

## 論文

## 데이터 마이닝을 통한 공력설계공간 지식습득

정신규\*, 사사키 다이스케\*\*, 오바야시 시게루\*

Knowledge Discovery in Aerodynamic Design Space  
using Data Mining

Shinkyu Jeong\*, Sasaki Daisuke\*\* and Shigeru Obayashi\*

## ABSTRACT

Two data mining techniques, analysis of variance (ANOVA) and self-organizing map (SOM), are applied to knowledge discovery in aerodynamic design space. These methods make it possible to identify the effect of each design variable on the objective functions. Furthermore, ANOVA shows the effect of interaction between design variables on the objective function and SOM visualizes the trade-off among objective functions. Present methods are applied to the result of the supersonic wing design which includes 72 design variables and 4 objective functions.

## 초 록

본 연구에서는 공력설계공간의 지식습득을 위해 분산해석법과 자기조직화지도의 이용을 제안하였다. 이 기법들은 각각의 설계변수가 목적함수에 미치는 영향을 예측 가능하게 한다. 더욱이, 분산해석법은 설계변수들의 상호관계가 목적함수에 미치는 영향도 예측 가능하게 하며, 자기조직화지도는 목적함수들 사이에 어떠한 trade-off 관계가 있는지도 예측 가능하게 한다. 본 논문에서 72개의 설계변수와 4개의 목적함수를 가진 초음속 날개 설계의 결과에 대하여 이들의 데이터 마이닝 기법들을 적용하였다.

**Key Words** : Data Mining(데이터 마이닝), ANOVA(분산해석법), SOM(자기조직화지도)

## 1. 서 론

최근 계산기발달과 수치해석 기법의 발달에 힘입어 CFD를 이용한 최적 설계기법은 항공기의 설계 있어서 없어서는 안 될 하나의 수단이 되었다. 구배법을 이용하는 Adjoint method는 많은 설계변수를 다루는 문제에 대해 효율적으로 적용되어 왔고, 집단탐색법에 의거한 유전자 알고리즘은 복수의 목적함수를 다루는 설계문제에

대해 널리 적용되어 왔다.

하지만 지금까지의 최적화기법에 관한 대부분의 연구들은 최적점 혹은 비지배해들(non-dominated solutions)을 효율적이고 정확하게 찾아내는 방법에 관해 집중되어 왔다. 하지만 실제의 엔지니어링 디자인의 경우, 비지배해 같은 몇몇의 디자인 후보들 중에서 실제로 제품화될 최종 디자인을 선택하는 과정도 최적화만큼이나 중요한 과정이다. 그러므로 설계자가 단순히 다목적 최적화 문제의 비지배해만을 제공하는 것보다는 최종 디자인의 선택에 도움이 되는 정보들을 함께 제공하는 것이 더욱 더 이상적인 최적화 설계라고 할 수 있을 것이다. 목적함수들 간의 trade-off 관계, 목적함수와 설계변수들 사이의 상관관계 같은 설

† 2005년 7월 20일 접수 ~ 2005년 11월 14일 심사완료

\* 정회원, 일본 동북대학교 유체과학연구소

연락처, E-mail : jeong@edge.ifs.tohoku.ac.jp

2-1-1 Katahira, Aoba-ku, Sendai, Japan

\*\* 정회원, Southampton University

제공간에 관한 지식들이 최종 디자인 결정에 도움이 되는 정보들이라고도 말할 수 있을 것이다. 이러한 정보들로부터 좋은 디자인이 되기 위한 조건 등을 이해하거나 나아가 목적함수에 큰 영향을 미치지 않는 설계변수들을 제거함으로써 설계문제 자체를 간략화 하는 것도 가능할 것이다[1].

설계 결과로부터 이러한 정보를 찾아내는 과정을 일종의 데이터 마이닝이라고 부른다. 이 연구에서는 데이터 마이닝을 위하여 분산해석법(Analysis of variance)[2] 과 자기조직화지도(Self-organizing map)[3, 4]의 이용을 제안하였다. 전자는 response surface 상의 분산을 이용하여 목적함수와 설계변수들간의 정보를 정량적으로 예측하고, 후자는 고차원 데이터의 저차원 상으로의 비선형적 투영과 크러스터링 기법을 이용하여 목적함수와 설계변수 사이의 정보는 물론 목적함수들 사이의 trade-off 관계 등을 시각화를 통하여 정성적으로 보여준다.

