

의사결정트리 기법을 이용한 스틸 파이프 생산 수율 및 불량률 측정에 관한 연구

(A Study of Measuring Yield Rate and Error Rate in
Steel Pipe Production using Decision Tree Technique)

김웅경*, 김종완**, 김수연**, 남인길**

(Woong-Kyung Kim, Jongwan Kim, Su-Yeon Kim, In-Gil Nam)

요약 본 연구는 스틸 파이프 생산의 과거 이력을 분석하여 주요 특성별 높은 수율을 갖는 제품을 선별하고 각 공정별 최소 불량률을 가진 제품 모델을 구현함으로써 제품 생산의 효율성을 제고하기 위한 목적으로 수행되었다. 이를 위해 본 논문에서는 주문을 통해 생산되는 각종 스틸 파이프 제품들의 특성을 비교 및 분석하여 어떠한 특성을 가지는 제품들이 가장 높은 수율을 창출해내고 적은 불량률 발생시키는지 의사결정트리 데이터마이닝 기법을 적용하여 분석하였다. 실험 결과로부터 중소구경이 많은 ERW 스틸 파이프는 기호에 의해, 그 외 주로 대구경 범위인 롤벤더, 스파이럴 스틸 파이프 경우에는 외경 범위에 따라 수율과 불량률이 분류되고 있다는 사실을 확인하였다. 본 연구는 주요 특성별 수율과 불량률이 어떠한 형태를 나타내는지 수치적으로 분류 및 구체화하여 그 영향정도를 구분하였다.

핵심주제어 : 스틸 파이프, 생산 수율, 불량률, 의사결정트리

Abstract This research aims to improve the efficiency of production by selecting production configuration with high yield rate and lower error rate based on production history of steel pipe. To achieve this, we identify the properties of various types of MTO(make-to-order) steel pipe products and determine properties affecting yield rate and error rate using decision tree technique. From experimental results, we find out that specification is critical to determine yield rate and error rate of ERW steel pipes with mostly small and medium caliber, and an external diameter range in case of roll benders or spiral steel pipes with mostly large caliber. This research classified and embodied the patterns of yield rate and error rate mathematically by product properties.

Key Words : steel pipe, production yield rate, error rate, decision tree

1. 서론

스틸 파이프(강관)는 아래와 같이 정의된다. 철 중에 탄소함량이 0.04~0.6%인 것을 강(steel)이

라 하고, 강으로 만든 관(pipe)을 강관(steel pipe)라 한다[12]. 대부분 열연코일(hot rolled coil) 또는 철판(plate)을 원소재로 제작한다. 현재 강관, 즉 스틸 파이프(steel pipe)는 굴뚝 산업 특성상 하향 국면으로 접어들고 있으며, 중국의 값싼 제품으로 인하여 가격 경쟁력에서 어려움을 겪고

* 대구대학교 석사과정, 세아네트웍스

** 대구대학교 컴퓨터·IT공학부 교수

있다. 또한 원자재 가격의 급등으로 전체적인 난항이 지속되고 있다. 이에 저부가가치 스틸 파이프 제품의 생산을 지양하고 고부가가치 및 고수익을 창출하는 제품으로의 전환이 필요한 시점이다. 세계 경제위기와 불황속에서 스틸 파이프는 설비능력의 급증과 가동률 하락, 생산과 소비의 입지 불균형, 건설의존적인 수요구조, 중국으로부터 수입재 증가와 수출경쟁력 약화, 해외 라인 파이프(line pipe) 프로젝트 수주능력 취약 등의 문제점을 안고 있다[1].

이러한 문제점을 해결하기 위해서는 향후 저성장이 불가피한 건설용 강관, 특히 공급과잉이 심각한 중소구경 강관부문을 통합하고, 통합화 이후 잉여설비의 해외매각과 폐쇄 등을 추진, 공급과잉 완화 및 가동률 제고를 통해 국내 강관시장에서의 교섭력을 제고시키는 것이 필요하다. 또한 국내 강관 시장의 정체 극복을 위한 수출 확대를 위해서는 경쟁력을 갖춘 고부가가치 제품의 연구개발이 이루어져야 한다. 아울러 원소재인 열연강관의 공급부족을 해소하기 위해 상대적으로 국내 소재 수급여건 및 시황이 좋은 고부가가치 강관 생산 비중을 확대하는 Product Mix를 더욱 강화하고, 고부가가치 강관 개발은 소재 단계에서부터 결정된다는 점에서 강관사와 소재 공급사간의 협력을 강화해야 한다[10].

이러한 스틸 파이프 산업의 문제점과 대응방안을 모색하기 위한 방안으로, 과거 스틸 파이프 생산의 이력을 분석하여 주요 특성별 고부가가치 제품을 선별하고 각 공정별 최소 불량률을 가진 제품 모델을 구현하여, 고수익의 제품 생산을 지향하고 설비증설, 설비보수, 작업자 교육의 자료로 활용하고자 한다. 구체적인 방법으로 주문을 통해 생산되는 각종 강관 제품들의 특성을 비교, 분석하여 어떠한 특성을 갖는 제품들이 가장 높은 수율을 창출해 내고, 적은 불량을 발생시키는지를 고찰한다. 이렇게 수집되고 비교, 분석된 자료들을 의사결정트리 기법으로 모델링함으로써 향후 생산관리를 위한 소중한 자료로 사용하는데 본 연구의 목적이 있다.

2. 관련 연구

2.1 스틸 파이프 산업 현황과 제조공정

국내 스틸 파이프 산업은 1990년대 이후 설비 증가율이 소비증가율을 넘어서 구조적 공급과잉이 심화되었고, 2000년대 들어 중국산 저가재가 수입되어 국내 시장 잠식이 늘어나고 있다. 또한 타 철강업종 대비 시장구조가 파편화되어 불황시 과잉경쟁이 불가피한 상황이다. 이러한 스틸 파이프 산업의 경영실적 부진이 지속되는 가운데, 제품별/규모별 경기 양극화가 심화되고, 건설 및 유통거래 의존도가 높아 가격 변동성 확대가 불가피하다[11].

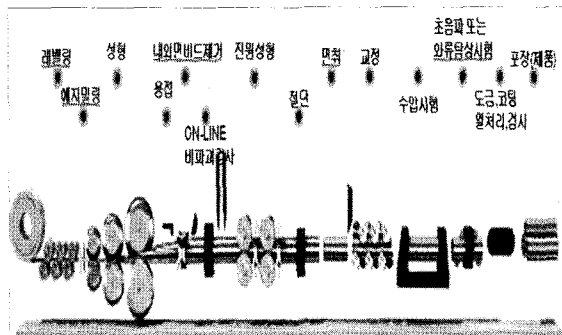
국내 스틸 파이프 산업의 문제점은 과열 경쟁으로 인한 설비능력 급증과 가동률 하락, 대구경 설비는 경상지역에 집중되고 국내 수요 60%는 경인지역에 집중된 생산과 소비의 입지 불균형, 수요가보다 소재에 초점을 맞춘 공장건설 및 건설의존적인 수요구조 등이 지적되고 있다[5].

