데이터마이닝을 이용한 철강기업의 시간당 수익성 탐색 및 예측
A study on time based performance evaluation of the steel industry using data mining

유성찬*, 안영효**, 진광우***, 박명섭****
*고려대학교강사(scyoo@korea.ac.kr), **포스코경영연구소(yhahn@posri.re.kr),
***포스코(jinkw@posco.co.kr), ****고려대학교 경영학과 교수(mspark@korea.ac.kr)

Abstract

A performance measure is the critical work of business processes. Especially it is necessary to explore and forecast the profits per hour through this activity. However there have been rarely studied on specific fields, like the steel industry. Therefore this study analyzes and evaluates the time based performance of the steel industry using data mining.

1. 머리말

기업은 생존을 위하여 지속적인 노력을 하고 있으며, 당면한 과제를 해결하기 위해 적절한 전략을 선택하여 실행에 옮기고 있다. 이러한 기업이 추구하는 가장 근본적인 목표로 수익창출을 꼽을 수 있으며, 수익창출에 관련된 활동 중 수익성에 대한 평가 및 예측 또한 중요한 의미를 가지고 있다. 제조기업은 생산물에 대하여, 서비스 기업은 고객서비스에 대하여 기업특성에 맞는 방법을 이용하여 기업 성과 또는 수익성을 평가하고 예측하고 있다. 수익성 평가가 필요한 이유는 과거와 현재에 대한 분석을 통하여 보다 향상된 미래의 설계 방향을 제시할 수 있기 때문이다.

철강기업의 경우도 예외가 될 수는 없다. 철강기업은 제조업의 한 분야로서 한 국가 또는 한 사회의 기간산업 역할을 해오고 있다. 산업사회에서 중요한 역할을 담당하고 있는 만큼 고유 특수성을 반영한 수익성 평가 및 예측은 개별기업의 차원을 넘어서는 의미있는 연구라 할 수 있다.

철강기업의 수익성 분석은 과거에는 단순하게 업무 담당자의 수작업을 통해 제품별 분석 또는 단위 공정별 분석을 수행하는 정도였다. 그러나 다량의 데이터 분석작업, 인원부족, 시간적인 제약으로 인하여 수익성 분석은 결과를 도출하는데 급급했으며, 실질적으로 업무에 사용될 만한 결과 값을 얻기가 상당히 어려웠다.

본 연구에서는 기존의 단순한 제품별, 공정별로 행해지던 철강기업 수익성 평가방법과는 차별화된, 데이터 마이닝을 이용한 수익성 평가 및 예측에 대하여 연구하고자 한다. 데이터 마이닝을 이용한 시간당 수익성 분석, 수익성에 영향을 미치는 인자간의 영향력 정도 측정은 기존 분석에서는 얻을 수 없었던 보다 정교하고 구체적인 결과 값을 제시할 수 있다. 특히 사례분석을 통하여 데이터 마이닝이 철강기업의 수익성을 평가 및 예측하는데 어떻게 사용될 수 있는지 실제 데이터를 이용하여 분석하고자 한다.

본 연구는 다음과 같이 진행하였다. 2장에서는 문헌연구를 통하여 철강기업 수익성 평가에 대한 기존연구에 대하여 살펴보고 데이터 마이닝에 대한 기본적인 개념 및 사용분야에 대하여 설명한다. 3장 사례분석에서는 철강기업의 시간당 수익성 탐색을 위하여 마이닝 기법을 도입, 모델링, 결과 도출하는 순서로 연구하였다. 4장에서는 3장을 통해서 얻은 결과를 분석하였으며, 5장 결론에서는 연구의 최종 분석결과에 대한 결론 및 요약, 향후 연구과제에 대하여 논하였다.

2. 시간당 수익성 및 데이터 마이닝

2.1 시간당 수익성 분석

철강기업의 수익성을 분석하는 기존 연구 방법은 크게 재무적 평가방법과 비재무적 성과측정 방법으로 나눌 수 있다.

재무적 방법을 이용하여 기업의 수익성을 분석한다는 것은 일정기간 또는 일정시점에서 기업의 재무제표 또는 회계수치를 이용하여 수익성을 비교 및 분석하는 것이라 할 수 있으며, 때로는 통계기법을 사용하여 미래의 수익성을 예측하기도 한다. 정진우(1993)는 철강기업의 재무구조에 관한 연구를 통하여 한국철강기업의 합리화 방안을 제시하였다[5]. 이 연구에서는 한국의 철강산업에 대한 재

무구조를 성장성, 수익성, 활동성, 유동성, 안전성, 생산성의 6가지 카테고리로 나눈 후 각각의 구조를 조사하였다. 이를 바탕으로 타 산업 및 외국의 주요 철강 기업과의 비교를 통하여 우리나라 철강기업의 재무구조 합리화 과제 및 방안에 대하여 논하였다.

