

퍼지회귀분석을 이용한 프로젝트 성과예측

Estimation of Project Performance Using Fuzzy Linear Regression

박영만* · 박광박**

Young-Man Park* and Kwang-Bak Park**

* 경남대학교 경영학부 교수

** 경남대학교 응용수리학부 교수

Division of Business Administration, Kyungnam University, Korea

요 약

퍼지회귀분석은 독립변수들과 종속변수간의 관계를 평가하는데 사용된다. 만약 언어적 표현으로 된 자료를 처리할 때 일반적인 회귀분석을 사용한다면 과도한 단순화 때문에 어느 정도 한계를 가진다. 본 논문에서는 프로젝트의 성과를 예측하기 위해 퍼지 입출력을 갖는 퍼지회귀분석을 사용한다.

Abstract

Fuzzy regression model is used in evaluating relationship between the dependent and independent variables. If linguistic data are obtained, ordinary regression have limitation due to oversimplification of data. In this paper, fuzzy regression model with fuzzy input-output data for estimation of project performance is used.

Key Words : Fuzzy Regression Model, Project Performance

1. 서 론

회귀분석은 다양한 분야에서 유용하게 사용되는 통계분석 방법 중 하나이다. 보통 불확실성을 가지고 있는 자료들을 수집하여 각 자료간의 관계를 수립하고 예측할 수 있는 식을 만들어 준다.

통계 프로그램을 이용하여 회귀분석을 수행하는 경우 자료 속에 정성적인 자료나 주관적인 평가 자료가 포함되어 있는 경우에는 이들 자료를 빼고 처리하거나 이 자료를 포함하여 처리하고자 할 때는 이 정성적인 자료를 수치화된 자료로 변환해야 처리가 가능하다. 어떤 수준을 상중하로 반응한 경우에 이 자료를 1, 2, 3이라는 수치로 변환한다면 이 변환에 의해 중요한 정보를 잃어버릴 수도 있다. 이럴 경우 퍼지회귀분석은 하나의 해결책으로 사용되어 질 수 있다.

일반적인 회귀분석은 회귀식에서 추정된 값과 실제 관측 값의 차이인 잔차의 제곱합을 최소화하는 회귀계수를 결정한다. 이때 잔차는 평균 0이고, 동일 분산의 정규분포를 따른다고 가정된다. 그러나 퍼지회귀분석은 회귀 추정치와 관측치가 시스템의 애매성을 갖는 퍼지수로 취급한다. 그렇기 때문에 정보의 누수는 어느 정도 방지하지만 일반 회귀분석에서의 잔차와 비슷한 개념은 여러가지가 있기에 퍼지회귀 모형 간의 비교는 어려울 수 있다.

퍼지회귀모형은 추정 퍼지수의 폭을 최소화하는 방법과 추정 퍼지수와와의 거리를 최소화하는 방법으로 대별할 수 있

다. Tanaka et al.[1]에 의해 퍼지회귀분석 모형이 제시되었고 그들은 모든 관측 자료를 모두 포함할 수 있는 추정 영역의 폭의 합을 최소화하는 선형계획법의 문제로 회귀계수를 구했다. 즉 그 회귀모형의 fuzziness를 최소화하는 회귀계수를 선택했다. Diamond & Korner[2], Chang & Ayyub[3], Kim & Bishu[4] 등은 관측과 추정 퍼지수와의 거리 또는 차이를 최소화하는 회귀계수를 구했다. 이럴 경우 모형은 비선형문제로 변환되고, 최적화 문제로 해결하거나 신경망이나 유전적기법 등을 이용하여 회귀모형 식을 구성한다.

본 연구에서는 실제의 업무에 종사하고 있는 사무/현장직의 회사원을 대상으로 프로젝트의 성과를 투입 자원인 인력, 자금의 충분성과 경영자의 관심 정도, 프로젝트의 성공에 따른 보상과의 관련성, 효과 등에 대해 프로젝트에 참여한 사람들의 주관적인 반응을 조사하고 이 자료를 이용하여 퍼지회귀식을 형성하여 이들 간의 관계를 분석하고자 한다. 퍼지회귀식의 형성은 비퍼지 계수를 갖는 퍼지 입출력 모형을 이용했다.

