

유전적 프로그래밍과 SOM을 결합한 개선된 선박 설계용 데이터 마이닝 시스템 개발

이경호*, 박종훈***, 한영수**, 최시영**

Development of Data Mining System for Ship Design using Combined Genetic Programming with Self Organizing Map

Kyungho Lee*, Jonghoon Park***, Youngsoo Han** and Siyoung Choi**

ABSTRACT

Recently, knowledge management has been required in companies as a tool of competitiveness. Companies have constructed Enterprise Resource Planning(ERP) system in order to manage huge knowledge. But, it is not easy to formalize knowledge in organization. We focused on data mining system by genetic programming(GP). Data mining system by genetic programming can be useful tools to derive and extract the necessary information and knowledge from the huge accumulated data. However when we don't have enough amounts of data to perform the learning process of genetic programming, we have to reduce input parameter(s) or increase number of learning or training data. In this study, an enhanced data mining method combining Genetic Programming with Self organizing map, that reduces the number of input parameters, is suggested. Experiment results through a prototype implementation are also discussed.

Key words : Data Mining, Knowledge Management, Shipbuilding, Genetic Programming, Engineering Knowledge, Neural Network, Self Organizing Map(SOM)

1. 서 론

정보화 기술의 발달로 인해서 과거와는 다르게 인간이 하는 일은 인간만이 할 수 있는 분야로 압축되고 있다. 다시 말하면 정보의 수집과 제품생산과 같은 단순하고 반복적인 일은 컴퓨터와 기계에게 맡기고 사람은 사람만이 가장 잘 할 수 있는 '생각하는 것'과 같은 분야에 조직원의 역량을 집중하는 등의 기업의 지식을 관리하여 효율성과 경쟁력을 높일 수 있게 하고 있다.

세계를 단일시장으로 하고 있는 조선 산업의 경우에도 관련 지식과 데이터의 활용이 큰 이슈가 되고 있다. 따라서 우리나라 조선소들도 ERP(Enterprise

Resource Planning)의 구축과 함께 지식관리에도 힘을 쓰고 있다. 하지만 형식화 되어있지 않은 지식을 관리 하는 것은 매우 어려운 일이며, 지식의 관리가 이루어지고 있다 하더라도 지식의 내용을 단순히 문서화하여 공유하고 있는 수준에 그치는 것이 현실이다.

지식의 분류에 여러 종류가 있다^[1]. 하지만 우리의 관심은 기술적 지식(Engineering knowledge)에 있으며, 기술지식은 축적된 공학 데이터에서 나올 수 있다. 이러한 공학데이터들 속에는 전문가의 경험과 노하우가 녹아 들어 있다^[2,3].

지식관리 관점에서 보자면 지식은 사람 및 컴퓨터 시스템이 이해하기 용이하도록 간략한 문장의 형태가 되어야 하는 것이 이상적이지만, 대다수의 기술지식은 한 마디로 정의하기 어렵다. 일반적으로 기술지식은 지식의 분류 측면에서 형식적 지식과 암묵적 지식, 경험적 지식과 분석적 지식을 모두 다 포함하고 있기 때문이기도 하다. 그러나 본 논문에서 관심을 갖고 있는 것은 형식화된 지식보다는 명시적으로 나타

*종신회원, 징희원, 인하대학교 선박해양공학과

**학생회원, 인하대학교 대학원 선박공학과

***학생회원, 삼성중공업

- 논문투고일: 2009. 02. 25

- 논문수정일: 2009. 08. 02

- 심사완료일: 2009. 09. 22

나 있지 않는 암묵적 지식과 기술지식이 녹아있는 데이터, 구조화되지 못한 지식요소 등, 데이터 마이닝^[1]을 통하여 얻어낼 수 있는 분석적 지식을 대상으로 한다.

현재 세계시장의 선두에 있는 우리나라의 조선 산업은 지금까지 선박을 건조해오며 많은 데이터와 경험을 축적하고 있다. 하지만 지금까지의 데이터에서 구해진 경험식들은 과거의 실적선들에서는 식용이 효율적이나 앞으로 건조되어야 하는 신개념 선박들에 적용하기는 어려운 것이 사실이다. 이러한 관점에서 우리나라의 조선소들은 지금까지 축적해온 데이터를 잘 활용, 가공하여 향후 건조될 신개념 선박에 적용하기 위한 새로운 경험식을 생성해야 할 큰 과제를 안고 있다.

