

고객관계관리에서 신경망을 이용한 제품-고객군의 형성에 관한 연구*

황 인 수**

A Study on Product-Customer Group Formation Using Neural Networks for CRM

Hwang, Insoo

CRM is at the core of any customer-focused business strategy and includes the people, processes, and technology questions associated with marketing, sales, and service. In today's hyper-competitive world, organizations looking to implement successful CRM strategies need to focus on a common view of the customer using integrated information systems and contact center implementations that allow the customer to communicate via any desired communication channel. A CRM solution contains a number of sophisticated tools that enable to extract detailed information about customers. This information can be used to gain a better understanding of customers. From this we can determine trends, and so refine business toward customers' needs and target new products to particular customer groups.

This paper presents an approach for forming the product-customer groups using neural networks for customer relationship management. The Carpenter-Grossberg's neural network, which has been used for manufacturing cell formation in group technology, is modified and applied for product-customer group formation. As a result of numerical experiments, it is also useful for more complex problems in which customers have different preferences for each product.

* 이 논문은 전주대학교 학술연구비 지원에 의해 연구되었음.

** 전주대학교 정보기술학부 조교수

I. 서론

기업간의 경쟁이 심화되고 고객의 욕구가 다양화됨에 따라, 많은 기업들이 정보통신 기술을 이용한 고객관계관리(customer relationship management, CRM)를 기업의 가장 중요한 전략의 하나로 추진하고 있다. 이는 개인 고객이나 일련의 고객그룹을 유치하거나 기존 고객의 개발 및 우수 고객의 유지를 위한 전사적인 활동으로서 고객을 중심으로 한 마케팅, 세일즈, 서비스 전략을 실행하는 일련의 과정이다[Kalakota & Robinson, 1999].

CRM을 실현함에 있어서 데이터웨어하우스, 데이터마이닝 등의 새로운 정보기술은 고객을 특성에 따라 세분화함으로써 고객 그룹별, 혹은 개인별로 차별화된 마케팅을 가능하게 하고 있다. 선행 연구에 따르면 신규고객을 획득하는데 소요되는 비용은 기존고객을 유지하는데 소요되는 비용의 5배에 이르며, 고객유지율을 5% 향상시키면 기업의 이윤을 20%에서 125%까지 증가시킬 수 있는 것으로 조사되었다. 따라서, 기존 고객에 대한 데이터를 분석하여 의미있는 정보를 제공하는 데이터마이닝은 CRM에 있어서 가장 중요한 요소의 하나가 되고 있다.

본 연구는 생산공장에서 유사한 기계를 사용하는 제품 혹은 부품들을 그룹으로 묶고 이의 생산에 필요한 기계들을 한 곳에 모아 제조셀을 구성하는 그룹테크놀로지(group technology)를 제품-고객군 형성에 도입함으로써, 제품-고객군별 차별화된 마케팅을 지원하는 데 그 목적을 두고 있다. 그룹 테크놀로지는 고객 요구의 다양화로 인해 생산시스템이 다품종 소량생산체제로 전환됨에 따라, 목크기(lot size)가 감소되어 차재취급비용이 증가하고 빈번한 교체준비(set-up)로 인해 생산비용이 상승함에 따라, 제품의 유연성을 증가시키면서도 대량생

산체제의 장점인 규모의 경제를 극대화할 수 있도록 부품의 유사성을 최대한 활용한다[박명섭 & 황인수, 1996]. 따라서, 기계-부품군의 형성방법을 제품-고객군 형성에 도입할 경우 그룹테크놀로지에서도 얻을 수 있는 장점을 CRM에서도 그대로 유지할 수 있을 것이다.

제조 셀 구성문제에 대한 기존의 연구는 부품의 속성(attribute)을 이용하여 분류하는 연구와 기계-부품간에 행렬을 구성한 후 행과 열을 재배치하면서 제조 셀을 형성하는 연구로 구분할 수 있다. 근래에 이루어지고 있는 대부분의 연구는 행렬을 이용하여 그룹을 구성하는 알고리즘이나 휴리스틱의 개발에 초점을 맞추고 있다. McAuley[1972]의 연구나 Seifoddini와 Wolfe[1986]의 연구는 유사한 부품을 생산하는 기계들로 제조 셀을 구성한 후 각 부품들을 적절히 할당하는 순차적인 방법을 제시하였으며, Chan과 Milner[1982]의 DCA(direct clustering algorithm), King[1980]의 ROC(rank order clustering), 그리고 Kusiak과 Chow[1987]의 CIA(cluster identification algorithm) 등은 기계와 부품의 그룹을 동시에 발견하는 방법을 제시하였다.

이들 연구는 대부분이 최적해 혹은 만족해를 찾아내기 위한 알고리즘이나 휴리스틱을 제시하고 있다. 그러나 그룹을 구성하는 문제는 문제의 크기가 증가함에 따라 탐색공간이 지수적으로 증가되는 NP-complete 문제이기 때문에 현실적인 대규모의 문제에서 최적해를 찾는 것은 거의 불가능하다[Kusiak, 1987]. 뿐만 아니라, 문제의 크기가 비교적 큰 경우에는 휴리스틱 기법도 많은 계산시간을 필요로 하기 때문에 적용이 용이하지 않을 수 있다. 예를 들어, King[1980]이 제시한 ROC의 경우, 기계의 수를 m , 부품의 수를 n 이라 할 때, 알고리즘의 복잡도는 $O(mn^2 + m^2n)$ 이며, 기타의 휴리스틱도 유사한 복잡도를 갖는 것으로 알려

져 있다[Kaparthi, 1992].

근래에는 인공지능분야에서 관심의 대상이 되고 있는 신경망(neural networks)이론을 제조 셀 구성 문제에 적용하는 연구가 많이 이루어지고 있다. Kao[1991]와 Karparthi[1991]는 제조 셀 구성을 위해서 역전과 신경망을 이용하였으며, Lee[1992]와 Venugopal[1992]은 시뮬레이티드 어닐링을 이용하였다. Carpenter와 Grossberg[1986, 1988]가 제안한 ART(adaptive resonance theory) 신경망 이론도 Kusiak[1991]과 Kaparachi[1992] 등 많은 연구에서 제조 셀 구성문제의 해법으로 활용되어 좋은 성과를 나타냈다. 이에 따라, 본 논문에서는 고객들의 속성을 이용하여 그룹을 설정하는 전통적인 방법을 이용하지 않고, 제품과 고객간의 관계만을 이용하여 제품-고객군을 형성하도록 ART 신경망을 적용하는 방안에 대한 연구결과를 기술한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 제품-고객군 구성의 개념 및 성과척도에 대해 기술하며, 제 3장에서는 제품-고객군의 형성을 위한 신경망 알고리즘과 컴퓨터 시뮬레이션 결과를 기술한다. 그리고, 제 4장에서는 본 연구의 결과에 대해 정리하면서 향후의 연구방향을 제시하였다.