2008년 하반기부터 부각된 글로벌 금융위기 및 경기침체로 인해 국내 철강 수요는 급감하고 있으며 철강재 가격 역시 급락한 상황이다. 강관 부문 역시 국내 수요 위축으로 인해 실적이 둔화되고 있는 상황으로 경기불황의 골이 깊어질 것으로 보이는 2009년 1분기부터 실적 하락이 본격화되고 있다[11].

스틸 파이프는 제조방법에 따라 ERW(전기저항용접) 방식과 SAW(아크용접) 방식으로 구분된다. ERW 방식은 별도의 용가제 없이 자체용접을 행함으로써 용접성이 우수하며 제조가능범위가 넓고 치수정밀도가 높다. SAW 방식 중 롤벤딩(roll bending) 방식은 모재와 용접부 사이에서 서로 다른 극을 두어 그 사이에서 발생하는 아크열에 의해 용접을 하여 대형 후속 및 고급관의 생산이 가능하므로 원유 및 천연가스 수송용의 라인 파이프를 비롯한 일반배관용, 구조용 등의 용도로 쓰이며, 스파이럴(spiral) 방식은 용접부가 나선형으로 수요자의 주문대로 다양한 외경의 관을 제조할 수 있으며, 주로 상수도용, 토목기초용(pile), 건축용 강관으로 사용된다[12].

스틸 파이프 중 국내에서 주로 생산되는 용접강관은 코일 상태로 감긴 장척의 대강(skelp)을 사용하고, 이것을 관상으로 성형하여 전기저항 또는 전기 유도에 의해 양쪽 에지(edge)에 발생

하는 저항열을 이용 용접한 강관을 말하며, 제조 공정은 다음과 같다. 먼저 원소재가 되는 열연코일을 투입하여 이를 평탄화하는 레벨링 작업을 거친후, 양 끝면을 잘라내는 에지 밀링처리를 한다. 롤러를 통해 파이프의 둥근 모양으로 열연코일을 성형하고, 끝부분을 용접해서 완전한 파이프 모양을 형성한다. 용접으로 생기는 비드를 제거한 후 비파괴검사를 통해 파이프의 용접부위에 대한 이상 유무를 검사한다. 용접된 부위를 열처리하고 냉각후 완전한 파이프 규격으로 성형하는 진원성형을 한다. 제품 길이에 맞도록 길이를 절단하고 관단별로 맞도록 파이프의 끝 단면을 절단하는 면취 과정을 거친다. 완전한 제품 규격을 갖추도록 교정을 한 후 수압, 초음파, 와류탐상시험을 통해 파이프의 이상 유무를 최종파악하고, 도금이나 나사공정을 거친 후 포장공정을 통해 최종 제품으로 생산된다. 그림 1은 스틸 파이프 중 ERW 스틸 파이프의 제조공정을 보여준다[4].



<그림 1> ERW 스틸 파이프의 제조공정

2.2 의사결정트리 기법

의사결정트리는 의사결정규칙(decision rule)을 트리구조로 도표화하여 분류(classification)와 예측(prediction)을 수행하는 분석방법이다. 의사결정트리는 대용량 데이터 해석으로 각광을 받고 있는 데이터마이닝(data mining) 기법의 대표적 기술 가운데 하나이다. 데이터마이닝이란 대용량의 데이터에서 존재하는 숨겨져 있던 데이터간의 관계, 예기치 않았던 패턴, 새로운 규칙 등을 탐색기법으로 찾아내고 모형화해서 유용한 정보로 변환하는 일련의 과정이다[2]. 의사결정트리 기법은 의사결정이란 이름이 나타내듯이 생산 현장이나 기

업 경영자의 의사결정에 도움을 줄 수 있어서 업계에서 많이 사용하는 데이터마이닝 방법이다.

의사결정트리는 다음과 같은 장단점이 있다. 장점으로는 첫째, 트리구조에 의해서 모형이 표현되기 때문에 모형을 사용자가 쉽게 이해할 수 있다. 둘째, 새로운 개체에 대한 판별 또는 예측을 하기 위해서 뿌리노드로부터 끝노드까지를 단순히 따라가면 되기 때문에 새로운 자료에 모형을 적합시키기가 매우 쉽다 셋째, 두 개 이상의 변수가 결합하여 목표변수에 어떻게 영향을 주는지를 쉽게 알 수 있다. 넷째, 선형성이나 정규성 또는 등분산성 등의 가정을 필요로 하지 않는 비모수적인 방법이다. 다섯째, 의사결정트리에서는 순서형 또는 연속형 변수는 단지 순위만 분석에 영향을 주기 때문에 이상치에 민감하지 않다. 단점으로는 첫째, 연속형 변수를 비연속적인 값으로 취급하기 때문에 분리의 경계점 근방에서는 예측오류가 클 가능성이 있다. 둘째, 선형성 또는 주 효과의 결여가 될 수 있다. 셋째, 분석용 자료에만 의존하는 의사결정트리는 새로운 자료의 예측에서는 불안정할 가능성이 높다[9].

의사결정트리는 분류 또는 예측의 과정이 트리 구조에 의한 추론 규칙(induction)에 의해서 표현되기 때문에 분석자가 그 과정을 쉽게 이해하고 설명할 수 있다는 점과 그 자체가 분류 또는 예측모형으로 사용될 수도 있다는 점에서 본 연구의 분석방법으로 채택한다.

3. 스틸 파이프 생산 수율과 불량률 연구 모형

3.1 연구모형 기준

수율관리는 보편화된 용어지만 연구문헌 자체가 매우 적어서 수율관리의 방법론에 대해 실무적으로 적용하거나 학문적으로 접근된 내용은 매우 부족한 편이다. 수율관리의 연구들은 대부분 관리기술적 측면의 수율관리 방안 보다는 고유기술적 측면의 수율향상 방안 또는 알고리즘에 기반한 수율 향상 그리고 모니터링 측면에 집중하고 있다[3,6,8]. 특히, 연구 대상이 특정 분야에 치중되어 있는데 대부분 반도체 및 LCD와 같이 특

수 산업에 관련된 내용이 주를 이룬다[7,13]. 스틸 파이프 관련 연구는 주로 강관 말뚝이라는 스틸 파이프 중 일부 특정 용도에 편중되어 있는 것으로 나타나며, 스틸 파이프 산업의 수율과 관련된 것이 아니라 스틸 파이프 고유의 기술적 기능에 대한 연구에 편중되고 있다. 따라서 본 연구에서는 수율에 관한 부분을 스틸 파이프라는 부분에 한정하며 스틸 파이프의 기술적 설계나 시공법 관련 부분이 아닌, 스틸 파이프의 생산시 나타나는 수율 및 불량률을 주요 특성 항목으로 분류하고자 한다.

