

# 퍼지균등화와 러프집합을 이용한 선박설계 지식기반 구축

서규열\*

\*동명대학교 조선공학과

## Knowledge Base Construction of Ship Design Using Fuzzy Equalization and Rough Sets

KYU-YOUL SUH\*

\*Department of Naval Architecture, Dong-Myung Univ., Busan, Korea

**KEY WORDS:** Automatic rule derivation 자동규칙추출, Fuzzy equalization 퍼지균등화, Rough sets 러프집합, Consistent measure 일관성척도, Expert system 전문가시스템

**ABSTRACT:** Inference rules of the knowledge base, generated by experts or optimization, may be often inconsistent and incomplete. This paper suggests a systematic and automatic method which extracts inference rules not from experts' subject but from data. First, input/output linguistic variables are partitioned into several properties by the fuzzy equalization algorithm and each combination of their properties comes to premise of inference rule. Then, the conclusion which is the most suitable for the premise is selected by evaluating consistent measure. This method, automatically from data, derives inference rules from experience. It is shown through application that extracts new inference rules between hull dimensions and hull performance.

### 1. 서 론

### 2. 데이터로부터의 퍼지 규칙 생성 기법

전문가 시스템은 특정 분야의 문제를 해결하기 위해 해당 분야의 전문가가 가지고 있는 지식을 인위적으로 컴퓨터에 부여한 시스템이다.

현실성 있는 전문가 시스템을 개발하기 위해서는 방대한 양의 정확한 사실과 규칙을 지식기반(Knowledge-base)으로 구축하여야 한다. 이러한 사실과 규칙은 특정 분야의 전문가로부터 획득된다. 그러나 전문가의 지식은 정확한 수치로 되어 있지 않고 정성적인 언어표현으로 되어 있어서 이를 정량화하여 규칙으로 생성하는 규칙 추출 과정을 필요로 한다.

전문가에 의한 규칙 추출 기법(김화수 등, 1999)은 방대한 양의 규칙을 생성하기 위한 많은 전문가의 확보에 따른 막대한 비용뿐만 아니라 전문가들의 주관에 개입되어 생성된 규칙의 일관성을 확보하는 것이 어렵다. 또한 데이터의 학습이나 최적화에 의한 규칙 추출 기법(곽근창 등, 2001)은 방대한 양의 데이터를 일관적으로 다룰 수 있는 장점이 있으나 생성된 규칙이 직관적 이해가 불가능하고 언어적 지식의 형태가 아닌 단점이 있다.

본 논문에서는 선박설계 분야에 대한 전문가 시스템의 지식기반 구축을 위한 체계적이고 자동적인 규칙 추출 기법을 제안한다.

공학 설계 과정에서 설계자의 지식은 설계 경험으로부터 얻은 노하우(Know-how)인데 이것은 규칙화 되어 정리 되어 있는 것이 아니라 대부분 설계 변수의 값을 정성적인 언어변수의 형태로 표현한다.

그러나 풍부한 경험을 가진 설계자라 할지라도 해당 분야의 설계 지식을 순차적으로 나열하거나 정리하지는 못한다. 이는 공학 설계가 일련의 문제 해결 과정이며 모든 가능한 경우를 고려하는 것이 아니라 매번 관련 데이터 및 경험을 바탕으로 한 새로운 의사결정을 하는 과정이기 때문이다.

본 연구에서는 기존의 설계 결과, 즉 실적 데이터에 전문가의 경험 지식이 녹아 있다고 판단하여 데이터로부터 퍼지 규칙을 추출한다. 다만, 실적 데이터는 각 상황에서의 항상 최적해는 아니고 우수한 허용해이므로 데이터에 기반한 최적화(곽근창 등, 2001) 아닌 데이터 분포를 바탕으로 정성적 언어 변수로 표현된 퍼지 규칙을 생성하였다.

본 연구의 목표는 인간의 주관적 판단을 배제하고 단지 데이터로부터 컴퓨터가 스스로 설계자의 경험지식에 가까운 형태의 퍼지 규칙을 추출하는 체계적인 방법을 개발하는 것이다. 이를 위해 Fig. 1과 같은 과정을 거친다.

#### · Step 1: Selecting I/O linguistic variables

설계자가 생성하고자 하는 규칙의 전제부, 결론부를 구성하는 언어변수를 선정한다. 전제부와 결론부의 언어변수는 상호연관성이 높은 변수를 선정한다.