비재무적 방법은 최적화 기법 도입 또는 평가 모델을 개발하는 방향으로 연구가 진행되었다. 박정원(1999)은 생산계획 수익성 평가를 위한 최적화 모델을 연구하였다[4]. 먼저 철강기업인 POSCO를 대상 모델로 선정하였으며, 대상항목은 POSCO의 전체 생산 물량의 대부분을 차지하고 있는 열연, 냉연, 후판 부문을 설정하였다. 이후 각 부문별로 최적화 기법을 이용하여 생산성 및 수익성 최대화 모델을 구성하였으며, 민감도 분석을 통하여 제약조건 변동시 목적함수에 미치는 영향, 즉 생산성 및 수익성의 차이를 분석하였다.

이상과 같은 연구가 있지만 철강산업에 대한 직접적인 수익과 관련된 연구는 부족한 실정이며 학계를 통한 연구가 활발히 이루지지 못한 것도 사실이다. 따라서 기존에 연구되었던 재무적/비재무적 방법의 틀을 벗어난, 직접적으로 수익성에 영향을 미치는 인자들에 대한 분석은 의사결정을 보다 효과적으로 지원할 수 있을 것이다.

2.2 데이터마이닝

데이터 마이닝은 데이터베이스로부터 추출 가능한 어떠한 패턴을 감지하는 것을 자동화하는 기법으로 간단히 정의될 수 있다[6]. 기업의 시각에서 데이터 마이닝에 대한 개념을 정리한다면, 대용량의 자료로부터 기업의 경쟁력 확보를 위한 의사결정을 돕는 유용한 정보를 찾아내는 일련의 분석과정이라 할 수 있다[1]. 이를 위해 통계학, 인공지능, 계량경제학 등 다양한 학문분야의 분석 방법들이 데이터 마이닝에 사용되고 있다. 데이터 마이닝에 사용되는 분석기법으로는 군집화(clustering analysis), 의사결정나무(decision tree), 시장 바구니 분석(market basket analysis), 신경망 분석 등이 대표적인 방법이라 할 수 있다[8].

분석 주제 및 목적에 따라 다양한 데이터 마이닝 기법이 개발되고 있으며, 사용자는 원하는 분석 결과 값을 얻기 위하여 단일 방법이 아닌 두 가지 이상의 분석방법을 사용한 복합적인 형태의 데이터 마이닝 기법을 사용하기도 한다. 본 연구에서도 세분화된 수익성 분석을 위하여 단일 방법이 아닌 2가지 이상의 분석기법을 사용하였다.

주어진 문제에 대하여 데이터 마이닝 기법을 도입하여 문제를 해결하기 위해서는 단계별 작업이 필요하며, 각 단계별로 철저한 분석과 검토가 필요하다. 따라서 기존의 데이터 마이닝프로세스 단계별 분석 절차에 근거하여 철강 산업의 수익성 분석 및 예측 프로세스를 다섯 단계로 재구성하였다[7].

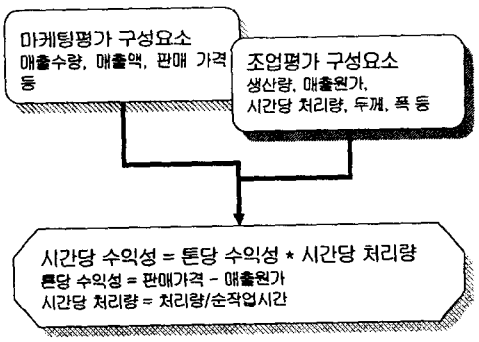
Step 1은 분석 필요성과 방향을 설정하고 분석기법에 대한 지식을 습득하며, Step 2는 현업의 업무상황을 정확히 반영할수 있는 대상 데이터를 확인한다. Step 3는 분석주제 및 방향에 적합한 데이터를 수집하고 데이터에 대한 pre-processing 작업을 한다. Step 4는 주요변수, 즉, 입력변수 및 목표변수를 정의하고 분석목적에 적합한 데이터 마이닝 기법을 이용하여 모델을 구축한다. 또한 모델이 구축된 후에는 모델 테스트 및 효과 검증을 한다. Step 5는 실제 업무에 적용하여 최종적인 분석 및 평가를 실시한 후 업무에 대한 문제점 및 개선점을 찾는다. 끝으로 초기단계로 피드백하여 향후 업무에 대한 준비를 하게 된다.

