

데이터마이닝을 이용한 수주생산시스템의 공정계획방안

오경모¹ · 박창권^{2*}

¹제론정보기술(주) / ²울산대학교 산업정보경영공학부

Process Planning Method under Make-to-Order Production System using Data Mining

Kyung-Mo Oh¹ · Chang-Kwon Park²

¹Zerone Information Technology co. Ltd., Ulsan, 680-749

²Department of Industrial Engineering, University of Ulsan, Ulsan, 680-749

The manufacturing industry with Make-to-Order production system is difficult to decide the standard information for the product and the demand is variable to estimate. In this paper, we concerned with the process planning method using data mining in the manufacturing industry with Make-to-Order environment. The subject of our study is the industry transformer plant which is received an diverse order of customer and then produced the product. Currently, process planning method is classified the standard information by hand based on the acquired knowledge through the experience. The standard information stored the various information, such as work sequence, time and so on. This process planning method needs an experts which possesses the field experience for several years. For the product specification which is varied in each order, current process planning method is not efficient due to need many times

To solve this problem, we extract the information using data mining process for each processing time, and then construct the knowledge base. We propose a method which is the process planning of the industry transformer product in Make-to-Order environment using the knowledge base.

Keywords: process planning, Make-to-Order production, data mining, standard information

1. 서론

최근의 제조업환경은 치열해진 기업 간의 경쟁, 원자재 가격의 상승, 제품의 수명주기 단축, 고객의 불규칙한 주문시기 등을 수용해야 하는 많은 어려움에 처해 있다. 특히, 수주생산환경에서는 제품에 대한 표준을 정하기가 어렵고, 가변적인 수요에 대한 예측에 어려움이 많다. 그러나 고객의 주문에 대하여 빠른 생산소요시간을 제공하여야 하며, 빈번한 주문의 변경에 신속하게 대처해야 한다. 이러한 환경에서는 고객의 주문에 따라

제조공정이나 작업방법도 제품의 사양에 따라 달라지며, 제품을 제작하는 작업시간도 매우 다양하여 공정의 정보를 표준화 하는 것이 현실적으로 어렵다. 제품 사양을 만족하는 신속한 공정계획(process planning) 수립은 수주생산형태를 갖는 제조업에서는 매우 절실하다.

고객의 주문에 의해 제품의 사양이 다양하게 발생되어, 표준 제품을 구성하기가 힘든 산업용 변압기 공장의 경우도 대표적인 수주생산형태를 갖는다. <그림 1>은 다양한 고객 주문을 받아 생산을 하는 산업용 변압기 공장의 생산관리절차이다. 현

이 논문은 2004년 울산대학교 연구비에 의하여 연구되었음.

*연락처 : 박창권 교수, 680-749 울산광역시 남구 무거2동 산29번지 울산대학교 산업정보경영공학부, Fax : 052-259-2180,

E-mail : ckpark@mail.ulsan.ac.kr

2004년 12월 29일 접수, 1회 수정 후 2005년 5월 13일 게재 확정.

제의 공정계획방법은 영업에서 수주정보가 입력되면 오랜 경험을 통하여 획득된 지식을 바탕으로 하여 수작업으로 표준정보를 분류하고 있다. 표준정보에는 각 작업의 순서, 작업시간 등 다양한 정보가 포함되어 있다. 이러한 공정계획은 항상 수년 간의 현장경험을 가진 전문가를 필요로 한다. 매 주문마다 달라지는 제품사양에 대해 수작업으로 표준정보를 산출하고 공정계획을 수작업으로 수행하므로 소요시간이 많이 걸리는 등의 비효율적인 요소를 갖고 있다.

본 연구에서는 공정계획수립에 필요한 공정별 소요시간을 데이터마이닝 기법을 활용하여 정보를 추출, 지식 베이스(knowledge base)를 구축하고, 이것을 활용하여 주문생산환경에서 수주되는 산업용 변압기제품의 공정계획을 신속하게 지원할 수 있는 방안을 제시하고자 한다. 이러한 문제를 해결하기 위해 과거 수년 간 진행된 실적 데이터를 근간으로 하여 데이터마이닝 기법을 적용하여 공정별 소요시간을 분류, 정제하여 지식 베이스를 만들고, 사례기반추론 기법을 결합하여 주문된 제품과 유사한 제품의 공정계획정보를 신속히 추출하여 공정계획수립을 지원하는 하이브리드형 데이터마이닝 기법을 적용하는 것이다. 2장에서는 데이터마이닝의 이론적 배경을 다루고, 3장에서는 공정계획방안을 제시하고, 4장에서는 실험 및 결과를 제시하며, 마지막 5장에서는 결론 및 향후 연구방향을 다루고자 한다.

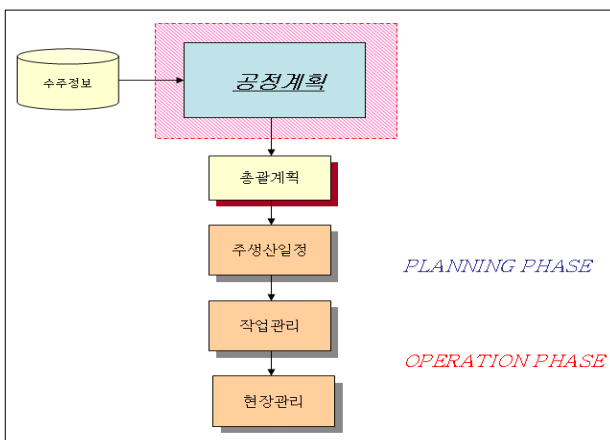


그림 1. 수주생산형태의 생산관리절차.

2. 데이터마이닝에 대한 이론적 배경

2.1 사례기반추론

사례기반추론(case-based reasoning; CBR)은 새로운 문제가 주어지면 이러한 문제를 해결하기 위해 과거의 실적 데이터 중에서 유사한 경험을 검색한다. 검색된 과거 실적 데이터에 직접 또는 부분적으로 수정을 하여 주어진 문제를 해결하는 과정이다. 이러한 문제해결과정은 과거의 경험한 사례를 이용함으로

써 의사결정에 효율성을 증대시킨다. 사례기반추론 과정은 학자에 따라 여러 단계로 세분화되고 있다.