이 논문에서는 위에서 소개한 두 데이터 마이닝 기법을 초음속 여객기의 날개 설계결과에 적용하였다.

## II. 본 론

### 2.1 분산해석법 (ANOVA)

본 연구에서는 분산해석법을 위한 response surface model으로서 Kriging 모델을 이용하였다. Kriging 모델은 지구통계학 및 지질통계학 분야에서 발달한 모델로서 일반적으로 널리 사용되는 다항식 모델보다 다봉성 (Multimodal) 함수 등을 정확하게 표현할 수 있는 장점이 있다. 본 연구에서 사용된 Kriging 모델은 아래의 식과 같이 표현된다.

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \mathbf{r}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}) \quad (1)$$

$\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ ,  $\mathbf{y}$ 는 모델 피팅을 위한 샘플 데이터값의 벡터, 그리고  $\mathbf{1}$ 은 유니트 벡터를 나타낸다.  $\mathbf{R}$ 은 correlation 행렬이며 이 행렬의  $(i, j)$  성분은 다음과 같다.

$$R(\mathbf{x}^i, \mathbf{x}^j) = \exp\left[-\sum_{k=1}^n \theta_k |x_k^i - x_k^j|\right] \quad (2)$$

미지의 점  $\mathbf{x}$ 와  $m$ 개의 샘플점 사이의 correlation vector는 다음과 같이 주어진다.

$$\mathbf{r}' = [R(\mathbf{x}, \mathbf{x}^1), R(\mathbf{x}, \mathbf{x}^2), \dots, R(\mathbf{x}, \mathbf{x}^m)] \quad (3)$$

식(1)의  $\hat{\mu}$ 는 일정한 벡터 파라미터  $\Theta$ 의 경우 다음 식에 의해 계산되어진다.

$$\hat{\mu} = \frac{\mathbf{1}'\mathbf{R}^{-1}\mathbf{y}}{\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}} \quad (4)$$

모델 피팅을 위한 벡터 파라미터  $\Theta$ 는 다음과 같은 likelihood function을 최대화함으로써 정해진다.

$$\text{Ln}(\hat{\mu}, \hat{\sigma}^2, \Theta) = -\frac{n}{2} \ln \hat{\sigma}^2 - \frac{1}{2} \ln(|\mathbf{R}|) \quad (5)$$

여기서  $\hat{\sigma}$ 는 아래와 같다.

$$\hat{\sigma} = \frac{(\mathbf{y} - \mathbf{1}\hat{\mu})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu})}{n} \quad (6)$$

일단 Kriging 파라미터  $\Theta$ 가 정해지면 모델 상에서의 전체 분산에 대한 각 설계변수에 의한 분산의 비를 구함으로써 각 설계변수가 목적함수에 얼마만큼의 영향을 미치는지 알 수가 있다. 전체 분산 값과 각각의 설계변수에 의한 분산은 다음 식들에 의해서 구해진다.

$$\hat{\mu}_{total} = \int \dots \int \hat{y}(x_1, x_2, \dots, x_n) dx_1 \dots dx_n \quad (7)$$

$$\hat{\sigma}_{total}^2 = \int \dots \int [\hat{y}(x_1, \dots, x_n) - \hat{\mu}]^2 dx_1 \dots dx_n \quad (8)$$

식(7)과 (8)는 각각 모델의 전평균과 전분산을 나타낸다.

$$\hat{\mu}_i(x_i) = \int \dots \int \hat{y}(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n - \hat{\mu}_{total} \quad (9)$$

$$\hat{\mu}_{i,j}(x_i, x_j) = \int \dots \int \hat{y}(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_{j-1} dx_{j+1} \dots dx_n - \hat{\mu}_i(x_i) - \hat{\mu}_j(x_j) - \hat{\mu}_{total} \quad (10)$$

식(9)은 설계변수  $x_i$ 에 의한 목적함수의 변동, 식(10)은 설계변수  $x_i$ 와  $x_j$ 의 상호작용에 의한 목적함수의 변동을 의미한다.

설계변수  $x_i$ 에 의한 분산은 아래 식과 같이 계산되어진다.

$$\int [\hat{\mu}_i(x_i)]^2 dx_i \quad (11)$$

전분산에 대한 설계변수  $x_i$ 에 의한 분산비는 식(11)을 식(8)으로 나눔으로써 계산되어진다.

$$\frac{\int [\hat{\mu}_i(x_i)]^2 dx_i}{\hat{\sigma}_{total}^2} \quad (12)$$

이 값이 설계변수  $x_i$ 가 목적함수에 미치는 영향을 나타낸다.