교신저자 서규열: 부산광역시 남구 용당동 535번지

051-620-3401 kyuyoul@tu.ac.kr

· Step 2: Fuzzy equalization of linguistic variables

시스템은 데이터의 확률분포로부터 언어변수의 각 속성이 균등한 퍼지확률을 갖도록 분할한다. 데이터의 확률 분포는 데이터의 경향에 따라 다르지만, 일반적으로는 인간의 언어 습관에 적합하도록 정규분포를 사용한다.

· Step 3: Constructing premises of rules

규칙의 전제부 언어변수의 분할은 설계자가 임의로 수를 정할 수 있다. 우선 전제부 언어변수의 분할 수에 따라 가능한 모든 규칙 전제부 후보를 생성한다.

· Step 4: Calculating consistent measure of the consequences

각 전제부에 가장 일관성 척도가 높은 결론부 속성을 선택하여 추론 규칙을 추출한다. 각 전제부마다 결론부 언어변수의 속성의 수에 따른 결론부를 생성할 수 있으며, 일관성 척도는 규칙 전제부에 속하는 데이터의 수와 각 결론부에 속하는 데이터의 수의 비로 정의한다.

· Step 5: Selecting consequences property of fuzzy rules

각 전제부에 가장 높은 일관성 척도를 갖는 결론부를 선택하고, 실적 데이터의 오차를 감안하여 전제부에 속하는 데이터의 수가 일정 수 이하일 경우 규칙 선택에서 제외한다.

이러한 과정을 통하여 인간의 언어표현에 근접한 속성 분할을 가지는 동시에 과거 실적자료에 기반한 반영한 퍼지 규칙을 체계적으로 생성할 수 있다.

3. 퍼지균등화 분할

퍼지 균등화(Fuzzy equalization)는 임의의 전체공간  $X$ 가 주어지고 이 공간상에서 취득된 수치 데이터가 있을 경우, 이 들로부터 확률분포 함수를 구한 후 전체공간  $X$ 에서 이들을 적절히 표현할 수 있는 소속함수들을 원하는 개수만큼 체계적으로 만들어 낼 수 있는 기법이다.

이러한 퍼지 균등화 이론(Pedrycz, 2001)은 Zadeh 교수에 의해 정의된 퍼지 사건의 개념(Zadeh, 1965)에 기인한다. 기존의 확률개념에서 사용되는 사건(Crisp event)의 개념에서는  $A$  라는 사건과 이의 확률분포함수  $p(x)$ 가 주어졌을 경우 사건  $A$ 의 확률은 다음과 같이 구해진다.

$$P(A) = \int_A p(x)dx \tag{1}$$

그러나 만약 사건  $A$ 에 대하여 확률분포함수  $p(x)$ 와 더불어 소속함수  $A(x)$ 값이 주어지면,

$$P(A) = \int_X A(x)p(x)dx \tag{2}$$

위와 같은 개념에 기초하여 전체공간  $X$ 에 대하여 확률밀도 함수  $p(x)$ 가 주어져 있을 때, 이 공간을 퍼지 집합  $\{A_1, A_2, \dots, A_n\}$ 로 나누는 퍼지사건을 정의한다고 하자. 그러면, 이 때, 가장 바람직한 분할 방법은 각각의 퍼지사건의 확률이 같도록 아래와 같이 균등화하는 방법이다.

$$P(A_1) = P(A_2) = \dots = P(A_n) = \frac{1}{n} \tag{3}$$

위의 개념을 적용하면 Fig. 2와 같이, 데이터가 많은 부분에서는 조밀한 소속함수의 값을 갖고 그 반대의 경우에는 듚성 한 소속함수의 배치를 갖게 되어, 인간의 직관과도 잘 일치하는 퍼지 분할이 가능하다.