3. 사례분석

3.1 분석 주제

사례분석에서는 우리나라 철강기업 중 POSCO사를 대상으로 하였다. 철강기업의 시간당 수익성 분석 및 예측을 위해서 먼저 분석 특성에 맞게 수익성 개념을 재조정할 필요가 있다. 한 기업이 다수의 제조공장을 운영하는 경우 다수의 공장을 통합할 수 있는 전체 수익성을 평가하는 모델이 필요하다. 이러한 특성에 따라 기존 재무적/비재무적 분석 방법과는 차별화된 철강기업 특성을 잘 반영할 수 있는 수익성 개념을 도입하고자 한다. 먼저 철강기업의 마케팅부문과 조업부문으로 나눈 후 시간당 수익성 구성인자를 도출하였다. 마케팅부문 평가와 조업부문 평가의 주요 구성인자들을 고려하여 마케팅부문의 매출수량, 매출액, 판매가격과 조업부문의 매출원가, 시간당 처리량 등을 고려한 수익성 계산법을 사용하였다.

<그림 1> 철강기업의 시간당 수익성 재구성



3.2 모델링

분석 주제인 수익성 분석을 효과적으로 처리할 수 있는 모델을 구성하기 위하여 수익성에 대한 인과관계 및 예측 분석에 용이한 의사결정나무

기법과 유사한 레코드들을 함께 그룹핑할 수 있는 군집분석기법을 사용하였다.

먼저 시간당 수익성에 영향을 미치는 핵심 인자 도출 및 공장별 우위와 열위를 분석하기 위하여 의사결정나무 기법을 사용하였다. 제품을 수익별로 다수의 소그룹으로 나누기 위하여 군집분석을 사용하였으며 역시 우위, 열위 그룹을 군집화한 후 공장별 수익성을 도출하였다.

데이터 마이닝 기법은 분석목적에 맞게 다양한 알고리즘들이 개발되어 연구되고 있으며, 본 연구에서는 알고리즘들의 장점을 결합한 SAS E-Miner 프로그램을 이용하여 수익성 분석을 하기로 하였다. 또한 철강기업의 수익성 분석을 위한 데이터 마이닝 분석 방법 개발 또는 알고리즘 개발이 아니라 기존에 개발된 마이닝 기법을 이용하여 현실 문제를 해결하는 데 초점을 맞추고 있다.

3.3 분석 변수

철강기업의 시간당 수익성은 판매량, 생산량, 원가, 처리시간 등 여러 인자의 영향을 받는 것으로 판단되어 <표 1>과 같이 데이터 마이닝을 이용한 분석을 위하여 입력변수를 채택하였으며, 목표변수로 POSCO사의 2002년 4/4분기의 시간당 수익성을 사용하였다.

<표 1> 시간당 수익성 분석 자료

구분	변수명	변수내용
독립변수	Sale Amt	매출액
	Cost Amt	비용
	Sale Qty	판매량
	Pro Qty	생산량
	Sale Unit	판매원가
	Cost Unit	제조원가
	Pro Unit	톤당수익성
	T/H	시간당 처리량
목표변수	Pro T/H	시간당 수익성

3.5 분석 자료 및 가공

데이터 마이닝을 위해 사용되는 원천자료는 데이터 정제과정을 거친 자료를 사용한다면 분석시간 단축 및 분석의 효율성을 높일 수 있다. 따라서 분석 초기단계에 다음과 같이 자료의 정제과정을 수행하였다.

분석자료는 기존에 구축되어 있는 데이터 웨어하우스로부터 추출한 후 원천자료를 재조정하였다. 분석 데이터는 2002년 10월부터 2002년 12월까지 2002년도 4/4분기 실적 자료를 사용하였으며, POSCO사의 제품 전체를 대상으로 하는 것이 아니라 매출액 비중이 가장 높은 열연 5,880건에 대한 실적자료를 사용하였다. 2002년 4/4분기 현재 열연(HR)은 POSCO사 제품 매출액의 28%를 차지하고 있다. 분석 대상 공장은 포항공장의 1, 2 열연(HR) 공장, 광양공장의 1, 2, 3 열연공장으로 하였다. 분석 데이터를 4/4분기(3개월)로 한정한 것은 데이터 간의 분기별 판매생산계획수립, 예산편성, 표준원가 설정으로 발생할 수 있는 우연원인 차이발생을 배제할 수 있기 때문이다.