II. 제품-고객군 구성에 대한 고찰

2.1 제품-고객군 구성의 개념

제품과 고객간의 상관관계를 통하여 제품-고객군을 구성하는 것은 생산시스템에서 기계와 부품군을 형성하기 위해 그룹테크놀로지를 활용하는 과정으로 표현될 수 있는데, 기계-부품군이 적절하게 구성되지 못하면 동일한 부품을 생산하는 기계들이 여러 곳에 산재되어 기계 혹은 부품의 생산에 필요한 이동거리가 길어질

뿐만 아니라 작업을 위한 대기시간이 길어져서 오히려 역효과를 가져올 수도 있다. 이와 같이, 제품-고객군을 형성함에 있어서도 적절하지 못한 군이 형성되면, 특정제품이나 고객이 마케팅 활동에서 누락되거나 별도의 관리가 요구되어 이를 위한 부가적인 비용을 유발시키게 된다. 따라서 다른 그룹간의 의존도(coupling)는 낮추고, 동일한 그룹내에서의 응집력(cohesion)은 높이도록 제품-고객군을 구성하여야 한다.

2.2 제품-고객 테이블

제품과 고객간의 관계를 시각적으로 쉽게 파악하여 제품-고객군의 구성을 용이하게 하는 대표적인 도구로 <표 1>과 같은 제품-고객 테이블(table)을 들 수 있다. 이는 각 제품별로 구입한 경험이 있거나 관심을 갖고 있는 고객들을 테이블의 형태로 표현 한 것으로, 제품과 고객간의 관계는 0 혹은 1의 값을 갖거나, 고객들이 제품에 대해 갖고 있는 선호도에 따라서 가중치를 부여하기도 한다.

<표 1> 제품과 고객의 관계를 나타내는 테이블

고객 제품	1	2	3	4	5	6	7	8	9	10
1		1			1		1			
2					1	1	1			1
3		1				1	1			1
4	1			1		1				
5	1			1					1	
6			1					1		
7			1					1		
8	1								1	

<표 1>은 어떤 고객이 어떤 제품을 구입했는지 혹은 관심이 있는지의 관계를 보여주고 있는데, 이 테이블의 행(row)은 그대로 두고 열(column)만을 재배치할 경우 <표 2>와 같이

몇 개의 제품-고객군을 발견할 수 있다.

<표 2> 열의 위치를 변경하여 생성한 제품-고객군의 예

고객 제품	1	2	3	4	5	6	7	8	9	10
1	1	1		1						
2		1	1	1	1					
3	1		1	1	1					
4			1			1	1			
5						1	1	1		
6									1	1
7									1	1
8						1	1			

<표 2>는 (4,3), (8,6), (8,8) 등 세 개의 원소를 제외한 모든 제품-고객 원소들이 각 그룹에 포함되고 있음을 보여주고 있는데, 이와 같이 각 그룹에서 제외된 원소들은 전사적인 차원에서 효과적으로 관리가 용이하지 않게 된다. 특히, 위에서 제품 8은 어느 그룹에도 속하지 않기 때문에 그 심각성은 더할 수 있다. 따라서, 몇 개의 제품-고객군을 생성할 것이며, 얼마나 많은 제품과 고객들이 각 그룹에 포함되도록 할 것인가가 제품-고객군 형성에 있어서 가장 중요한 의사결정사항으로서, 이는 다음에서 기술하는 성과척도에 따라 달라질 수 있을 것이다.

2.3 성과척도

2.3.1 EE(exceptional elements)

EE는 셀 밖에 위치하는 원소의 수 혹은 이들의 합으로 성과를 측정한다. 예를 들어, <표 2>에서는 세 개의 원소가 셀 밖에 위치하고 있으므로 EE로 측정된 성과는 3이 된다. 그러나 문제의 크기나 유형에 따라 셀 밖에 위치하는 원소의 수가 크게 달라질 수 있으므로, 전체 원소의 수에 대한 비율로 변환하여 평가하는 것이 바람직하다.

2.3.2 ME (measure of effectiveness)

ME는 제품-고객군의 형성이 완료된 후의 제품-고객 테이블에서 각 행과 열의 상하좌우에 위치한 값들의 합으로 평가한다. 이는 값들이 분산되지 않고 집중될수록 ME의 값은 커진다는 원리에 따른 것으로, 그 계산식은 다음과 같다.

$$ME = \frac{1}{2} \cdot \sum_{p=1}^{N_p} \sum_{i=1}^{N_c} x_{p,i} \cdot (x_{p-1,i} + x_{p+1,i} + x_{p,i-1} + x_{p,i+1})$$

여기서, N_p 는 제품의 수, N_c 는 고객의 수, 그리고 $x_{p,i}$ 는 p -행, i -열의 원소값으로 제품 p 에 대한 고객 i 의 선호도로서 1과 0, 혹은 상대적인 가중치를 부여한다. 이 성과척도는 성과측정이 용이하다는 장점을 갖고 있으나, 동일한 제품-고객군 내에서도 제품과 고객의 배치순서에 따라서 ME 값이 달라진다는 단점이 있다.