2. 퍼지입출력 회귀모형

독립변수 $X=(X_1, X_2, \dots, X_n)$ 와 종속변수 Y 가 모두 퍼지수인 다음과 같은 퍼지모형을 고려해 보자.

$$Y_i = b_1 X_{i1} + b_2 X_{i2} + \dots + b_n X_{in}$$

단 여기서 b_j 는 회귀계수로 실수이다. 퍼지수 Y 와 X 는 좌우 같은 폭을 갖는 삼각함수로 가정하고 $Y_i=(y_i, e_i)$, $X_{ij}=(m_{ij}, c_{ij})$ 로 표현한다. 여기서 y_i, m_{ij} 는 중심이고 e_i, c_{ij} 는 폭을 의미한다. 위의 회귀식에서 만들어지는 추정치를

접수일자 : 2008년 2월 19일

완료일자 : 2008년 9월 3일

본 연구는 2008학년도 경남대학교 학술연구장려금 지원으로 이루어졌음.

$\hat{Y}_i=(\hat{y}_i, \hat{\epsilon}_i)$ 라면 추정 중심 $\hat{y}_i = \sum_j b_j m_{ij}$, 이고 추정 폭 $\hat{\epsilon}_i = \sum_j |b_j| c_{ij}$ 이 된다.

이 회귀식을 풀다는 것은 관측치 Y_i 와 추정치 \hat{Y}_i 간의 오차를 최소화 시키는 계수 b_j 를 구하는 것이다. 독립변수와 종속변수가 모두 실수라면, 잔차 제곱합 $\sum(Y_i - \hat{Y}_i)^2$ 을 최소화하는 계수를 구한다. 그러나 관측치와 추정치가 퍼지수로 표현되는 경우는 여러 연구들이 여러 가지 형성 기준을 제시하고 있다.

2.1 퍼지선형 회귀모형(Fuzzy Linear Regression Model)

Tanaka et al.[1]은 퍼지시스템에서 최초로 선형회귀분석 모형을 제시했다. 이 모형은 선형계획법을 이용하여 회귀계수를 구한다. 그들이 제시한 회귀모형은 회귀계수가 퍼지수이고, 입출력 변수가 실수형 일 때이지만 퍼지입출력 모형에도 적용 가능하다. 추정치의 하한, 상한의 범위 내에 관측치가 포함돼야 한다는 조건하에서 추정치 폭의 합을 최소화하는 회귀계수를 아래의 선형계획법 문제를 해결함으로써 구할 수 있다.

$$\begin{aligned} \text{Min } S &= \sum \hat{\epsilon}_i \\ \text{s.t } \hat{y}_i - (1-h) \hat{\epsilon}_i &\leq y_i - (1-h) \epsilon_i \\ \hat{y}_i + (1-h) \hat{\epsilon}_i &\geq y_i + (1-h) \epsilon_i \end{aligned}$$

이때 h 는 0와 1사이의 조정가능한 수로 자료와 회귀모형의 적합성의 척도로 사용된다. h 수준에서 Fig.1처럼 추정치는 관측치를 포함해야 한다.

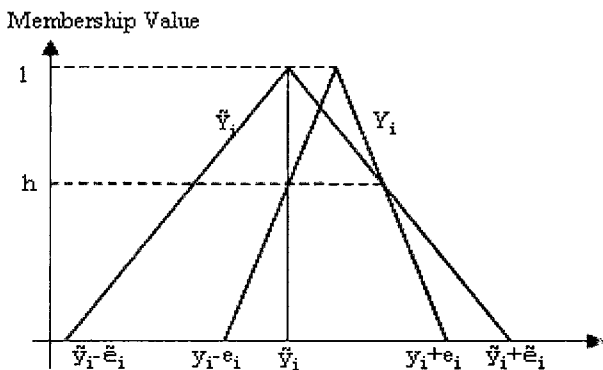


그림 1. 퍼지선형회귀에서의 Y_i, \hat{Y}_i, h
Fig.1 Relation of Y_i, \hat{Y}_i and h in fuzzy linear regression model

2.2 퍼지 최소자승 회귀모형(Fuzzy Least Square Regression Model)

실수형 자료처럼 관측치와 추정치 사이의 오차 제곱을 최소화하는 회귀계수를 구하는 모형이다. 이 오차를 어떻게 보느냐에 따라 회귀계수를 구하는 방법이 달라진다.