한편 자료에 대한 매출원가, 톤당이익, 시간당 이익을 산출하는 과정에서 미니밀을 포함 시키는 경우 인자별 분포형태에 이상치가 발생하였다. 즉, 미니밀 생산을 포함시키는 경우 자료 분포가 일관성을 잃어버릴 가능성이 존재하여 모델을 구성할 때 미니밀을 제외하였다. 따라서 공장전체의 열연 처리 건수 5,880건 중 미니밀을 제외한 5,434건이 실질적인 분석대상 건수가 된다.

목표변수인 시간당 수익성과 관련된 입력변수의 중복을 제거하기 위하여 입력변수 상관분석을 하였다. 분석결과를 통하여 변수간의 상관계수가 0.85 이상 나오는 변수 즉, 상관정도가 높은 변수는 두 인자를 모두 사용하지 않고 1개의 인자만을 입력변수로 채택하였다. 상관 분석을 통하여 입력변수간의 중복성을 제거한 후 제조원가, 판매단가, 시간당 처리량(T/Hr)을 최종 입력변수로 사용하기로 한다.

4. 분석 결과

시간당 수익성에 대한 분석결과는 의사결정나무, 군집화를 이용한 결과값으로 각각 나누어 살펴볼 수 있다. 의사결정나무 분석 결과 시간당 수익성에 가장 큰 영향을 미치는 인자는 시간당 처리량(T/H), 판매단가(SALE_UNIT), 제조원가(COST_UNIT)로 나타났다. 그 결과 세 인자의 영향에 따라 시간당 수익성은 8개 그룹으로 분류 되었다. 목표변수 즉, 열연의 거래건수 5,434건과 시간당 평균수익성 72,206.58원을 대상으로 독립변수들을 분석한 결과 수익성이 가장 높은 그룹은 TH>487.5/ton , SALE_UIT>327.5/원 , COST_UNI <216.5/원 인 경우로 나타났다. 또한 수익성이가장 낮은 그룹은 TH<395.5/ton, COST_UNI>228.5원으로 분석 되었다. 수익성이 가장 높은 그룹과 가장 낮은 그룹은 처리 빈도수가 각각 256회, 595회로 나타났으며, 시간당 수익은 각각 131,949.8원, 30,183.49원으로 결과 값이 도출되어 그룹간 시간당 수익성에 상당한 차이를 나타내고 있는 것을 알 수 있다

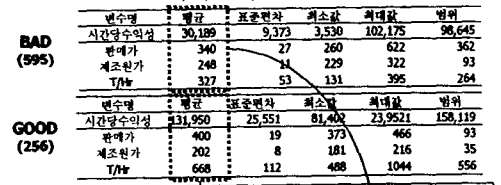
공장별로 분석결과가 어떻게 적용되는지 살펴 보았다. 그 결과 시간당 높은 수익성을 보이는 공장은 광양공장 1열연과 2열연이며, 가장 낮은 수익 공장은 포항공장 1열연과 2열연으로 나타났다 (<표 2> 참조). 또한 같은 조건하에 군집 분석을 실행한 결과 의사결정 나무와는 달리 11개 그룹으로 군집화되었으며, 최대 수익성을 보이는 공장과

최저 수익성을 나타내는 공장에 대한 분석한 결과 광양공장 1열연과 2열연은 높은 수익성을 나타냈으며, 포항공장 2열연은 상대적으로 수익성이 낮은 공장으로 나타났다(<표 3> 참조).

결론적으로 의사결정나무분석과 군집분석에 따라 분석내용을 정리하면, 시간당 수익성에 가장 크게 영향을 주는 인자는 시간당처리량, 판매단가, 제조원가이며, 공장별로는 광양공장 1열연과 2열연은 높은 수익성을, 포항공장 2열연은 낮은 수익성을 보이는 것으로 나타났다.

분석결과 전반적으로 포항공장 보다는 광양공장의 시간당 수익성이 높은 것으로 나타났으며, 기업은 포항공장의 수익성이 낮은 원인을 파악한 후 개선안을 제시하는, 즉 기업의 수익성을 높일 수 있는 방법을 모색할 필요가 있다.