3.2 연구모형 대상

본 연구에서 사용된 데이터는 국내 스틸 파이프업계 S사에서 실제 생산된 스틸 파이프에 대한 것으로 약 35,000건의 생산오더 중 대상으로 선정된 것은 약 28,000여 건이다. 이는 2006년 4월부터 2009년 1월까지의 생산된 전체 데이터를 기준으로 하여, 일부 부적합한 데이터는 다음과 같은 기준으로 배제하였다. 첫째, 재작업, 시조관, 중간 공정부터 시작되는 데이터는 전체 공정의 불량 및 정확한 수율을 산정하기 어려워 제외하였다. 둘째, 작업수량이 10개 미만인 데이터는 데이터의 정확도 및 신뢰도 측면에서 제외하였다. 셋째, 특수 공정인 SRM이라는 공정은 불량을 수량으로 관리하지 않고, 중량으로 관리하여 불량수량 집계가 불가능하여 제외하였다. 넷째, 투입량이 없거나 정품량이 없는 데이터는 작업자의 실수나 데이터의 오류로 간주하여 제외하였다.

3.3 연구모형 변수

이 절에서는 스틸 파이프 생산과 관련한 특성별 수율과 불량률을 모델링하여 연구 목적을 달성하기 위한 기초를 마련하고자 한다. 본 연구는 다음과 같은 연구 질문을 가지고 시작하였다. 첫째, 기존 생산된 스틸 파이프의 주요 특성별 수율과 불량률은 어떠한 형태로 모형화되며, 이로 인해 예측할 수 있는 결과는 무엇일까? 둘째, 수율과 불량률의 각 특성의 상관관계 및 연관성은 어떠한가? 가장 큰 영향을 주는 특성(인자)은 어떤 것인가? 두 가지의 질문을 해결하기 위하여 과거

스틸 파이프의 주요 특성별 생산 자료와 수율, 불량률의 자료를 수집하여 분석하였으며, 스틸 파이프 생산에 영향을 끼치는 주요 특성은 표 1과 같이 파악되었다. 이는 표 1에 나타난 8가지의 특성이 제품을 이루는 핵심 요소이며, 이러한 8가지의 특성중 하나만 달라도 다른 제품으로 구분되기 때문이다.

<표 1> 스틸 파이프의 주요 특성

변수명	형식	변수 설명
물종	문자	제품의 종류
기호	문자	제품의 제원 혹은 설계명세서
관단	문자	제품의 단면처리 방법
외경	숫자	제품의 바깥 지름
두께	숫자	제품의 두꺼운 정도
길이	숫자	제품의 길이
패킹	문자	제품의 포장방법
코팅	문자	제품의 표면처리 방법

3.4 연구모형별 측정방법

3.4.1 수율과 불량률의 분포모형

표 1과 같이 스틸 파이프의 주요 특성(인자)으로 선정된 8가지 특성(물종, 기호, 관단, 외경, 두께, 길이, 패킹, 코팅)에 대해서 표 2와 같이 독립변수와 종속변수로 선정하고, 각 특성별 수율과 불량률을 그래프로 표시하여 각 특성별 수율과 불량률에 대한 분포모형을 구현하고자 한다. 특성값이 문자인 물종, 기호, 관단, 패킹, 코팅은 순위를 매겨 어떤 특성값이 가장 수율과 불량률이 높은지를 나타내고, 특성값이 숫자인 외경, 두께, 길이는 특성값의 변화에 따른 수율과 불량률의 추이를 나타내도록 구현하여 이러한 분포모형의 결과를 분석하고 그 결과를 제시한다.

<표 2> 수율과 불량률의 분포모형 변수 구성

독립변수	종속변수
물종, 기호, 관단, 외경, 두께, 길이, 패킹, 코팅	수율
	불량률

3.4.2 의사결정트리를 이용한 분류모형

스틸 파이프의 주요 특성(인자)으로 선정된 8 가지 특성을 위의 표 2와 같이 독립변수로 구성하고, 이에 따른 수율과 불량률을 2가지 종속변수로 하여 의사결정트리 기법을 이용하여 분류모형을 구현하고자 한다. 분류모형의 결과를 분석하여 루트노드와 주요노드, 그리고 링크(link)에 해당하는 주요 특성들이 어떤 것인지를 규명하여 스틸 파이프 생산 시 주요 특성들의 영향정도와 상관관계를 도출한다.

