

결측값이 포함된 실험자료의 민감도 해석

Sensitivity Analysis of Experimental Data with Missing Value

*백석¹, #홍순혁², 김창기³, 정유엽³, 장득열⁴, 주원식⁵

*S. H. Baek¹, #S. H. Hong(shhong@pknu.ac.kr)², C. K. Kim³, Y. Y. Jeong³, D. Y. Jang⁴, W. S. Joo⁵

¹동아대학교 대학원 기계공학과, ²부경대학교 산업과학기술연구소, ³(주)동아이엔씨, ⁴강원대학교 기계공학과, ⁵동아대학교 기계공학과

Key words : Missing Value(결측값), Taguchi Method(다구찌 방법), Signal to Noise Ratio(SN비), Successive Approximation(축차근사법)

1. 서론

실험계획법(design of experiment)에 따라 수행한 실험이나 모사(simulation) 결과의 자료가 예정대로 얻어지지 않는 경우가 있다. 예를 들어, 실험중단자료(censored data), 구간자료(grouped data), 실험장치 고장 등과 모사의 불확실성(uncertainty) 문제가 발생한 경우이다. 최근 비용과 효율성으로 인해 실험계획법을 이용한 민감도 해석, 최적화를 수행하는 경우가 많아지는데 이때 자료에 결측값(missing value)¹이 있는 경우 해석의 어려움이 있다.

본 논문은 해석 자료에 결측값이 있는 경우 처리 방법을 제안한다. 결측값의 정의는 얻어야 할 자료가 실험중의 문제로 얻어지지 않는 것을 의미한다. 그리고 자료가 얻어지더라도 오차가 지나치게 커서 실험 자료로 사용 하지 못하는 경우도 포함된다.

결측값 처리의 개념은 다구찌 방법(Taguchi method)에 기인한다. 다구찌 방법의 두 가지 중요한 수단은 응답을 평가하는 직교배열표(orthogonal array)와 신호 대 잡음비(signal-to-noise ratio, SN비)의 이용이다.²

직교배열표는 다른 인자가 변할 때 제어인자의 평균효과와 교호작용의 유무를 확인하기 위해서 실험계획법에 의한 설계에 이용되었다. SN비는 제어인자의 효과들에 대한 가법성(additivity)을 개선하기 위해 일반적으로 평균과 표준편차의 비 μ^2/σ^2 에 log를 취하는 형태로 표현한다.

$$\eta = 10 \log_{10} \left(\frac{\mu^2}{\sigma^2} \right) \quad (1)$$

설계에서 식 (1)의 의미는 입출력의 정확한 인과 관계를 구하기 위한 것이 아니라 전출력의 효과와 나머지 효과로 분해하여 두 효과의 비를 설계 평가에 이용하기 위한 것이다. 이 SN비는 산출 응답에 예측할 수 없는 요소의 효과를 감소시킨다. 따라서 응답의 평균과 무관하게 최적화를 이룰 수 있고, 이 점은 결측값에 따라 목표값이 달라지는 경우에도 최적조건은 타당성을 지닐 가능성이 많아진다.

적용 예제는 밀스케일(mill scale)을 이용한 환원철(direct reduction iron, DRI) 제조공정 최적화 문제이다. 공정변수 최적화를 위한 공정 실험은 고가이고 직교배열표가 제시하는 최소한의 실험마저도 불가능한 경우가 많아 결측값이 존재한다. 결측값은 제안된 SN비 근방의 관찰 자료(observational data)^{2,3}를 이용하여 효과가 큰 공정변수의 민감도를 제시하고 다구찌 재현성 실험을 통해 결측값 처리의 타당성을 검증하였다.

2. 축차근사법

종래의 실험계획법은 출력 y에 대한 여러 설계변수(제어인자, 신호인자, 오차인자)를 포함)의 영향을 해석하는 방법이다. 이 방법은 일반적으로 반응표면(response surface)을 구하는 것이 목적이므로 어떻게 하면 되도록 정확한 반응표면을 얻을 수 있는지만이 문제이다. 이와 반대로, 제어변수의 효과를 유지하면서 오차인자의 종합적인 영향을 최소로 하고자 하는 손실함수(loss function)의 개념은 반응표면과 비교해서 크게 다르다.

SN비를 구하기 위해 신호인자와 오차인자를 할당해서 분산분석(variance of analysis)을 하는 것은 정확한 인과관계를 구하기 위한 것이 아니라 전출력의 효과와 나머지 효과로 분해하여 두 효과의 비를 설계 평가에 이용하기 위한 것이며, 이것은 종래의 실험계획법과 크게 다른 점이다.

