

데이터 마이닝 기법을 이용한 특별고압 파급고장 발생가능 고객 예측모델 구축 및 신뢰도 향상방안에 관한 연구

논문
58-10-6

A study on Reliability Enhancement Method and the Prediction Model Construction of Medium-Voltage Customers Causing Distribution Line Fault Using Data Mining Techniques

배성환* · 김자희* · 홍정식** · 임한승***
(Sung Hwan Bae · Ja-Hee Kim · Jung-Sik Hong · Han-Seung Lim)

Abstract - Distribution line fault has been reduced gradually by the efforts on improving the quality of electrical materials and distribution system maintenance. However faults caused by medium voltage customers have been increased gradually even though we have done many efforts. The problem is that we don't know which customer will cause the fault. This paper presents the concept to find these customers using data mining techniques, which is based on accumulated fault records of medium voltage customers in the past. It also suggests the prediction model construction of medium voltage customers causing distribution line fault and methods to enhance the reliability of distribution system. We expect that we can effectively reduce faults resulted from medium voltage customers, which is 30% of total faults

Key Words : Customer Fault, Data Mining, Decision Tree, Regression, Neural Network, Distribution Fault, Fault Prediction Model, Reliability

1. 서론

변전소에서 고객에게 이르는 배전계통은 태풍, 폭우, 폭설 등 가혹한 자연재해에 노출되어 운전되고 있으며, 건설된 시점부터 외기 온도, 습도, 자외선 등에 의해 설비가 열화되기 때문에 일정한 시점이 경과되면 불가피하게 설비의 고장이 발생하고 있다[1]. 이러한 고장은 배전분야 종사자들의 꾸준한 예방정비와 설비개선의 노력으로 지속적인 설비증가에도 불구하고 2000년 13,719건에서 2008년 8,024건으로 41.5% 감소하였다[2]. 그러나 특별고압고객의 설비에서 발생하는 고장이 고객 내부에서 차단되지 못하고 전기적으로 연결된 한전의 배전선로에 고장전류가 파급되어, 배전선로에 설치된 보호 장치가 동작하여 고장전류를 차단함으로써, 건설한 배전계통이 정전되는 고장(이하 '특별고압 파급고장'이라 한다)은 증가하고 있다. 한국전력의 고장정전 통계에 의하면 특별고압 파급고장은 2000년 1,590건에서 2008년 2,405

건으로 오히려 51.3% 증가하였다[3].

지금까지 배전분야의 고장원인 분석은 주로 일원분석방법으로 전주, 전선, 애자, 변압기, 개폐기, 피뢰기, 고객설비, 지중설비, 기타로 분류하는 정전설비별 분석과 외물접촉, 자연현상, 기자재관련, 일반인 과실, 고객파급, 원인불명, 시공불량, 기타로 분류하는 정전원인별 분석이 사용되었다. 그리고 이원분석 방법으로 정전설비와 정전원인을 함께 분석하는 방법을 사용하여 왔다. 이러한 전통적인 고장분석 방법에 의한 고장 예방대책만으로는 특별고압 파급고장을 감소시키는데 한계에 도달하였다. 그러므로 새로운 다변량 분석 방법에 의한 고장 분석과 예방에 대한 연구가 필요하다. 따라서 본 연구에서는 다변량 분석을 위해 데이터 마이닝 기법을 제안한다. 외국에서는 이미 데이터 마이닝을 이용하여 배전계통의 고장분석 및 감소를 위한 연구가 활발히 진행되고 있다. 배전계통의 고장 분류, 원인분석 및 규명, 고장원인 인지, 고장진단 및 예방 등에 데이터 마이닝 기법이 사용되고 있다[4]~[11]. 국내 전력분야에서 데이터 마이닝 기법의 사용은 시작단계로 배전변압기 고장유형별 복구예측 모델 연구에 도입되었다[12]. 지금까지 국내외 전력분야에서 고장 발생 가능성 예측과 특별고압고객의 파급고장 연구에 데이터 마이닝 기법을 도입한 사례는 찾아볼 수 없었다. 따라서 선행 연구사례를 참조하여 과거 3년간 특별고압고객의 파급고장 발생 자료를 토대로 배전계통의 특별고압 파급고장 분석에 데이터 마이닝 기법을 적용하여 고장발생 가능 고객군을 예측하고 효과적인 고장예방 방안을 제시하고자 다음과 같은 단계로 연구를 진행하고자 한다. 첫째는 여러 가지 데이터 마이닝 기법 중에서 특별고압 파급고장 분석에 적합한

* 시니어회원 : 국립서울산업대학교 IT정책전문대학원
산업정보시스템전공 박사과정, 한국전력공사
기술기획처장

† 교신저자, 정회원 : 국립서울산업대학교 IT정책전문대학원
교수·공학박사.