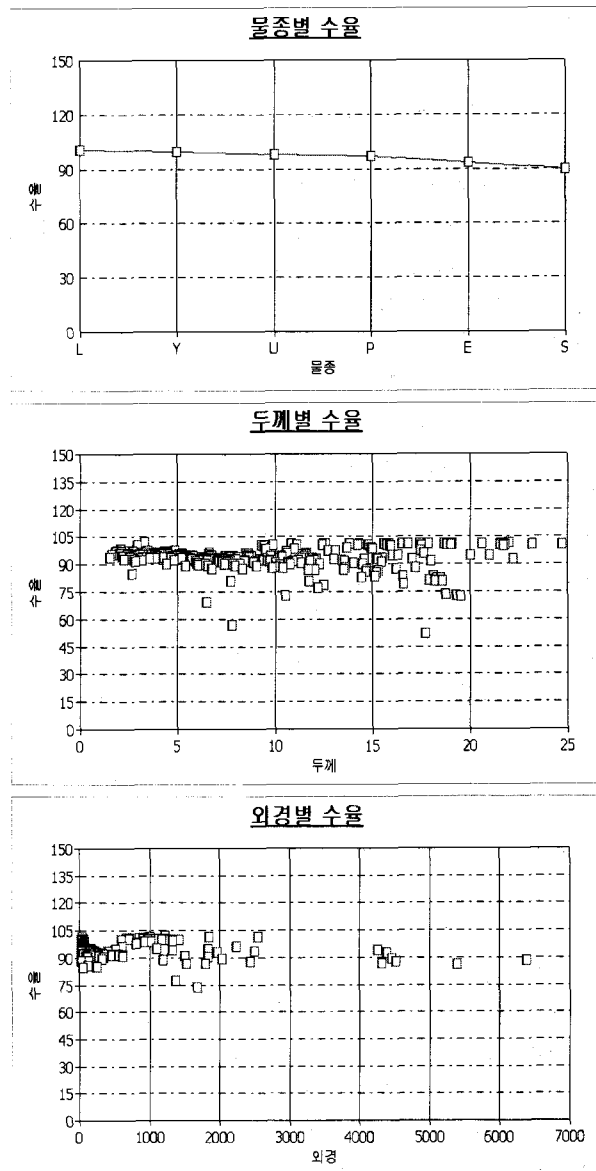
4. 생산 수율 및 불량률 측정결과

4.1 연구모형별 측정결과

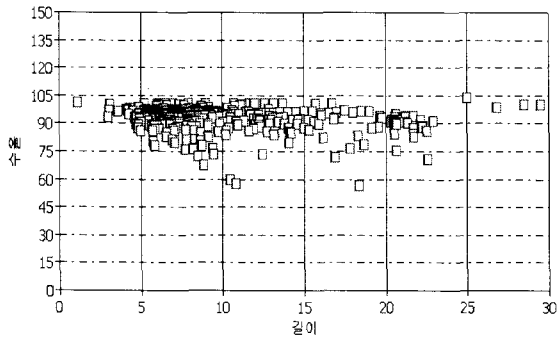
4.1.1 수율과 불량률의 분포모형 측정결과

주요 특성별 불량률 및 수율에 대한 모형은 그림 2, 그림 3과 같으며, 이러한 모형의 결과로 볼 때, 불량률의 경우, 물종별로는 L(롤벤더강관), Y(강관전주) 물종은 불량률이 0, 다음으로는 E(전기저항용접강관), P(스파이럴강관), U(스테인레스강관)의 순으로 나타났다. 기호별로는 LIFE600이라는 기호가 특히 높은 것으로 나타났으며, 그 외 기호들은 대체로 비슷한 불량률을 보인다. 관단의 경우, BME 관단이 다소 높은 불량률을 보이며, 나머지 관단들은 유사한 불량률을 보인다. 외경은 약 610mm 이하구간인 중소구경일 때 대구경에 비해 대체로 발생률이 높으며, 두께는 두꺼워질수록 불량률이 높아지고, 길이에 의한 영향정도는 크게 없으나 일부 제품은 길이가 길수록 불량률이 높아진다는 것을 알 수 있다. 나머지 패키징, 코팅에 대한 특성별로는 대체로 유사한 불량률을 나타낸다. 수율의 경우, 전반적으로 고른 분포를 나타내는데, 대체로 90~105% 사이의 분포를 보이며, 일부 두께가 두꺼운 제품에서 낮은 수율을 나타내고, 길이에 의해서는 크게 영향을 받지 않지만 길이가 약 6~12m 구간에서 낮은 수율을 나타내고 있다. 수율이 100%가 넘는 경우가 발생하는 이유는 스틸 파이프의 투입중량의 경우 주 원자재인 코일(Coil)은 무게를 계근한

실중량이고, 제품중량은 코일을 가공하여 파이프 형태가 이루어질 때 매번 무게를 계근할 수 없어 중량계산식($(\text{외경}-\text{두께}) \times \text{두께} \times 0.22466 \times \text{길이}$)을 통해 이론중량으로 관리가 되므로 투입중량 대비 제품중량인 수율은 실중량과 이론중량의 차이에 의해서 100%를 초과할 수도 있다. 이는 주로 생산되는 제품이 6~12m 사이의 크기가 많기 때문이며, 길이에 의해서 분포모형의 결과가 크게 좌우된다고 할 수는 없다.