이에 반해 공학적 지식이 녹아있는 실적 데이터들을 활용하여 이로부터 유용한 정보/지식을 추출하기 위한 적절한 도구를 보유하고 있지 못한 것도 사실이다. 본 연구에서는 선행연구로서 “조선설계에서의 데이터 해석 및 활용을 위한 데이터 마이닝 도구 개발”을 수행하였다. 앞선 연구에서 설계지식이 녹아있는 데이터들을 활용한 유전적 프로그래밍(Genetic Programming, 이하 GP)을 이용한 데이터 마이닝 도구(Data mining Tool)를 개발하였으며, Fig. 1에 나타나 있다.

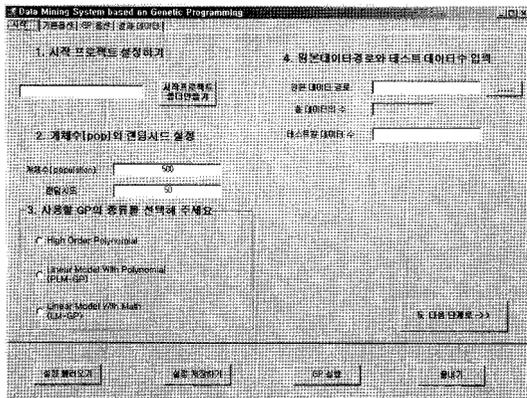


Fig. 1. 유전적 프로그래밍 기반의 데이터 마이닝.

조선분야의 실적 데이터는 우리가 개발한 데이터 마이닝 도구를 활용할 수 있을 만큼 충분하지 못한 양이다. 오히려 부족한 데이터의 사용으로 인해서 정확도가 떨어지는 문제인 오버피팅이 발생하는 역효과가 나타나기도 한다.

또한 입력 파라미터의 개수가 많으면 많을수록 결

과 와 입력 파라미터와의 관계를 수식화하기 힘들기 때문에 더욱 많은 학습데이터가 필요하다.

이러한 조선 분야의 특성에 기인한 문제점들을 개선하기 위해 학습에 필요한 입력 파라미터들 가운데, 결과에 대한 영향력이 떨어지는 파라미터의 개수를 감소시키는 과정이 요구된다. 본 논문에서는 기존에 개발한 유전적 프로그래밍으로 학습하던 것을 입력 파라미터의 영향도를 고려할 수 있도록 인공지능망의 한 종류인 SOM(Self Organizing Map, 이하 SOM)을 도입하여 실험을 진행하였다. 본 연구에서 SOM을 통하여 입력 파라미터들을 평가하고, 그 중에서 영향도가 높은 파라미터들을 선택하여 학습데이터를 만들어 낼 수 있도록 개선된 데이터 마이닝 도구를 개발하였다.

이 도구로 학습한 결과와 영향도 평가 없이 학습데이터를 데이터 마이닝 도구를 사용하여 나온 결과와의 비교를 통해 SOM의 영향도 평가가 데이터 마이닝의 정확도에 얼마나 큰 영향을 미치는가에 대하여 알아보고자 한다.

2. Self organizing map을 활용한 데이터마이닝의 성능 개선

2.1 선행 연구의 문제점 및 해결방안

지난 데이터 마이닝 도구개발^[1] 연구 과정에서 나타난 문제점은 다음과 같으며 아래의 Table 1과 같이 정리할 수 있다.

첫 번째, 데이터 마이닝의 결과로 나온 수식은 알아보기 힘든 정도의 복잡하기 때문에 바로 사용하기 어렵다는 것이다. 하지만 이 문제는 데이터 마이닝 도구 안에 수식을 자동으로 컴퓨터 코드(C Code)로 생성하도록 하였고, 이를 통해 다른 설계 프로그램과 인터페이스를 할 수 있도록 하여 해결하였다.

Table 1. GP 기반 데이터 마이닝의 문제점

문 제 점	원 인
계산시간의 증가	개체 및 세대 수 증가
오버피팅의 발생	학습데이터 부족
활용도의 감소	복잡한 수식의 생성

두 번째, 계산시간의 과도한 증가이다. 본 데이터 마이닝 도구는 GP를 사용하여 연산을 수행하는데, 트리구조로 연산이 수행되는 GP의 특성상 개체수와 세대수가 증가할 수록 급격하게 계산 시간이 증가하는 문제점을 발생시키고 또한 복잡한 수식의 생성으로

인해 활용도가 떨어지는 악영향을 미칠 수 있다.

세 번째, 입력 데이터의 개수가 부족하거나 입력 파라미터의 개수가 데이터의 개수에 비해 상대적으로 많을 경우 데이터 마이닝 결과의 정확성이 많이 떨어지는 문제이다. 학습데이터가 부족할 경우 오버 피팅이 발생하여 정확성이 감소할 수도 있기 때문이다.

