

데이터 마이닝을 위한 생산공정 데이터 추출

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

Data Extraction of Manufacturing Process for Data Mining

H. K. Park(Space Solution), G. A. Lee(Digital Production Processing Team, KITECH), S. Choi(Digital Production Processing Team, KITECH), H. W. Lee(Digital Production Processing Team, KITECH),
S. M. Bae(Industrial and Management Eng. Dept., HNU)

ABSTRACT

Data mining is the process of autonomously extracting useful information or knowledge from large data stores or sets. For analyzing data of manufacturing processes obtained from database using data mining, source data should be collected from production process and transformed to appropriate form. To extract those data from database, a computer program should be made for each database. This paper presents a program to extract easily data from database in industry. The advantage of this program is that user can extract data from all types of database and database table and interface with Teamcenter Manufacturing.

Key Words : data mining(데이터 마이닝), database(데이터베이스), data extraction(데이터 추출)

1. 서론

생산 공정에서 불량이 생기면 생산 공정에서 수집되는 여러 가지 데이터와 이러한 불량이 어떤 상관 관계에 있는지를 분석하려는 노력은 오랫동안 지속되어왔다. 컴퓨터 기술이 발전되면서부터 이러한 분석은 데이터 마이닝(Data mining)이라는 기법을 활용해 왔다[1].

데이터 마이닝을 위해서는 생산 공정에서 수집되는 데이터들을 적절한 형태의 데이터로 변환하여야 한다. 데이터 마이닝을 위한 이러한 데이터들은, 일반적으로 생산 공정과 연결된 데이터 베이스에 저장되어 있다. 데이터베이스에 저장되어 있는 데이터는 컴퓨터 프로그램에 의해 파일 형태로 만들어 내고, 다시 엑셀과 같은 프로그램으로 읽어 들여 필요한 데이터를 추출하거나 다른 형식으로 변환하여, 데이터 마이닝을 위한 데이터를 만든다.

이러한 과정은, 또 다른 데이터에 대해 작

업을 하려면 같은 과정을 반복해야 하므로, 시간이나 비용측면에서 효율적이지 못하다. 또한 이런 작업을 컴퓨터 프로그램을 이용해서 하더라도, 다른 시스템이나 다른 데이터베이스에서는 이 프로그램을 수정하거나 다시 개발해야 한다.

본 연구에서는 이러한 작업을 쉽고 간단하게 할 수 있는 프로그램의 개발에 대하여 논의하고자 한다. 이 프로그램의 기능을 살펴보면 다음과 같은 기능이 있다. 개발된 프로그램이 기업내의 정보시스템에 있는 데이터베이스가 가지고 있는 시스템 테이블로부터 저장된 데이터베이스에 대한 정보를 읽어오고, 사용자는 대화형 방식으로, 이중 필요한 데이터만 선택하면, 자동으로 데이터 마이닝을 위한 데이터가 만들어진다.

또한 생산 공정의 각 프로세스와 연결하여 특정 프로세스에 문제가 있을 경우 그 프로세스와 관련된 데이터를 자동으로 추출할 수 있

도록, 생산 공정을 정의하는 Teamcenter Manufacturing 프로그램과 연결하는 방법도 논의하고자 한다. 또한 데이터 마이닝을 위하여 기업의 데이터베이스 시스템에 접근하여 프로그램의 타당성 검증을 수행하고자 하였다.

2. 데이터베이스의 시스템 테이블

2.1 MS*SQL의 시스템 테이블

Microsoft 사에서 만든 MS*SQL은 Windows 시스템에서 작동하는 데이터베이스 시스템이다. 주로 중소회사나 닷넷(.NET)을 사용하는 회사에서 사용하는 데이터베이스 시스템이다.

MS*SQL은 데이터베이스를 생성할 때 Sysindex와 Syscolumn이라는 데이터베이스 테이블이 자동으로 생성된다.

Sysindex 테이블은, 데이터베이스에서 사용자가 만드는 모든 테이블에 대한 정보를 담고 있다. 이 테이블에는 데이터베이스 내에 있는 테이블의 개수만큼의 행(Row)이 있고, 각 행에는 각 테이블에 대한 정보를 가지고 있다. 이 테이블에 들어가는 중요한 필드는 다음과 같다.