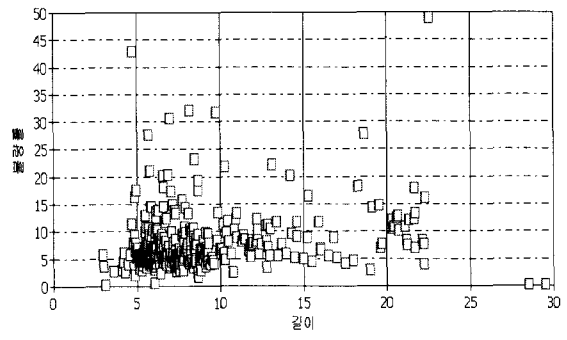


길이별 수율



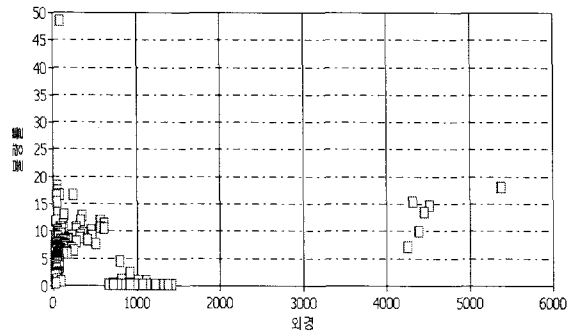
<그림 2> 주요 특성별 수율

길이별 불량률

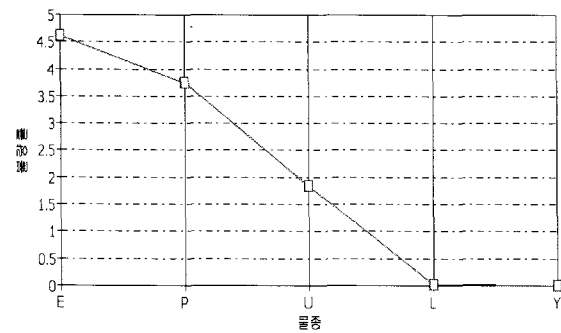


<그림 3> 주요 특성별 불량률

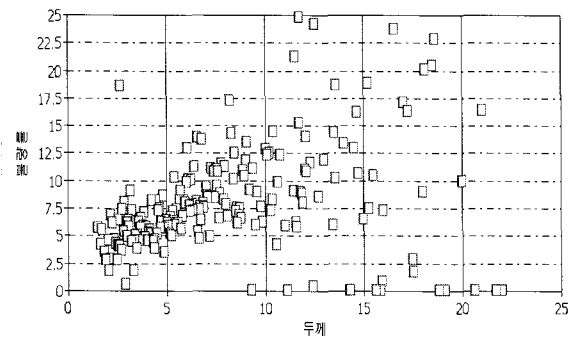
외경별 불량률



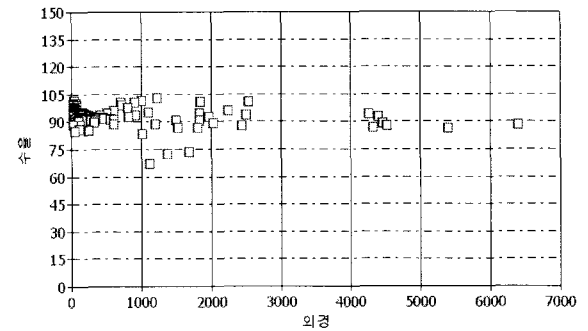
물중별 불량률



두께별 불량률



외경별 수율



두께별 수율

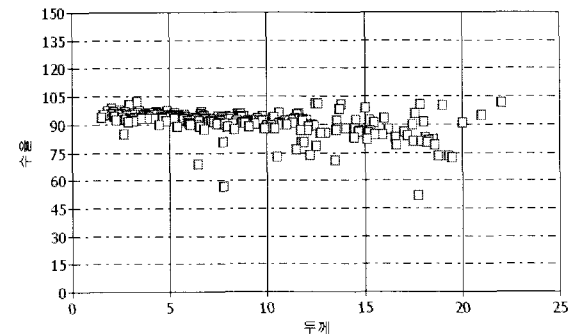
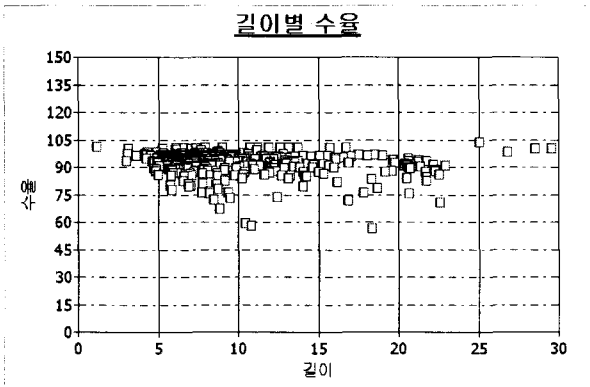
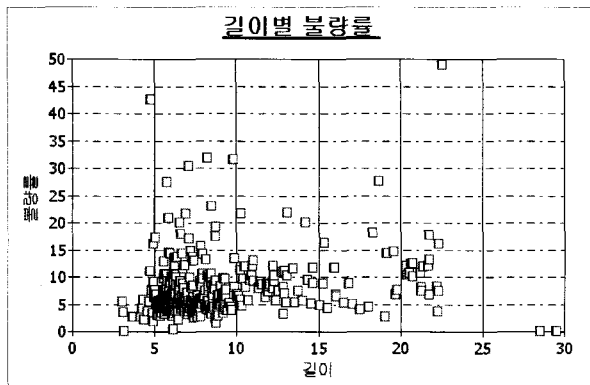
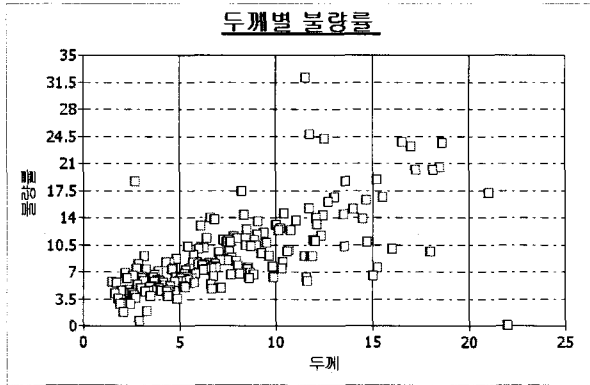
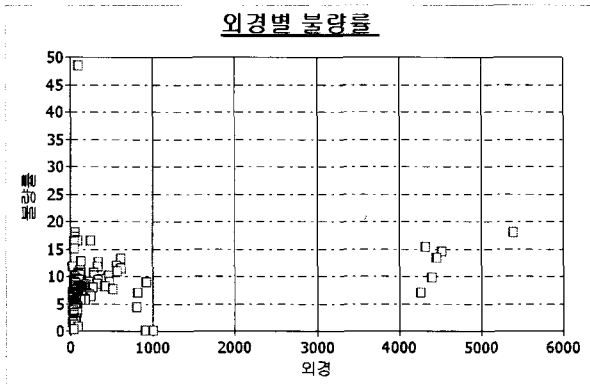


그림 2, 그림 3에 나타난 것처럼 물중이 L, Y 일때,

수율은 거의 100%를 넘어서고, 불량률은 0% 이기 때문에 분포모형의 정확도를 높이기 위해 물중 L, Y를 제외한 데이터를 선별하여 분포모형을 구현하였으며, 그 결과는 그림 4, 그림 5과 같다.



<그림 4> 물종 L, Y를 제외한 주요 특성별 수율



<그림 5> 물종 L, Y를 제외한 주요 특성별 불량률

물종 L, Y를 제외한 분포모형인 그림 4, 그림 5의 결과를 보면 물종 L, Y를 제외하기 전인 그림 2, 그림 3과 큰 차이를 보이지 않는다. 이는 분포모형이나 분석의 오류가 아니라, 물종 L, Y인 데이터의 건수가 전체 분석 건수에 비해 차지하는 비중이 크지 않은 것으로 분포모형의 신뢰도에는 영향을 미치지 않는다.

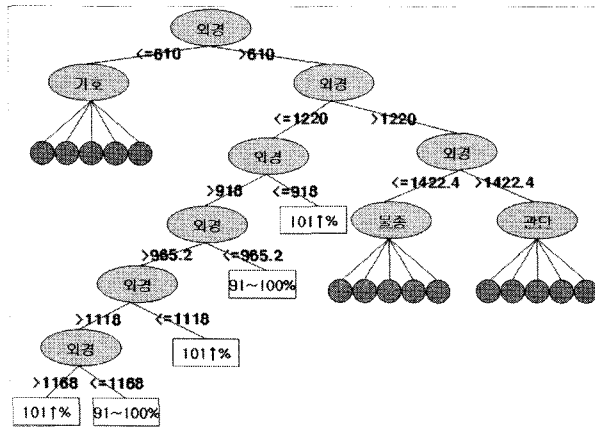
4.1.2 의사결정트리를 이용한 분류모형 측정 결과

수율과 불량률의 각 특성의 상관관계 및 연관성은 어떠하며 가장 큰 영향을 주는 인자는 어떤 것인가라는 두 번째 연구 질문에 대한 측정은 Witten과 Frank[14]가 보급한 데이터마이닝 개발 도구인 WEKA 틀에서 제공되는 J48 의사결정트리 기법을 사용한다. WEKA(Waikato Environment for Knowledge Analysis)는 자바 언어로 개발된 기계학습 및 데이터마이닝을 위한 소프트웨어로서 주요 기능으로는 데이터 전처리 기능, 학습 알고리즘 및 평가 방법, 학습 알고리즘 비교 기능, 처리 결과를 시각적으로 보여주는 시각화(visualization) 기능 등이 제공된다. 이해의 편의를 위하여 종속변수인 수율과 불량률 값을 표 3과 같이 구간 범위에 따른 대표값으로 정의하였다.

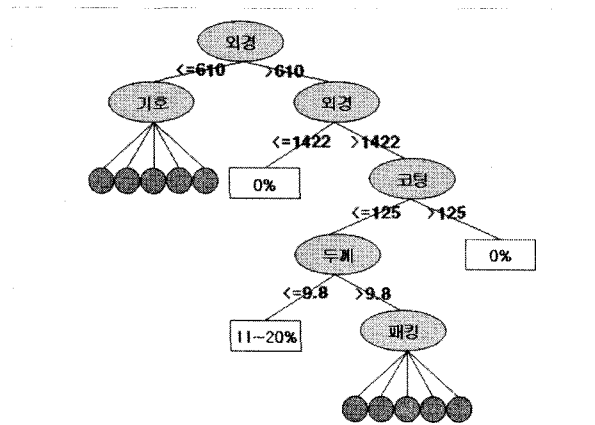
<표 3> 종속변수 구간별 변환표

수율 구간	불량률 구간	종속변수
> 100	= 0	A
100 ≤ AND > 90	1 ≥ AND ≤ 3	B
90 ≤ AND > 80	3 > AND ≤ 6	C
80 ≤ AND > 70	6 > AND ≤ 10	D
70 ≤ AND > 60	10 > AND ≤ 20	E
60 ≤ AND > 50	20 > AND ≤ 30	F
50 ≤ AND > 40	30 > AND ≤ 40	G
40 ≤ AND > 30	40 > AND ≤ 50	H
30 ≤ AND > 20	50 > AND ≤ 60	I
20 ≤ AND > 10	60 > AND ≤ 70	J
10 ≤ AND > 6	70 > AND ≤ 80	K
6 ≤ AND > 1	80 > AND ≤ 90	L
= 0	90 > AND ≤ 100	M

이러한 의사결정트리 기법의 분석결과, 수율은 약 69%, 불량률은 약 61%의 분류 정확도를 나타내며 트리모형은 그림 6, 그림 7와 같다.