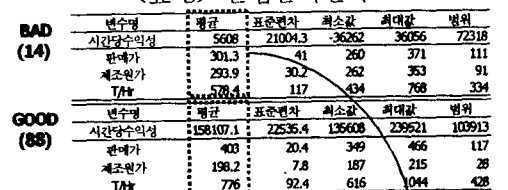
<표 2> 의사결정나무 분석결과



그룹(8) \ 공장		광 1열	광 2열	광 3열	포 1열	포 2열	계
BAD	빈도	2	4	24	500	65	595
	비율(%)	0.04	0.1	0.4	9.2	1.2	11.0
GOOD	빈도	112	79	65	0	0	256
	비율(%)	2.1	1.5	1.2	0	0	4.7
기타(6)	빈도	1107	997	1054	174	1251	4583
	비율(%)	20.4	18.4	19.4	3.2	23.0	84.3
Total	빈도	1221	1080	1143	674	1316	5434
	비율(%)	22.5	19.9	21.0	12.4	24.2	100

광: 광양공장, 포: 포항공장

<표 5> 군집분석결과



그룹(11) \ 공장		광 1열	광 2열	광 3열	포 1열	포 2열	계
BAD	빈도	0	0	5	0	9	14
	비율(%)	0	0	0.09	0	0.17	0.26
GOOD	빈도	60	18	10	0	0	88
	비율(%)	1.1	0.33	0.18	0	0	1.62
기타	빈도	1161	1062	1128	674	1307	5332
	비율(%)	21.37	19.54	20.76	12.4	24.05	98.12
Total	빈도	1221	1080	1143	674	1316	5434
	비율(%)	22.47	19.87	21.03	12.4	24.22	100

광: 광양공장, 포: 포항공장

5. 결론 및 추후 연구과제

기존 기업 수익성 분석 방법들이 개별산업의 특성을 충분히 고려하지 못하며, 과거 및 현재 상황에 대한 분석은 용이하되 미래에 대한 예측 부분에 있어서는 효과적으로 의사결정을 지원하기 쉽지 않다는 지적을 받아왔다. 본 연구에서는 이러한 기업 수익성 분석과는 차별화된 데이터마이닝 분석기법을 이용한 철강기업의 수익성 탐색 및 예측에 관한 분석을 하였다. 그 결과 사례 분석을 통하여 데이터 마이닝기법이 철강산업의 특성을 어떻게 반영할 수 있으며, 수익성을 탐색 및 예측하는 과정 및 결과 값을 도출하였다. 의사결정나무와 군집화 기법을 통해 철강기업의 시간당 수익성에 가장 크게 영향을 주는 인자는 시간당처리량, 판매단가, 제조원가이며, 공장별로 수익성 우위와 열위의 공장을 탐색할 경우, 광양 1,2공장은 우위, 포항 2공장은 열위의 수익성을 나타내는 것으로 확인되었다. 이러한 분석결과는 수주공정 주문 접수 후 공장별 물량 배분시 기준 정보로 사용될 수 있으며, 마케팅 판매전략 수립, 조업부서간 원가절감 수립을 위한 정보로도 활용될 수 있다.

본 연구는 철강기업의 일부 제품에 대한 데이터를 사용하여 수익성을 분석한 것으로, 향후 보다 의미있는 분석결과를 얻기 위해서는 제품의 범위 확대와 더불어 다양한 데이터 마이닝 기법을 사용한 연구가 요구된다.

6. 참고 문헌

[1]. 강현철, 한상태, 최정후, interpretation of Data Ming Prediction Model Using Decision Tree, Ther Korean Communication in statistics Vol.7, No.3, 2000

[2]. 강현철 외, 데이터 마이닝-방법론 및 활용, 자유아카데미, 1999

[3]. 김창오, 안광일, 김성집, 김재련, 트리구조를 이용한 연관규칙의 효율적 탐색, Journal of the Korean Institute of Industrial Enginnering Vol.27, 2001

[4]. 박정원, 포항 제철소 생산계획 수립을 위한 수익성 판단 Model 연구, 포항공과대학 철강대학원 석사학위논문, 1999

[5]. 정진우, 한국 철강공업의 재무구조 합리화 방안에 관한 연구, 서울시립대학교 대학원석사학위논문, 1993

[6]. Alex Berson, Stephen Smith, Kert Therling, Building Data Mining Application for CRM, 2000, McGraw Hill

[7]. A. Feelders, H. Daniels, M. Holsheimer, Methodological and practice aspects of data mining, Information & Management Vol.37, 2000

[8]. Berry.M, Linoff.G, Data Mining Techniques for Market sales, and Customer Support, John Wiley & Sons, 1997