위의 두 번째 세 번째, 문제점이 해결해야 할 과제가 되는 것으로 이에 대한 해결방안으로 본 연구에서는 SOM을 이용한 영향도 평가를 통해 영향도가 낮은 입력 파라미터를 제거하여 파라미터의 개수를 감소시켜 문제점을 해결하였다.

2.2 SOM (Self Organizing Map)의 도입

부족한 데이터의 개수로도 좋은 데이터 마이닝 결과를 얻기 위하여 영향도 평가를 통해 입력 파라미터의 개수를 줄여나가는 과정이 필요하다. 조창기의 경쟁 학습 알고리즘은 단순 경쟁 학습 알고리즘으로서 항상 승자 뉴런만을 학습 시키므로 초기 가중치 벡터들의 분포에 따라 전혀 학습이 이루어지지 않는 출력 뉴런들이 생기는 문제점이 발생했다. 이를 해결하기 위해서 여러 가지 해결안들이 제공되었으며, 가장 대표적인 경쟁학습법 중의 하나가 Kohonen의 SOM이다¹¹⁾. SOM 알고리즘은 승자 뉴런뿐만 아니라 위상적으로 이웃한 뉴런들도 함께 학습시킨다. SOM은 입력 및 출력 자료가 쌍의 형태로 주어지는 감독 제어형 학습(supervised learning)이 아니라 입력 자료의 규칙성이나 패턴을 찾는 비감독 제어형 학습(unsupervised learning)의 한 예이다.

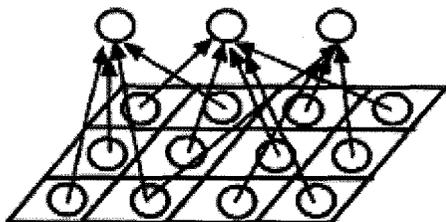


Fig. 2. SOM의 클러스터링.

즉 입력 벡터는 연속적으로 제시되지만, 출력 값은 제시되지 않는다는 것이다. 또한 SOM은 Fig. 2와 같이 입력 자료들을 여러 클러스터로 그룹 짓는 클러스터링 망의 하나이다. 클러스터링 망의 출력 노드들은 각 클러스터를 나타내며 입력 노드들은 입력 벡터의 구성원들을 나타낸다. 각 출력 노드를 위한 가중치 벡터는 망이 각 클러스터를 위해 설정하는 것으로 입력

패턴들에 대한 표본 벡터의 역할을 하며 이는 Kohonen 학습 알고리즘을 사용하는 SOM의 특성 중 하나이다. 가중치 벡터가 입력벡터와 가장 유사한 출력 노드(winner) 또는 그와 그의 이웃 노드들도 학습시킨다. 즉 경쟁학습을 통하여 입력 벡터와 가장 유사한 출력노드만이 살아남아 학습에 참여 할 수 있는 것이다. 이 특성을 이용하면 SOM을 통해 데이터마이닝 도구의 학습에 필요한 데이터에서 영향도가 높은 파라미터들을 선택하여 학습을 하게 되고, 데이터마이닝 도구에 입력되는 파라미터의 개수를 줄일 수 있을 뿐만 아니라, 충분한 학습에 필요한 데이터의 수량도 줄일 수 있을 것이다.

2.3 영향도 평가의 진행 및 가시화

영향도 평가 방법은 아래와 같은 순서로 수행된다.

- 1 단계: 연결 가중치를 초기화 한다. N개의 입력과 M개의 출력을 연결하는 가중치들은 아주 작은 임의의 값으로 설정한다.
- 2 단계: 새로운 입력 패턴을 입력 뉴런에 제시한다.
- 3 단계: 입력 벡터와 출력 뉴런들과의 거리를 계산한다.

$$d_i = \sum_{i=0}^{n-2} [x_i(t) - W_{ij}(t)]$$

$X_i(t)$ 는 시간 t 에서 뉴런 i 로의 입력, $W_{ij}(t)$ 는 시간 t 에서 입력 뉴런 i 로부터의 출력 뉴런 j 로의 가중치이다.

- 4 단계: 최소 거리를 가지는 승자 뉴런 c 를 구한다.

$$c = j^{\min} d_j$$

- 5 단계: 승자 뉴런 c 와 이웃 출력 뉴런의 연결 가중치들을 갱신한다.

$$W_{ij}(t+1) = W_{ij}(t) + \alpha [X_i - W_{ij}]$$

for $j \in NE_c(t)$

$\alpha(t)$ 는 0과 1 사이의 값을 가지는 학습률로 시간에 따라 감소하는 함수이며, $NE_c(t)$ 는 승자 뉴런 c 로부터 일정한 범위 내에 있는 출력 뉴런을 포함한다.