<그림 6> 의사결정트리를 이용한 수율 트리

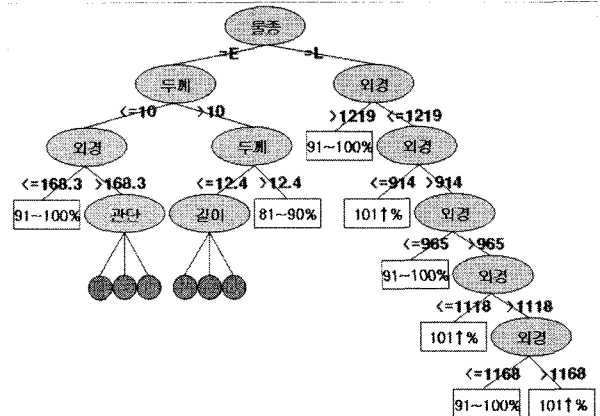


<그림 7> 의사결정트리를 이용한 불량률 트리

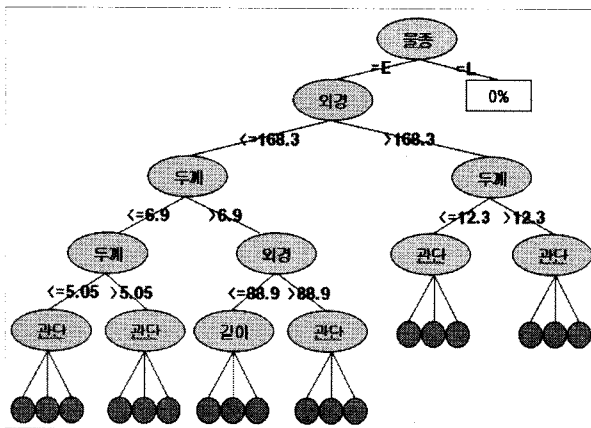
이러한 의사결정트리 모형의 결과는 수율과 불량률 모두 독립변수 중 외경이라는 특성이 트리의 루트노드가 되며, 외경이 610mm 보다 큰지 아닌 지에 따라 1차 분류트리가 나뉘지고, 610mm 보다 같거나 작을 경우에는 기호에 따라, 클 경우에는 다시 수율은 외경이 1220mm 보다 큰지 작은지에 따라, 불량률은 외경이 1422.4mm 보다 큰지 작은지에 따라 분류되고 있음을 보여 준다. 좀더 하위 레벨의 트리를 살펴보면, 수율의 경우 외경이 610mm를 초과하면 외경 1220mm를 기준으로 그 이하면 외경 918mm를 기준으로 세분되고, 1220mm를 초과하면 다시 외경 1422.4mm를 기준으로 초과면 관단 특성에 의해

서, 이하면 각 물종별로 수율이 세분화된다. 불량률의 경우, 외경이 610mm를 초과할 때, 다시 외경 1422.4mm를 기준으로 그 이하면 모두 불량률이 0%이고, 그보다 크면 코팅특성 125를 기준으로 초과면 불량률이 0%이고, 그 이하면 두께 9.8mm를 기준으로 이하면 불량률은 약 11~20%이고, 초과면 각 패킹특성에 의해 세분화됨을 알 수 있다. 일반적으로 스틸 파이프는 통상 외경 610mm를 기준으로 그 이상이면 대구경, 미만이면 중소구경 파이프로 구분된다. ERW 스틸 파이프는 외경의 규격 범위가 최대 609.6mm으로 중소구경에 해당되는데, 분류트리의 결과로 볼 때, ERW 스틸 파이프는 기호에 의해, 그 외 주로 대구경 범위의 롤벤더, 스파이럴 스틸 파이프의 경우에는 외경범위에 따라 수율과 불량률이 분류되고 있음을 확인할 수 있다.

전체 특성을 독립변수로 하고 분류한 수율 및 불량률의 의사결정트리는 그림 6, 그림 7과 같이 분류 모형이 복잡하여 이해하기 어려우므로 주요 특성 및 독립변수의 하나인 기호의 값 중 5LBL1인 것만 대상으로 의사결정트리를 이용하여 다시 분류하였다. 5LBL1인 데이터 대상으로 의사결정트리 기법을 통해 나타난 분석결과는 수율은 약 72%, 불량률은 약 68%의 분류 정확도를 나타내며, 트리모형은 그림 8, 그림 9와 같다. 이 결과는 전체 기호 대상으로 분류 작업한 결과 보다는 정확도가 개선되었음을 의미하고, 현업 실무자가 본 연구결과를 해석하기 쉽도록 트리모형이 단순해진다는 장점이 있다.



<그림 8> 기호 5LBL1의 수율 트리

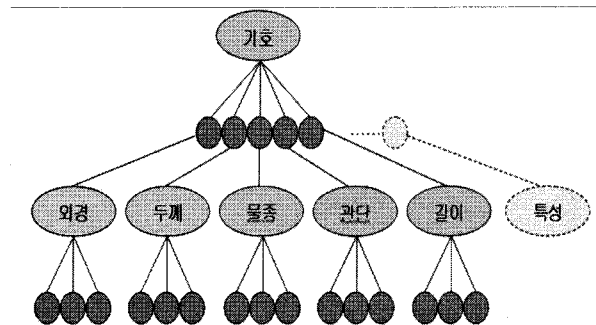


<그림 9> 기호 5LBL1의 불량률 트리

이렇게 주요 특성 가운데 기호가 5LBL1인 데이터를 의사결정트리로 분류한 모형의 결과는, 수율은 물종이 L인 경우는 외경에 의해서, 물종이 E인 경우는 두께에 의해서 하위 특성으로 분류되는데, 물종이 L인 경우, 외경이 1219mm가 초과될 경우는 수율이 약 91~100%, 외경이 1219mm이하일 경우에는 구간별 외경범위에 의해서 수율이 영향을 받는다는 것을 알 수 있다. 수율이 100%를 초과하는 고수율 외경범위는 1168mm 초과 구간, 966~1118mm 구간, 914mm 이하 구간인 것으로 나타났다. 물종이 E인 경우는 두께가 10mm초과이면 두께 12.4mm를 기준으로 12.4mm 초과이면 수율은 81~90%이고, 12.4mm 이하면 길이 6.157m를 기준으로 수율이 분류되고, 10mm이하면 외경 168.3mm를 기준으로 그 이하면 91~100%이고, 초과면 관단에 의해서 세분되고 있음을 보여준다. 불량률은 물종이 L인 경우는 불량률이 거의 발생하지 않는 것으로 나타났으며, 물종이 E인 경우는 외경 168.3mm를 기준으로 그 이하면 두께 6.9mm를 기준으로 분류되고, 초과면 두께 12.3mm를 기준으로 불량률이 분류된다고 모형은 제시한다. 이는 두께가 두꺼울수록 용접 등의 처리공정이 어렵고, 두께가 12.3mm를 초과할 경우 고두께 사양이 되어, 불량률이 발생할 소지가 많다는 것을 수치적으로 보여준다.