Aamodt and Plaza(1994)은 사례기반추론 과정을 검색, 재사용, 수정, 유지(보관)라는 4단계로 나누었다. Slade(1991)는 사례기반추론 과정을 과정모형(process model)을 통해 새로운 문제의 속성에 대한 색인부여, 검색, 수정, 테스트 및 저장이라는 5단계로 나눈 후 실패에 대한 설명단계를 추가해서 총 6단계로 나누었다. McGovern, et al.(1994)은 사례기반추론 과정을 문제의 명세화, 검색, 선택, 해결안의 구축, 해결안의 실험, 해결안의 시행, 해결안의 평가 및 사례 기억장치의 갱신이라는 8단계로 나누었다. 이렇게 사례기반추론 과정은 학자에 따라 세분화 정도에 차이는 있으나 상호 매우 유사한 과정으로 나타나고 있다.

2.2 데이터마이닝

데이터마이닝은 수집된 데이터에서 아직 발견되지 않은 가치 있는 패턴을 찾는 것을 목적으로 하는 기술의 집합을 지칭하는 것이다. 데이터마이닝의 목적은 과거 행위의 분석을 바탕으로 향후 행위를 예측하는 의사결정지원을 위한 모델을 만드는 것이다. CRM(customer relationship management)을 위한 고객관계관리 데이터마이닝은 주로 마케팅 분야나 영업/판매 분야, 금융 서비스에 주로 사용되고 있으나, 제조분야에 사용된 경우는 매우 빈약한 실정이다.

최근 들어, 제조업분야에 적용된 사례를 보면, 품질분야에 대한 연구로는 Aho et al.(1999)는 제품의 품질특성치를 이용하여 데이터마이닝 기법을 적용한 최적 공정조건탐색을 수행하였다. Jhee(2003)은 데이터마이닝에 의한 생산공정의 모형화를 통하여 신경망과 사례기반추론을 결합하여 효과적인 품질설계 시뮬레이션 기능을 구현함으로써 산업현장에 직접적으로 활용 가능하도록 하였다. Lee(2001)은 의사결정모형 가운데 하나인 CHAID(Chi-Square automatic interaction detection), 로지스틱 회귀모형, 이들을 이용한 베깅모형 등 4가지의 예측모형에 대한 오분류율과 훈련시간을 모의실험 비교를 통하여 주어진 알고리즘들의 효율성을 평가하였다. Jhee et al.(1998)는 공급사슬 관리의 마케팅, 생산 및 정보시스템 측면에서 데이터마이닝을 활용한 주요 연구흐름 및 방법론을 살펴보고, 이를 응용한 의사결정지원 시스템(decision support system; DSS) 구조를 제시하였다. Back et al.(2003)는 실시간으로 수집되는 기계상태의 데이터로부터 기계의 고장원인을 점진적으로 발견하는 데이터마이닝 기법을 적용한 적응형 의사결정 트리(adaptive decision tree) 알고리즘을 제시하였다.

2.3 데이터마이닝과 사례기반추론의 결합

Kim(2003)은 인터넷 기반하에서 과거 고객의 구매 패턴을 연관성 규칙으로 분석 적용하여 규칙을 찾아내고, 사례에 대한 학습을 통하여 지능적인 구매지원이 가능하도록 사례기반

추론 기법을 결합한 하이브리드 구매지원 메커니즘을 제시하였다. 이상과 같은 기존 연구들을 적용기법으로 정리하면 <표 1>과 같다. 본 연구의 경우 <*>는 데이터마이닝 기법과 사례기반추론 기법을 결합하여 공정계획수립을 지원하는 적용사례로 구분되며, 수주생산형태를 갖는 생산시스템에 적용하고자 한다.

업용 변압기의 생산공정을 대상으로 하고자 한다. <그림 2>에 전형적인 산업용 변압기 생산공정흐름을 나타내고 있다. 계약을 통하여 고객과의 수주가 성립되면 수주의 세부내용에 따라 생산공정에 대한 설계와 자재준비 등 일련의 공정들이 우선순위에 따라 이루어지는 전형적인 주문생산방식을 따르고 있다.

생산공정의 순서에 따라 공정별 우선순위가 정해지게 되므로 선행 공정의 작업지연이 후속 공정에 연속적으로 영향을 미치게 되어 공정별 목표납기의 준수가 엄격히 요구된다.

<그림 3>은 산업용 변압기 생산공정의 세부공정도이다. 각 공정은 절연물이 투입되어 총 조립을 거쳐 완성된 제품을 시험을 통하고, 이동하기 위한 해체작업을 거치는 총 25개의 대공정으로 분류된다. 대공정이란 단위작업의 묶음을 지칭하며, 대

3. 수주생산시스템의 공정계획 지원방안

3.1 산업용 변압기의 생산공정

본 연구는 전형적인 수주생산시스템의 특성을 갖고 있는 산

표 1. 데이터마이닝과 사례기반추론 관련 연구의 분류

구분	적용기법	데이터마이닝	사례기반추론	데이터마이닝 & 사례기반추론
알고리즘 설계		Kang <i>et al.</i> (2000) Lee(2001)	Aamodt and Plaza(1994) McGovern <i>et al.</i> (1994) Slade(1991)	
적용사례 (제조업/금융/서비스)		Aho <i>et al.</i> (1999) Back <i>et al.</i> (2003) Han <i>et al.</i> (2002) Jhee <i>et al.</i> (1998) Jhee(2003)		Kim(2003) <*>

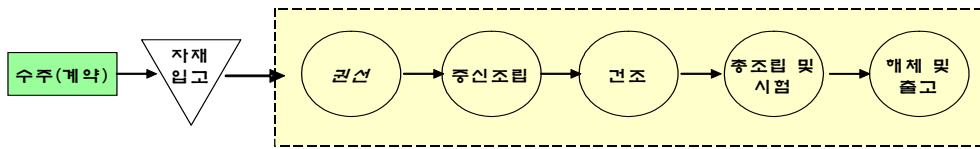


그림 2. 산업용 변압기 생산공정의 대공정 흐름도.