(a) 관측치와 추정치간의 거리를 다음과 같이 정의한다.

$$d(Y_i, \hat{Y}_i)^2 = (y_i - \hat{y}_i)^2 + ((y_i + \epsilon_i) - (\hat{y}_i + \hat{\epsilon}_i))^2 + ((y_i - \epsilon_i) - (\hat{y}_i - \hat{\epsilon}_i))^2$$

P. Diamond[5], Yang & Lin[6] 등이 이 거리 합을 최소화하는 계수를 구했다.

(b) Celmins[7]는 자료와 회귀모형간의 적합도

(compatibility measure) $r(Y_i, \hat{Y}_i)$ 를 기초로 최소자승 회귀모형을 제시했다. 적합도는 다음과 같이 두 퍼지수간의 멤버십 함수의 maximin 값으로 정의된다.

$$r(Y_i, \hat{Y}_i) = \max \min \{ \mu_{Y_i}, \mu_{\hat{Y}_i} \}$$

적합도 r 은 0에서 1사이의 값이며 Fig. 2처럼 유사한 퍼지수 간에는 1에 가깝고, 둘간의 차이가 커지면 0에 가까워지며 전혀 공통 영역이 없는 경우는 0값을 갖는다. 중심점이 일치한다면 r 은 1이 된다. Celmins[7]는 비적합도 제곱합을 최소로 하는 즉 $\text{Min } W = \sum (1-r(Y_i, \hat{Y}_i))^2$ 을 목적함수로 두고 비선형계획법을 이용해 회귀계수를 구했다.

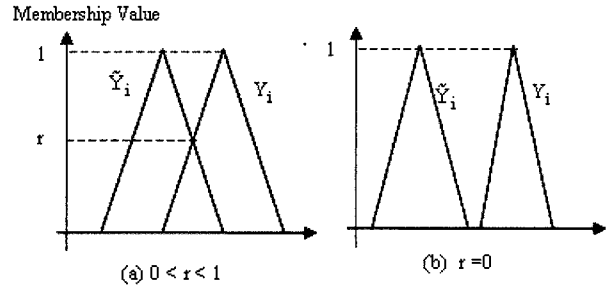


그림 2. 적합도 $r(Y_i, \hat{Y}_i)$
Fig. 2 compatibility measure $r(Y_i, \hat{Y}_i)$

(c) Savic & Pedrycz[8]는 퍼지수의 중심값을 이용해 퍼지회귀가 아닌 일반 회귀분석을 통해 중심값을 찾고, 폭은 중심값이 결정된 가운데 퍼지선형 모형의 제한조건하에서 최소 폭의 기준으로 찾는 혼합된 모형을 제시했다.

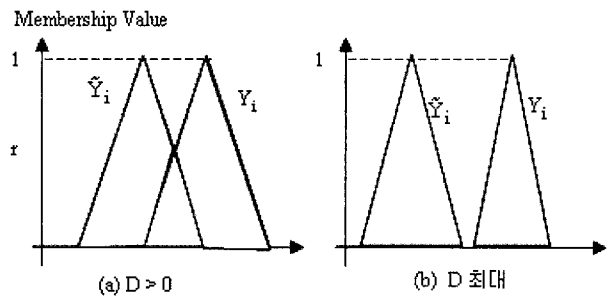


그림 3. Y_i, \hat{Y}_i 의 멤버십함수 간 Difference
Fig. 3 Difference between membership functions of Y_i and \hat{Y}_i

(d) Kim & Bishu[4]는 두 퍼지수의 멤버십 함수를 비교하는 척도를 제시했다. 두 삼각퍼지수의 차이 합 D 는 다음 식으로 정의된다.