2.3.3 GE (grouping efficiency)

GE는 제품-고객군에 포함되지 못한 제품-고객 원소들의 개수뿐만 아니라 제품의 이용율(utilization)까지도 고려하여 만든 성과척도로서, 제품-고객군 내에 포함되는 원소들의 비율과 제품-고객군에서 제품이 선호되는 비율을 종합하여 평가한다.

$$\eta = w\eta_1 + (1-w)\eta_2$$

여기서, η_1 과 η_2 는 다음 식에서 보는 바와 같이 제품의 이용율과 각 원소가 제품-고객군 내에 포함되는 비율을 측정된 값이며, w 는 η_1 에 주어진 가중치로 0.0 ~ 1.0의 값을 가질 수 있으나, 일반적으로 0.5로 설정한다[Chen, 1995].

$$\eta_1 = \frac{e_i}{\sum_{i=1}^k d_i c_i}$$

$$\eta_2 = 1 - \left[\frac{e_o}{N_p N_c - \sum_{i=1}^k d_i c_i} \right]$$

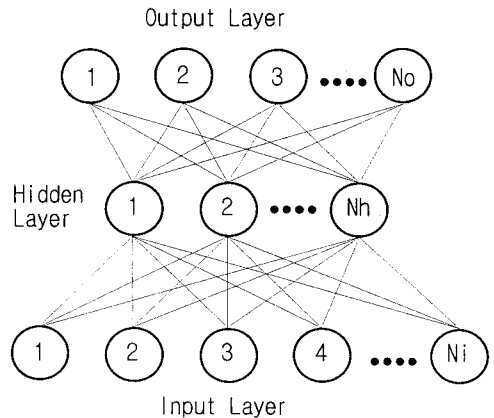
여기에서 각각의 기호가 의미하는 바는 다음과 같다.

- N_p : 제품의 수(행의 수)
- N_c : 고객의 수(열의 수)
- p_i : i 번째 제품-고객군에 속한 제품의 수
- c_i : i 번째 제품-고객군에 속한 고객의 수
- k : 전체 제품-고객군의 수
- e_i : 제품-고객군 안에 있는 원소들의 합
- e_o : 제품-고객군 밖에 있는 원소들의 합

Ⅲ. 신경망을 이용한 제품-고객군의 형성

3.1 신경망(neural networks)의 개념

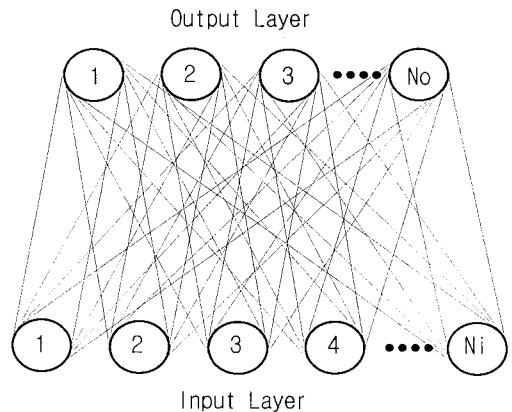
신경망은 <그림 1>에서 보는 바와 같이, 뉴런(neuron)이라 불리우는 처리요소들로 이루어진 하나의 네트워크로, 각 뉴런은 몇 개의 층(layer)을 형성한다. 네트워크 상의 각 층에 위치한 뉴런들은 인접한 뉴런이나 상하위에 있는 다른 뉴런들과 연결되는데, 이 때 각각의 연결강도는 수치형태의 가중치로 표현된다. 또한, 이들 가중치는 특정 학습 알고리즘에 의해 갱신되어 일정한 값으로 수렴해 가게 된다. 일반적으로 신경망은 뉴런들을 연결하는 물리적인 구조, 입력 및 출력을 변환하기 위한 활성화 함수, 그리고 학습 알고리즘 등으로 구성된다.



<그림 1> 역전파 신경망의 구조

3.2 Carpenter-Grossberg 신경망의 구조

Carpenter-Grossberg Neural Networks(CGNN)은 Carpenter와 Grossberg[1986, 1988]가 ART에 대한 연구과정에서 개발한 신경망으로서, 입력값이 주어지면 내부적으로 출력값을 계산하여 적용하는 관찰에 의한 학습방법(learning by observation, unsupervised learning)을 이용한다. 즉, 네트워크 자체에서 출력의 정확성을 평가하여 뉴런간의 가중치를 자동적으로 갱신하기 때문에, 출력에 대한 사전지식이 없는 제품-고객군 형성 문제에 효과적으로 적용될 수 있다.



<그림 2> Carpenter-Grossberg 신경망의 구조

CGNN은 <그림 2>에서 보는 바와 같이 각각 한 개씩의 입력층과 출력층으로만 이루어져 있으며, 입력층과 출력층간의 뉴런들은 서로 2중으로 연결되어 있다. 입력층에는 각 제품에 대한 고객들의 선호도가 입력되므로 입력층의 뉴런수는 고객의 수와 같아야 하며, 출력층은 각 제품이 포함될 제품-고객군을 의미하므로 출력층의 뉴런수는 예상되는 제품-고객군의 최대수가 되어야 한다.

3.3 제품-고객군의 형성을 위한 신경망 알고리즘

Karparthi[1992]는 CGNN을 이용하여 그룹 테크놀로지에서 기계-부품군을 형성하기 위한 알고리즘을 제시하였다. 본 연구에서는 제품-고객군을 형성함에 있어서, 고객들이 각 제품에 대해 구입한 경험이나 관심을 갖는 정도를 가중치로 표현할 수 있도록 CGNN을 수정하였다.

- ① 출력층에서 입력층으로, 그리고 입력층에서 출력층으로 연결된 가중치를 다음과 같이 초기화하고, 입력과 출력의 일치정도를 나타내는 vigilance threshold(ρ)를 설정한다.
 - 출력노드 j 에서 입력노드 i 로의 연결강도: $\forall i, j \quad t_{ji}(0) = 0.5$
 - 입력노드 i 에서 출력노드 j 로의 연결강도: $\forall i, j \quad b_{ij}(0) = 1/(1 + N_c)$
 여기서, N_c 는 고객의 수로서 입력층의 노드수를 결정하며, N_g 는 예상되는 그룹의 최대개수로서 출력층의 노드수를 결정하며, 이는 제품의 개수 N_p 보다 작거나 같은 값을 갖게 된다. 따라서, i 와 j 는 각각 $1 \leq i \leq N_c$ 와 $1 \leq j \leq N_g$ 를 만족하여야 한다.

• ρ 의 값은 0.0~1.0의 범위에서 결정되는데, ρ 가 너무 크면 일치하는 그룹을 발견하기가 어려워져서 많은 수의 그룹이 형성되며, ρ 가 너무 작으면 대부분의 경우에 이를 만족하기 때문에 작은 수의 그룹이 만들어지게 된다.