E-mail : jahee@snu.ac.kr

** 정 회원 : 국립서울산업대학교 산업정보시스템공학과
교수·공학박사

*** 정 회원 : 한국전력공사 차장대리

접수일자 : 2009년 8월 17일

최종완료 : 2009년 8월 31일

기법과 알고리즘을 선택한다. 둘째는 선택한 데이터 마이닝 기법과 알고리즘을 적용하기 위해서 특별고압 파급고장에 영향을 주는 요인들을 선정하고 분석방법을 구체화 한다. 셋째는 특별고압 파급고장과 고장요인간의 관계성을 분석하기 위해서 계약전력, 사용전력량, 안전관리자, 수전설비 설치위치, 기자재의 사용연수, 제작회사 등을 파악한다. 넷째는 수집된 자료를 활용하여 변수선정 및 데이터 코드화를 하고 데이터 변환과 정제 과정을 거쳐 데이터 마이닝용 마스터데이터를 작성한다. 다섯째 상용 데이터 마이닝 도구를 사용하여 특별고압 고객 및 수전설비 관련 데이터를 분석하고 그 결과를 토대로 고장발생 가능고객 예측모델과 신뢰도 향상대책을 제시한다.

이를 위한 본 논문의 구성은 다음과 같다. 제2장에서는 특별고압 파급고장의 발생 개요, 주요 고장요인과 현황을 분석하고 제3장에서는 데이터 마이닝 기법과 알고리즘에 대한 이론적 배경을 검토한다. 다음으로 제4장에서는 상용 데이터 마이닝 툴을 사용하기 위해 필요한 자료 및 마스터데이터의 작성과정을 설명하고 데이터 마이닝 결과를 분석한다. 제5장에서 특별고압 파급고장 발생가능 고객 예측 모델을 설정하고 제6장에서 특별고압 파급고장 예방대책을 통한 배전계통 신뢰도 향상에 대해 논한다. 다음으로 제7장에서 연구결과의 결론을 종합 정리하였다.

2. 특별고압 파급고장 발생 개요 및 현황

이 장에서는 특별고압고객 수전설비의 구성과 배전계통의 고압연장 100km 당 고장발생 추이, 특별고압 파급고장의 점유율 추이, 최근 3년간의 고장발생 현황 및 고장요인에 대해 알아본다.

2.1 특별고압고객 수전설비의 구성

특별고압고객의 수전설비는 단로기(DS), 피뢰기(LA), 계기용 변류기(CT), 과전류 계전기(OCR), 전류계(A), 지락과전류 계전기(OCGR), 차단기(CB), 트립코일(TC), 계기용 변성기(MOF), 전력량계(WHM), 컷아웃 스위치(COS), 전력퓨즈(PF), 계기용 변압기(PT), 전압계(V), 수전용 변압기(Tr) 등으로 구성되어 있으며 수전설비 결선도는 그림 1과 같다[13].

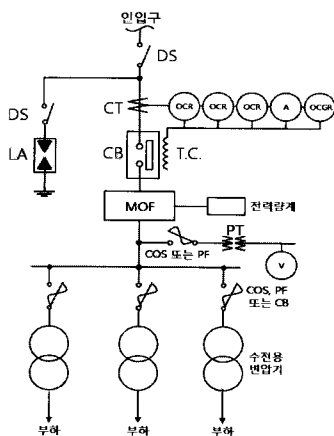


그림 1 특별고압 수전설비 결선도

Fig. 1 Electric Facility Connecting Diagram of Medium-Voltage Customer

2.2 고장발생 추이

배전계통의 고압선로 연장은 2000년 545,167km에서 2008년 671,777km로 23.2% 증가[14] 하였으며 고압선로 연장 100km당 고장 발생률은 2000년 2.5건에서 2008년 1.2건으로 52% 감소하였다. 그러나 특별고압 파급고장의 점유율은 2000년 11.6%에서 2008년 30%로 1.6배 증가한 것으로 나타났다. 연도별 배전계통 고장발생률과 특별고압 파급고장 점유율 추이는 그림 2와 같다. 그림 2에서 보는 바와 같이 배전계통의 신뢰도를 지속적으로 향상시키기 위해서 특별고압 파급고장의 감축에 관한 연구가 절실히 필요함을 알 수 있다.

