

데이터마이닝 기법을 이용한 생산데이터 분석시스템 설계

이형욱(한국생산기술연구원), 이근안(한국생산기술연구원), 최석우(한국생산기술연구원),
박홍균(스페이스 솔루션), 배성민* (한밭대학교 산업경영공학과)

Design of Manufacturing Data Analysis System using Data Mining Techniques

H. W. Lee (KITECH), G. A. Lee (KITECH), S. Choi (KITECH),
H. K. Park (SpaceSolution Co. Ltd.), S. M. Bae (Dept. of IME, HANBAT National Univ.)

ABSTRACT

Many data mining techniques have been proved useful in revealing important patterns from large data sets. Especially, data mining techniques play an important role in a customer data analysis in a financial industry and an electronic commerce. Also, there are many data mining related research papers in a semiconductor industry and an automotive industry. In addition, data mining techniques are applied to the bioinformatics area. To satisfy customers' various requirements, each industry should develop new processes with more accurate production criteria. Also, they spend more money to guarantee their products' quality. In this manner, we apply data mining techniques to the production-related data such as a test data, a field claim data, and POP (point of production) data in the automotive parts industry. Data collection and transformation techniques should be applied to enhance the analysis results. Also, we classify various types of manufacturing processes and proposed an analysis scheme according to the type of manufacturing process. As a result, we could find inter- or intra-process relationships and critical features to monitor the current status of the each process. Finally, it helps an industry to raise their profit and reduce their failure cost.

Key Words : Data Mining (데이터마이닝), Type of Manufacturing Process(생산공정 형태), Inter-process relationship(공정 간 관계)

1. Introduction

본 논문에서는 생산현장에서 생성되는 데이터를 데이터마이닝 기법을 이용해서 분석하는 방법에 대해서 설명하고자 한다. 데이터마이닝 기법은 대용량의 데이터로부터 의미 있는 정보를 추출하는데 유용하다고 알려져 있으며, 생산현장에서 생성되는 공정 데이터 분석에도 많이 적용되어왔다. 또, 최근 들어 각광받고 있는 생물정보학(bioinformatics) 영역에서도 활발하게 이용되고 있는 기법이다. 본 논문에서는 많이 사용되고 있는 데이터마이닝 기법에 대해 간단히 설명하고, 이것이 라인생산방식을 따르는 생산현장에 적용되었을 때, 어떻게 이용이 되어야 할 것인지에 대해서 자세히 설명하도록 한다.

2. 데이터마이닝의 생산공정에의 적용

2.1 제조업에서의 생산형태

제조업은 여러 가지 생산형태를 가지게 된다. 특히, 대표적인 생산형태는 3 가지로 구분이 되는데, 흐름(flow)생산, 라인(line)생산, 잡샵(jobshop)생산방식이 그것이다. 흐름생산은 연속적인 공정을 가지는 장치산업으로 철강, 반도체, 화학분야가 해당된다. 특히, 반도체 분야의 공정은 수많은 공정과 테스트작업이 연속적으로 이루어지기 때문에 데이터마이닝을 적용한 사례가 많다 [1] [2]. 라인생산방식은 가장 일반적인 제조업에서의 생산공정으로써 자동차부품, 기계부품 등의 가장 광범위한 영역에 적용된다.

잡삽형식의 생산은 단품이나 단량위주의 생산 방식으로 금형, 중공업, 조선 등의 특화된 제품에 적용되는 방식이다.

2.2 생산 공정 정형화

공정데이터를 분석하기 위해서는 분석하고자 하는 공정(process)을 정확히 파악하고, 각 공정에서 생성되는 데이터와 그에 대한 판정기준(criteria), 그리고 공정간의 상관관계에 대한 정보를 수집하는 것이 가장 첫 번째 단계이다.

생산공정에 대한 상세한 이해는 데이터 분석 및 해석에 많은 도움을 줄 수 있다. 예를 들어, 현재 수집되고 있는 데이터들 가운데 어떤 데이터들을 추출(extraction)해야 할 것이며, 변환(transformation) 과정에서 정규화(normalization)는 어떻게 해야 할 것인지, 또 분석 결과들에 대한 해석(interpretation)에 도움을 줄 수 있다.