- ② 제품별로 고객들의 초기 선호도 x_{pi}^0 을 다음과 같이 선행처리하여 정규화한다. 여기서 n_i 는 고객 i 가 구입 혹은 관심을 갖고 있는 제품들의 개수를 나타낸다.

$$x_{pi} = \frac{x_{pi}^0}{\sum_{j=1}^{N_c} x_{pj}^0} \cdot n_i \quad \text{for all } p \text{ \& } i$$

- ③ 제품 p 에 대한 고객들의 선호도로 구성된 입력벡터 X_p 를 신경망의 입력층에 입력한다.
- ④ 출력층에 있는 모든 선택 가능한 노드(j)에 대해서 출력 $O_j(k)$ 를 계산한다.

$$O_j(k) = \sum_{i=1}^{N_c} b_{ij}(k) \cdot x_{pi}$$

여기서, x_{pi} 는 제품 p 에 대한 입력노드 i 의 입력값이며, $b_{ij}(k)$ 는 k 번째 입력에서 입력노드 i 로부터 출력노드 j 로 연결되는 연결강도를 의미한다.

- ⑤ 선택 가능한 모든 출력노드 중에서 가장 큰 출력을 갖는 노드를 선택한다. 이 때, 단계 ⑦에서 일치도가 ρ 보다 작아서 선택이 불가능(disable)하도록 설정된 출력노드는 제외한다.

$$O_{j^*}(k) = \text{Max}_j \{O_j(k)\}$$

- ⑥ 선택된 j^* 노드가 입력노드의 값과 얼마나 잘 일치하는지를 평가하는 vigilance test를 실시한다. 다음과 같이 선택된 출력노드와 입력이 일치된 정도의 합 $O_{j^*}(k)$ 에 대한 입력 값들의 합 I_p 의 비율을 계산한다.

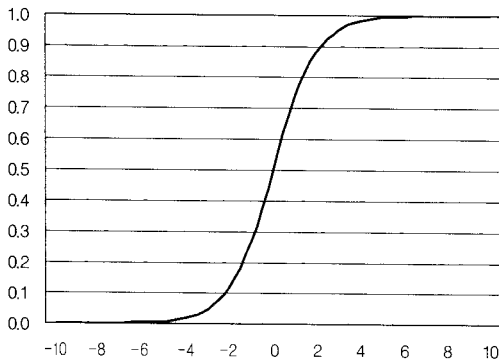
$$I_p = \sum_{i=1}^{N_c} x_{pi}$$

$$O_{j^*}(k) = \sum_{i=1}^{N_c} t_{j^*i}(k) \cdot x_{pi}$$

이 값이 단계 ①에서 설정한 vigilance threshold(ρ) 값보다 크면 설정된 일치도를 만족함을 의미하므로 ⑧로 가서 선택된 출력노드와 모든 입력노드간의 연결 강도를 갱신해 주며, ρ 보다 작으면 ⑦로 이동한다.

- ⑦ 위의 vigilance test에서 최대 출력노드에 대한 일치도가 낮게 나타난 경우, 이 노드가 선택되지 않도록 설정한 후 ④로 돌아가서 다른 출력노드를 검색한다.
- ⑧ 선택된 출력노드(j^*)로부터 입력노드로 연결되는 모든 연결강도를 시그모이드 함수 $1/(1+e^{-t})$ 에 따라서 증감시킨다. 여기서, x_{pi} 의 값이 0이면 t 를 감소시키고, 0이 아니면 t 를 증가시켜서 t_{j^*i} 의 값을 갱신한다. 또한, 입력노드로부터 출력노드로의 연결강도는 다음과 같이 갱신한다.

$$b_{j^*i}(k+1) = \frac{t_{j^*i}(k) \cdot x_{pi}}{0.5 + \sum_{i=1}^{N_c} t_{j^*i}(k) \cdot x_{pi}}$$



<그림 3> 시그모이드 함수의 유형

- ⑨ 각 노드를 연결하는 연결강도 $t_{ji}(k)$ 와 $b_{ji}(k)$ 가 안정되거나 사전에 설정한 횟수

만큼 학습이 이루어질 때까지 단계 ③으로부터 위의 과정을 반복한다.

CGNN에서 학습을 통해 연결강도를 조정하는 과정을 간략히 기술하면 다음과 같다. 예를 들어, k 번째 입력에 따른 결과로 $t_{j^*i}(k)$ 가 0.5의 값을 갖고 있을 때, $(k+1)$ 번째 입력으로 1이 입력되면 $t_{j^*i}(k+1)$ 는 단계 ⑧에서 제시한 계산식인 $1/(1+e^{-1})$ 에 따라서 0.731로 증가되며, $(k+2)$ 번째 입력으로 1이 다시 입력되면 $t_{j^*i}(k+2)$ 는 0.881로 증가된다. 반면에 $(k+3)$ 번째 입력에서 0이 입력되면, $t_{j^*i}(k+3)$ 은 다시 0.731로 감소하게 된다. 즉, 입력노드에 1이 입력되면 t_{j^*i} 의 값은 1.0의 방향으로 접근하고, 0이 입력되면 0.0의 방향으로 접근하므로, 입력데이터를 이용하여 학습을 계속하게 되면 0 혹은 1로 수렴하게 된다.

본 연구에서 CGNN을 수정하여 적용한 알고리즘은 고객들이 각 제품에 대해 갖고 있는 선호도를 선행처리하여 정규화하기 때문에, 고객의 제품에 대한 선호도 값으로서 정수 혹은 실수를 모두 사용할 수 있다는 장점이 있다. 또한, 시뮬레이션 결과 학습횟수를 10회 이하로 하더라도 충분한 학습이 이루어지는 것으로 나타났는데, 이는 많은 계산시간을 필요로 하는 대규모의 문제에서도 비교적 작은 시간내에 제품-고객군을 형성할 수 있을 것이다. <그림 4>는 CGNN 신경망에서 학습이 이루어지는 과정의 일부를 java 프로그래밍 언어의 문법에 따라 기술한 것이다.