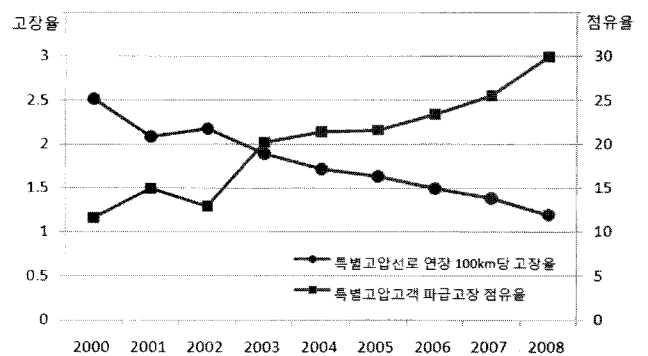


그림 2 배전계통 고장발생률과 특별고압 파급고장 점유율 추이

Fig. 2 Distribution System Fault Rate and Medium-Voltage Customer Fault Occupation Rate

2.3 특별고압 파급고장의 발생 요인

최근 발생하는 특별고압 파급고장의 주요 고장발생 요인을 환경적 요인과 설비의 열화요인으로 파악해 보면 그림 3과 같다. 환경적 요인으로는 사업소, 계약종별, 계약전력, 사용전력량, 수전설비 위치, 고장횟수, 설비관리자 형태로 구분할 수 있다. 설비의 열화요인으로는 선로의 종류, 규격, 인입거리, 개폐기의 종류, 차단기의 차단용량, MOF의 전류강도, 변압기의 용량 및 개폐기, 차단기, 피뢰기, MOF, 변압기에 공통으로 해당하는 사용연수와 제작회사로 구분할 수 있다.

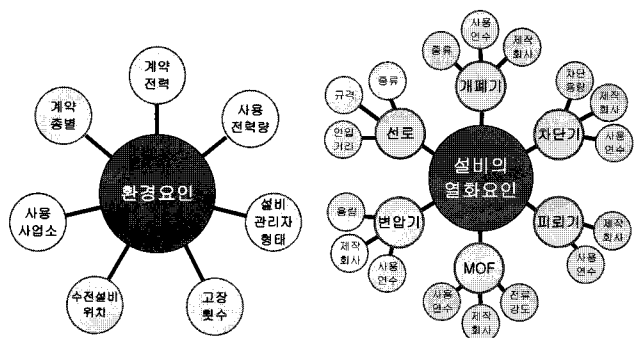


그림 3 특별고압 파급고장의 주요 고장발생 요인

Fig. 3 Main Causing Factors of Medium-Voltage Customer Fault

2.4 특별고압 파급고장 발생 현황

2005년부터 2008년까지 발생한 특별고압 파급고장의 설비별 고장발생 현황은 표 1과 같고 고장발생 순위는 주로 차단기, 피뢰기, 개폐기, MOF에서 고장이 발생하고 있는 것으로 나타나고 있다.

표 1 특별고압 파급고장 설비별 발생현황

Table 1 Fault Status of Medium-Voltage Customer by Materials

(단위: 건, %)

설비 구분	2005	2006	2007	2008
개폐기	393 (17.8)	384 (17.1)	367 (15.9)	449 (18.7)
차단기	497 (22.5)	543 (24.2)	500 (21.7)	381 (15.8)
변압기	168 (7.6)	190 (8.5)	185 (8.0)	262 (10.9)
MOF	365 (16.5)	354 (15.8)	332 (14.4)	334 (13.9)
케이블 및 헤드	168 (7.6)	105 (4.7)	145 (6.3)	169 (7.0)
가공 인입선	231 (10.5)	56 (2.5)	198 (8.6)	183 (7.6)
피뢰기	315 (14.2)	367 (16.3)	484 (21.0)	526 (21.9)
기타	70 (3.1)	247 (11.0)	95 (4.1)	101 (4.2)
계	2,207 (100)	2,246 (100)	2,306 (100)	2,405 (100)

원인별 고장발생 현황은 표 2와 같고 고장발생 순위는 주로 시공·보수불량, 외물접촉, 자연열화, 자연현상으로 고장이 발생하는 것으로 나타나고 있다.

표 2 특별고압 파급고장 원인별 발생현황

Table 2 Fault Status of Medium-Voltage Customer by Cause

(단위: 건, %)

설비 구분	2005	2006	2007	2008
자연열화	683 (30.9)	955 (42.5)	1,036 (44.9)	623 (25.9)
외물접촉	414 (18.8)	408 (18.2)	526 (22.8)	663 (27.6)
화재	-	-	41 (1.8)	23 (1.0)
일반인(직업자과실)	-	-	129 (5.6)	119 (4.9)
시공보수불량	151 (6.8)	217 (9.7)	204 (8.8)	694 (28.9)
제작불량	78 (3.5)	100 (4.5)	120 (5.2)	150 (6.2)
자연현상	125 (5.7)	180 (8.0)	212 (9.2)	129 (5.4)
기타	756 (34.3)	386 (17.1)	38 (1.6)	4 (0.2)
계	2,207 (100)	2,246 (100)	2,306 (100)	2,405 (100)

3. 데이터 마이닝 기법과 알고리즘

3.1 데이터 마이닝 기법

데이터 마이닝은 대량의 데이터로부터 새롭고 의미 있는 정보를 추출하여 의사결정에 활용하는 작업을 말한다. 현재 주로 사용되고 있는 분야는 은행의 대출업무 수행 시 고객의 신용불량 유무판별, 신용카드회사의 고객 이탈을 방지하기 위한 가입고객의 특성분석, 통신회사의 고객 서비스 가입해지 예측 등에 활용되고 있다. 보편적인 데이터 마이닝 기법으로는 의사결정나무(Decision Tree), 회귀분석(Regression), 인공신경망(Artificial Neural Networks)이 주로 사용되고 있다.