- 6 단계: 2단계의 새로운 입력 벡터를 처리한다.
- 7 단계: 2단계에서 6단계까지의 과정을 지정해놓은 학습 회수만큼 반복 수행한다.

이러한 순서로 영향도 평가가 진행되고 그 결과를 가시화하기 위해 U-Matrix방법을 사용하였다. 일반적

으로 2차원 입력에 대하여 그 결과를 2차원 맵으로 표현하기는 쉽지 않기 때문에 점층적 색상 대비를 이용하는 U-Matrix 가시화 방법을 사용하는 것이 효과적이다.

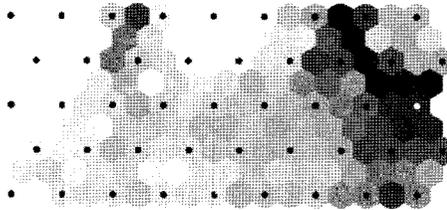


Fig. 3. U-Matrix 가시화.

Fig. 3은 U-Matrix 방법으로 가시화한 예이다. 각각의 인접한 노드사이의 거리에 따라 인접한 뉴런들이 다른 색으로 나타나며 짙은 색으로 갈수록 뉴런간의 거리가 멀어 진다는 의미이다.

SOM 알고리즘을 적용하여 얻게 되는 장점들은 다음과 같다.

첫 번째, 비감독 제어형 학습인 SOM은 패턴 분석 및 클러스터링, 양쪽 모두에게 적용성을 제공한다. 또한 U-Matrix 기법을 이용한 클러스터 가시화가 가능하므로 알려져 있지 않는 문제에서도 직관적인 이해가 가능하다는 장점을 제공한다.

두 번째, 신경망 학습에서 감독 제어형 학습은 알려져 있는 문제에 대하여 학습 데이터들의 레이블링 작업이 수반되어야 한다. 특히 생성되는 데이터들의 양이 많을 경우 그들을 분석하고, 레이블링하는 작업은 많은 비용과 시간이 필요한 반면, 비감독 제어형 학습은 레이블이 필요치 않으며, 또한 알려져 있지 않은 문제에 대해서도 특징들 간의 관계만을 가지고 학습이 가능한 장점을 제공한다.

3. SOM을 활용한 데이터 마이닝 성능 실험

앞에서 말한 바와 같이, SOM을 이용하여 데이터 마이닝 도구의 학습에 필요한 데이터의 입력 파라미터의 개수를 줄일 수 있다면 데이터의 수를 줄이는 효과를 내거나 보다 더 정확한 결과의 예측 모델을 만들 수 있을 것이다.

먼저 SOM의 도입과 영향도 평가를 위해서 Matlab을 이용하였다. SOM Toolbox를 구성하고 데이터마이닝 도구의 학습에 쓰일 데이터 두 종류를 준비하였다. 첫 번째는 인하대학교 선형구조에서 수행한

trimaran 모형선의 저항계수(Ct) 실험 데이터 50건을 학습에 활용하여 실험을 진행하였으며, 두 번째 실험은 7가지 주요복을 기초로 설계선의 방형계수를 구하는 실험을 진행하였다. 두 실험 모두 각각의 입력 데이터를 선형연구에서 개발한 데이터 마이닝 도구에 입력하여 추정식을 생성해내고 생성된 추정식으로 다수의 추정 값을 산출하였다. 또한 SOM 을 통하여 영향도 평가를 수행하여 영향도가 낮은 입력 파라미터를 제거한 데이터를 데이터마이닝 도구에 학습시켜 개발된 추정식을 만들어 다수의 추정 값을 산출하였다. 이렇게 생성된 두 종류의 추정 값들을 RMSE (Root mean square error) 방법을 사용하여 실제 데이터와 비교하는 방법으로 SOM을 활용한 영향도 평가가 데이터 마이닝 도구의 효과를 개선할 수 있는지 에 대하여 실험을 진행하였다.

3.1 trimaran 모형선의 저항계수를 구하는 실험

3.1.1 trimaran 모형선의 저항계수를 구하는 실험 데이터

Table 2는 인하대학교 선형구조에서 실험한 trimaran 모형선의 저항계수(Ct)를 구하는 실험 데이터 50건을 나타낸 것이다. 입력 파라미터는 Rn(Reynolds수), Fn(Froude 수), R_f(모형선에 발생한 저항), Z_f(모형선의 선수 침하량), Z_h(모형선의 선미 침하량), EHP(유효 마력), H_t(모형선의 선수/미 침하량의 평균), Sink(모형선의 운동 중 sinkage), Trim(모형선의 운동 중 트림) 등으로 총 10개이며 출력 파라미터는 Ct이다.