3.4 제품-고객군 형성과정 및 결과

신경망에 각 제품에 대한 고객들의 선호를 나타내는 제품-고객행렬을 입력하여 학습시키면, 입력노드와 출력노드를 연결하는 연결강도

```

// 신경망 학습을 통해 그룹을 생성하는 메소드
public void training(int p, double vigilance) {
    int jstar; // 최대 출력값 노드번호
    double Ojstar, Ip; // 입력 및 출력값

    for (int j = 0; j < Np; j++) { // 초기화 및 출력값 계산
        disabled[j] = false;
        for (int i = 0, output[j] = 0; i < Nc; i++)
            output[j] += bij[i][j] * input[p][i];
    }

    do { // 최대 출력노드 선택
        for (int j = 0, int max = 0; j < Np; j++)
            if (max < output[j] && !disabled[j]) {
                max = output[j];
                jstar = j;
            }
        // 입출력 값 계산
        for (int i = 0, Ojstar = 0, Ip = 0; i < Nc; i++) {
            Ojstar += sigmoid(tij[jstar][i]) * input[p][i];
            Ip += input[p][i];
        }

        if (Ojstar / Ip >= vigilance) {
            group[p] = jstar; // 제품을 선택한 그룹에 배정
            update(jstar, p); // 연결강도 갱신 메소드 호출
        } else
            disabled[jstar] = true; // 불능처리
    } while (group[p] == 0); // 그룹이 결정될 때까지 반복
}

// 시그모이드 함수 메소드
public double sigmoid(double k) {
    return (1 / (1 + Math.exp((-1) * k)));
}

```

<그림 4> 자바로 구현한 CGNN 학습 알고리즘

가 갱신되며, 이에 따라 각 제품이 배치될 그룹이 결정된다. 예를 들어, <표 1>에서 신경망에 가장 먼저 입력되는 제품 1에 대한 고객들의 선호도 벡터 (0, 1, 0, 0, 1, 0, 1, 0, 0, 0)에 대해 충분한 학습이 이루어진 상태에서의 연결강도를 이용하여 CGNN의 출력값을 계산하면 다음과 같다.

$$O_1 = 0.22 + 0.22 = 0.44$$

$$O_2 = 0.00$$

$$O_3 = 0.00$$

여기서, 제품 1은 그룹 1에서 가장 높은 출력값을 가지며, 이는 ρ 값으로 설정된 0.3보다 높기 때문에 그룹 1에 할당된다. 이와 같이, 나머지 제품들에 대해서도 동일한 과정을 수행함으로써 각 제품이 속하게 되는 그룹을 결정할 수 있다. 뿐만 아니라, 이미 구성되어 있는 제

품-고객군을 그대로 둔 채로, 새로 도입된 제품을 특정 제품-고객군에 배치할 경우에도 위와 동일한 과정을 거쳐 제품-고객군을 결정할 수 있다. 위의 과정에 따라서 유사제품에 대한 제품군을 형성이 끝나면, 2단계로서 각 고객들이 선호하는 제품들을 가장 많이 포함하고 있는 제품군에 고객들을 할당함으로써 제품-고객군의 형성을 마치게 된다.

<표 3> CGNN을 이용한 제품-고객군 형성 결과

고객 제품	2	5	6	7	10	1	4	9	3	8
1	1	1		1						
2		1	1	1	1					
3	1		1	1	1					
4			1			1	1			
5						1	1	1		
8						1		1		
6									1	1
7									1	1

<표 3>은 앞에서 기술한 <표 1>의 예제에 대해 vigilance threshold(ρ)의 값을 0.3으로 설정한 후 10회의 학습을 실시하여 형성한 제품-고객군을 보여주는데, 이 결과를 앞에서 기술한 세 가지 성과척도에 따라 평가하면 다음과 같다.

① EE (exceptional elements)

<표 3>에서는 제품 4, 고객 6에 해당하는 원소 한 개만이 구성된 제품-고객군의 밖에 존재하므로 EE는 1이 된다.

② ME (measure of effectiveness)

ME를 구하기 위해 행렬에서 1의 값을 갖고 있는 각 원소에 대해 상하좌우에 인접해 있는 1의 개수를 세어 표로 구성하면 <표 4>와 같으며, ME는 각 원소가 갖고 있는 값들의 합을

2로 나눈 23.0이 된다.

<표 4> ME로 측정된 제품-고객군 형성의 성과

고객 제품	2	5	6	7	10	1	4	9	3	8
1	1	2		1						
2		2	3	4	2					
3	0		3	3	2					
4			1			2	2			
5						3	3	2		
8						1		1		
6									2	2
7									2	2

③ GE (grouping efficiency)

GE를 구하기 위해서는 먼저 η_1 과 η_2 에 부여할 가중치를 결정해야 하는데, 각각 0.5의 가중치를 갖고 있는 것을 가정하여 GE를 구하면 다음과 같다.

$$\eta_1 = \frac{e_i}{\sum_{i=1}^k p_i c_i} = \frac{22}{3*5 + 3*3 + 2*2} = \frac{22}{28}$$

$$\eta_2 = 1 - \left[\frac{e_o}{N_p N_c - \sum_{i=1}^k p_i c_i} \right] = 1 - \frac{1}{8*10 - 28} = \frac{51}{52}$$

$$\eta = w \eta_1 + (1-w) \eta_2 = 0.5 * \frac{22}{28} + 0.5 * \frac{51}{52} = 0.8832$$

3.5 가중치를 갖는 선호도에 대한 문제

고객들이 각 제품에 대해 갖고 있는 선호도가 다를 경우 그 중요도에 따라서 가중치를 부여하여 신경망에 입력하면, 신경망에서는 입력된 값들을 선행처리한 후 학습을 수행한다. <표 5>는 각 셀이 가중치를 갖도록 앞의 <표 1>을 수정한 것이다.

<표 5> 제품과 고객의 관계가 가중치를 갖는 예

고객 제품	1	2	3	4	5	6	7	8	9	10
1		1			1		1			
2					2	1	3			1
3		3				1	2			1
4	3			2		3				
5	1			1					2	
6			3					1		
7			1					1		
8	1								1	

즉, 고객 1은 제품 4, 5, 8에 모두 관심을 갖고는 있으나, 제품 4에 대해 상대적으로 많은 관심을 갖고 있음을 의미하는데, 개인의 선호도를 가중치로 변환하는 방법은 본 연구의 범위를 벗어나기 때문에 상세한 설명은 생략한다. 위의 테이블에 나타난 데이터를 이용하여 신경망을 학습시키면 <표 6>과 같은 제품-고객군이 형성된다.