3.1.1 의사결정나무

의사결정나무는 수집된 데이터의 레코드를 분석하여 이들 사이에 존재하는 부류별 특성(패턴)을 속성의 조합으로 나타내는 분류모형을 나무형태로 만든다. 그 다음 이 분류모형으로 새로운 레코드를 분류하며 해당 부류의 값을 예측하는 것으로 순환적 분할(Recursive Partitioning) 방식을 이용하여 나무를 구축하는 기법이다. 의사결정나무는 데이터를 구성하는 속성의 수가 많은 경우에도 모형구축 시 분류에 영향을 미치지 않는 속성들을 자동적으로 제외시키기 때문에 데이터 선정이 용이하다. 또한 연속형과 명목형 데이터 값을 기록된 그대로 처리할 수 있기 때문에 데이터 변환에 소요되는 기간과 노력을 단축시킬 수 있는 장점이 있다. 반면에 연속형 데이터를 처리하는 능력이 신경망이나 전통적인 통계 기법에 비해 떨어지고 예측력이 감소하는 단점이 있다[15].

3.1.2 회귀분석

회귀분석(Regression)은 목표변수가 입력변수들에 의해서 어떻게 설명 또는 예측 되는 지를 알아보기 위해 데이터를 적절한 함수식으로 표현하여 분석하는 통계적 방법이다. 회귀분석은 선형방정식에 의해서 데이터를 표현하는 선형회귀분석, 비선형 방정식에 의해서 데이터를 표현하는 비선형회귀분석, 입력변수가 하나인 단순회귀분석, 입력변수가 여러개인 다중회귀분석으로 나누어진다. 회귀분석은 일반적으로 널리 알려져 여러 분야에 사용되며 통계소프트웨어를 이용하여 해석이 편리한 반면에 적절한 입력변수를 선택하여야 하고 비선형성의 경우 제약이 있으며 교호작용의 탐색에 어려움이 있다.

3.1.3 인공신경망

인공신경망(Artificial Neural Networks)은 인간의 세포를 모방화한 은닉마디(Hidden Units)를 기본 구성단위로 입력 변수들의 선형결합을 수신하고 활성화함수를 이용하여 입력을 변환해서 이를 입력으로 사용하는 다른 마디로 출력하는 구조로 구성된다. 인공신경망 중에서 자료분석에 가장 널리 사용되는 모형은 다계층 인식인자(Multilayer Perceptron)이며 입력계층, 은닉계층, 출력계층으로 구성된 전방향(Feed-forward) 신경망으로 범용근사법(Universal Approximator)을 사용하여 복잡한 관계식을 여러 관계식의 선형결합으로 표현한다. 모수 100개까지는 레벤버그-마르퀴르트(Levenberg-Marquardt), 모수 500개까지는 준뉴턴(Quasi-Newton), 모수 500개 이상은 결합기울기(Conjugate Gradient) 알고리즘을 사용하여 수렴할 때 까지 반복적으로 수행하는 반복적 최적화 방법을 사용한다. 그러나 1995년 Sarle가 인공신경망의 훈련도중, 즉 최적화 알고리즘이 반복되고 있는 동안에 모형 확인용 데이터를 사용하여 복잡한 신경망의 적합성을 감시하여 적합성이 나빠지는 시점에서 추정(훈련)을 중단하는 추정중단(Stopped Training)을 개발하였다. 이로써 시행착오(Trial and Error)에 의해 수렴할 때 까지 반복하지 않고 중단된 때의 모수 추정 값을 최종 값으로 사용하여 모형 복잡도의 딜레마를 해결하였다. 인공신경망은 의사결정나무보다 정확도는 높지만 어떤 입력변수가 중요한지와 어떻게 상호작용을 하는 지에 대한 해석결과의 설명력(Comprehensibility)이 부족한 것이 단점이다.

3.2 데이터 마이닝 기법별 알고리즘

3.2.1 의사결정나무

의사결정나무의 대표적인 알고리즘은 1980년 Kass가 개발한 CHAID, 1984년 Breiman et al.이 개발한 CART, 1993년 Quinlan이 개발한 C4.5, 1997년 Loh 와 Shih 이 개발한 QUEST 등이 있다. 이들 알고리즘을 사용한 데이터 마이닝 도구는 SAS, SPSS, IBM사에서 상용화하여 보급하고 있다.