### 2.2 자기조직화지도 (SOM)

SOM은 artificial neural network 알고리즘으로서 고차원 데이터를 저차원 상으로 비선형 투영하여 시각화하는 방법이다. 여기서의 비선형 투영은 저차원 상에 배열된 뉴우런의 자율학습 (unsupervised learning)에 근거하여 이루어지며, 입력벡터와 뉴우런 사이의 특성 나타내는 특성벡터는 고차원 데이터의 특성을 저차원 지도상에 표현하기 위해서 학습되어진다. 이러한 투영에 의해서 고차원 공간상에서 밀접한 위치에 있는 데이터들은 저차원 지도상에서도 인접한 뉴우런 내에 위치하게 된다.

실제로 각각의 뉴우런과 관련된 특성벡터는 입력벡터(고차원 데이터)가  $n$ 차일 경우 다음과 같이 표현된다.

$$m_i = [m_{i1}, m_{i2}, \dots, m_{in}] \quad (i = 1, \dots, M) \quad (13)$$

여기서  $M$ 은 뉴우런의 수를 나타내며, 각각의 뉴우런들은 인접한 뉴우런들과 이웃관계 (neighborhood relation)가지고 연결되어있다. 일반적으로 이용되는 2차원상의 사각형과 육각형 위상은 Fig. 1과 같다.

SOM의 학습 알고리즘은 먼저 입력 벡터  $x$ 로부터 가장 인접한 특성벡터를 찾는 것에서부터 시작한다. 여기서는 가장 인접한 벡터( $m$ )를

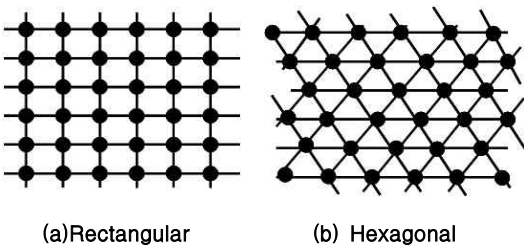


Fig. 1. Topology used in SOM

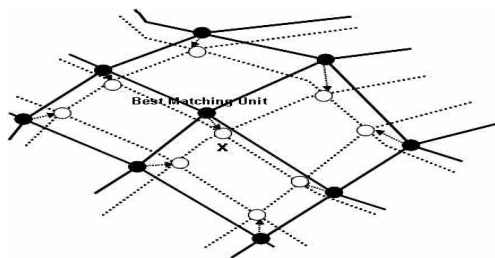


Fig. 2. Adjustment of the best-matching unit and its neighbors

'best-matching unit (BMU)'라고 부르며 다음과 같이 정의된다.

$$\|x - m_c\| = \min \|x - m_k\| \quad (k = 1, \dots, M) \quad (14)$$

일단 BMU가 정해지면, BMU와 이에 인접한 뉴우런들의 특성벡터에 대한 수정이 이루어진다. 이 수정은 입력벡터와 뉴우런간의 거리(유사성)에 의존한다. 이 거리에 의거하여 BMU와 인접한 뉴우런들은 Fig. 2와 같이 입력 벡터에 더욱 근접한 곳으로 이동하게 된다. Fig. 2에서 실선들이 만나는 곳에 위치한 검정색 원들은 특성벡터를 나타내고 입력벡터  $x$ 로부터 가장 가까운 곳에 위치한 검정색 원이 BMU를 나타낸다. 흰색 원들은 이동후의 특성벡터를 나타낸다.

이러한 학습을 반복함으로써 다차원 공간상에서 밀접한 거리에 위치한 입력 데이터들은 2차원 지도상에서도 인접한 배열을 가지게 된다. 이 연구에서는 상용 프로그램인 SOMine plus 4.0 [5]을 사용하여 SOM을 작성하였다. SOMine은 입력 데이터의 초기화와 Batch-SOM 기법[3]을 이용하여 생성되는 SOM의 유일성(uniqueness)을 보장하고 있다.

### 2.3 데이터 마이닝기법의 적용

이 논문에서는 4개의 목적함수를 가진 초음속 여객기 (SST)의 날개 최적화설계 결과[6]에 대하여 데이터 마이닝 기법을 적용하였다. 최적설계에 사용된 4개의 목적함수는 다음과 같다.