결측값이 발생한 경우 결측값을 추정에는 Fisher-Yates 방법

Table 1 Process parameters and their levels

Process parameter	Description	Level 1	Level 2	Level 3
X1	C content	20%	25%	30%
X2	Thickness	10 mm	15 mm	20 mm
X3	Granularity	1 mm	2 mm	3 mm
X4	Heat treatment temperature	1100℃	1200℃	1300℃

Table 2 Results of L18 array with missing data

Exp.	X1	X2	X3	X4	Metallization (%)	SN ratio
1	1	1	1	1	80.25	38.089
2	2	2	2	2	92.01	39.277
3	3	3	3	3	91.13	39.193
4	2	2	3	3	87.21	38.811
5	3	3	1	1		
6	1	1	2	2	61.67	35.801
7	1	3	2	3	95.47	39.597
8	2	1	3	1	96.11	37.129
9	3	2	1	2	96.85	39.722
10	3	2	2	1	95.13	39.566
11	1	3	3	2	95.52	39.602
12	2	1	1	3	85.8	38.670
13	3	1	3	2	88.52	38.941
14	1	2	1	3	88.34	38.923
15	2	3	2	1	64.85	36.238
16	2	3	1	2	69.12	38.625
17	3	1	2	3		
18	1	2	3	1	57.45	32.104

등이 있는데 본 논문은 SN비 기반한 축차근사법(successive approximation)을 제안한다. 축차근사법의 순서는 다음과 같다.

[단계1] 결측 근방에 SN비 평균값을 대입하여 이를 제 0차 근사값으로 한다.

[단계2] 결측 이외의 큰 효과의 요인을 사용하여 결측 근방을 SN비를 추정하고 이를 제 1차 근사값으로 한다.

[단계3] 결측 추정값이 수렴하기까지 [단계2]를 반복한다.

3. 결측값 해석 및 요약

3.1 실험결과와 가정

본 논문의 밀스케일 환원특성에 대한 실험결과는 참고문헌⁴에 나타나 있다. Table 1은 4개의 공정변수(X1=고정탄소량, X2=성형 두께, X3=입도, X4=열처리 온도)와 3 수준 범위를 나타낸다. Table 2는 결측값이 포함된 L18 실험결과이다.

이 실험은 대단히 고가이고 많은 시간이 소요되어 예정대로 실험 자료를 얻지 못하는 결측값이 No. 17에 나타났다. 결측값 추정에 대한 검증을 위해 No. 5의 실험자료(또는 SN비)를 얻을 수 없다고 가정한다. 또한, 실제로는 84.25%(SN비 38.511)이다.

3.2 결측값의 추정

결측이 발생한 경우 실험의 조합별 SN비에 상당한 산포가

Table 3 Analysis of variance

Variable	Dimension	DOF	Variance	F-ratio	Effective ratio(%)
X1	1	1	0.0062	0.03	0.035
	2	1	1.2525	1.7648	2.064
X2	1	1	0.7228	3.31	3.870
	2	1	1.2471	5.71	6.678
X3	1	1	0.6338	2.9	3.392
	2	1	0.2374	1.09	1.275
X4	1	1	9.1235	41.81	48.898
	2	1	4.9929	22.88	26.759
X1X2	1	1	0.9617	4.41	5.158
X1X4	1	1	0.35	1.6	1.871
Error		7	1.5275		
Total		17	21.0555		100

Table 4 Process parameters affecting the metallization

Process parameter	Effect			Deviation	Percentage contribution (%)	Rank
	Level 1	Level 2	Level 3			
X1	38.59	38.39	38.85	0.46	8.3	4
X2	39.72	38.18	39.45	1.54	27.8	2
X3	38.71	38.32	38.09	0.62	11.2	3
X4	36.02	38.89	38.94	2.92	52.7	1
Total				5.54	100	

있는 경우가 많다. 따라서 금속화율(metallization)에 대한 SN비에 효과가 있는 변수를 발견하는 것만을 목적으로 한다면 회귀분석으로 관계식을 구할 수 있다. 17개 자료의 SN비 평균은 38.165이다. 여기서 1차항과 2차항을 분해해서 고려한 분산분석은 Table 3과 같다. 일반적으로 변수 x 가 등간격일 때 2차항까지 고려한 Chebyshev 직교다항식⁵은 다음과 같다.

$$y = b_0 + b_1(x - \bar{x}) + b_2 \left[(x - \bar{x})^2 - \frac{k^2 - 1}{12} h^2 \right] + b_3 \left[(x - \bar{x})^3 - \frac{3a^2 - 7}{20} (x - \bar{x}) h^2 \right] \dots \quad (2)$$

여기서 b_0 는 응답 y 의 평균값, k 는 수준수, h 는 수준간의 간격이다. 분산분석에서 유의한 차수의 항을 선택하고 b_1, b_2, b_3 는 직교다항식의 계수표에 의해 계산하여 회귀식을 작성하면 식 (3)과 같다.

$$y = -218.54 + 47.56X_1 - 4.485X_1^2 - 3.546X_2 + 0.05895X_2^2 + 0.3239X_3 - 0.1161X_3^2 + 0.35649X_4 - 0.000114X_4^2 + 0.5884X_1X_2 - 0.02654X_1X_4 \quad (3)$$

Table 1의 No. 17의 실험조건은 X1(3 수준)X2(1 수준)X3(2 수준)X4(3 수준)을 식 (3)에 대입하면 37.805을 얻는다. 따라서 예측값의 1차 근사값은 37.805이다. 이 과정을 반복하여 4회 이후는 39.316로 수렴했으므로 No. 16의 예측값을 39.316로 한다.