$$D = \int_{S_Y \cup S_{\hat{Y}}} |\mu_Y(x) - \mu_{\hat{Y}}(x)| dx$$

여기서 $S_Y, S_{\hat{Y}}$ 는 각각의 멤버십함수의 support이다. D 는 그림 3에 표시된 부분의 넓이가 된다. 그림 3에서 보는 것처럼 멀어질수록 퍼지수 간에는 차이합 D 가 커지고, 두 퍼지수 간 겹치는 부분이 없다면 최대값을 갖는다. 중심과 폭이 완전히 일치될 때만 0값을 갖는다. 추정 퍼지수는 관측 퍼지수에 가까운 것으로 기대되기 때문에 회귀 오차 E 는 관측치 멤버십 함수에 대한 추정 멤버십함수 값과의

차이의 비 E로 나타낼 수 있다[4].

$$E = D / \int_{S_Y} \mu_Y(x) dx$$

Min $\sum_i E_i$ 를 목적함수로 두고 회귀 계수를 구할 수 있고, 모형의 유효성을 밝히는 지표로도 사용한다.

3. 사례 적용

경남에 위치한 중소기업의 관리/기능직 사원을 대상으로 본인이 경험한 프로젝트의 성공, 실패에 관한 설문을 110명을 대상으로 조사했다.

설문에는 프로젝트의 투입 자원인 인력과 자금, 기간의 충분성, 프로젝트 달성의 어려움, 달성의 효과, 경영진의 관심의 정도, 조직내부 관심도, 참여자의 전문성, 프로젝트의 경험, 적극성, 프로젝트의 성공/실패에 따른 보상/질책과의 관련성, 프로젝트의 기대 효과의 크기, 기타 내부 지원의 충분성, 프로젝트 성과 정도 등에 대해 5점 척도로 물었다. 퍼지회귀분석에 사용할 자료는 그 중 응답 누락 등의 이유로 16명을 제외한 94명의 자료를 대상으로 실시한다.

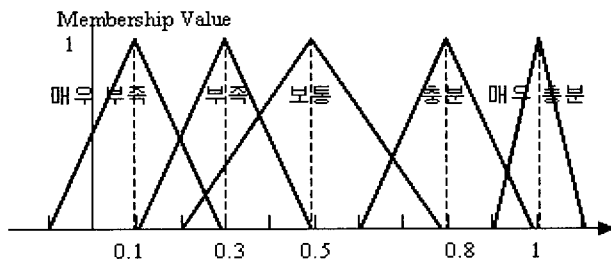


그림 4. 설문 반응에 대한 멤버십 함수
Fig. 4 Membership functions of response of questionnaire

각 항목에 대한 응답을 그림 4와 같이 삼각 퍼지수로 변환하여 퍼지회귀분석을 실시했다. 중앙값만을 이용하여 프로젝트 성과에 대해 일반적인 회귀분석을 SPSS를 이용하여 수행했다.

표 1. 일반 회귀분석의 결과

Table 1. Results of ordinary linear regression

	회귀 계수	표준화계수	t	유의확률
(상수)	-0.003		-0.028	.977
인력의 충분성	0.295	0.256	2.743	.007
자금의 충분성	0.243	0.198	2.185	.032
경영자 관심	0.333	0.257	2.681	.009
성공 보상	-0.239	-0.234	-2.695	.008
기대 효과	0.239	0.197	2.019	.047

SPSS 회귀분석의 진입법으로 처리한 결과 표 1에 나와 있는 5개의 독립 변수들만이 선택되어 차후 퍼지회귀분석에서도 이 5개의 변수만을 이용하기로 했다. 5개의 독립변수들로 SPSS를 이용하여 처리한 결과는 표 1에 나와 있다. 이때 설명력 $R^2 = 0.354$ 이고 잔차제곱합은 3.795였다.

종속변수로 사용할 프로젝트 성과와 5개 독립변수에 대한 설문 응답의 자료가 표 2에 나타나 있다.