- ① 천음속 순항시의 저항계수의 최소화(CD<sub>Trans</sub>)
- ② 초음속 순항시의 저항계수의 최소화(CD<sub>Super</sub>)
- ③ 초음속 순항시의 벤딩 모멘트 최소화(CM<sub>Bending</sub>)
- ④ 초음속 순항시의 피칭 모멘트 최소화(CM<sub>Pitching</sub>)

처음 두 개의 목적함수는 대양 위에서의 초음속 순항 시와 대륙 위에서의 천음속순항시의 성능을 높이기 위한 것이고 나머지 두 개의 목적함수는 구조와 제어성능 향상을 위해서 지정되었다. 날개의 형상을 정의하기 위한 설계변수는 Fig. 3과 같이 평면 형상을 지정하기 위한 변수

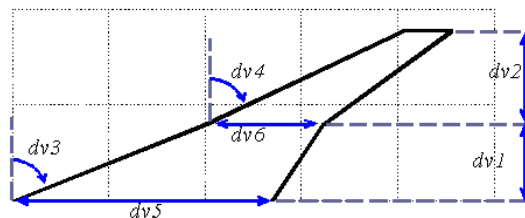


Fig. 3. Planform definition

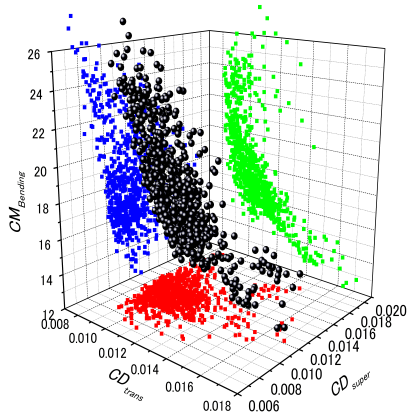


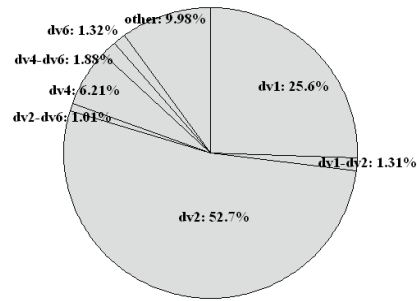
Fig. 4. 3D plot of Pareto solutions

6개 이외에 루트, 긴크, 팁의 3단면에서의 캠버와 두께를 지정하기 위한 변수 59개 그리고 트위스트를 지정하기 위한 변수 7개 등 총 72개 변수가 사용되었다.

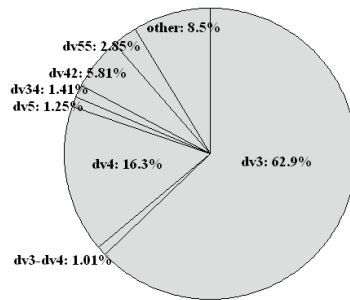
이 4개의 목적함수를 가진 최적화설계를 위해 영역적응형 다목적 유전알고리즘 (Range Adaptive Multi-Objective Genetic Algorithm) [7]이 이용되었다. 한 세대당 개체 수는 64이고 75세대까지의 진화를 실시하여 총 766개의 Pareto해를 얻었다. 각 개체의 평가는 Euler Code를 가지고 실시하였으며 동북 대학교 유체 과학연구소의 ORIGIN2000의 128PE를 이용하여 총 18일간의 계산 시간이 소요되었다. Fig. 4에는 776개의 Pareto해를 3개의 목적함수에 관한 3차원 공간에 표시한 결과이다. 검은색 점들은 3차원 공간상의 점들이고 그 외의 색을 가진 점들은 각각의 2차원 공간상에 투영을 실시한 점들이다. 하지만 이와 같은 고차원 데이터의 경우 기존의 3차원 플롯과 2차원 플롯으로부터 목적함수들간의 trade-off 관계를 비롯한 설계 공간의 지식을 얻어내는 것은 쉽지가 않다.

설계공간의 지식을 좀 더 효율적으로 습득하기 위해서 Pareto해들에 대해 ANOVA와 SOM을 적용하였다. Fig. 5는 각각의 목적함수에 대한 설계 변수의 기여도를 ANOVA를 이용해 분석한 결과이다.