Table 2. GP 학습에 사용된 trimaran data set

Reynolds	Fn	R _f	Z _f	Z _h	EHP	H _t	Sink	Trim	Ct
2.75E+05	0.283	0.0225	0.304	0.4978	2.194	1.572	0.269	0.449	0.125
1.24E+06	0.293	0.026	0.3157	5.9616	15.265	9.623	0.236	0.242	0.1455
1.67E+06	0.4027	0.0285	0.2719	12.9427	42.362	30.294	0.184	0.278	0.1717
3.16E+06	0.514	0.0294	0.21129	22.2689	85.209	63.568	0.138	0.202	0.1936
2.49E+06	0.392	0.0493	0.17792	21.2728	214.624	133.496	0.236	0.235	0.12729
2.84E+06	0.677	0.0361	0.19252	21.6232	107.934	110.388	0.253	0.201	0.1606
3.07E+06	0.732	0.0477	0.19482	18.1472	176.147	129.927	0.219	0.204	0.16413
2.12E+06	0.758	0.0397	0.19699	16.7442	184.776	134.447	0.142	0.204	0.16413
2.51E+06	0.791	0.0398	0.1804	0.9712	2.206	1.643	0.169	0.049	0.069
1.25E+06	0.299	0.149	0.2122	0.9918	12.332	9.214	0.252	0.042	0.088
1.66E+06	0.403	0.2599	0.21719	17.5407	41.198	30.732	0.234	0.036	0.077
1.19E+06	0.214	0.28226	0.24322	23.2693	36.181	0.443	0.918	0.031	0.064
2.12E+06	0.752	0.03999	0.19797	0.2525	136.806	88.764	0.294	0.01E	0.1231
2.78E+06	0.473	0.04923	0.195523	21.8232	157.349	115.79	0.295	0.005	0.127
2.99E+06	0.232	0.22275	0.246527	10.2478	173.97	129.141	0.298	0.001	0.1792
3.39E+06	0.756	0.03897	0.19699	16.7941	187.472	129.254	0.142	0.004	0.067
2.79E+06	0.54	0.26	0.2129	17.24	37.194	24.363	0.217	0.002	0.104
3.49E+06	0.675	0.041	0.1959	0.2	223.242	95.695	0.493	0.016	0.131
4.19E+06	0.756	0.034	0.1934	244.538	327.324	4.945	0.017	0.046	0.1046
4.51E+06	0.872	0.11	0.1784	0.464	302.593	242.181	0.126	0.023	0.149
3.49E+06	0.675	0.0345	0.1928	0.0306	1.113	1.129	0.142	0.029	0.023

이 실험은 10개의 입력 파라미터들의 Ct에 대한 영향도를 평가하여 영향도가 낮은 파라미터를 찾아내어, 그 파라미터를 제거하여 다시 GP 학습을 진행한다. 영향도가 낮은 입력 파라미터가 제거된 학습결과를

다시 평가하여 파라미터가 제거되기 전과의 결과와 비교하는 방법으로 실험을 진행하였다.

3.1.2 trimaran 모형선 실험에서의 영향도 평가 전의 데이터 마이닝 도구의 학습결과

추후 진행 할 영향도 평가 결과와의 비교를 위하여 입력 파라미터의 영향도를 측정하기 전의 학습 데이터를 데이터 마이닝 도구를 통하여 학습시켰다.

trimaran 모형선 실험의 경우 총 50건의 데이터 중 학습 40건과 테스트 10건, 세대 수 1,500으로 수행하였다.

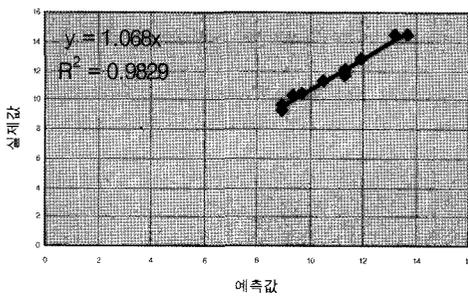


Fig. 4. 영향도 평가 전의 trimaran 데이터 학습 결과.

실험 시행 후 테스트 데이터의 결과는 Fig. 4에 보여지고 있으며 이것은 RMSE 방법으로 GP 학습 실험의 결과를 평가한 것으로 예측 값과 실제 값의 차이가 나는 정도를 나타내며 그 값은 0.9829로 나타났다.

