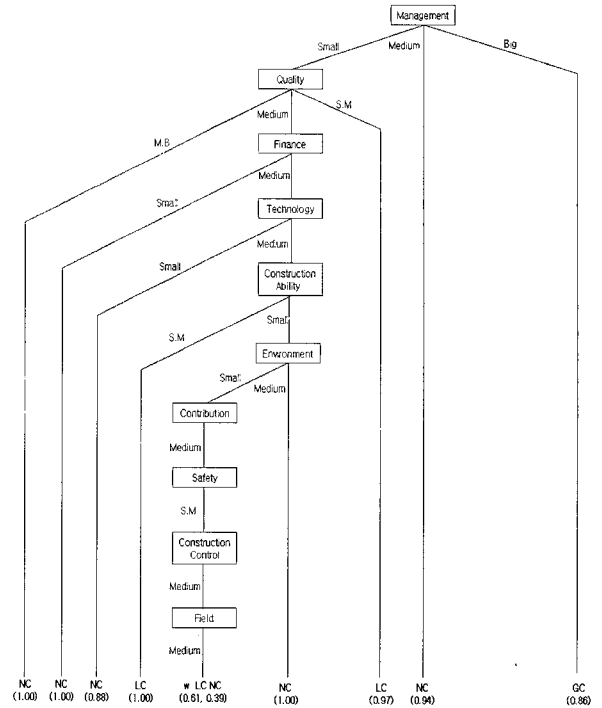


그림 1. 의사결정트리
Fig. 1 Decision Tree

5. 결 론

본 논문에서는 퍼지ID3를 이용하여, 건설협력업체의 핵심역량모델에 대한 평가시의 if-then 퍼지룰의 획득에 관한 연구를 하였다.

건설업에서 우수한 협력업체를 확보하기 위해서는 협력업체에 대한 종합적인 평가가 필요한데 이들 평가에 대한 룰을 사전에 알 수 있다면 협력업체를 관리하는데 매우 효과적으로 사용될 수 있다.

본 논문에서 구한 연구결과는 주어진 데이터로부터 if-then 퍼지룰을 획득하는 방법이므로 데이터의 정확성이 매우 중요하다고 할 수 있다. 이러한 관점으로 본 논문에서는 데이터의 정확성을 전문가를 통하여 다시 검증한 후 퍼지 ID3을 이용하여 건설협력업체의 평가에 대한 룰을 구하고 의사결정트리로서 그 구조를 보였다.

참 고 문 헌

- [1] 김성은, 황승국, "퍼지 TAM 네트워크를 이용한 건설협력업체 핵심역량모델의 패턴분석", *퍼지 및 지능시스템학회 논문지*, Vol. 16, No. 1, pp. 86-93, 2006.
- [2] 김성은, *퍼지 TAM 네트워크를 이용한 건설협력업체 핵심역량모델의 패턴분석*, 경남대학교 대학원 박사학위논문, 2006.
- [3] J. R. Quinlan : *Discovering Rules by Induction from Large Collections of Examples, Expert Systems in the Micro Electronics Age*, Edinburgh University Press, 1979.
- [4] 馬野元秀, 岡本宏隆, 鳩野逸生, 田村坦之, 河内二三夫, 敏津祐久, 木下淳一, "ID3に基づくファジィ決定木の油中ガス分析診断への適用について", 第4回インテリジェントシンポジウム予稿集, pp. 201-204, 1993.
- [5] Isao Hayashi, Toshiyuki Maeda, Jun Ozawa, "A Proposal of Fuzzy ID3 with Ability of Tuning for AND Connectives," *Journal of Japan Society for Fuzzy Theory and Systems*, Vol. 11, No. 4, pp. 677-683, 1999.



황승국(Seung-Gook Hwang)
 1981년 : 동아대학교 산업공학 학사
 1983년 : 동아대학교 산업공학 석사
 1991년 : Osaka Prefecture University
 경영공학 박사
 1991년~현재 : 경남대학교 산업공학과 교수

관심분야 : 퍼지모델링 및 평가

Phone : +82-55-249-2705
 Fax : +82-55-249-2463
 E-mail : hwangsg@kyungnam.ac.kr



원유동(You-Dong Won)
 1978년 : 한양대학교 산업공학 학사
 1981년 : 미국 Columbia대학교
 경영학 석사
 1994년 : 한남대학교 산업공학 박사
 1987년~현재 : 경남대학교 경영학과 교수

관심분야 : 생산관리
 Phone : +82-55-249-2430
 Fax : +82-55-223-1655
 E-mail : wonyd@kyungnam.ac.kr

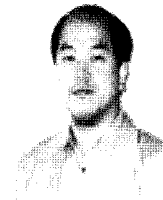
저 자 소개



양상열(Sang-Yul Yang)
 1979년 : 조선대학교 전기공학 학사
 2002년 : 경남대학교 경영학 석사
 2005년~현재 : 경남대학교 경영학과
 박사과정

관심분야 : 건설품질경영

Phone : +82-55-249-2705
 Fax : +82-55-249-2463
 E-mail : ysyf21@nate.com



林 勳(Isao Hayashi)
 1981년 : Osaka Prefecture University
 경영공학 학사
 1983년 : Osaka Prefecture University
 경영공학 석사
 1991년 : Osaka Prefecture University
 경영공학 박사
 2004년~현재 : Kansai University
 총합정보학부 교수

관심분야 : 지능정보학(신경회로모델, 퍼지모델, 생체로봇)
 Phone : +81-72-690-2448
 Fax : +81-72-690-2491
 E-mail : ihaya@kcn.res.kutc.kansai-u.ac.jp



김성은(Sung-Eun Kim)
 2000년 : 경남대학교 산업공학 학사
 2002년 : 경남대학교 경영학 석사
 2006년 : 경남대학교 산업공학 박사
 2007년 : 경남대학교 산업공학과 강사
 2007 현재 : 덕진종합건설(주) 대표이사

관심분야 : 건설품질경영

Phone : +82-55-268-6611
 Fax : +82-55-268-6622
 E-mail : dukjinn@yahoo.co.kr