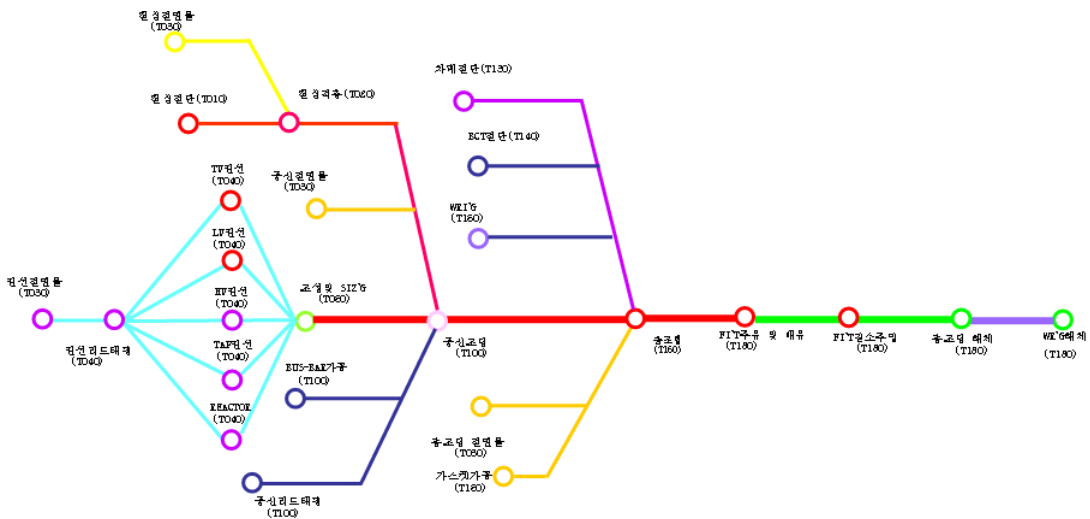


그림 3. 산업용 변압기 생산공정의 세부공정 흐름도.

공정은 세부단위작업 2~20개 정도로 구성되는 공정단위이다. 수주된 제품은 사양에 따라 1개월~10개월 정도의 공정 소요시간을 거쳐서 제품이 완성된다. 제품의 사양에 따라 공정별 소요시간의 차이가 많으며, 이러한 공정별 소요시간정보를 사전에 추출하여 부하분석 등의 의사결정을 하는 것이 매우 중요하다. 현재 적용하고 있는 공정소요시간의 추출방법은 과거 실적을 근간으로 수작업에 의해 진행되고 있다. 공정계획 담당자가 과거의 경험을 토대로 수주된 각 제품의 특성을 파악하여 분석을 하다 보니 주관적인 의사결정요소가 많이 반영되어 공정계획정보의 신뢰성이 떨어지는 실정이다. 또한, 수작업으로 인한 데이터 분류에 과도한 시간이 소요되고 있다. 이러한 문제점을 해결하기 위해서 보다 과학적이면서 체계적으로 공정계획정보를 추출하는 방법의 개발이 요구되고 있다.

3.2 공정계획 지원절차

공정계획을 신속하게 지원할 수 있는 공정계획 지원절차는 <그림 4>와 같은 전체 6단계로 구분된다. 각 과정에 대하여 내용을 자세히 살펴보면 다음과 같다.

1단계 : 데이터 추출 및 탐색

과거 실적 데이터를 근간으로 하여 프로젝트 정보를 취합하여 구성한다. 추출하는 방법은 비복원 랜덤 샘플링하여 프로젝트를 대공정기준으로 실적을 구성하였다. 실적정보에는 수많

은 요소의 실적이 존재하나 본 연구에 맞는 항목만을 추출하여 리드타임 원시 데이터베이스에 저장한다. 그리고 본격적인 모형구축에 앞서 분석하고자 하는 데이터 세트에 대한 면밀한 검토가 수행되어야 한다.

즉, 데이터 세트를 구성하는 변수 중 결측치가 많은 데이터는 없는지, 관측값들이 어느 한 부분으로 치우치게 몰려 있지는 않은지, 이름은 서로 다르나 동일한 내용을 갖는 변수는 없는지 등의 사항이 자세히 검토되어야 한다. 또한 이러한 문제들 없이 변수들이 잘 정리되었다 하더라도, 효과적인 모형구축을 하기 위해서는 변수 상호간의 관계에 대하여 충분한 탐색을 실시해야 한다. 변수 상호간의 관계파악을 위한 통계적인 개념을 생각하기에 앞서, 직관적으로 이해하기 쉬우면서도 많은 내용을 파악할 수 있는 방법으로 그래프를 이용하는 방법이 있다. 본 연구에서의 탐색은 통계분석도구로 알려진 SAS Insight를 이용하여 변수 간의 탐색적 분석을 실시하였다.

2단계 : 데이터 정제 및 변환

샘플링한 데이터는 과거에 수행한 데이터를 그대로 추출한 결과이므로 본 연구에 맞는 데이터 세트를 변환하는 과정이 필요하다. 추출된 데이터는 실제로 대공정별로 산출된 데이터이며 리드타임 정보로 변환하기 위해서는 각 대공정별 투입인원으로 나누어 실제 걸리는 시간을 산출해야 한다. 그래서 목적에 맞는 모형의 데이터로 구성하기 위해 각 대공정별 인원을 나누어 산정한 데이터 세트를 만들었고, 수주번호는 순번으로