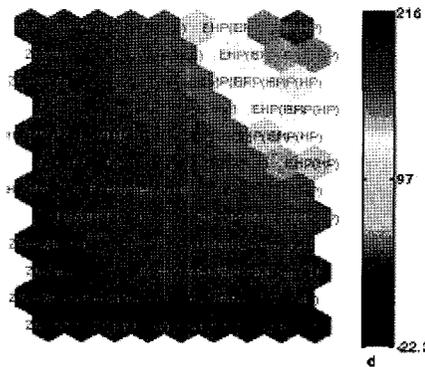


Fig. 5. Effect evaluation of Input parameters for trimaran Data by using SOM.

3.1.3 trimaran 모형선 실험에서의 SOM을 활용한 영향도 평가후의 데이터 마이닝 도구의 학습결과

SOM을 이용한 영향도 평가를 진행하여 Fig. 5와

같은 U-Matrix를 얻었으며 이것을 분석하여 영향도가 낮은 파라미터로 Z_r , H_s , Sink를 찾을 수 있었다.

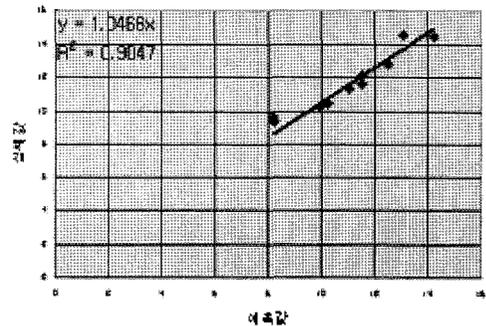


Fig. 6. 영향도 평가 후의 trimaran 데이터 학습 결과.

이 결과를 토대로 학습데이터를 변경하여 실험을 진행하였다. 새로운 학습 데이터는 기존의 것에서 Z_r , H_s , Sink를 제외한 R_n , F_n , R_c , Z_u , EHP, Trim 등의 7개의 입력 파라미터를 데이터로 갖도록 구성하여 데이터 마이닝 도구를 통하여 학습 후의 결과를 도출하였다(3.1.2절의 실험방법과 동일한 조건으로 실험 수행).

Fig. 6은 영향도 평가후의 데이터 마이닝 도구의 학습 결과를 나타낸 것이다.

이 그림으로부터 GP 학습 후 영향도 평가를 수행, 학습 데이터에서 입력 파라미터의 개수를 줄여주는 것으로 실제 값과 예측 값의 차이를 나타내는 RMSE가 Fig. 4에 비해 0.9829에서 0.9047로 작아진 것을 확인할 수 있었다.

Table 3. GP학습에 사용된 Cb data set

	A	B	C	D	E	F	G	H
	DWT	LBP	Seese	Breadth	Drift	Depth	Fn	Cb
1	109146.00	245.89	15.00	42.03	14.81	22.50	0.18	0.83
2	103214.00	244.00	15.50	42.00	14.81	23.94	0.16	0.82
3	159900.00	274.00	15.20	48.00	16.00	23.20	0.15	0.85
4	127000.00	272.00	14.80	46.00	14.80	22.80	0.15	0.86
5	146300.00	267.00	14.90	44.40	16.50	24.10	0.15	0.85
7	182000.00	274.18	15.50	50.00	18.00	23.10	0.15	0.84
7	159000.00	274.00	15.20	48.00	15.00	23.70	0.16	0.84
9	138121.00	287.00	14.70	44.40	18.50	24.10	0.15	0.86
10	126500.00	263.00	14.50	46.00	14.50	23.70	0.15	0.86
11	150000.00	270.45	15.00	44.80	16.10	24.20	0.15	0.85
12	144348.00	270.04	14.80	43.06	17.33	23.80	0.15	0.86
13	147916.00	274.00	15.00	47.80	18.02	22.80	0.15	0.85
14	146270.00	274.30	15.20	45.20	18.71	23.20	0.15	0.85
15	147500.00	289.00	14.50	43.24	17.53	23.50	0.15	0.88
16	146194.00	277.00	15.50	44.40	16.55	24.10	0.15	0.95
17	155980.00	291.00	15.20	53.10	15.77	22.80	0.15	0.85
18	155200.00	285.00	15.00	44.20	17.15	22.50	0.15	0.86
19	145698.00	267.00	14.50	45.20	16.81	22.00	0.15	0.86
20	36700.00	221.00	15.50	30.00	11.55	16.90	0.17	0.80
21	37000.00	182.56	14.00	27.54	11.50	16.70	0.17	0.80
22	45000.00	183.00	14.20	32.20	11.50	19.00	0.17	0.80
23	46000.00	183.00	14.00	32.20	7.50	12.20	0.17	0.80
24	36000.00	176.00	14.00	31.00	11.50	17.00	0.17	0.80
25	37000.00	182.80	14.10	27.30	11.30	18.70	0.17	0.80
26	35970.00	183.00	14.50	32.20	11.50	17.00	0.18	0.79
27	37602.99	176.00	14.50	31.00	11.50	17.00	0.18	0.78
28	48020.99	183.00	14.00	32.20	7.50	12.50	0.17	0.80
29	36330.99	221.00	14.50	30.00	11.70	16.80	0.16	0.83