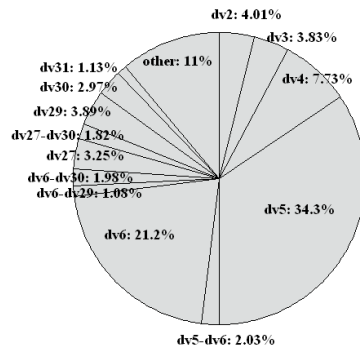
위의 결과에 의하면  $CD_{Trans}$ 의 경우 스펠의 길이에,  $CD_{Super}$ 의 경우는 후퇴각에,  $CM_B$ 의 경우는 날개의 코드길이에 그리고  $CM_P$ 의 경우는 후퇴각에 가장 큰 영향을 받는 것으로 나타났다. 이 결과는 천음속시 저항을 줄이기 위해서는 가로세로비 (aspect ratio)가 큰 날개를, 초음속시 저항을 줄이기 위해서는 큰 후퇴각을 가진 날개를,



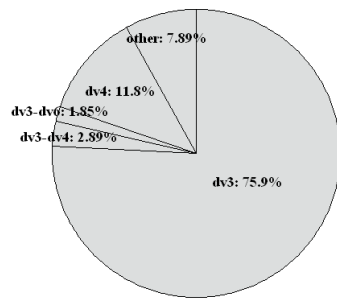
(a)  $CD_{Trans}$



(b)  $CD_{Super}$



(c)  $CM_{Bending}$



(d)  $CM_{Pitching}$

Fig. 5. Results of ANOVA

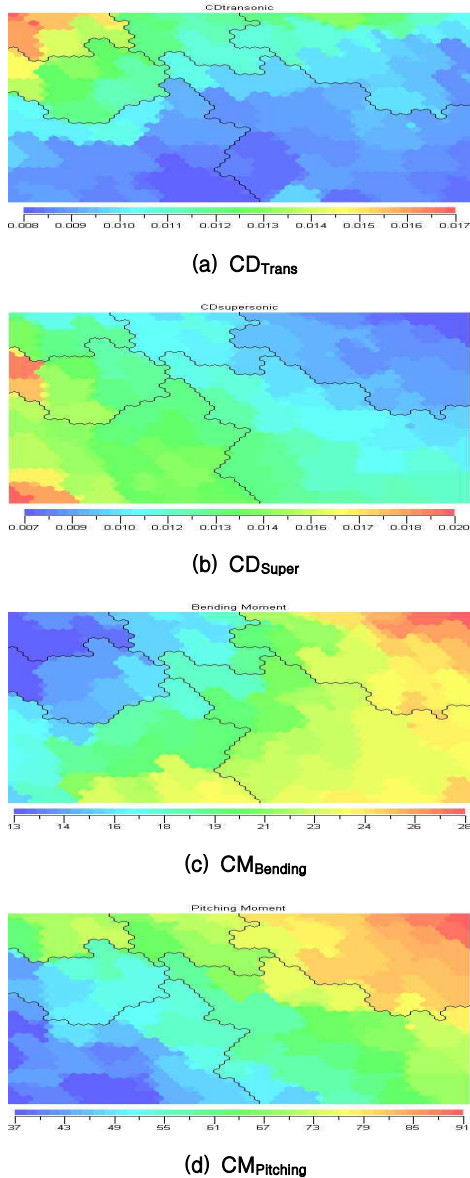


Fig. 6. SOM colored by objective function

벤딩 모멘트를 줄이기 위해서는 가로세로비가 작은 날개를 그리고 피칭모멘트를 줄이기 위해서는 후퇴각이 작은 날개가 이상적이라는 일반적인 공기 역학지식과 일치한다.

목적함수들 간의 trade-off 관계를 가시화 하기 위해서 Pareto해들에 대해서 SOM을 이용한 클러스터링을 실시하였다. Fig. 6은 얻어진 SOM을 각 목적함수의 값으로 칼라링을 한 결과이다.