표 2. 설문 결과 자료

Table 2. Responses of Questionnaire

no.	성	과	인	력	자	금	관	심	보	상	효	과	no.	성	과	인	력	자	금	관	심	보	상	효	과
1	4	4	3	4	3	5	48	4	2	3	4	2	4												
2	4	3	4	5	3	3	49	2	2	2	3	3	4												
3	3	3	3	2	1	4	50	2	4	2	3	2	4												
4	4	3	3	3	1	4	51	2	3	1	4	2	4												
5	4	3	3	4	4	4	52	3	4	3	3	3	3												
6	3	2	3	4	2	4	53	4	3	3	4	3	5												
7	4	4	3	5	2	4	54	5	4	4	4	3	3												
8	4	3	4	5	1	2	55	3	4	4	3	3	3												
9	5	4	4	4	2	4	56	2	3	3	3	1	4												
10	2	2	2	4	2	4	57	4	4	3	3	4	5												
11	2	4	2	4	3	2	58	4	4	3	4	2	5												
12	3	2	2	5	1	4	59	3	3	3	4	2	4												
13	3	3	3	4	2	4	60	1	3	1	1	1	5												
14	3	3	2	5	3	3	61	3	3	2	3	3	3												
15	5	5	3	4	3	4	62	5	5	1	4	2	4												
16	3	3	2	3	3	3	63	3	4	3	3	3	3												
17	2	2	2	3	3	3	64	4	4	3	4	1	5												
18	2	4	2	4	2	4	65	4	2	2	4	3	3												
19	2	4	2	3	3	3	66	4	4	3	3	2	3												
20	5	4	3	3	1	4	67	4	4	4	4	2	5												
21	4	3	1	4	1	3	68	3	2	3	5	5	1												
22	4	3	2	4	3	3	69	4	2	2	4	2	1												
23	3	4	4	4	1	5	70	4	4	1	3	1	5												
24	3	3	3	4	3	3	71	3	4	3	4	4	2												
25	2	4	3	4	1	5	72	2	3	3	3	3	3												
26	3	4	2	3	1	4	73	5	4	4	5	5	1												
27	2	3	3	4	4	5	74	4	5	5	5	2	4												
28	2	2	3	3	1	5	75	2	3	4	3	4	5												
29	2	2	4	3	1	5	76	4	5	3	3	3	3												
30	3	4	3	4	4	4	77	1	4	1	4	4	2												
31	3	1	3	4	3	3	78	3	2	3	4	2	5												
32	4	4	3	4	3	2	79	3	4	3	4	3	4												
33	2	4	2	3	4	4	80	4	3	3	3	3	5												
34	1	3	1	4	4	2	81	4	4	4	4	2	5												
35	1	4	2	2	5	1	82	3	3	3	4	3	3												
36	4	3	3	4	4	4	83	2	3	2	2	2	5												
37	2	3	3	3	3	3	84	4	3	3	4	1	4												
38	5	4	3	3	3	5	85	3	4	3	4	3	3												
39	4	4	4	4	3	5	86	4	3	1	4	1	4												
40	3	2	3	4	2	5	87	4	3	4	5	4	2												
41	4	4	3	4	3	3	88	3	3	2	3	2	4												
42	5	5	3	3	1	5	89	2	1	3	5	5	2												
43	4	4	4	4	1	4	90	2	3	4	4	3	3												
44	2	2	3	3	3	3	91	3	3	3	5	3	3												
45	3	3	2	3	3	3	92	4	3	1	4	1	4												
46	5	3	4	4	3	4	93	3	2	3	4	3	0												
47	3	3	3	3	3	3	94	4	3	3	4	1	4												

관측치와 추정치 간의 거리를 최소화하는 퍼지 최소자승 회귀모형(2.2 a)은 2차식을 최소화하는 문제로 Excel의 해 찾기 기능을 이용하여 구했다. 그리고 비적합도 제곱의 합을 최소로 하는 모형(2.2 b)에 대한 해는 정형적인 비선형 문제가 아니기 때문에 전역최적해(global solution)을 구하기가 쉽다. 이 모형은 수리계획법에서 계산 효율이 높지 않지만 가장 간단한 탐색법인 순환법(Cyclic Coordinate Method)과 일양탐색법(Uniform search)를 사용했다. 즉 각 변수를 순서적으로 한 변수씩 탐색해 나가는 방법을 사용하여 목적함수를 최소화 했다.