<표 6> 가중치를 갖는 문제에서의 제품-고객군 형성결과

고객 제품	2	5	7	10	1	4	6	9	8	3
1	1	1	1							
2		2	3	1			1			
3	3		2	1			1			
4					3	2	3			
5					1	1		2		
8					1			1		
6									1	3
7									1	1

가중치를 부여했을 때의 제품-고객군의 형성 결과는 가중치를 갖지 않은 문제에서 첫 번째 그룹에 속해 있었던 고객 6이 두 번째 그룹으로 이동되었음을 볼 수 있는데, 이는 고객 6이 제품 2와 3에 갖고 있는 선호도의 합보다 제품 4에 대해 갖고 있는 선호도가 더 크기 때문에

나타난 결과로서, 타당한 결과를 도출하고 있음을 볼 수 있다.

3.6 컴퓨터 시뮬레이션

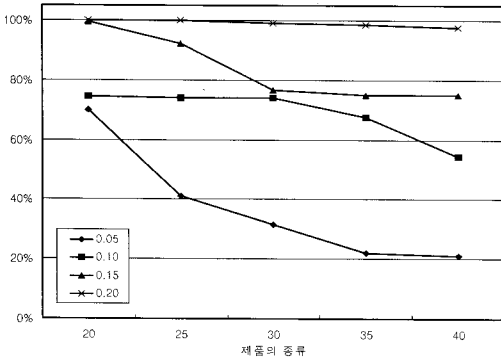
3.6.1 시뮬레이션 문제의 구성

본 연구에서 제시한 수정된 CGNN 알고리즘이 문제의 특성에 따라 어떠한 행태를 보이며, 대규모 문제에 적용이 가능한지를 평가하기 위해 다음과 같은 컴퓨터 시뮬레이션을 실시하였다. 예비 시뮬레이션으로부터 고객의 수가 10,000명을 넘는 경우에도 펜티엄-1GHz 급 PC에서 충분히 처리될 수 있는 것으로 나타났다. 다양한 특성의 문제를 생성하여 반복적인 시뮬레이션을 수행하기 위해서는 많은 시간이 요구되기 때문에 고객의 수는 1,000명을 가정하였다. 또한, 제품은 20~40개, 각 고객들이 기업이 취급하는 전체 제품중에서 관심을 갖고 있는 제품의 비율을 의미하는 계산되는 밀도(density)는 5~20%, 그리고 최적해에서 그룹의 개수는 2~5개가 되도록 시뮬레이션 프로그램을 이용하여 각각 50개씩의 예제를 생성하였다. 여기서 각 시뮬레이션 예제들은 제품-고객군 구성성과의 측정이 용이하도록 제품-고객 관계테이블에서 0이 아닌 원소들이 임의로 설정한 2~5개의 그룹에 모두 포함되도록 하였다. 즉, 최적해에서 각 그룹을 벗어나는 원소는 없는 것으로 가정하였다. 따라서, 본 시뮬레이션의 성과적도는 앞에서 기술한 EE 성과적도를 변형하여 전체 원소중에서 그룹에 포함되는 원소들의 비율로 측정하였으며, 이 비율이 100%가 되면 최적해를 발견하였음을 의미한다.

3.6.2 밀도에 따른 성과비교

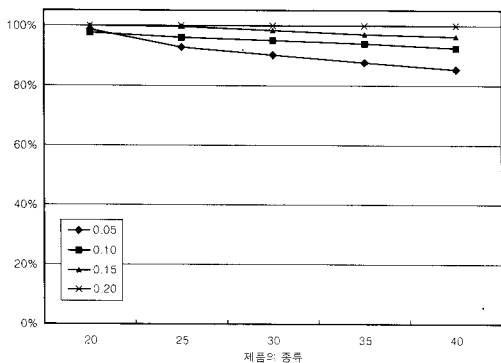
<그림 5>는 제품의 종류가 20개로부터 40개까지 증가할 때 제품-고객 관계 테이블의 밀도에 따라서 최적해를 발견한 비율이 어떻게 변

화하고 있는지를 보여주고 있다. 여기서, 밀도는 전체 테이블에서 0이 아닌 값이 차지하는 비율 혹은 한 명의 고객이 전체 제품에 대해서 구입 혹은 관심을 갖고 있는 제품의 비율을 의미하는 것으로, 0.05~0.20 범위에서 시뮬레이션을 실시하였다.



<그림 5> 밀도에 따른 최적해 발견 비율 추이

<그림 5>는 위에서와 동일하게 제품의 종류가 20개로부터 40개까지 증가할 때 제품-고객 관계테이블의 밀도에 따라서 각 그룹에 포함되는 원소들의 비율이 어떻게 변화하는지 보여주고 있다. 여기서, 성과척도의 값이 100%라는 것은 제품-고객 관계 테이블에서 0이 아닌 모든 원소들이 각 제품-고객군 포함됨을 의미하는 것으로 최적의 그룹을 형성하였음을 나타낸다.

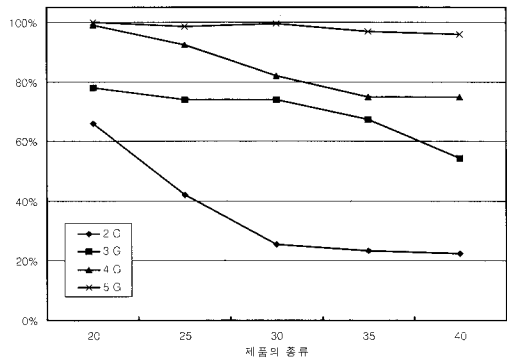


<그림 6> 밀도에 따른 EE 성과 추이

<그림 5>와 <그림 6>으로부터 제품-고객 관계테이블의 밀도가 높을수록, 즉 한 명의 고객이 관심을 갖고 있는 제품의 개수가 많을수록 더욱 정확한 그룹이 생성됨을 보여준다. 그러나, 기업에서 취급하는 제품의 종류가 증가할수록 문제의 복잡도가 증가되어 최적의 그룹을 생성하는 비율은 감소하는 것으로 나타났다.

3.6.3 그룹의 개수에 따른 성과비교

<그림 7>은 제품의 종류가 20개로부터 40개까지 증가할 때 예제가 갖는 그룹의 개수에 따라서 최적해를 발견한 비율이 어떻게 변화하고 있는지를 보여주고 있다. 시뮬레이션 결과, 그룹의 개수가 많을수록 보다 정확한 그룹을 생성하는 것으로 나타났으며, 위에서와 마찬가지로 제품의 종류가 증가할수록 생성되는 그룹의 정확도는 저하되는 것으로 나타났다.