Fig. 6(a)의 경우, 비교적 작은 CD<sub>Trans</sub> 값을 가진 클러스터들은 맵의 하단에 위치하고 있다. 하

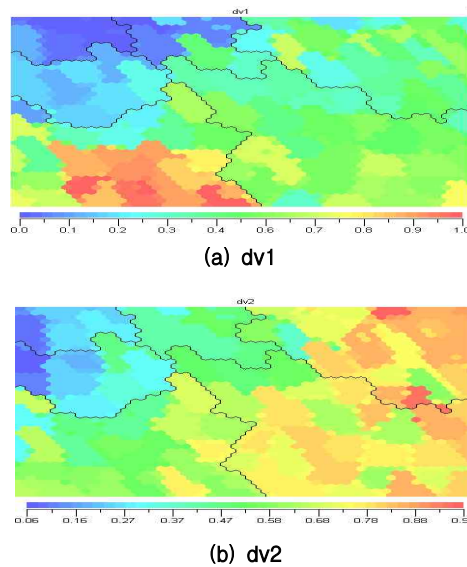
지만 Fig. 6(c)의 경우, 이 클러스터들은 비교적 큰 CM<sub>Bending</sub> 값을 가진 것을 알 수 있다. 반대로 CM<sub>Bending</sub> 이 작은 값을 가진 좌측상단의 클러스터들은 큰 CD<sub>Tran</sub> 값을 가진다. 그러므로 CD<sub>Trans</sub>와 CM<sub>Bending</sub>은 trade-off관계에 있다고 볼 수 있다.

Fig. 6(b)의 경우에는 비교적 작은 CD<sub>Super</sub>의 값을 가진 클러스터들은 우측 상단에 모여있다. 하지만 Fig. 6(d)을 보면 이 클러스터들은 큰 CM<sub>Pitching</sub> 값을 가지고 있다. 반대로 작은 CM<sub>Pitching</sub> 값을 가진 좌측 하단의 클러스터들은 비교적 큰 CD<sub>Super</sub>의 값을 가지고 있다. 그러므로 CD<sub>Super</sub>와 CM<sub>Pitching</sub>도 trade-off관계에 있다는 것을 SOM을 통해서 확인 할 수가 있다.

Fig. 7은 SOM을 몇몇의 설계변수의 값으로 칼라링을 실시한 결과이다. Fig 7(a)와 7(b)의 경우 dv1과 dv2가 큰 값을 가진 클러스터들은 작은 CD<sub>Trans</sub> 값을 가지고 있다는 것을 알 수 있다. 그러므로 좋은 CD<sub>Trans</sub> 성능을 가지기 위해서는 dv1과 dv2가 큰 값을 가져야한다는 것을 알 수 있다. 즉 가로세로비가 큰 날개를 설계해야한다는 것을 알 수 있다.

Fig 7(c)와 7(d)의 경우 dv3와 dv4가 작은 값을 가진 클러스터들을 좌측 하단에 위치하고 있다. 이 부분은 CM<sub>Pitching</sub> 성능이 좋은 클러스터들이 위치한 지역이다. 그러므로 dv3와 dv4가 작은 값을 가지면, 즉 후퇴각이 적으면, CM<sub>Pitching</sub> 성능이 좋아진다는 것을 알 수 있다. 반대로 dv3와 dv4가 큰 클러스터들은 좋은 CD<sub>Super</sub> 성능을 가진다는 것도 SOM을 통해서 확인할 수 있다.