- Id(int): 테이블의 ID. 이 값이 256 미만인 경우는 시스템에서 사용하고, 256보다 크면 사용자가 사용하는 테이블이다.
- Status(int): 내부 시스템의 상태에 대한 정보. 이 값이 0이면 정상적으로 사용하는 테이블이다.
- Rowcnt(bigint): 테이블 행 개수
- Name(sysname): 테이블의 이름

또한 Syscolumn 테이블에 들어가는 중요한 필드는 다음과 같다.

- Name(sysname): 열(Column)의 이름
- Id(int): 테이블의 개체 ID.
- Xtype(tinyint): 데이터 유형. 이 값이 48이면 정수(int), 59이면 실수(float)이다.

2.2 Oracle의 데이터 Dictionary

Oracle 사에서 만든 Oracle 데이터베이스는 Unix와 Windows 시스템에서 작동하는 데이터베이스 시스템이다[2]. 대기업에 가장 많이 사용하는 시스템이며, 국내에서도 가장 많이

사용되고 있다.

Oracle에서는, 데이터베이스에 대한 정보를 저장해두는 시스템 테이블을 데이터 딕셔너리(Data dictionary)라고 하는데, 테이블에 대한 정보는 USER_TABLES, 각 테이블의 열에 대한 정보는 USER_TAB_COLUMNS에 저장되어 있다.

USER_TABLES에는 다음과 같은 필드가 들어 있다.

- Table_name: 테이블의 이름
- Num_Rows: 테이블의 행 개수

USER_TAB_COLUMNS에는 다음과 같은 필드가 들어 있다.

- Table_name: 테이블의 이름.
- Column_name: 테이블의 열 이름
- Data_type: 데이터 유형. 이 값이 number이면 정수나 실수, char이면 문자가 들어간다.
- Low_value: 최소값
- High_value: 최대값

3. 데이터 추출 프로그램

3.1 프로그램 실행 환경

이 프로그램은 웹(Web) 환경에서 작동되며 웹 서버가 설치되어 있는 Windows NT나 Windows XP에서 작동한다. Oracle 데이터베이스에서 데이터를 추출하기 위해서는 웹 서버에 Oracle client가 설치되어 있어야 한다.

3.2 데이터베이스 접속

프로그램을 실행하면 Fig. 1과 같은 화면이 나타나게 된다. 사용자는 데이터베이스에 Login하기 위한 정보를 입력하고 접속하려는 데이터 베이스 이름, 데이터베이스에 Login을 위한 사용자 ID 및 암호를 입력한다. 만약 MS*SQL을 사용하는 경우에는 데이터 베이스가 설치되어 있는 컴퓨터의 URL이나 IP address를 입력한다. Oracle인 경우에는 Oracle client에 미리 데이터 베이스에 대한 정보를 설정해 두었으므로, URL이나 IP address를 입력할 필요가 없다.

3.3 테이블 정보 선택

데이터 베이스 접속에 관련되는 정보를 입

력하면 자동으로 그 데이터베이스가 가지고 있는 모든 테이블에 대한 목록이 Fig. 2와 같은 화면이 나타난다.

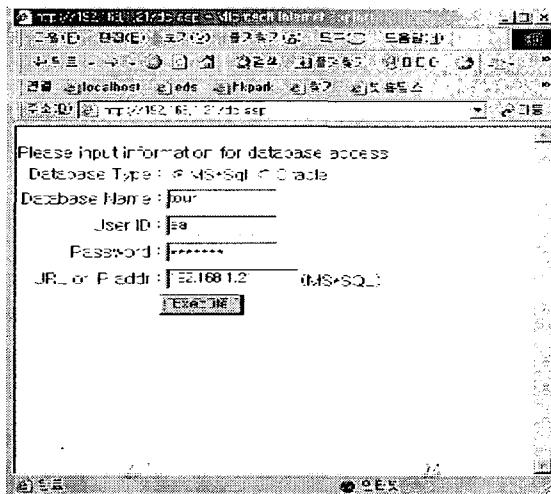


Fig. 1 Input for connecting database.