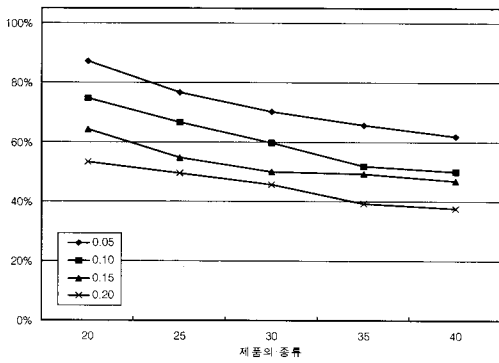


<그림 7> 그룹의 개수에 따른 최적해 발견 비율 추이

3.6.4 CGNN의 ρ 값에 따른 성과비교

<그림 8>은 CGNN 신경망에서 그룹의 개수를 결정하는 유일한 파라미터인 ρ 의 값에 따라 최적해를 발견한 비율이 어떻게 변화하고 있는지를 보여주고 있다. 본 연구의 시뮬레이션 결과에서는 ρ 의 값이 높아질수록 성과가 저하되는 것으로 나타났는데, 이는 앞에서도 기술한 바와 같이 최적의 그룹에서 0이 아닌

모든 원소들이 그룹에 포함되도록 문제를 구성한 후 시뮬레이션을 실시했기 때문에 나타난 현상으로 분석된다. 즉, 시뮬레이션의 성과측정을 용이하게 하기 위한 문제의 구성에 따른 것으로서 본 시뮬레이션의 결과를 일반화시킬 수는 없으며, 문제의 특성 및 생성하고자 하는 그룹의 개수에 적합한 ρ 값을 제안하기 위한 추가적인 연구가 필요할 것이다.

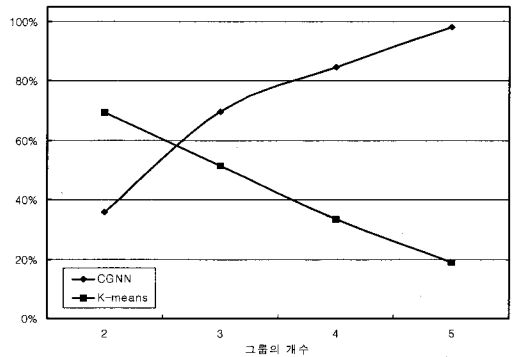


<그림 8> CGNN의 ρ 에 따른 최적해 발견 비율 추이

3.6.5 k-means 클러스터링과의 성과비교

본 연구에서 제시한 수정된 CGNN은 형성하고자 하는 그룹의 개수가 정해져 있지 않은 경우에도 ρ 값에 따라서 자동적으로 그룹을 형성하는 특징을 갖고 있다. 따라서, 주어진 문제에 대해 특정한 개수의 제품-고객군을 형성하기 위해서는 ρ 값을 반복적으로 변화시키면서 그룹을 형성하여야 한다. 이와 같이 그룹의 개수가 사전에 정해져 있는 경우에는 일반적으로 k-means 알고리즘을 많이 사용하고 있다. k-means 알고리즘에서는 먼저 임의로 k개의 개체를 선택하여 각각을 하나의 그룹으로 설정한 후 나머지 개체들을 가장 근접한 그룹에 할당한다. 다음으로, 각 그룹의 중심점을 다시 계산한 후에 모든 개체를 가장 근접한 그룹에 다시 할당하는 과정을 반복한다.

따라서, 본 연구에서 제시한 수정된 CGNN과 k-means 알고리즘의 성과를 직접적으로 비교하는 것은 사실상 불가능하다. 그러나, 본 연구에서는 사전에 2~5개의 그룹을 갖는 예제를 생성하여 시뮬레이션을 실시했기 때문에 k-means 알고리즘을 적용하는 것이 가능하였다. 이에 따라, 동일한 예제에 대해 시뮬레이션을 실시한 결과 예제가 갖고 있는 그룹의 개수에 따라 두 가지 알고리즘을 적용했을 때 최적해를 찾는 비율을 그림으로 나타내면 <그림 9>와 같다. 여기서, 그룹의 개수가 증가할수록 k-means 알고리즘의 성과는 낮아지지만 본 연구에서 제시하는 CGNN의 성과는 향상되는 것으로 나타났는데, 이와 같이 그룹의 개수가 증가함에 따라서 성과가 반대방향으로 움직이는 현상에 대한 상세한 분석은 향후 연구에서 계속하고자 한다.



<그림 9> CGNN과 k-means의 성과비교

IV. 결론 및 향후 연구계획

알고리즘이나 휴리스틱을 이용하여 기계-부품군을 구성하는 문제를 풀고자 하는 노력이 많이 이루어져 왔으나, 관찰에 의한 학습방식을 이용하는 CGNN은 단순하면서도 강력한 접근법으로 알려져 있다. 이에 따라, 본 연구에서는 CGNN을 제품-고객군의 형성문제에 도입

하여 이에 적합하도록 수정하여 적용하는 방안을 제시하였다. 연구결과, 제품과 고객간의 관계에 있어서 0과 1의 이진값을 갖는 문제뿐만 아니라 제품별로 선호도가 다를 경우의 제품-고객군의 형성문제에 있어서도 타당한 결과를 도출하는 것으로 나타났다. 또한, 컴퓨터 시뮬레이션 결과 비교적 대규모의 문제에서도 좋은 성과를 나타냈다.

본 논문의 예에서는 제품-고객군을 구성할 때 제품들의 그룹을 먼저 형성한 후에 고객을 배치하였는데, 고객을 먼저 배치하면 각 고객의 그룹의 형성하기 위해 사용되는 속성에 해당하는 제품의 수가 고객의 수보다 적은 것이 일반적이므로 적은 수의 속성을 이용하여 그룹을 형성하는 결과를 가져와서 제품-고객군 구성의 성과가 더 나빠질 수 있기 때문이다. 그러나, 제품을 이용하여 먼저 그룹을 설정할 경우 새로운 고객을 각 그룹에 배치하기 위해서는 추가적인 학습과정이 요구되는 단점이 있기

때문에, 신경망의 입력층에 고객별 제품벡터를 입력하여 고객들의 그룹을 먼저 형성한 후에 제품을 배치하는 방안을 연구하고자 한다. 신규고객이나 제품에 대해서는 사례기반 추론을 이용하여 각 그룹에 배치하는 방안도 함께 연구될 것이다.