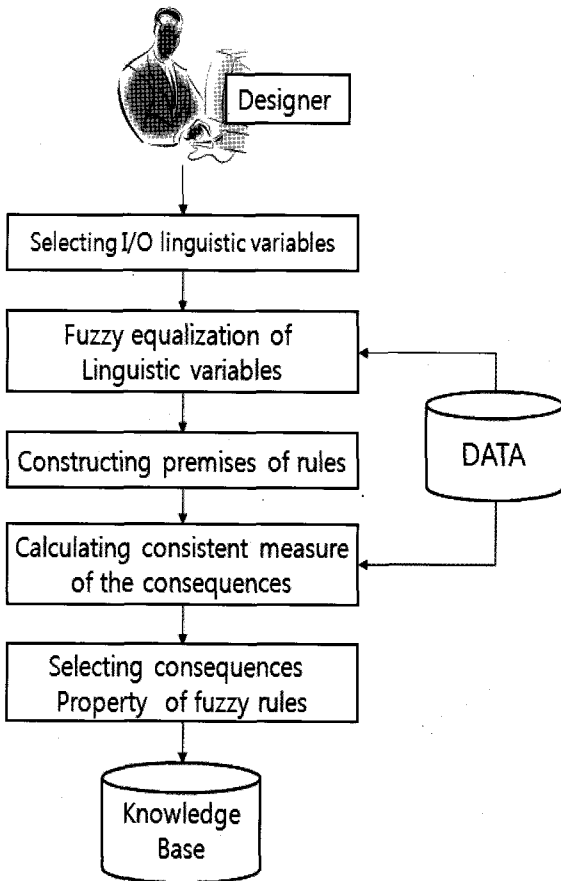


Fig. 1 Fuzzy rule derivation from data

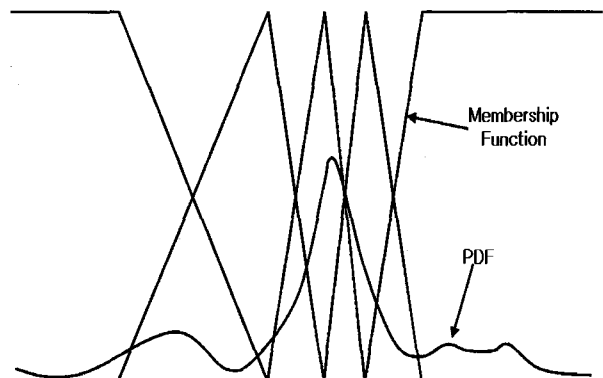


Fig. 2 The idea of Fuzzy equalization

### 4. 러프집합과 일관성척도

#### 4.1 러프집합의 개념

러프집합(Rough set) 이론(Pawlak, 1992)은 논의 영역에 있는 모든 개체는 그것을 기술하기 위하여 정보와 관련지어 생각할 수 있다는 가정에 근거를 두고 있으며, 임의의 집합  $X$ 를 속성으로 묘사하기 위하여 상한 근사와 하한 근사를 다음과 같이 정의한다.

전체 집합  $U$ 의 부분 집합  $X$ 에 대하여 다음의 연산으로 정의되는 두 집합  $B_*(X)$ ,  $B^*(X)$ 를 각각  $X$ 의  $B$ -하한,  $B$ -상한 근사라고 한다.

$$\begin{aligned} B_*(X) &= \{x \in U \mid B(x) \subseteq X\} \\ B^*(X) &= \{x \in U \mid B(x) \cap X \neq \phi\} \end{aligned} \quad (4)$$

$BR_B(X) = B^*(X) - B_*(X)$ 를  $X$ 의  $B$ -경계 영역이라고 하는데 이는 집합  $X$ 가 속성  $B$ 의 관점에서 어느 정도 정확하게 관련되는가의 정도를 나타내는 의미를 가지고 있으므로  $BR_B(X) = \phi$ 이면 집합  $X$ 는 속성  $B$ 에 대하여 크리스프(Crisp)하다고 하고  $BR_B(X) \neq \phi$ 이면 집합  $X$ 는 속성  $B$ 에 대해 러프(Rough)하다고 한다. 다시 말해서 러프집합  $X$ 는 근사 정확도(Accuracy of approximation)라고 불리는 다음의  $\alpha_B(X)$ 로 그 특성이 나타난다고 말할 수 있다.

$$\alpha_B(X) = \frac{|B_*(X)|}{|B^*(X)|} \quad (5)$$

이 때,  $|B^*(X)|$ 와  $|B_*(X)|$ 는 각각  $B^*(X)$ 와  $B_*(X)$ 의 원소의 개수를 나타낸다.

#### 4.2 일관성 척도

퍼지 규칙을 생성하기 위하여 퍼지 규칙의 전제부에 해당하는 속성에 대한 결론부에 해당하는 속성의 적합한 정도를 정량화하기 위한 척도가 필요하다. 이를 위해 조영완 등(1996)이 제안한 퍼지 규칙의 일관성 척도를 사용한다.

$X_i$ 와  $Y_j$ 가 전제부 속성,  $Z_k$ 가 결론부 속성일 때,

$$\gamma_{X_i \times Y_j}(Z_k) = \frac{|C(d) \cap X_i \times Y_j|}{|X_i \times Y_j|} \quad (6)$$

를 일관성 척도(Consistent measure)라고 한다.

이 때,  $Z_k$ 는 결론부 속성  $Z$ 에 의해 분할된 클래스 즉,  $Z_k \in U/I(Z)$ 이고,  $|X_i \times Y_j|$ 는 입력 공간  $X_i \times Y_j$ 에 존재하는 데이터의 수,  $C(d)$ 는 데이터  $d \in Z_k$ 를 포함하는 전제부 속성 집합  $C$ 에 의한 등가 클래스이다.