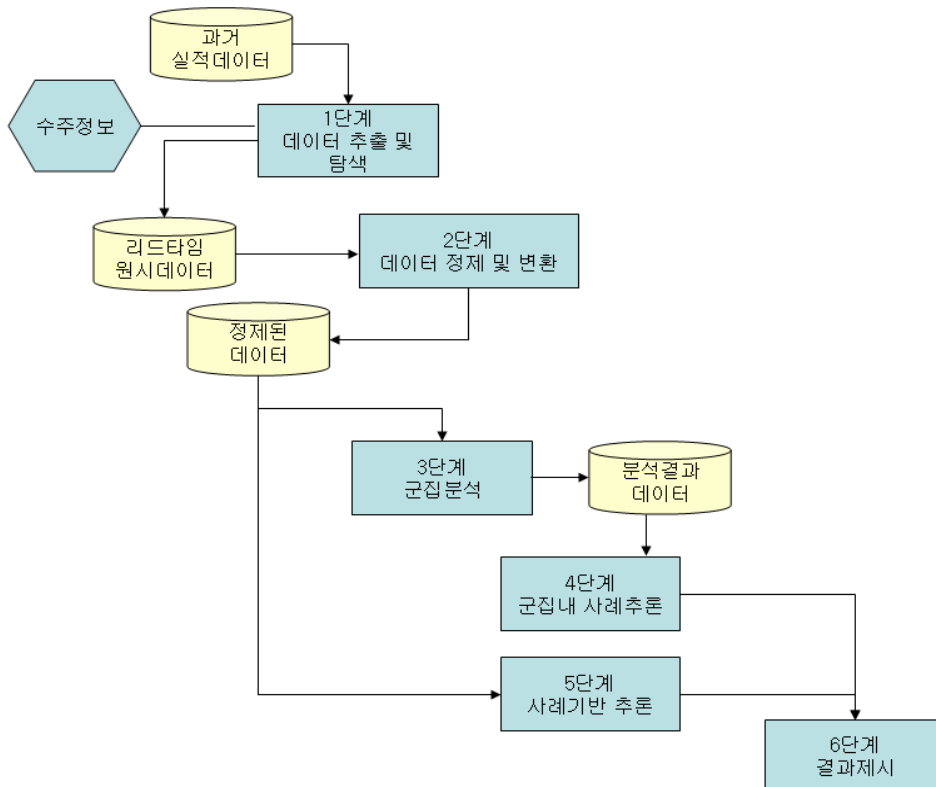


그림 4. 공정계획 지원절차.

일괄 처리하였다.

3단계 : 군집분석

정제된 데이터를 근간으로 하여 군집분석을 실시하는 단계이다. 군집분석은 주어진 관찰치 중에서 유사한 것들을 몇몇의 집단으로 그룹화하여, 각 집단의 성격을 파악함으로써 데이터의 전체 구조에 대한 이해를 돕고자 하는 분석방법이다. 본 연구에서는 최적 분리군집방법인 K-평균 군집방법을 사용한다. 즉, 데이터의 구조를 다변량 통계분석법을 이용해 관찰치 위치를 시각화해 사전에 결정된 군집 수 k에 기초하여 전체 데이터를 상대적으로 유사한 k개의 군집으로 구분하는 방법이다.

- Step 1) 군집 수 k를 정한다.
- Step 2) 초기 k개 군집의 중심을 선택한다.
- Step 3) 각 관찰치를 그 중심과 가장 가까운 거리에 있는 군집에 할당한다.
- Step 4) 각 군집별로 그에 속하는 관찰치를 이용해 새로운 중심을 계산한다.
- Step 5) Step 3~Step 4의 과정을 기존의 중심과 새로운 중심의 차이가 없을 때 까지 반복한다.

이러한 과정을 그림으로 살펴보면 <그림 5>는 거리를 표현하는 변수가 2차원으로 표현된 각 관찰치의 위치이다. 그림을 관찰하여 군집 수는 k(<그림 5>의 데이터에서는 14)로 결정하고 각 군집의 중심을 전체 데이터를 임의로 k등분해 각각의 평균으로 정하여 <그림 6>과 같이 가장 가까운 군집에 할당한다. 다음으로, 새로이 할당된 관찰치를 기준으로 그 군집의 중심을 다시 계산한다. 이 결과에 의해 평균이 이동된 것을 볼 수 있고 이동된 군집의 평균을 통해 관찰치를 재할당한 후 다시 새로운 군집의 평균을 계산한다. 이러한 과정을 군집의 평균이 거의 변화하지 않을 때까지 계속 반복하면 k개의 군집을 얻을 수 있다. 이러한 방법을 통하여 얻어진 분석결과를 분류하여 각 그룹별 대공정 항목별로 평균을 구하고 그룹별 평균정보를 분석결과 데이터베이스에 저장한다. 이렇게 하는 이유는 유사

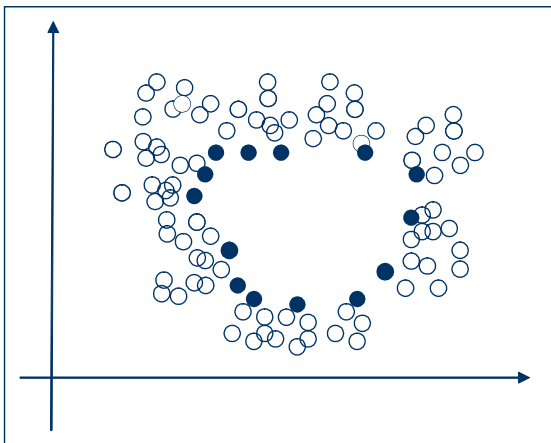


그림 5. 2차원으로 표현한 관찰치.

제품군이 입력되면 가장 빨리 쉽게 데이터에 접근하여 신속한 의사결정을 내리기 위해서이다.

4단계 : 군집 내 사례추론

본 연구에서는 현재 많이 이용되고 있는 Aamodt 등이 분류한 <그림 7>과 같은 4단계(4Re : Retrieve, Reuse, Revise, Retain)의 순환과정을 적용하여 군집 내 사례추론을 진행하였다. 사례기반 추론 과정에서 가장 많이 사용하는 사례 간 유사도 측정기법은 k-최근접 이웃방식(k-nearest neighbors)이다. 유사도를 구하는 목적은 새로운 사례가 입력되면 이를 통해 문제의 주요 속성을 추출한 다음 이들 속성을 이용해서 과거의 유사한 사례를 검색하기 위해 그 측정치로 유사도를 이용한다. 수치형 자료인 경우의 계산식은 다음과 같다.

$$\text{유사도} = \left(1 - \frac{|\text{비교 사례의 속성값} - \text{과거 사례의 속성값}|}{\text{해당 속성의 최대값}} \right) \times 100$$

수주정보가 입력되면 분석결과 데이터베이스로부터 유사제품을 찾아주는 단계이다. 앞에서 언급한 인공지능 분야 중 하나의 방법인 사례기반추론 기법을 활용하여 유사도를 측정한다. 다음 가장 가까운 유사제품을 쉽게 찾을 수 있게 한다.

5단계 : 사례기반추론

군집 내 사례추론 결과의 내용이 부족하거나, 미흡할 경우 군집분석 전의 정제된 데이터베이스로부터 유사제품을 찾아내는 단계이다. 이러한 방법론은 가장 유사한 사례를 순위별로 제공하며, 분석결과 데이터로부터 사례를 추론하는 것보다 정제된 데이터로부터 사례를 추론하면 더욱 근접하게 유사제품을 찾을 수 있다.