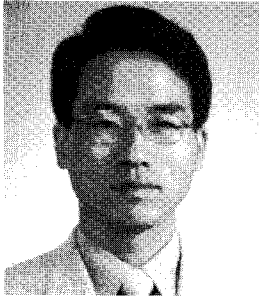
본 연구의 한계점으로서, 본 연구에서는 시뮬레이션 프로그램이 자동적으로 생성한 예제를 이용하여 제품-고객군을 형성하는 시뮬레이션을 실시하였으나, 향후에는 이를 현실의 문제에 적용하는 연구를 수행하고자 한다. 아울러, 일반적으로 CRM에서는 고객의 속성정보를 이용하여 고객군을 형성함으로써 안정적인 고객군을 유지하고 있지만, 본 연구에서는 고객들이 관심을 갖고 있는 제품들에 대한 정보를 이용하기 때문에 시간의 흐름에 따라 고객군이 변화하는 문제점이 발생할 수 있을 것으로 사료된다. 따라서, 지속적인 연구를 통해 본 연구의 적용효과성을 검증할 필요가 있을 것이다.

〈참 고 문 헌〉

- [1] 김기운, 김성근, 김유경, "데이터 웨어하우스 기반의 고객관계관리 모델링: 프로세스 및 데이터 관점," *Information System Review*, 제2권 2호, 2000, pp. 283-299.
- [2] 박명섭, 황인수, "그룹 테크놀로지에서의 제조 셀 구성을 위한 신경망기반 프로토타입 의사결정지원 시스템," *경영학연구*, 제25권 3호, 한국경영학회, 1996. 8, pp. 55-81.
- [3] Alexander Hinneburg, Daniel A. Keim, "Clustering Techniques for Large Data Sets-- From the Past to the Future," *In Proc. of ACM SIGMOD international Conferences on KDD*, San Diego, CA, USA, August 1999.
- [3] Burbidge, J.L., "Production flow analysis," *Production Engineer*, Vol. 42, 1963, pp. 742.
- [4] Carpenter, G.A. and S. Grossberg, "Neural dynamics of category learning and recognition: attention, memory consolidation, and amnesia," *In Brain Structure, Learning, and Memory*, R.N. Davis and E. Wegman (eds) (AAAS Symposium Series).
- [5] Carpenter, G.A. and S. Grossberg, "ART2: self-organization of stable category recognition codes for analog input patterns," *Applied Optics*, Vol. 26, 1987, pp. 4919-4930.
- [6] Carpenter, G.A. and S. Grossberg, "The ART of adaptive pattern recognition by a self-organizing neural network," *Computer*,

- Vol. 21, No. 3, 1988, pp. 77-88.
- [7] Carpenter, G.A. and S. Grossberg, "ART3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures," *Neural Networks*, Vol. 3, 1990, pp. 129-152.
- [8] Chan, H.M. and D.A. Milner, "Direct clustering algorithm for group formation in cellular manufacturing," *Journal of Manufacturing Systems*, Vol. 1, 1982, pp. 65-75.
- [9] Chandrasekharan, M.P. and R. Rajagopalan, "An ideal seed non-hierarchical clustering algorithm for cellular manufacturing," *International Journal of Production Research*, Vol. 24, 1986, pp. 451-464.
- [10] Chen, S.J. and C.S. Cheng, "A neural network-based cell formation algorithm in cellular manufacturing," *International Journal of Production Research*, Vol. 33, No. 2, 1995, pp. 293-318.
- [11] Kalakota, R. and M. Robinson, *e-business: Roadmap for Success*, Addison Wesley, 1999.
- [12] Kao, Y. and Y.B. Moon, "A unified group technology implementation using the backpropagation learning rule of neural networks," *Computer and Industrial Engineering*, Vol. 20, 1991, pp. 425-437.
- [13] Kaparthi, S. and N.C. Suresh, "A neural network system for shaped-based classification and coding of rotational parts," *International Journal of Production Research*, Vol. 29, 1991, pp. 1771-1784.
- [14] Kaparthi, S. and N.C. Suresh, "Machine-component cell formation in group technology: a neural network approach," *International Journal of Production Research*, Vol. 18, 1992, pp. 213-232.
- [15] King, J.R., "Machine-component grouping in production flow analysis: an approach using a rank order clustering algorithm," *International Journal of Production Research*, Vol. 18, 1980, pp. 213-232.
- [16] Kusiak, A. and W.S. Chow, "Efficient solving of the group technology problem," *Journal of Manufacturing Systems*, Vol. 6, No. 2, 1987, pp. 117-124.
- [17] Kusiak, A. and Y. Chung, "GT/ART: Using neural network to form machine cells," *Manufacturing Reviews*, Vol. 4, 1991, pp. 293-301.
- [18] Lee, S. and H.P. Wang, "Manufacturing cell formation: a dual-objective simulated annealing approach," *International Journal of Advanced Manufacturing Technology*, Vol. 7, 1992, pp. 314-320.
- [19] McAuley, J., "Machine grouping for efficient production," *The Production Engineer*, Vol. 52, 1972, pp. 53-57.
- [20] Michel J., A. Berry, and Gordon Linoff, *Data Mining Techniques: For Marketing, Sales, and Customer Support*, John Wiley & Sons, Inc., 1997.
- [21] Seifoddini, H. and P.M. Wolfe, "Application of the similarity coefficient method in group technology," *IEEE Transactions*, Vol. 18, 1986, pp. 271-277.
- [22] Venugopal, V. and T.T. Narendran, "Cell formation in manufacturing systems through simulated annealing: an experimental evaluation," *European Journal of Operational Research*, Vol. 63, 1992, pp. 409-422.

◆ 저자소개 ◆



황인수 (Hwang, Insoo)

현재 전주대학교 정보기술학부 조교수로 재직중이다. 고려대학교 경영학과를 졸업하고 동 대학원에서 경영정보시스템을 전공하여 석사 및 박사학위를 취득하였으며, 산업연구원(KIET) 물류·유통연구센터의 연구원을 역임하였다. 주요 관심분야는 e-Business, CRM, 데이터 마이닝, 경영문제 해결을 위한 인공지능기법의 활용 등이다.

◆ 이 논문은 2001년 6월 4일 접수하여 1차 수정을 거쳐 2001년 10월 4일 게재확정되었습니다.