본 논문에서는 분리개수가 2개 이상인 다지분리를 사용하는 CHAID(Chi-squared Automatic Interaction Detector) 알고리즘에 대해 알아본다. CHAID 알고리즘은 카이제곱-검정(이산형 목표변수) 또는 F-검정(연속형 목표변수)을 이용하여 분리와 병합을 반복하면서 부모마디에서 두개 이상의 자식마디의 분리가 일어나는 다지분리(Multiway Split)를 수행하는 알고리즘이다.

Person의 카이제곱량은 (1)

여기서 n 은 관측빈도수(Frequency), n_{ij} 은 분포의 동일성 또는 독립성의 가설 하에서 계산된 기대빈도수(Expected Frequency)를 말하며 다음과 같다[16].

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

n_{ij} 은 i 범주 행열에서 j 열의 합계
 $n_{.j}$ 은 j 범주 행열에서 열의 합계
 n 은 i 범주 행열에서 행과 열의 총합계

3.2.2 회귀분석

회귀분석에서 많이 사용되는 로지스틱 판별분석에 대해 알아본다. 로지스틱 판별분석은 목표변수가 이항형(Binary Type) 또는 순서형(Ordinal Type)으로 나타나 선형회귀모형으로 분석할 수 없는 경우에 사용하며 확률에 대한 로짓변환(Logit Transformation)을 고려하여 분석하는 것이다. 다중 로지스틱회귀모형은 다음과 같이 표현된다.

$$\ln \left(\frac{p_j}{1 - p_j} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad \dots \quad (2)$$

여기서 사후확률 추정식은 다음과 같으며 p_j 는 추정된 회귀계수이다.

$$p_j = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}$$

3.2.3 인공신경망

인공신경망의 최적화 알고리즘은 수정뉴턴(Modified Newton), 레벤버그-마르퀴르트(Levenberg-Marquardt), 준뉴턴(Quasi-Newton), 결합기울기(Conjugate-Gradient), 역전파(Back Propagation)가 있다. 수정뉴턴(Modified Newton)은 가장 신뢰할만한 반복적 최적화알고리즘으로 현재 값이 한

포물선 위에 위치해 있고 최소값 방향으로 이동하고 있다고 가정 한다. 그 다음 기울기(1차도함수)와 곡률(2차도함수)에 관한 정보를 파악하여 각 반복 작업마다 2차도함수의 행렬을 분석적으로 계산하고 역 과정을 거쳐야하는 복잡한 계산 작업을 수행하여 최적화한다. 레벤버그-마르퀴르트(Levenberg-Marquardt)와 준뉴턴(Quasi-Newton)은 수정뉴턴 방법이지만 헤시안(Hessian, 2차도함수 행렬)을 분석적으로 계산할 필요가 없고 1차도함수를 사용하여 근사화하는 방법이다. 결합기울기(Conjugate-Gradient)는 곡률에 관한 정보를 사용하지 않아 계산과정이 간단하다. 기본적으로 기울기에 관한 정보를 이용하여 계산되지만 이전과정으로 되돌아가지 않아도 되도록 이전 단계들을 미리 기록해 계산시간을 줄이는 방법이다. 역전파(Back Propagation)는 인공신경망에서 가장 널리 사용되는 알고리즘으로 임의의 초기강도(weight)값을 지정하여 초기 모형을 설정한다. 그 다음 모형구축용 데이터의 레코드를 하나씩 모형에 입력시켜 모형이 제시하는 부류의 값과 실제 부류의 값을 비교한다. 이 두 값에 차이가 있으면 차이를 최소화하기 위해서 입력값이 전달되는 방향과 역방향으로 차이를 최소화하는 강도값의 변화량을 계산하여 적용한다. 이런 학습과정을 강도값의 변화가 미미해지거나 미리 정해진 횟수만큼 모형 구축용 레코드들을 입력시킬 때까지 계속하는 방법이다.

4. 데이터 마이닝 처리절차와 분석결과

본 논문에서는 데이터 마이닝 기법을 전개하는데 필요한 자료의 수집에서 최종 결과분석까지를 다음 4단계로 나누어 진행한다. 제1단계는 분석용 자료를 수집하여 마스터 테이블을 작성하고 제2단계는 데이터를 정제 및 통합하고 제3단계는 데이터를 모형화 및 변환하며 제4단계는 모형을 해석하고 결과분석을 시행한다.

4.1 분석용 자료수집과 마스터 테이블 작성

특별고압고객의 과급고장 발생을 예측하고자 2005년부터 2007년까지 전국에서 발생한 특별고압고객의 과급고장이 발생한 경험 데이터와 자료검증을 위해 고장이 발생하지 않은 특별고압고객의 경험 데이터를 한전의 데이터베이스에서 랜덤 샘플링으로 확보하였다. 전체 확보한 데이터는 9,651개 표 3에서 보는 바와 같이 구성되어 있다.