=0.008에 비해 감소되어 예측 값과 실제 값이 더 가
 까워진 결과를 얻을 수 있었다.

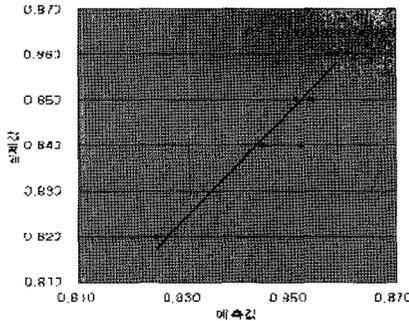


Fig. 10. 영향도 평가를 적용한 후의 학습결과.

3.3 실험결과 및 고찰

앞선 연구를 통해 데이터 마이닝 도구를 개발하였
 고, 입력 파라미터의 많은 개수와 입력 데이터의 적은
 수로 인해 발생하는 문제점을 해결하기 위해 입력 파
 라미터의 영향도를 평가하여 파라미터를 개수를 제어
 하는 방법을 도입하였다. 이러한 방법으로 3.1절과 3.2
 절의 실험을 진행하여 각각 Ct와 Cb를 얻기 위해 초
 기 입력 파라미터의 영향도를 평가하였다. 영향도 평
 가의 결과는 U-Matrix가시화 방법을 사용하여 표현하
 였으며, 그 결과 3.1절의 실험에서는 Zf, IIs, Sink 가
 제거되었고 3.2절의 실험에서는 Fn을 제거하여 실험
 을 진행하였다. 영향도가 낮은 파라미터들을 제거한
 후에 진행한 실험에서 RMSE(Root Mean Square
 Error)가 3.1, 3.2절의 실험에서 각각 약 0.004~0.09
 작게 나타나 실제 값과 예측 값과의 차이가 줄어든 것
 을 확인할 수 있었다.

4. 결론 및 향후 과제

조선 산업에서 경험식은 매우 중요하다. 조선 산업
 의 특성상 대형 선박한 척당 가격이 수 천만 달러에서
 십 수억 달러에 이르기에도 하는 선박의 설계에 가장 큰
 영향을 미치는 것이 바로 경험식에 의해 얻어진 선박
 설계 파라미터들이다. 초기 설계 단계에서 경험식에
 의해 결정된 주요 파라미터를 통해 설계를 진행하여
 선박 등의 해양구조물을 생산한다. 만약 설계 혹은 생
 산을 진행하는 과정에서 파라미터를 수정해야 할 경
 우 그 시간과 비용이 엄청나게 클뿐더러 대외적 신임
 도 또한 크게 영향을 받기 때문에 초기 파라미터 선정
 이 매우 중요하다.

현재 선박의 초기 설계는 과거의 실적 선박을 토대
 로 만들어진 경험식에 의존하여 설계 파라미터를 추
 정한 후 그 파라미터들을 기준으로 설계를 진행한다.
 하지만 과거에는 없던 고부가가치를 가진 신개념의
 선박들이 개발되어 건조되고 있는 상황에서 과거의
 실적선박을 기반으로 한 기존의 경험식을 새로운 선
 종 및 선형에 적용하여 사용하기 매우 어렵다.

그렇기 때문에 새로운 선종의 선박 설계 시, 설계
 파라미터를 추정하기 위해 새로운 경험식이 요구되며
 실적 데이터를 이용하여 새로운 경험식을 얻을 수 있
 는 방법론이 꾸준히 요구되어 왔다.

지난 연구에서는 데이터 마이닝 개념을 적용하여
 설계 파라미터의 추정을 돕는 도구를 개발하기 위해
 지난 연구에서는 Genetic Programming에 기반한 조
 선 설계용 데이터 마이닝을 개발하였다.

하지만 조선 설계현장에 적용하였을 때 그 설계 특
 성상 적은 양의 학습 데이터로도 경험식 생성이 가능
 해야 하는 것과 적은 양의 데이터 사용으로 인한 오버
 피팅에 대한 문제가 예상되어 본 연구에서는 그 대책
 으로 SOM(Self Organizing Map)을 도입하였다.