Table 4는 구한 예측값을 대입하여 금속화율에 대한 공정변수의 수준별 SN비 주효과(main effect)를 계산한 것이다. SN비에 대한 주효과는 분산분석과 비교해서 정확한 인과관계를 구하기 위한 것이 아니라 반응값의 입력효과와 나머지 효과로 분해하여 두 효과의 비를 설계평가에 이용하기 위한 것이다. 예측값이 있을 경우 SN비에 대한 주효과는 공정변수의 설계 방향을 결정하는데 효과적인 정보를 얻을 수 있다.

금속화율의 평균 SN비는 38.538, 반응값은 85.62%이다. 공정변수의 최적조건은 고정탄소량이 30%(3 수준), 성형두께 10 mm(1 수준), 입도1 mm(1 수준), 열처리 온도 1300℃(3 수준)이다.

3.3 재현성 평가

재현성 평가는 선택된 공정변수로부터 기대되는 반응값을 예측하는 것이다.

SN비에 가장 좋은 효과를 주는 공정변수를 선택하고 다른

Table 5 Comparison between prediction and optimal results

Result	Optimal value	Predicted value	Error (%)
SN ratio	38.538	39.722	3.07
Mean	85.62	96.85	13.12

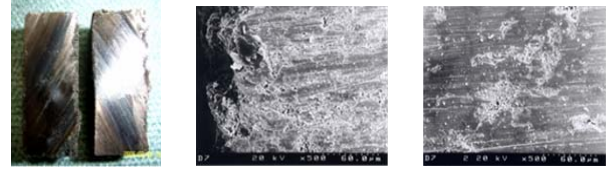


Fig. 1 SEM photographs for verification experiment result

공정변수의 몇몇 조합으로부터 SN비와 평균 반응을 예측한다. 그러므로 예측값 처리가 분석결과에 얼마나 영향을 미치는지 평가할 수 있다. 다음으로 선택된 수준으로 실험을 수행하여 예측된 결과와 일치되는지 확인한다.

만약 예측한 반응값과 실험값 사이에 불일치하면 예측값 및 교호작용, 순수 실험의 오차가 있다고 가정할 수 있다.

SN비 예측값을 μ 로 나타내면 식 (4)와 같다.

$$\mu = \bar{T} + (X_i^{Level} - \bar{T}) \quad (4)$$

여기서 \bar{T} 는 전체 평균, $X_i^{Level} - \bar{T}$ 는 각 변수의 최적수준에서 좋아지는 효과를 의미한다. 예측의 기본개념은 효과를 더해 나가는 것이다. 이를 변수 효과의 가법성이라고 한다. SN비는 단위 변화량 산포의 역수이다. 즉, 제곱의 형태로 나타내는 의미는 가법성 때문이다.

Table 5는 Table 4에서 얻은 금속화율의 SN비 결과를 이용해서 최적조건 값과 예측된 값을 비교한 것이다. 비교적 일치하므로 재현성이 있다고 판단된다. 얻은 조건으로 확인실험을 수행하여 실제 실험결과와 분석한다.

Fig 1은 재현성 평가 결과에 대한 확인실험을 한 경우로 주사전 자현미경(scanning electronic microscopy)을 이용하여 단면을 관찰한 사진이다. 직접환원철의 금속화율은 92%이다. 이것은 재현성 평가에서 예측된 96.85%와 비교하면 오차는 5.3%로 유의하다.

4. 결론

본 논문의 실험 결과는 직교배열표의 실험계획에 따라 예정대로 실험자료를 얻지 못하는 예측값이 존재한다. 이 경우는 정확한 통계량 추정이 가능하지 않다. 따라서 반응값의 입력효과와 나머지 효과로 분해하여 두 효과의 비를 설계평가에 이용하는 SN비와 측차근사법을 이용해서 공정변수의 민감도를 해석하고 최적수준을 결정하였다. 최적수준의 재현성은 오차 5.3% 이내로 일치하였다.

참고문헌

- Godolphin, J. D., "Reducing the Impact of Missing Values in Factorial Experiments Arranged in Blocks," Quality and Reliability Engineering International, Vol. 22, No. 6, pp. 669-682, 2006.
- Ramsey, F. L., and Schafer, D. W., The Statistical Sleuth: A Course in Methods of Data Analysis, Second Edition, Duxbury Press, 2001.
- Peace, G. S., Taguchi Methods, Addison-Wesley Publishing Company, 1993.
- 부경대학교 산업과학기술연구소, Mill Scale을 Lump상태로 성형하여 Pig Iron을 회수하기 위한 용융 전처리 기술개발 중간 보고서, 산업자원부 지역산업기술개발사업 지원과제(과제번호: 10024327), 2006.
- Baek, S. H., Cho, S. S., Kim, H. S., and Joo, W. S., "Trade-off Analysis in Multi-objective Optimization using Chebyshev Orthogonal Polynomials," Journal of Mechanical Science and Technology, Vol. 20, No. 3, pp. 366~375, 2006.