6단계 : 결과 제시

군집 내 사례추론 결과와 사례기반추론의 결과는 수주정보가 입력되면 둘 다 유사제품을 찾아주지만, 그 결과의 정확성

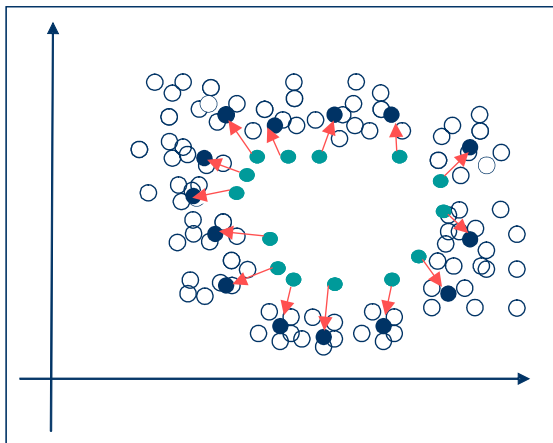


그림 6. 군집 중심의 이동.

은 기존의 데이터를 본 연구방법론으로 활용할 경우 얼마나 정확도가 높은지를 검증함으로써 이 방법론의 우수성을 평가할 수 있을 것이다. 본 연구에서는 연구방법의 결과로서 정보시스템을 활용하여 사용자에게 그 결과를 정보제공 화면으로 제시한다.

샘플링을 통하여 140개를 추출하였다. 이 정보를 리드타임 원시 데이터베이스에 저장하였고 수주번호는 순번으로 변환하여 공정별 인원 수로 나누어 정제된 데이터베이스에 저장한다.

4. 실험 및 결과 검증

4.1 실험

(1) 데이터 추출 및 탐색

대공정 25개의 과거 실적을 수주번호의 제품용량별로 랜덤

(2) 군집분석

데이터는 k-평균 군집방법을 통하여 데이터를 분석한 것이다. 이 모형개발을 위해 SAS E-Miner를 이용하였다. <그림 8>은 SAS E-Miner에서 구성한 분석흐름도이다. <그림 9> 파이 조각의 크기는 군집의 변이성을 나타내고, 파이 조각의 높이는 각 군집에 속하는 관찰치의 수를 나타내고 있으며, SAS Insight를 통하여 14개의 그룹으로 분류된 것을 알 수 있다. 군집분석은 탐색적인 분석방법으로의 장점을 가지고 있는 반면에, 사전에 주어진 목적이 없으므로 결과를 해석하는 데 있어

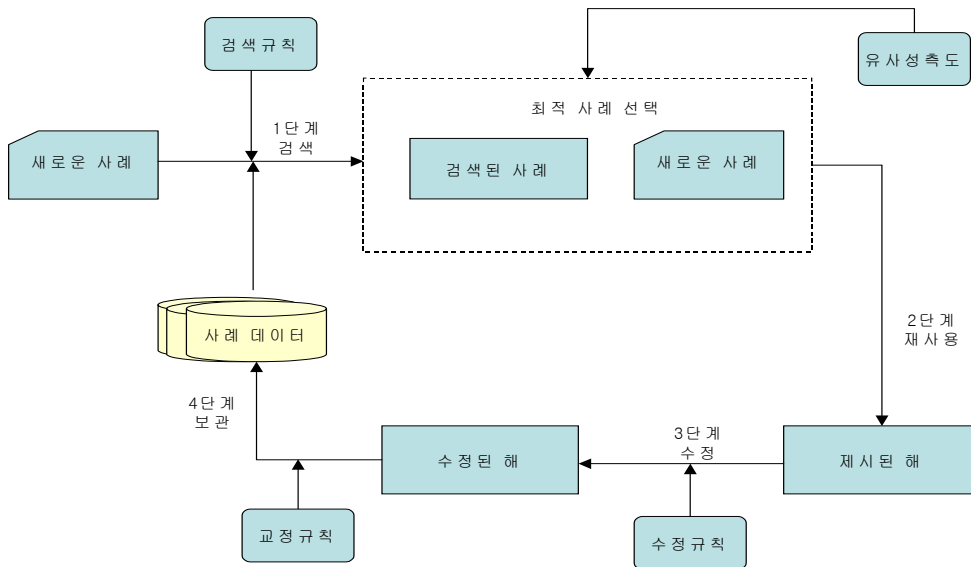


그림 7. Aamodt and Plaza가 제시한 4단계 사례기반추론의 순환과정.

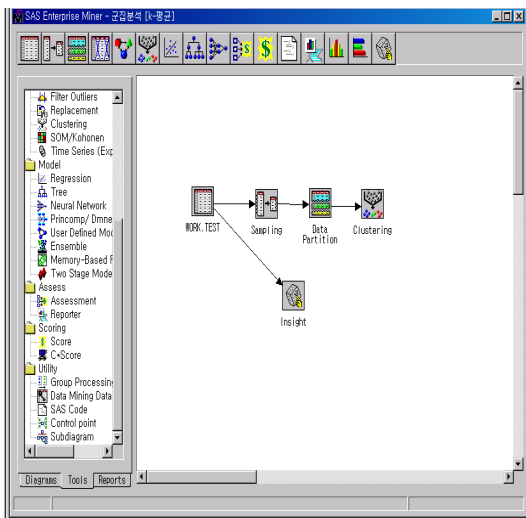


그림 8. E-miner에서 구성한 분석 흐름도.

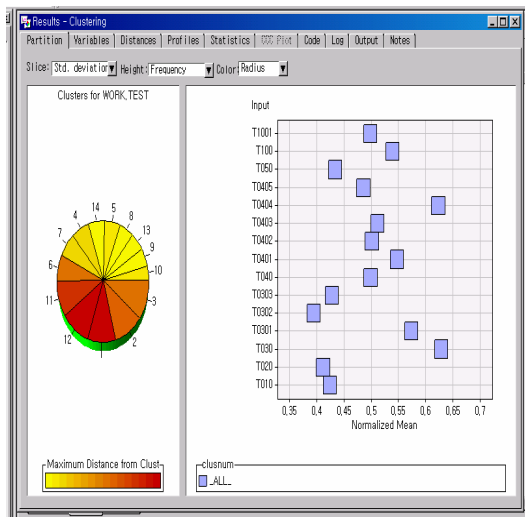


그림 9. 군집의 특성 파이 도표.