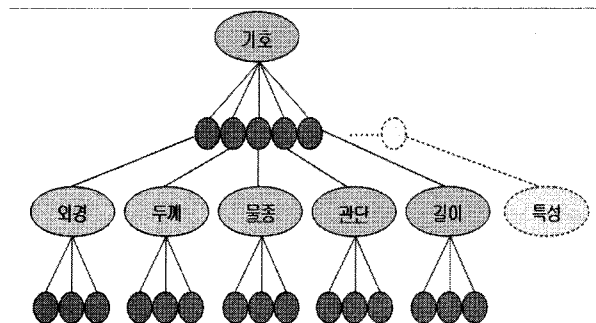
앞서 4.1.1절 수율과 불량률의 분포모형 추정결과에서 나타난 것처럼 물종이 L, Y일때, 수율은 거의 100%를 넘어서고, 불량률은 0%이기 때문에 분류모형의 신뢰도를 높이기 위해 물종 L, Y를

제외한 데이터를 선별하여 분류모형을 구현하였으며, 그 결과는 그림 10, 그림 11, 그림 12, 그림 13과 같다.



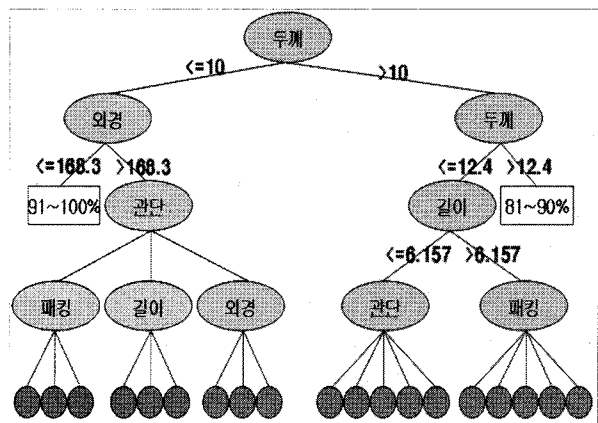
* 기호의 종류가 너무 다양하고, 기호에 따른 하위 특성이 불규칙적으로 분류되어 최상위 레벨인 기호이외에 하위 레벨에 대한 분류모형으로 구현하여 세부적으로 표현하기 불가능함.

<그림 10> 물종 L, Y를 제외한 수율 트리

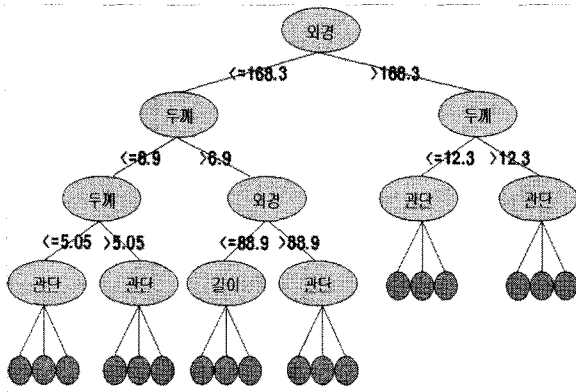


* 기호의 종류가 너무 다양하고, 기호에 따른 하위 특성이 불규칙적으로 분류되어 최상위 레벨인 기호이외에 하위 레벨에 대한 분류모형으로 구현하여 세부적으로 표현하기 불가능함.

<그림 11> 물종 L, Y를 제외한 불량률 트리



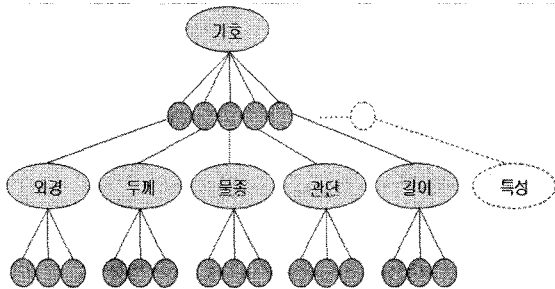
<그림 12> 물종 L, Y를 제외한 기호5LBL1의 수율 트리



<그림 13> 물종 L,Y를 제외한 기호5LBL1의 불량률 트리

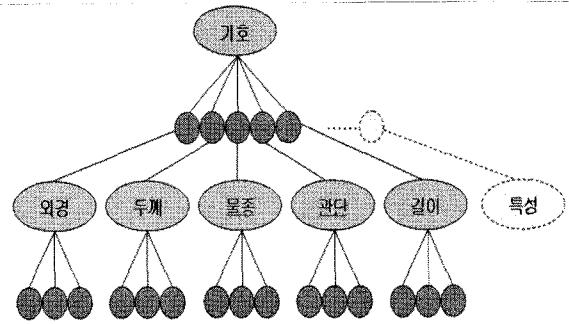
그림 10, 그림 11, 그림 12, 그림 13에서 물종 L, Y를 제외하고 분류한 분류모형은 물종 L, Y를 제외하지 않고 분류한 분류모형 그림 6, 그림 7, 그림 8, 그림 9와 서로 비교했을 때, 전체 데이터의 경우 수율과 불량률 모두, 외경범위가 610mm이하인 부분, 즉 중소구경은 기호에 의해서 세분화되어 분류되고 있음을 구체적으로 증명해 주며, 불량률은 분류모형의 좌측부분인 물종이 E일 경우에 나타나는 분류모형의 세분된 모형을 좀더 구체적으로 나타냈다. 이는 앞서 물종 L, Y를 제외하지 않고 분류했던 분류모형에서 중소구경에 대한 세분된 분류모형의 특징을 좀더 정확하고 구체적으로 나타낸 결과라고 할수 있다.

아울러 분류모형의 신뢰도를 높이기 위해 여러 물종중 중소구경의 대표물종인 E물종과 대구경의 대표물종인 L물종만 가지고 수율과 불량률에 대한 분류모형을 구현해 보았고 그 결과는 그림 14, 그림 15, 그림 16과 같다.



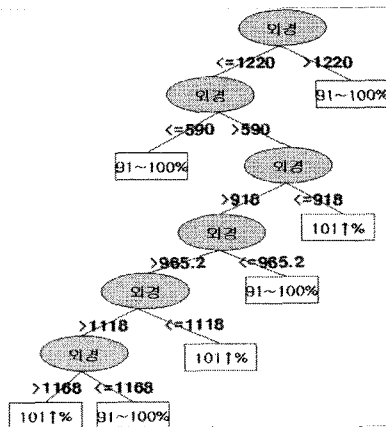
* 기호의 종류가 너무 다양하고, 기호에 따른 하위 특성이 불규칙적으로 분류되어 최상위 레벨인 기호이외에 하위 레벨에 대한 분류모형으로 구현하여 세부적으로 표현하기 불가능함

<그림 14> 물종 E의 수율 트리



* 기호의 종류가 너무 다양하고, 기호에 따른 하위 특성이 불규칙적으로 분류되어 최상위 레벨인 기호이외에 하위 레벨에 대한 분류모형으로 구현하여 세부적으로 표현하기 불가능함

<그림 15> 물종 E의 불량률 트리



<그림 16> 물종 L의 수율 트리

물종 L의 불량률 트리를 별도로 표기하지 않은 이유는 그림 3. 주요 특성별 불량률의 물종별 불량률과 그림 9. 기호 5LBL1의 불량률 트리에서 나타나듯이 불량률이 모두 0으로 트리모형자체가 무의미하기 때문이다. 이처럼 여러 형태의 분류모형을 구현하여 비교분석한 결과 모두 동일한 결과를 나타내었으며, 본 연구에서 수행한 분포모형과 분류모형의 측정결과는 S사에서 스틸 파이프 생산 업무를 10년 이상 담당하고 있는 현업 전문가 그룹을 대상으로 심층 인터뷰를 통하여 그 타당성을 확인하였다.