본 연구에서는 전제부 속성에 해당하는 데이터에 대한 각 결론부 속성에 해당하는 데이터의 수의 비로 나타내었다.

### 5. 선박설계 지식추출에의 적용

$L/B$ , 선수부 주형계수( $C_{p_f}$ )를 규칙의 전제부 언어변수로 하고, 조파저항계수( $C_w$ )를 결론부 언어변수로 하여 제안한 방법으로 규칙을 추출하여 보았다.

각 언어변수의 분할 속성의 수는 3개로 하고, 데이터의 확률 분포를 정규분포로 가정하였다. 49척의 저속비대선형( $Fn \leq 0.2$ ,  $C_B \geq 0.75$ )의 실적 자료로부터 평균과 표준편차를 계산하여 정규분포 식을 구하고, 퍼지균등화 분할 결과는 Fig. 3, 4, 5과 같다.

각 언어변수의 속성들의 조합으로 생성되는 규칙의 전제부 분할은 소속도가 0.5인 경계로 나누면 Fig. 6과 같이 분할된다.

각 전제부에 대하여 일관성 척도를 계산한 결과는 Table 1과 같다. 전제부에 해당되는 데이터가 5개 이상이고 일관성 척도가 0.5보다 큰 규칙을 선택하였다.

최종적으로 추출된 퍼지 규칙은 Table 2에 나타내었다.