또한 테이블 이름과 함께 이 테이블의 열과 행의 개수가 함께 표시된다. 사용자는 이러한 테이블 중 데이터를 추출하려는 테이블을 선택한다.

table List			
No	Table Name	Cols	Rows
1	address	30	6
2	album	3	8
3	appy	25	15
4	acter	21	1
5	boerc	22	25
6	book	20	6
7	calendar	24	1
8	chronology	12	6

Fig. 2 Table list of database information.

3.4 필드 정보 선택

데이터를 추출하기 위한 테이블을 선택하면, Fig. 3과 같이 프로그램은 이 테이블이 가지고 있는 열에 대한 정보를 표시한다. 화면에는 열 이름, 이 열의 데이터 유형, 이 열에 있는 데이터의 최소값과 최대값 등이 나온

다.

사용자는 각각의 열에 대한 분포도를 그래프로 볼 수 있으며, 데이터 마이닝을 위한 데이터를 파일로 변환할 수 있다.

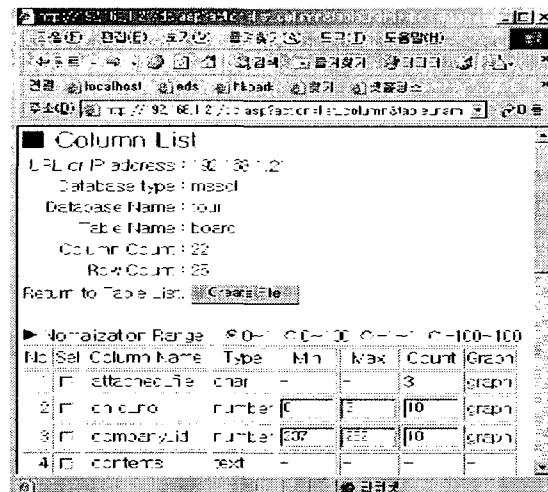


Fig. 3 Column list included table list of database information.

3.5 그래프로 데이터 분포도

지정한 열의 데이터의 분포도를 보기 위해서는 최대값, 최소값, 분포 간격을 지정하고 Fig. 3에서 Graph 표시를 선택하면 Fig. 4와 같이 열에 대한 데이터 분포가 나타나게 된다.

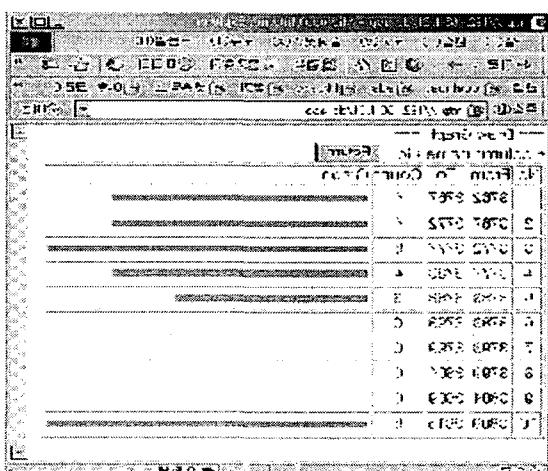


Fig. 4 Graph of data distribution in column list.

3.6 데이터 파일 만들기

파일을 생성할 경우, 데이터를 정규화(Normalization)하기 위한 범위를 선택하고,

파일에 들어갈 열을 선택하면 된다.

4. Teamcenter Manufacturing 과 연결

4.1 Teamcenter Manufacturing 의 개요

UGS사에서 만든 Teamcenter Manufacturing [3] 은 제품의 생산 계획 단계에서 제품을 생산하는 방법, 위치, 사용 리소스 등을 정의하기 위한 소프트웨어이다. Teamcenter Manufacturing에서 다음과 같은 것들을 정의 할 수 있다.

- 제품 구조(Product Structure): 제품 구조는 설계 엔지니어가 CAD시스템에서 생성한다. 이러한 제품 구조는 생산해야 할 대상(What)을 설명한다.
- 플랜트 구조(Plant Structure): 플랜트 구조는 제품이 생산되는 공장에 대해 정의한다. 이러한 공장에 대한 정의는 제품의 각 부분이 어디에서(Where) 제작되는지에 대한 정보를 제공한다.
- 프로세스 구조(Process Structure): 프로세스 구조는 제품이 제조되는 방법(How)에 대해 설명한다.