그리고 Fig 7(e)와 7(f)의 경우, dv5와 dv6가



(a) dv1

(b) dv2

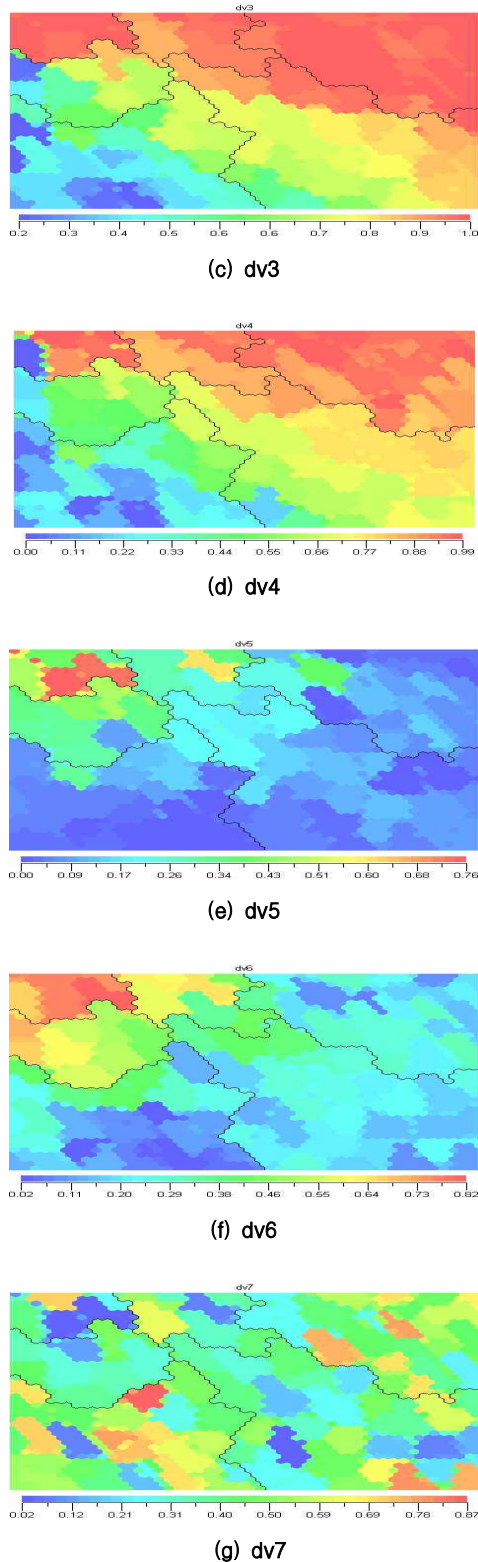


Fig. 7. SOM colored by design variables

큰 값을 가진 크러스터들은 좌측 상단에 모여있다. 이 지역에서는  $CD_{\text{Bending}}$  성능이 좋은 날개들이 모여있다. 그러므로  $dv5$ 와  $dv6$ 가 크면, 즉 가로세로비가 작으면,  $CD_{\text{Bending}}$  성능이 향상된다는 것을 알 수 있다.

그리고 Fig. 7(g)의 경우는, 색의 배열이 목적함수의 SOM과 특별한 연관성을 보이지 않으므로  $dv7$ 은 목적함수에 대하여 큰 영향을 미치지 않는 것을 알 수 있다.

SOM을 통해서 찾아낸 사실들은 ANOVA에 의해 얻어진 결과와도 일치하며 일반적인 공기역학 지식과도 일치한다. 이상의 결과를 통해서 논문에서 제안한 데이터 마이닝의 기법들인 ANOVA와 SOM의 유용성을 확인할 수 있다.

### III. 결 론

본 연구에서는 공력설계공간의 지식 습득을 위한 데이터 마이닝의 기법으로 analysis of variance(ANOVA) 와 self-organizing map(SOM) 기법을 소개하였다. ANOVA는 설계변수와 변수들 사이의 상호관계가 목적함수에 끼치는 기여도를 예측 가능하게 하고 SOM는 설계변수와 목적함수와의 관계는 물론 사이의 trade-off관계를 예측 가능하게 하였다.

데이터 마이닝 기법에 의해 얻어진 이와 같은 설계공간의 지식들은 최종디자인을 결정하는데 도움이 될 뿐만 아니라 설계문제의 간략화 및 효율화에 큰 도움이 될 것으로 기대된다.

### 참고문헌

- 1)Jeong, S., Murayama, M. and Yamamoto, K., "Efficient Optimization Design Method Using Kriging Model", Journal of Aircraft, Vol. 42, 2005, pp. 413-420.
- 2)Donald, R. J., Matthias S and William J. W, Efficient Global Optimization of Expensive Black-Box Function, Journal of global optimization, Vol. 13, 1998, pp. 455-492.
- 3)Kohonen, T., Self-Organizing Maps, Springer, Berlin, Heidelberg, 1995.
- 4)Krzysztof, J. C., Witold, P. and Roman, W. S., Data Mining Methods for Knowledge Discovery, Kluwer Academic Publisher, 1998.
- 5) Eudaptics software gmbh, <http://www.eudaptics.com/somine/index.php?sprache=en>, last access on April 14, 2005.

6) Sasaki, D., Obayashi, S. and Nakahashi, K., Navier-Stokes Optimization of Supersonic Wing with Four Objectives Using Evolutionary Algorithm, Journal of Aircraft, Vol. 39, No. 4, 2002, pp. 621-629.

7) Sasaki, D. and Obayashi, S., Efficient Search for Trade-Offs by Adaptive Range Multi-Objective Genetic Algorithms, Journal of Aerospace Computing, Information, And Communication, Vol. 2, 2005, pp. 44-64.