서 어려움이 있다. 그러나 14개의 그룹의 군집분석 결과 데이터를 살펴보면 군집된 그 결과를 충분히 이해할 수 있다. 과거 실적의 수주번호를 역추적하여 각 대공정별 공정소요시간의 변화를 살펴보면 수주된 정보의 용량에 따라 공정소요시간의 변화가 많음을 확인할 수 있다. 산업용 변압기는 입력되는 전압을 가지고 소요목적에 맞게 출력용량을 보내주는 기능을 하기 때문에 용량이 커질수록 제품의 크기도 비례적으로 변화하며 작업량도 많이 발생한다. 즉, 공정소요시간이 많이 걸린다는 결과이다. 군집분석 결과 변환 데이터를 구체적으로 설명하면 140개의 데이터를 정제과정을 거쳐 k-평균 군집분석을 한 결과, 총 14개의 그룹으로 분류가 가능했고 본 연구에서는 분류된 데이터를 각 그룹에 대한 대공정별로 평균을 구하여 분석한 결과를 데이터베이스에 저장한다.

이 14개의 데이터 그룹은 용량별로 대표정보를 구성하기 위한 사전절차이다. 그 이유는 새로운 정보가 수주되면 대표정보를 선택하여 사전에 공정소요시간을 추출하기 위함이다. 즉, 이 대표정보 14가지는 각 대공정별로 수행되는 시간을 포함하

고 있어 새로운 수주발생 시 대표정보를 선택하여 미리 공정소요시간을 추출할 수 있다. 현재는 14개의 군집으로 분석하여 대표정보를 만들었지만 시간이 흘러 새로운 데이터가 쌓이면 이 대표정보는 당연히 증가할 것이다.

다음은 SAS E-Miner로 생성한 데이터 세트를 데이터베이스에 저장한 결과가 아래의<그림 10>에 나타나 있다. 총 14개의 군집 140개의 원시 데이터를 얻을 수 있으며, 군집의 평균값을 구하여 분석결과를 데이터베이스에 저장한다. 이러한 데이터는 영업에서 수주된 정보를 통하여 설계하는 과정에서도 예산편성 및 견적자료로서 활용이 가능할 뿐만 아니라 그 활용가치는 매우 높다.

(3) 군집 내 사례추론

고객으로부터 수주정보를 검색한 결과가 <그림 11>에 나타난다. 중요한 정보는 각 수주별 용량정보이다. 이 정보는 분석결과 데이터 저장소의 모델번호 정보에서 유추할 수 있다. 모델번호의 3~5째 자리는 용량정보를 나타낸다. 수주된 용량

Chk	대표번호	절단	적출	총절	월절	중절	권절	W/L
<input type="checkbox"/>	1	60	90	6	8	28	34	1
<input type="checkbox"/>	2	100	192	8	10	46	60	1
<input type="checkbox"/>	3	138	242	10	12	66	88	30
<input type="checkbox"/>	4	182	298	12	14	88	118	42
<input type="checkbox"/>	5	210	361	14	16	114	132	54
<input type="checkbox"/>	6	248	414	16	16	144	161	62
<input type="checkbox"/>	7	293	467	16	16	176	187	73
<input type="checkbox"/>	8	340	523	16	16	198	229	87
<input type="checkbox"/>	9	439	576	16	18	224	265	93
<input type="checkbox"/>	10	593	699	18	18	258	303	105
<input type="checkbox"/>	11	640	928	18	18	295	345	115
<input type="checkbox"/>	12	707	1033	18	18	342	381	126
<input type="checkbox"/>	13	779	1145	18	21	422	435	138
<input type="checkbox"/>	14	864	1314	20	22	482	495	153

그림 10. 군집 원시 데이터.

chk	프로젝트번호	프로젝트번호	용량	수주금액
<input type="checkbox"/>	20040001	원광원자력	259	100,000,000
<input type="checkbox"/>	20030001	구리원자력	337	230,000,000
<input type="checkbox"/>	20040002	월성원자력	437	320,000,000

Chk	그룹번호	모델번호	유사도
<input type="checkbox"/>	1	T3010	39
<input type="checkbox"/>	2	T3020	40,429
<input type="checkbox"/>	3	T3040	43,286
<input type="checkbox"/>	4	T3060	46,143
<input type="checkbox"/>	5	T3090	50,429
<input type="checkbox"/>	6	T3120	54,714
<input type="checkbox"/>	7	T3160	60,429
<input type="checkbox"/>	8	T3200	66,143
<input type="checkbox"/>	9	T3250	73,286
<input type="checkbox"/>	10	T3300	80,429
<input type="checkbox"/>	11	T3400	94,714
<input type="checkbox"/>	12	T3500	91
<input type="checkbox"/>	13	T3600	76,714
<input type="checkbox"/>	14	T3700	62,429

그림 11. 수주정보 및 군집 내 유사도 검색.

정보를 가지고 대표정보로 생성된 분석결과 데이터베이스와 비교하여 군집 내의 유사도를 탐색한 결과이다. 그룹번호 11개의 유사도가 찾고자 하는 수주정보와 가장 유사한 결과를 보여주고 있다.

이러한 결과는 실제 생산현장에 적용해도 무방하나 좀더 근접된 정보를 과거 실적을 검색하여 찾아볼 수도 있다. 즉, 대표성을 가진 군집 내 데이터를 검색하는 것도 의미가 있지만, 잘 정제된 과거 데이터도 참조해 반영하면 더욱 좋은 결과를 얻을 수 있을 것이다. 다음은 좀더 근접된 사례기반 데이터를 참조하여 검색하는 과정을 살펴보겠다.

(4) 사례기반 추론

군집 내 사례추론을 통하여 나타난 결과의 정보가 부족할 경우 같은 수주정보에 대하여 사례기반 추론을 할 수 있다. <그림 12>는 고객으로부터 수주 정보를 검색한 결과와 수주된 용량 정보를 가지고 정제된 데이터베이스와 비교하여 사례기반 유사도를 탐색한 결과이다. 이 결과를 보면 군집 내 사례추론과 사례기반추론에서 나타난 결과와 같은 11번 군집에서 데이터를 추천할 수 있으며 사례기반추론의 결과가 더욱더 유사 제품과 근접한 데이터를 제공한다. 그 이유는 군집 내 사례추론은 분석결과 데이터베이스 내의 자료를 검색한 결과를 제공하나 이 데이터는 앞에서 언급했듯이 대표정보의 데이터로 각 군집별로 평균값을 제공하기 때문에 실제 사례를 저장한 정제된 데이터베이스에서 추출하는 것이 더 나은 정보를 제공한다.