프로세스 구조는 제품을 생산하는데 필요한 여러 단계로 구성되어 있으며, 이러한 각 단계는 다시 하위 프로세스로 나눌 수 있는 계층 구조로 되어 있어서, BOM(Bill of Material)과 비슷한 구조를 이루고 있다. 이런 이유로 프로세스 구조를 BOP(Bill of Process)이라고도 부른다. 이러한 프로세스 구조는 제품구조와 플랜트 구조와 연결되어 있다.

또한, 각각의 프로세스를 수행하는데 필요한 리소스와도 연결되어 있어서 프로세스에서 필요한 인원이나 장비 등을 알 수 있다. 경우에 따라서는 그 프로세스에 필요한 도면이나 3차원 모델을 제품 구조로부터 가져와 볼 수도 있다. 이와 같은 구성이 Fig. 5에 나타나 있다.

4.2 Teamcenter Manufacturing 과 연결

Teamcenter Manufacturing에서는 URL을 모든 개체에 연결할 수 있는 기능이 있다. 이러한 기능을 이용하여, 위에서 설명한 프로그램을 연결하여 데이터를 추출할 수 있도록 하였다.

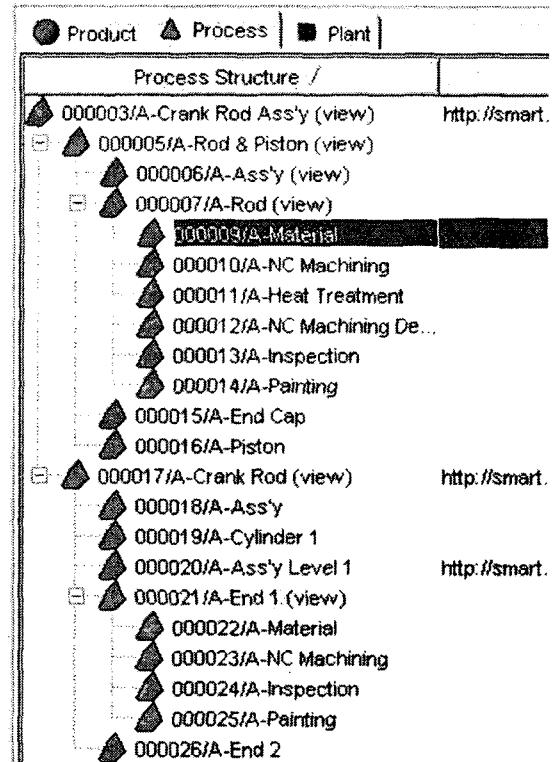


Fig. 5 Process structure of Teamcenter manufacturing.

즉, Teamcenter Manufacturing에서 프로세스를 만들 때, 이 프로세스와 관련되는 데이터베이스와 테이블 이름을 미리 입력하여 어떤 프로세스에 대한 공정분석 혹은 공정간의 분석을 위하여 그 프로세스와 관련되는 데이터베이스 테이블에서 데이터를 쉽게 추출해줄 수 있도록 하였다.

5. 결론

생산 공정에서 문제가 발견되는 경우, 공정에서 발생되는 데이터에서 데이터 마이닝을 위한 데이터로 쉽고 빠르게 변환할 수 있는 프로그램에 대해 살펴보았다. 이 프로그램은 생산 공정에서 발생되는 데이터뿐만 아니라, 다른 분야의 데이터 마이닝을 위해서도 사용할 수 있도록 만들었다.

현재에는 국내에서 가장 많이 사용하는 MS*SQL과 Oracle만을 지원하도록 개발되었고 이를 기업 시스템에 연결하여 검증을 수행하였다. 향후에는 접근할 수 있는 데이터베이스 시스템의 영역을 확장하고자 Infomix 와 Sybase에 대해서도 적용할 예정이며, 또한 검

증을 수행하여 가능하면 많은 데이터베이스 시스템에 접근할 수 있도록 확대할 예정이다.

후기

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

참고문현

1. Computational Intelligence: An Introduction, Andries P. Engelbrecht, 2002
2. Koch, George and Loney, Kevin. "The Complete Reference; ORACLE.", 1997.
3. UGS, "Teamcenter Manufacturing, Reference Manual", 2004