이 두 가지 회귀모형에 대한 결과는 표 3에 나와 있다. 최소거리 모형(2.2 a)의 계수보다 적합도 모형(2.2 b)의 회귀계수가 목표달성의 효과를 제외하고는 같거나 크게 나왔다. 이 결과를 보면 경영자의 관심과 인력의 충분성이 프로젝트의 성과에 다른 변수에 비해 큰 영향을 미친다. 인력이 부족하고, 경영자의 관심도가 적으면 프로젝트의 성과가 나오기 어렵고, 그리고 프로젝트가 성공 보상과의 관련성이 높으면 오히려 성과가 나빠지는 경향을 보인다. 인력, 자금 등 모든 조건이 최상이었을 때 최소거리모형의 추정치는 (0.893, 0.109)였고, 적합도 모형은 (1.656, 0.206)이 된다. 만약 모든 조건이 최악의 경우는 최소거리모형은 (0.089, 0.219), 적합도 모형은 (0.166, 0.412)로 예측한다.

최소거리모형의 94개 자료에 대한 거리 합은 12.667이었고, 적합도를 고려한 기준에는 47.751로 그 기준을 사용한 방법에 비해서는 약간 나빠지만 추정치와 관측치의 중심치의 제곱합은 3.918로 낮은 값을 보였다. 두 방법에 대한 Kim & Bishu[4]의 E값의 합은 각각 177.41, 269.68로 최소거리 모형이 좋은 결과를 보였다.

표 3. 회귀계수 및 결과

독립 변수	최소거리 모형	적합도 모형
인력의 충분성	0.272	0.612
자금의 충분성	0.191	0.605
경영자 관심	0.285	0.642
성공 보상	-0.100	0.100
기대 효과	0.245	0.102
$\sum d(Y_i, \hat{Y}_i)^2$	12.687	75.944
$\sum (1-r(Y_i, \hat{Y}_i))^2$	47.751	33.784
$\sum (y_i - \hat{y}_i)^2$	3.918	20.636

일부분의 자료(1-20번 자료)에 대한 관측치와 추정치의 구간에 대해 그림 5에서 보여주고 있다. 적합도를 이용한 모형의 구간이 관측치를 많이 포함하고 있지만 전반적으로 구간의 크기가 크다는 것을 알 수 있다. 점선으로 표시된 최소거리 모형이 관측치와 비슷한 구간을 점유하고 있음을 볼 수 있다.

이 프로젝트에 관한 자료를 이용하여 퍼지선형회귀 방법(2.1모형)을 적용해 보았으나 좋은 결과를 얻지 못했다. 원래의 문제와 쌍대문제로 변환시켜 처리했지만 제한을 만족하는 최적해를 구하지 못했다. 자료의 수가 적은 경우에는 그런대로 좋은 결과를 보이는 방법인데도 이 사례에서는 관측치를 포함하는 회귀식을 만들지 못했다. 자료의 갯수가 많기에 그들 간에 모순이 발생하여 선형계획법에서 해를 만들지 못했다고 생각된다. 또한 추정치와 관측치 간의 멤버

십 함수값의 차이를 이용하는 모형(2.2 d모형) 또한 회귀계수가 모두 0인 경우에 목적함수가 최소화 되어 의미있는 회귀식을 만들지 못했다.