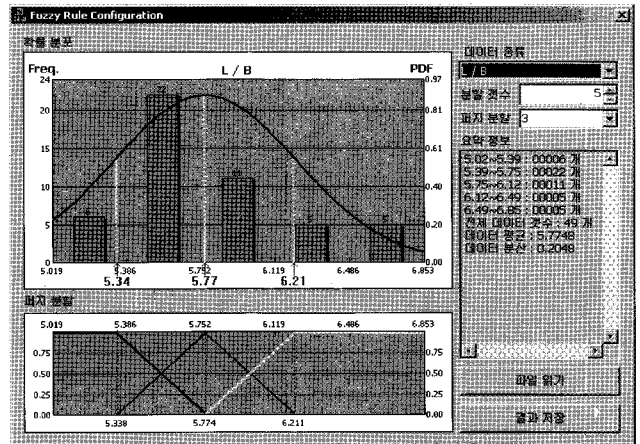


Fig. 3 The probability density function and Fuzzy partitions of  $L/B$

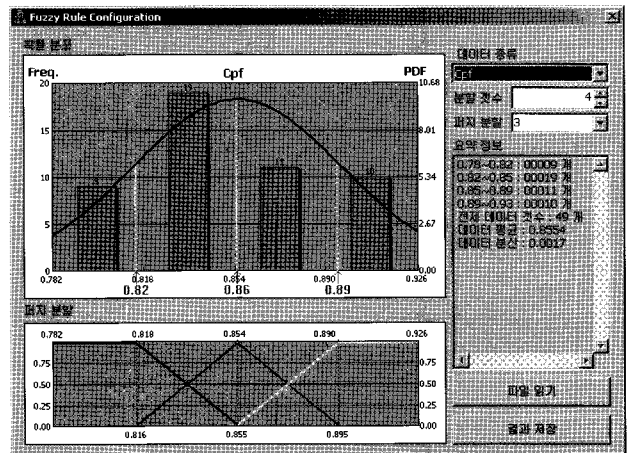


Fig. 4 The probability density function and Fuzzy partitions of  $C_{p_f}$

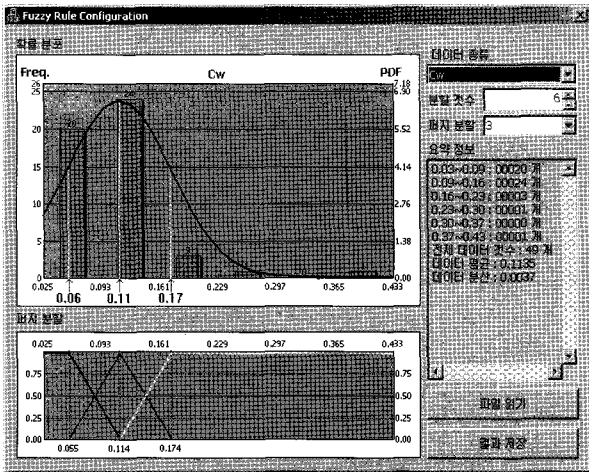


Fig. 5 The probability density function and Fuzzy partitions of  $C_w$

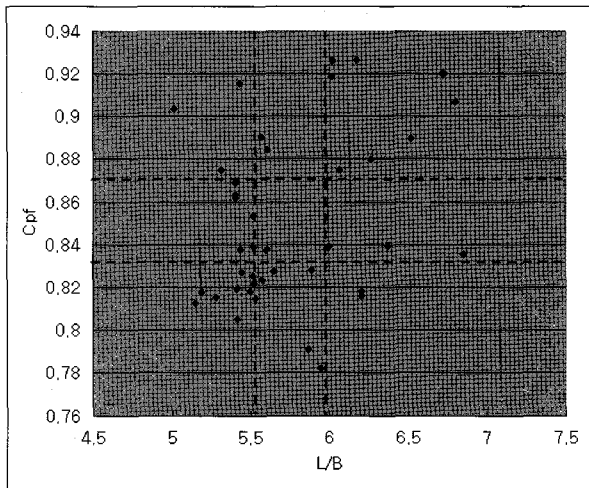


Fig. 6 The input space of the premise

Table 1 Consistent measures of the consequences

L/B	$C_{p_f}$	$C_w$			Select consequences
		Small	Middle	Large	
Small	Small	1/12	8/12	3/12	Middle
Small	Middle	4/6	1/6	1/6	Small
Small	Large	0/3	3/3	0/3	Middle
Middle	Small	0/5	5/5	0/5	Middle
Middle	Middle	0/1	0/1	1/1	Large
Middle	Large	2/4	1/4	1/4	Small
Large	Small	2/2	0/2	0/2	Small
Large	Middle	4/5	1/5	0/5	Small
Large	Big	4/11	5/11	2/11	Middle

Table 2 Derived Fuzzy rules

Derived Fuzzy rules	Consistent measure
If $L/B$ is small and $C_{p_f}$ is small, then $C_w$ is middle.	8/12
If $L/B$ is small and $C_{p_f}$ is middle, then $C_w$ is small.	4/6
If $L/B$ is middle and $C_{p_f}$ is small, then $C_w$ is middle.	5/5
If $L/B$ is large and $C_{p_f}$ is middle, then $C_w$ is small	4/5

### 6. 결 론

퍼지 균등화 이론과 Rough Set 일관성 척도를 이용한 자동 규칙 추출 기법을 사용하여 언어적 표현의 전문가 지식을 모델링 하였다. 전체부 언어변수의 분할의 수에 따라 총 9개의 규칙이 생성 가능하였으나, 일관성 척도에 적합한 4개의 규칙이 최종적으로 선택되었다.

본 연구에서 제안한 방법의 특징은 다음과 같다.

- (1) 추출된 규칙의 언어변수의 분할은 데이터가 밀집한 곳에서는 분할이 잘게 나누어지고 데이터 분포가 적은 곳에서는 분할이 크게 나누어져 우리의 직관과 유사함을 보이고 있다.
- (2) 생성된 규칙에 대하여 데이터에 근거한 수치적 평가가 가능하다.
- (3) 일련의 과정을 통하여 사전 지식이 없는 상태 즉, 주관을 배제한 상태에서도 데이터로부터 체계적으로 추론 규칙을 추출할 수 있다.

### 후 기

이 논문은 2006학년도 동명대학교 학술연구비 지원에 의하여 이루어진 것임을 밝히며, 연구비 지원에 감사드립니다.

### 참 고 문 헌

곽근창, 김승성, 유정웅, 전명근 (2001). "퍼지균등화와 유전알고리즘에 의한 자동적인 퍼지 규칙 생성", Proceedings of KFIS 2001 Spring Conference, pp 121-125.

김화수, 김세점, 조동래, 김용수 (1999). "전문가 시스템 개발을 위한 체계적인 규칙추출 프로세스 방안", 한국지능정보시스템학회 추계학술대회 - 지능형 정보기술과 미래조직 Information Technology and Future Organization, pp 79-88.

조영환, 노홍식, 위성윤. 이희진, 박민용 (1996). "Rough

- Set을 이용한 퍼지 규칙의 생성", Proceedings of KFIS Fall Conference 1996, pp 327-332.
- Pawlak, Z. (1992). Rough Sets: Theoretical Aspects of Reasoning About Data, Kluwer Academic Publishers
- Pedrycz, W. (2001). "Fuzzy equalization in the construction of fuzzy sets", Fuzzy Sets and Systems, Vol 119, pp 329-335.
- Zadeh, L. (1965). "Fuzzy sets", Information and control., Vol 8, pp 338-353.
- 
- 2007년 9월 3일 원고 접수  
2007년 12월 10일 최종 수정본 채택