2.3 Clustering 을 이용한 분석방법

공정에서 측정해야 되는 데이터의 종류는 매우 많다. 예를 들어 반도체 공정의 경우, 한 공정에서 테스트를 위해 측정하는 변수의 개수는 600 여개가 넘는 경우도 있다. 이런 경우 모든 변수에 대한 분석을 하기가 어렵기 때문에 비슷한 경향을 지닌 변수들을 그룹화하여 해당 그룹에서 가장 대표성을 가지는 상위 3~5 개의 변수들을 도출하여 이에 대한 분석을 하는 것이 유리하다. 이를 Feature construction 이라고 하는데, 이 때 유용하게 쓰일 수 있는 방법이 군집(clustering)분석이다.

군집분석을 통해 각 공정에서 분석의 대상이 될 수 있는 변수의 숫자를 획기적으로 줄일 수 있으며 이를 통해 좀 더 세밀한 분석을 수행할 수 있게 된다.

2.4 C4.5 를 이용한 분석방법

앞서 소개한 데이터마이닝 기법 가운데, 공정분석에서 가장 유용하게 쓰일 수 있는 기법 가운데 하나가 의사결정나무를 이용한 방법이다. 일반적으로 각 공정에서 측정되는 데이터의 항목은 그 수가 매우 많으며, 측정결과 또한 ‘양호’ 또는 ‘불량’의 명확히 구분된 2 개의 클래스를 결과값으로 가지게 된다. 이러한 데이터의 특성상 의사결정나무, 특히 C4.5 [3]를 이용한 공정 데이터의 분석은 각 공정에서 양호와 불량을 구분하는 가장 중요한 특성이 무엇인지 도출해 낼 수 있다.

즉, C4.5 의 수행결과로 생성된 의사결정나무에서 가장 상위 루트 노드(root node)에 나타난 변수가 양호와 불량을 구분하는데 가장 중요한 역할을 하는 변수가 된다.

이러한 관점에서 의사결정나무에서 루트 노드를 포함한 상위 3 개~5 개 정도의 변수들이 집중적으로 관리해야 할 필요가 있는 중요 변수들이 될 수 있다.

2.5 Neural Network 을 이용한 분석방법

신경망은 주로 에러율(error rate)에 대한 예측(forecasting)에 주로 사용된다. 각 공정에서 중요한 변수들을 도출하고 이러한 변수들이 최종 수율(yield)에 어떠한 영향을 미치는지를 파악할 수 있으며 이에 대한 학습을 통해 새로운 환경으로 바뀌었을 때 또는 새로운 변수가 추가되었을 때 최종 수율이 어떻게 바뀔 것인지에 대해 예측을 하는데 도움을 줄 수 있다.

신경망을 이용한 예측에서는 어떤 변수를 사용하여 예측하는지 얼마나 신뢰성이 있는 데이터들이 사용되는지에 대한 사항들이 결과에 큰 영향을 미치기 앞부분에서 적절한 변수를 추출하는 것이 매우 중요하다.

3. 결론 및 추후 연구과제

본 논문에서는 여러 가지 데이터마이닝 기법들을 소개하고 이들이 실제 생산공정의 분석에 사용되기 위해서는 어떻게 해야 할 것인지에 대해 논의하였다. 앞으로의 연구과제는 이러한 분석방법론을 우리나라의 대표산업 가운데 하나인 자동차 부품업계에 적용시켜 봄으로써 품질 및 성능개선에 유용한 정보들을 도출하고 이를 현장에 적용함으로써 도출된 정보에 대한 평가를 해보고자 한다.

후 기

본 연구는 산업자원부의 중기거점 개발사업인 “웹기반 SMART 제조시스템 개발” 과제의 지원으로 수행되었으며, 이에 도움을 주신 관계자 여러분들께 감사 드립니다.

참고문헌

1. 백동현, 한창희, “데이터마이닝을 이용한 반도체 FAB 공정의 수율 개선 및 예측,” 한국지능정보시스템학회지, 제 9 권, 제 1 호, pp. 157-177, 2003
2. J. H. Lee, S. J. Yu, and S. C. Park, “Design of Intelligent Data Sampling Methodology based on Data Mining Technology,” IEEE Transaction on Robotics and Automation, Vol. 17, No. 5, pp. 637-649, 2001
3. J. Ross Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publisher, 1993
4. <http://www.rulequest.com>