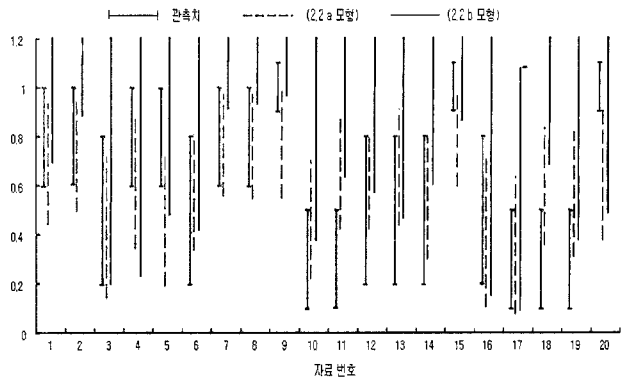


그림 5. 퍼지회귀분석의 결과 예

Fig. 5 An Example of interval of estimated fuzzy number by fuzzy regression

4. 결 론

실제의 업무에 종사하고 있는 사무/현장직의 회사를 대상으로 프로젝트의 성과를 투입 자원인 인력, 자금의 충분성과 경영자의 관심 정도, 프로젝트의 성공에 따른 보상과의 관련성, 기대 효과 등을 조사하여 이들 간의 관계에 대해 실수 계수를 갖는 퍼지 입출력 회귀식으로 추정했다.

프로젝트의 성과는 경영자의 관심과 인력의 충분성이 프로젝트의 성과에 다른 변수에 비해 큰 영향을 미친다. 인력이 부족하고, 경영자의 관심도가 적으면 프로젝트의 성과가 나오기 어렵고, 그리고 프로젝트가 성공 보상과의 관련성이 높으면 오히려 성과가 나빠지는 경향을 보였다.

퍼지회귀분석의 경우는 일반 선형회귀분석과는 달리 명쾌한 기준이 없기에 어느 방법이 가장 우수했는가를 판별하기가 어렵다. 여러 가지 기준으로 비교해 본 결과 이 사례에서는 최소거리 모형이 가장 좋은 결과를 보여 주었다. 또한 몇 가지 퍼지 회귀모형의 결과가 이 사례에서 의미없는 결과를 보여주기도 했다.

참 고 문 헌

- [1] H. Tanaka, S. Uejima, and K. Asai, "Linear Regression Analysis with Fuzzy Model", *IEEE Trans. Systems, Man and Cybernetics*, 12, 1982, 903-907
- [2] P. Diamond and R. K. Korner, "Extended Fuzzy Linear Models and Least-Squares Estimates", *Journal of Computational and Applied Mathematics*, 9, 1997, 15-32
- [3] Y. O. Chang and B. M. Ayyub, "Fuzzy Regression Methods - a Comparative Assessment", *Fuzzy Sets and Systems*, 119, 2001, 187-203.
- [4] B. Kim and R.R.Bishu, "Evaluation of Fuzzy

Linear Regression Models by Comparing Membership Functions", *Fuzzy Sets and Systems*, 100, 343-352

- [5] P. Diamond, "Correlation-based Feature Selection of Discrete and Numeric Class Machine Learning", *Proceedings of the International Conference on Machine Learning*, 2000, 359-366.
 - [6] M. S. Yang and T. S. Lin, "Fuzzy Least-Squares Linear Regression Analysis for Fuzzy Input-Output Data", *Fuzzy Sets and Systems*, 126, 2002, 389-399.
 - [7] A. Celminš, "Least Square Model Fitting to Fuzzy Vector Data", *Fuzzy Sets and Systems*, 32, 1987, 245-269
 - [8] D. Savic and W. Pedrycz, "Evaluation of Fuzzy Regression Models", *Fuzzy Sets and Systems*, 39, 1991, 51-63
-

저 자 소 개



박영만(Young-Man Park)

1980 : 서울대학교 산업공학 학사
1982 : 서울대학교 산업공학 석사
1999 : 일본 동아대학교 정보시스템 박사
현재 : 경남대학교 경영학부 교수

관심분야 : 경영과학, 의사결정
Phone : +82-55-249-2704
Fax : +82-55-223-1655
E-Mail : youngman@kyungnam.ac.kr



박광박(Kwang-Pak Park)

1969년 : 부산대학교 수학 학사
1977년 : 부산대학교 수학 석사
1986년 : 경상대학교 수학 박사
현재 : 경남대학교 응용수리학부 교수

관심분야 : 함수해석학
Phone : +82-55-249-2207
Fax : +82-55-244-6504
E-mail : kppark@kyungnam.ac.kr