4.2 실험결과에 대한 검증

본 연구에서 제시한 방법론의 신뢰성 여부를 검증하기 위하여 <그림 13> 같은 시나리오를 구성하여 본 연구의 신뢰성을 검증하고자 한다. 총 140개의 정제된 수주정보 데이터를 구간으로 하여 14개 그룹별로 5개, 총 70개의 데이터를 추출하여 검증 데이터로 사용하였다. <표 2>는 70개의 검증데이터를 요약한 것이다.

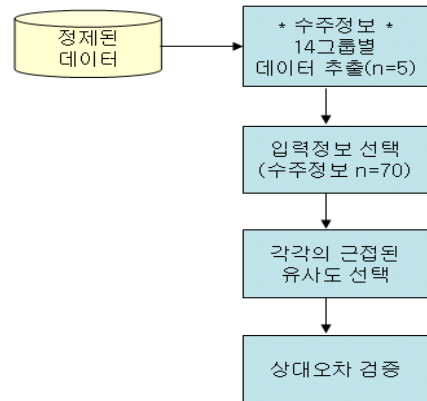


그림 13. 시나리오 흐름도.

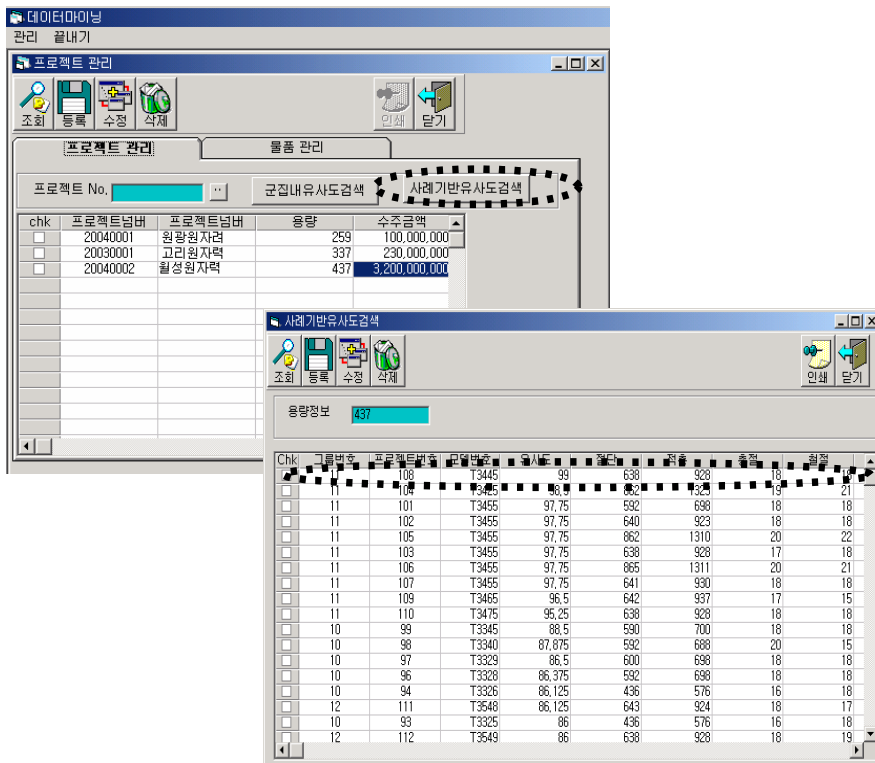


그림 12. 수주정보 사례기반 유사도 검색 선택.

예를 들어 그룹번호 1, 수주번호 8에 대한 철심절단 공정의 구체적 산출내용을 살펴보면 다음과 같다.

$$\text{상대오차} = \left(\frac{\text{수주정보 철심절단 실제 값} - \text{철심절단 사례기반 추론 값}}{\text{수주정보 철심절단 실제 값}} \right) \times 100$$

$$\text{상대오차 평균} = \frac{\sum \text{각 공정의 상대오차 값}}{\text{총 공정 수}}$$

이와 같이 산정한 결과를 정리해 보면 70개의 추출 데이터의 상대오차 평균값은 최대 26%의 차이가 발생된다는 사실을 알 수가 있었다. 또한, 표준편차는 각 공정별로 상대오차 값 n=70에 대하여 산출하였고 관리상한(UCL; Upper- Control Limit)은 3σ 관리한계를 사용하였다. 이렇게 산정한 결과에 대하여 \bar{x} 관리도를 적용하여 관리 상한값을 구하면 <표 2>의 마지막 행과 같다.

다음은 작성된 \bar{x} 관리도를 이용하여 총 140개의 정제된 수주정보 데이터에서 n=15개를 추출하여 사례기반추론 과정을

거쳐 상대오차 평균값을 비교한 결과는 <표 3>과 같이 구해진다. <그림 14>에서 확인할 수 있는 바와 같이 n=15 모두 관리상한값 범위 내에 분포되어 있으므로 본 연구결과가 신뢰성이 있다는 사실을 확인할 수가 있었다.

5. 결론 및 향후 연구방향

본 연구에서는 수주생산 형태를 갖는 산업용 변압기제품의 공정계획을 지원할 수 있는 방법으로 군집분석과 사례기반 추론의 형태를 결합한 하이브리드형 데이터마이닝 기법을 제시하였다. 이러한 방법은 산업용 변압기제품뿐만 아니라, 수주생산 형태의 제조업분야에 확장 적용될 수 있다고 본다. 본 연구의 효과를 정리하면 다음과 같다.