표 3 특별고압고객 자료 수집 현황

Table 3 Data Collection Status of Medium-Voltage Customer

구 분	고장발생 고객	고장미발생 고객	합계
2005	1,239	1,719	2,958
2006	1,325	1,853	3,178
2007	1,449	2,066	3,515
합 계	4,013	5,638	9,651

마스터 테이블의 작성은 분석용으로 실제 입수한 자료를 엑셀파일로 작성하여 데이터 마이닝 프로그램에서 사용하도록 하기 위해 데이터별로 변수이름을 정한다음 데이터의 부류를 기호화 하고 변수의 속성을 표 4와 같이 정의한다.

표 4 특별고압고객 파급고장 예측용 변수 정의

Table 4 Variable Define for Predicting Fault Caused by Medium-Voltage Customer

데이터	변수	변수 속성
(1) 데이터연도	YEAR	Nominal
(2) 사업소	SAUPSO	Nominal
(3) 고객번호	CNO	Id
(4) 계약종별	CTTYPE	Nominal
(5) 계약전력	CTPOWER	Interval
(6) 사용량배수	KWHBAESU	Interval
(7) 전기사용연수	UY	Interval
(8) 수전설비 위치	LOCATION	Nominal
(9) 고장횟수	OUTAGES	Interval
(10) 안전관리자 구분	ELECTRICIAN	Nominal
(11) 선로종류	WIRETYPE	Nominal
(12) 선로규격	WIRESIZE	Nominal
(13) 선로거리	WIREDST	Interval
(14) 개폐기종류	SWTYPE	Nominal
(15) 개폐기 사용연수	SWUY	Interval
(16) 개폐기 제작회사	SWCOM	Nominal
(17) 차단용량	BCAPACITY	Ordinal
(18) 차단기 제작회사	BCOM	Nominal
(19) 차단기 사용연수	BUY	Interval
(20) 피뢰기 제작회사	LACOM	Nominal
(21) 피뢰기 사용연수	LAUY	Interval
(22) MOF 제작회사	MOFCOM	Nominal
(23) MOF 사용연수	MOFU	Interval
(24) MOF 전류강도	MOFCS	Ordinal
(25) 변압기 용량	TRCAPACITY	Ordinal
(26) 변압기 제작회사	TRCOM	Nominal
(27) 변압기 사용연수	TRUY	Interval
(28) 고장여부	FAULT	Binary 목표변수

다음으로 구하고자 하는 목표변수와 데이터를 입력하고자 하는 입력변수를 정하고 정의한 변수 이름으로 목표변수와 입력변수들을 변환한 다음 각 변수들의 해당 입력 데이터를 내용별 분류기호를 사용하여 변환 코드화함으로써 마스터 테이블을 표 5와 같이 완성하였다.

표 5 마스터 테이블 작성사례

Table 5 Case of Master Table

YEAR	SAUPSO	CNO	CTTYPE	CTPOWER	KWHBAESU	UY	LOCATION	OUTAGES
2005	3970	0105276683	GENERAL	400	0.095	13	T	0
2005	3970	0135100593	GENERAL	275	0.219	17	T	1
2005	3970	0135100815	GENERAL	275	0.409	14	T	0
2005	3970	0135101994	GENERAL	900	0.163	13	U	1
2005	3970	0135102314	INDUSTRIAL	793	0.293	19	G	1

ELECTRICIAN	WIRETYPE	WIRESIZE	WIREDST	SWTYPE	SWUY	SWCOM	BCAPACITY	BCOM
4	CABLE	C	15	COS	1	OTHERS	12.5	MEDIUM
4	CABLE	C	15	LS	17	OTHERS	7.1	OTHERS
4	CABLE	C	50	ASS	14	OTHERS	7.1	OTHERS
4	CABLE	C	120	LBS	14	BIG	12.5	OTHERS
1	CABLE	C	50	IS	20	OTHERS	12.5	OTHERS

BUY	LACOM	LAUY	MOFCOM	MOFU	MOFCS	TRCAPACITY	TRCOM	TRUY	FAULT
13	MEDIUM	13	SINPUM	13	150	400	OTHERS	13	1
17	OTHERS	17	SINPUM	17	150	75	DOMESTIC	17	1
14	OTHERS	14	SINPUM	14	150	75	OTHERS	14	1
14	OTHERS	14	SINPUM	14	75	300	OTHERS	14	1
20	OTHERS	20	OTHERS	20	75	100	DOMESTIC	20	1