5. 결 론

본 연구는 스틸 파이프 생산에 영향을 미치는 주요 특성에 대한 연구모형을 제시하였으며, 주

요 특성으로는 표 1에 나타난 것처럼 물종, 기호, 관단, 외경, 두께, 길이, 패킹, 코팅을 독립변수로 하고 수율과 불량률을 종속변수로 하여, 독립변수가 종속변수에 어떠한 영향을 미치는가를 파악함으로써 향후 스틸 파이프 생산 시 저불량율, 고수율의 제품을 생산하는데 주요한 지표로 활용하고자 하였다. 구체적으로는 주요 특성별 수율과 불량률을 모델링하였고, WEKA 툴을 이용하여 주요 특성별 수율과 불량률의 분류 모형을 구현하였다. 그 결과 주요 특성별 수율과 불량률의 분포를 분석할 수 있게 되었으며, 주요 특성들이 수율과 불량률에 미치는 영향정도를 측정할 수 있게 되었다.

본 사례연구를 통해서 확인할 수 있었던 핵심 요소는 중소규격이 많은 ERW 스틸 파이프는 기호에 의해서, 그 외 주로 대구경 범위인 롤벤더, 스파이럴 스틸 파이프의 경우에는 외경범위에 따라서 불량률, 수율이 분류되고 있다는 사실을 알 수 있었다. 이러한 분석 결과를 토대로 향후 제품 생산시 제품의 유형에 따른 주요 핵심 특성을 고려한 설비증설이나 작업자의 작업 집중도 관리가 이루어져야 할 것이다. 과거 스틸 파이프 산업에 대한 주요 특성별 수율 측정에 대한 연구가 미흡하였으나, 본 논문에서 이를 구체화 하여 주요 특성별 수율과 불량률이 어떠한 형태를 나타내는지 수치적으로 분류 및 구체화하여 그 영향정도를 구분한 것은 의미있는 결과이다.

향후에는 본 연구를 기초로 특정 회사가 아닌 스틸 파이프 산업 공통의 표준 데이터 추출 및 선별 작업이 요구된다. 또한 통계학에서 주로 사용하는 주요 특성간의 인자분석(factor analysis), 다변량 분석(multivariate analysis) 등을 통하여 독립변수의 제거 및 조정을 통한 분류정확도를 높이는 연구를 수행하여 분류결과 및 모형에 대한 신뢰도를 높이는 과정이 필요하다.

참 고 문 헌

[1] 공원일, “강관: 2002년 철강산업 업종별 경기전망”, 한국철강협회, 철강보 제28권 제1호 통권317호, pp. 14-16, 2002.
 [2] 김종완, 진승훈, 김병익, 김태균, 김영순,

“전자상거래에 데이터마이닝 기술 적용”, 과학기술연구, pp. 145-153, 2001.
 [3] 백동현, “데이터마이닝을 활용한 반도체공정의 지능형 수율관리시스템”, 천안대학교, 2003.
 [4] 세아제강, 제품/기술정보-생산공정-ERW 강관”, http://www.seahsteel.co.kr/02_product/product03.asp.
 [5] 손영욱, “한국 강관산업 현황과 발전전략”, 제3회 강관산업 발전세미나, 2007.
 [6] 이동권, “수율관리 체계화를 통한 수율향상 방안 연구”, 아주대대학원, 2007.
 [7] 이윤식, “초고집적 환경에서의 반도체 수율 분석에 관한 연구”, 한국정보과학회 학술발표논문집, 29(1), pp. 733-735, 2002.
 [8] 이장희, “데이터마이닝 도구의 혼합적용 방법간 수율 예측 성능 비교 연구”, Journal of Business Research, 23(1), pp. 233-310, 2008.
 [9] 최종후 외, “데이터마이닝 예측 및 활용”, SPSS아카데미, 2002.
 [10] 한국철강협회, “강관보고서”, pp.1-12, 2007.
 [11] 한국철강협회, “세계 및 국내 철강산업 동향”, 2009-1호('09.1월), pp. 16-26, 2009.
 [12] 현대하이스코, 제품정보-강관제품-제품 개요”, <http://www.hysco.com/hrml/product/kangkwan/outline.aspx>.
 [13] 하정훈, “LCD생산시스템에서 repair와 rework을 고려한 수율과 원가분석모델”, Journal of the Korean Institute of Industrial Engineers, 33(3), pp. 364-372, 2007.
 [14] I. Witten and E. Frank, Data Mining: practical machine learning tools and techniques, 2nd ed. Morgan Kaufmann, 2005.



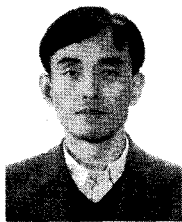
김 용 경 (Woong-Kyung Kim)

- 2004년 2월 : 위덕대학교 컴퓨터공학과(공학사)
- 2009년 8월 : 대구대학교 컴퓨터정보공학과(공학석사)
- 2004년 1월 ~ 현재 : 아네트웍스 SM사업부 재직중
- 관심분야 : 데이터베이스, 데이터마이닝, 유비쿼터스, 정보보안
- E-mail: kwk0910@hanmail.net
- Tel: +82-10-9336-0910



남 인 길 (In-Gil Nam)

- 1978년 2월 : 경북대학교 전자공학과(공학사)
- 1981년 2월 : 영남대학교 전자공학과(공학석사)
- 1992년 2월 : 경북대학교 전자공학과(공학박사)
- 1990년 3월 ~ 현재 : 대구대학교 컴퓨터·IT공학부(교수)
- 관심분야 : 데이터베이스, 데이터베이스 보안, 정보검색
- E-mail: ignam@daegu.ac.kr
- Tel: +82-53-850-6573



김 종 완 (Jong-wan Kim)

- 1987년 2월 : 서울대학교 전자계산기공학과(공학사)
- 1989년 2월 : 서울대학교 컴퓨터공학과(공학석사)
- 1994년 8월 : 서울대학교 컴퓨터공학과(공학박사)
- 1995년 3월 ~ 현재 : 대구대학교 컴퓨터·IT공학부(교수)
- 관심분야 : 인공지능, 데이터마이닝, IT융합서비스
- E-mail: jwkim@daegu.ac.kr
- Tel: +82-53-850-6575

논문접수일 : 2009년 8월 31일
 논문수정일 : 2009년 10월 13일
 게재확정일 : 2009년 10월 15일



김 수 연 (Su-Yeon Kim)

- 1991년 2월 : 포항공과대학교 수학과(이학사)
- 1997년 8월 : 숭실대학교 정보산업학과(이학석사)
- 2003년 3월 : 포항공과대학교 산업공학과(공학박사)
- 2004년 3월 ~ 현재 : 대구대학교 컴퓨터·IT공학부(교수)
- 관심분야 : 지식경영, e-비즈니스, 고객관계관리, 유비쿼터스 컴퓨팅
- E-mail: sykim@daegu.ac.kr
- Tel: +82-53-850-6585