첫째, 기존에는 과거의 실적기준으로 공정별 소요시간의 특성을 수작업으로 파악한 것을 데이터마이닝 기법의 하나인 군

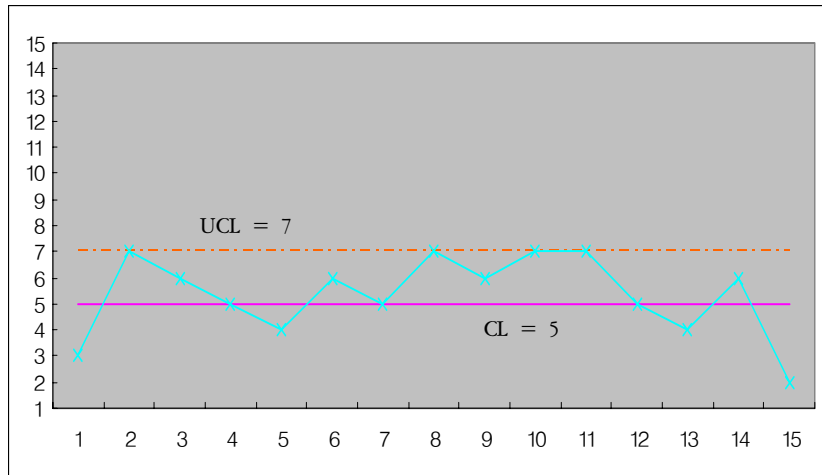


그림 14. 상대오차의 \bar{x} 관리도.

표 2. 검증 데이터 요약

구분	그룹 번호	수주 번호	용량	유사도	철심절단	철심적층	총조립절연	철심절연	중심절연	권선절연	조성 및 사이징	TV	LV	HV	TAP	RE'T	사이징	중심	B/B	중심리드태핑	차폐절단	BC'T	FI'T	가스켓가공	총조립	WT'G	FI'T	총조립해체	WR'G해체	상대오차평균
상대오차SUM					414	337	689	430	511	373	484	334	266	269	331	347	348	213	402	468	468	209	446	289	216	297	569	323	535	383
평균					6	5	10	6	7	5	7	5	4	4	5	5	5	3	6	7	7	3	6	4	3	4	8	5	8	5
표준편차					9	8	14	8	9	8	8	6	7	6	9	7	8	5	11	9	10	4	9	10	4	5	10	5	9	5
UCL					11	9	18	11	13	10	12	8	8	7	10	9	9	6	12	12	12	5	11	10	5	7	14	8	13	8

표 3. 검증결과

순서	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
상대오차 평균값	3	7	6	5	4	6	5	7	6	7	7	5	4	6	2

집분석을 통하여 각 집단의 성격을 빠르고 과학적으로 분석함으로써 데이터의 신뢰성을 높였다.

둘째, 대용량 과거 실적 데이터를 미리 분류하여 데이터베이스에 저장함으로써, 유사수주 발생 시 신속한 의사결정을 내릴 수 있는 체계를 마련하였다.

셋째, 기존의 데이터마이닝 연구는 주로 마케팅, 영업, 보험업, 금융서비스업에 적용된 사례가 많다. 본 연구는 수주생산 형태에서 데이터마이닝 기법을 적용하여 공정소요시간을 미리 추출함으로써 공정계획을 지원할 수 있도록 한 데 그 의의가 있다.

넷째, 본 연구의 방법은 관리자뿐만 아니라, 고객의 입장에서 정보시스템을 활용함으로써 아주 빠르고 편리한 정보의 제공을 가능케 할 수 있다.

향후 연구에서는 다양한 수주정보의 인자요소를 감안하여 인자요소별로 규칙을 찾아내고 그 정보가 사전에 견적정보를 제공하고 또한, 제품을 설계하는 데 반영될 수 있는가를 연구해 보고자 한다.

참고문헌

Aamodt, A., and E. Plaza (1994), Case Based Reasoning : Foundational Issues, Methodological Variations, and System Approaches, *Artificial Intelligence Communication*, 7(1), 39-59.

Ahn, J.S., Goh, Y.M., Jang, J.S. (1999), Searching Optimal Process Conditions Using Data Mining Techniques, *Journal of the Korean Institute of Plant Engineering*, 4(2), 129-144.

Back, J.G., Kim, K.H., Kim, S.S., Kim, C.O. (2003), Adaptive Decision Tree Algorithm for Data Mining in Real-Time Machine Status Database, *Journal of the Korean Institute of Industrial Engineers*, 26(2), 171-182.

Han, S.T., Lee, S.K., Kang, H.C., Ryu, D.K. (2002), Development of Scoring Model on Customer Attrition Probability by Using Data Mining Techniques(in Korean), *The Korean Communications in Statistics*, 9(1), 271-280.

Jhee, W.C., Suh, M.S. (1998), DSS Architectures to Support Data Mining Activities for Supply Chain Mnagement, *Journal of MIS Research*, 8(3), 51-73.

Jhee, W.C (2003), Quality Design Support System based on Data Mining Approach, *Journal of the Korean Operations Research and Management Science Society*, 28(3), 31-47, .

Kang, H.C., Han, S.T., Choi, J. H. (2000), Interpretation of Data Mining Prediction Model Using Decision Tree(in Korean), *The Korean Communications in Statistics*, 7(3), pp. 937-943.

Kim, J.S. (2003), A Study on the Development of Internet Purchase Support Systems Based on Data Mining and Case-Based Reasoning, *Journal of the Korean Operations Research and Management Science Society*, 28(3), 135-148.

Lee, S.B. (2001), Evaluations of predicted models fitted for data mining, *Journal of the Korean Data & Information Science Society*, 12(2), 113-124.

McGovern, J., D. Samson, and A. Wirth (1994), Using Case Based Reasoning for Basis Development in Intelligent Decision System, *European J. of Operational Research*, 74, 40-59.

Slade, S. (1991), Case-Based Reasoning : A Research Paradigm, *AI Magazine*, Spring, 42-55.



오 경 모

울산대학교 산업공학과 학사
 울산대학교 산업공학과 석사
 울산대학교 산업정보경영공학과 박사
 과정수료
 만도기계(주) 생산관리부
 현재: 제론정보기술(주) 컨설팅팀
 관심분야 : Data Mining, ERP, SCM, ISP, BI,
 PDM, Manufacturing Information System



박 창 권

울산대학교 산업공학과 학사
 한국과학기술원 산업공학과 석사
 한국과학기술원 산업공학과 박사
 현대정공(주) 공작기계 기술부
 현재: 울산대학교 산업정보경영공학부 부교수
 관심분야 : Production Scheduling, ERP, SCM,
 Data Mining, BI