4.2 데이터의 정제 및 통합

데이터의 불순도가 데이터 마이닝 해석의 정확도에 매우 민감하기 때문에 데이터의 질을 향상시키고 처리과정의 효율을 높이기 위하여 상관분석을 통해서 데이터간의 연관성을 분석하고 불필요한 자료를 제거하였다. 특히 특별고압고객의 수전설비는 한국전력의 설비가 아닌 관계로 하나의 레코드가 원하는 항목들이 다 있는 경우가 드물었다. 따라서 1차로 추출한 9,651건의 데이터에서 수전설비의 주요 기자재 사용연수가 마이너스인 데이터, 계약이 해지된 고객 데이터, 특별고압고객은 계약전력이 100kW이상임으로 계약전력이 99kW이하인 데이터를 일괄하여 레코드 삭제하였다. 다음으로 계약전력에 맞는 용량으로 설치하게 되어 있는 MOF의 용량이 계약전력과 맞지 않는 것은 계약전력의 규모에 따라 일괄 정비하였다. 그리고 간단히 고객에게 전화하여 알 수 있는 수전설비 위치, 안전관리자 구분은 데이터를 파악하여 입력하였다. 이렇게 데이터를 정제하는 데 약 6개월이 소요되었으며 결과는 표 6에서 보는 바와 같이 전체 확보 데이터의 73.8%인 7,034건의 데이터가 최종적으로 확정되었다.

표 6 데이터 정제 결과

Table 6 Result of Data Clearing

구분	고장발생 고객	고장미발생 고객	합계
2005	829	1,346	2,175
2006	858	1,416	2,274
2007	951	1,634	2,585
합계	2,638	4,396	7,034

다음으로 이렇게 정제한 전국의 14개 사업본부와 연도별로 작성된 데이터를 하나의 마스터 테이블로 통합하여 데이터 마이닝 분석 작업을 위한 준비를 완료하였다.

4.3 데이터 모형화 및 변환

4.3.1 의사결정나무

SAS사의 Enterprise Miner 9.1을 이용하여 분석용(Train) 데이터 60%, 검증용(Validation) 데이터 40%로 분할한다. 다음으로, 수집된 데이터에서 일부 결측치가 있는 피뢰기와 차단기의 제작회사를 'Replacement'틀을 사용하여 의사결정나무에서 결측치를 목표변수로 하여 Leaf Node의 사후확률이 가장 큰 것을 적용해서 결측치를 채운다. Leaf의 최소 데이터 개수를 70개, 나무깊이는 5단계, 다지분리 3으로 설정한 후 의사결정나무(Decision Tree)와 인터랙티브 의사결정나무(Interactive Decision Tree)를 이용하여 분석하고 ROC(Receiver Operating Characteristic)커브를 이용하여 평가한다[17][18].

4.3.2 회귀분석

회귀분석을 위해 SAS사의 Enterprise Miner 9.1로 'Transform Variable' 틀을 사용하여 연속형(Interval) 입력변수를 범주형으로 변환하고 히스토그램을 통해 분포를 조정 후 고장 미발생(OUTAGES=0) 데이터의 편중을 반영하기 위해 'Rare Class'로 데이터의 비중을 조정한다. 회귀분석의 'Type'은 'Logistic'으로, 'Selection Method'는 'Stepwise'로, 'Optimization Method'는 'Default(입의)'로 설정하였다.

4.3.3 인공신경망

SAS사의 Enterprise Miner 9.1의 'Data Partition' 툴을 사용하여 모형구축용 데이터를 모형추정용과 모형확인용으로 60% 대 40% 비율로 구분한다. 인공신경망의 은닉마디(Hidden Neurons)는 3으로, 'Network Architecture'는 다계층 인식인자(Multilayer Perception)로, 'Runtime Limit'는 2시간으로 설정하였다. 학습방식(Training Technique)은 임의(Default), 레벤버그-마르퀴르트(Levenberg-Marquardt), 결합기울기(Conjugate-Gradient), 준뉴턴(Quasi-Newton), 역전파(Standard Back-propagation)를 적용하여 최적의 해를 찾도록 한다.

4.4 데이터 마이닝 분석결과

4.4.1 의사결정나무 분석결과

SAS사의 Enterprise Miner 9.1로 표 6의 데이터를 분석하였다. 실제 현장에서는 전체 15만호의 고객 중에서 연간 2,200~2,400호 정도의 특별고압 파급고장이 발생하고 있으므로 표 6의 데이터는 고장발생 가능 고객군을 예측하기 위하여 고장발생 고객이 실제보다 오버샘플링(Over Sampling) 되었음을 밝혀준다. 이 데이터를 분석한 결과 그림 4와 같이 첫 단계에서 고장횟수(OUTAGES)를 기준으로 나무가 분리되었다. 즉 한번이상 고장이 발생한 경험이 있는 고객의 수전설비에서 다시 고장이 발생하는 확률이 오버샘플링