데이터의 수가 부족하고 그에 비해 상대적으로 입
 력 파라미터가 많은 조선 산업의 경우 SOM을 이용한
 영향도 평가를 거쳐 영향력이 적은 입력 파라미터를
 줄여 줌으로써 데이터 마이닝 도구가 좀 더 정확한 학
 습을 할 수 있도록 개선하였으며 앞서 진행한 실험들
 에서 보여지듯 데이터 마이닝 도구의 학습에 개선된
 결과를 줄 수 있다는 것을 확인할 수 있었다.

또한 현재 SOM을 구현하여 U-matrix를 보여주기
 위해 Matlab을 활용하여 별도로 구현하였지만, 이
 과정에서 복잡한 조사가 필요하게 된다. 그렇기 때문
 에 좀더 용이한 사용을 위해 SOM 프로세스를 데이터
 마이닝에 내장하여 시스템을 통합하는 연구를 진행할
 것이다.

감사의 글

이 논문은 2007년 정부(교육과학기술부)의 재원으
 로 한국연구재단의 지원을 받아 수행된 연구임(KRF-
 2007-331-D00558).

참고문헌

1. 박우창, 송현우, 용원승, 최기현, 데이터 마이닝, 자
 유아카데미, 서울, 2004.
2. 오준, 이경호, 박종훈, 최영복, 장영훈, "조선설계에

- 서의 데이터 해석 및 활용을 위한 데이터 마이닝 도구 개발”, 대한조선학회논문집, Vol. 43 No. 6, pp. 700-706, 2006.
3. 이경호, 손미애, “차세대 성장동력과 조선산업 (어떻게 해야 하나? How-to-do): 표준화와 기술지식관리”, 대한조선학회 학회지, Vol. 41 No. 3, pp. 15-26, 2004.
 4. 이경호, 연운석 “유전적 프로그래밍을 이용한 노이지 데이터의 Curve Fitting과 선박설계에서의 적용”, 한국CAD/CAM 학회 논문집, 제9권, 제3호, pp. 183-191, 2004.
 5. 이경호, 연운석, 양영순, “데이터 마이닝을 위한 다항식기반의 유전적 프로그래밍 기법과 조선분야 응용”, 대한조선학회 춘계학술대회 논문집, pp. 845-850, 2004.
 6. 이경호, 연운석, 양영순, “조선 기술지식 활용을 위한 데이터 마이닝 기법의 적용”, 한국해양과학기술협의회 2005 공동학술대회, pp. 375-380, 2005.
 7. 이경호, 연운석, 양영순, “조선분야의 축적된 데이터 활용을 위한 유전적 프로그래밍에서의 선형 모델 개발”, 대한조선학회 논문집, Vol. 42 No. 5, pp. 309-405, 2005.
 8. 임종수, MATLAB GUI PROGRAMMING, 아진출판사, pp. 260-350, 2005.
 9. Koza, J. R., “Genetic Programming: On the Programming of Computers by Means of Natural Selection”, *The MIT Press*, 1992.
 10. Coyne, R. D., Rosenman, M. A., Radford, A. D., Balachandran, M. and Gero, J. S., Knowledge-Based Design Systems, Addison-Wesley Publishing Company, 1989.
 11. Teuvo Kohonen, Self Organizing Maps, Springer, 2001.



이 경 호

1988년 서울대학교 조선해양공학과 학사
 1990년 서울대학교 조선해양공학과 석사
 1998년 서울대학교 조선해양공학과 박사
 1990년~2003년 한국해양연구원 선임 연구원

2002년~2003년 University of Maryland Visiting Researcher

2003년~현재 인하대학교 선박해양공학과 부교수

관심분야: Artificial Intelligence in Design, Simulation-Based Design, Data Mining, Evolutionary Computation, Ubiquitous, Augmented Reality, PLM



박 종 훈

2006년 2월 인하대학교 학사
 2008년 2월 인하대학교 석사
 2008년 3월~현재 삼성중공업
 관심분야: 3D CAD/CAM 시스템, 선박용 의장 CAD, 데이터 마이닝, 시뮬레이션



한 영 수

2008년 2월 인하대학교 학사
 2008년 3월~현재 인하대학교 석사과정
 관심분야: Ubiquitous, Augmented Reality, PLM, Data Mining, Modeling & Simulation



최 시 영

2007년 2월 광주대학교 학사
 2008년 3월~현재 인하대학교 석사과정
 관심분야: 선박해양구조물의 LSN, 위치인식