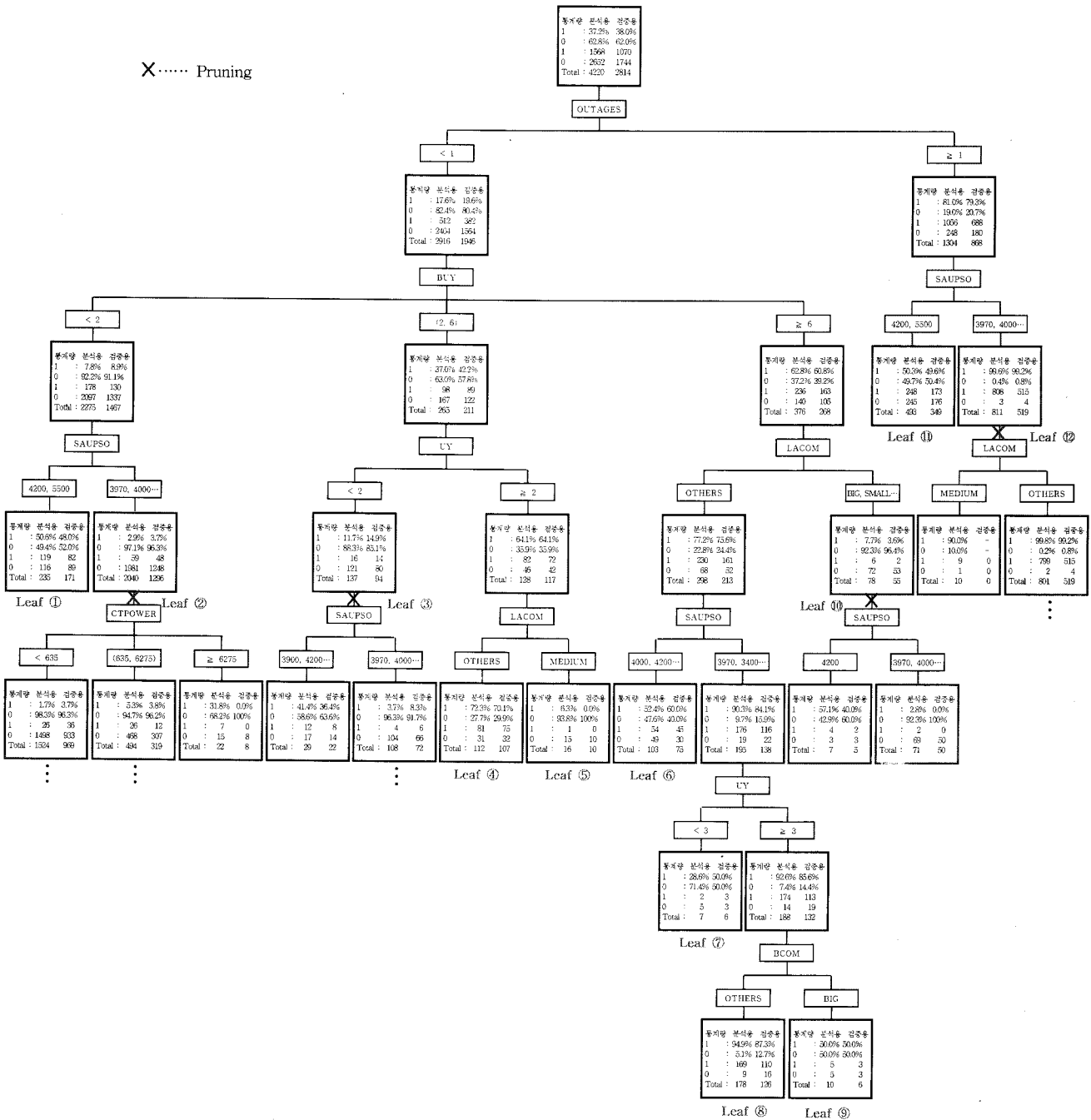


그림 4 특별고압 파급고장 발생유형 인터랙티브 의사결정나무 및 가지치기

Fig. 4 Interactive Decision Tree and Branch Pruning of Medium-Voltage Customer Fault Pattern

된 데이터를 기준으로 81%로 나타났다. 이는 결국 취약한 설비를 갖고 있는 고객에게서 한번 고장이 나면 다시 고장이 발생한다는 것을 의미한다. 일반적으로 특별고압 수전설비의 소유주는 전기설비에 대한 전문지식이 없어 전기설비가 노후 되었어도 현재 전기사용에 문제가 없다면 고장발생 때까지 비용투자가 수반되는 전기설비를 잘 교체하려고 하지 않는 경향이 있기 때문인 것으로 판단된다. 따라서 특별고압 파급고장 예방을 위해서는 이들 고장경험 고객에 대한 특별관리가 필요함을 알 수 있다. 두 번째 단계에서 고장경험이 없는 고객 경우 일반 트리는 차단기사용연수(BUY)로 분리되었다. 그러나 인터랙티브 의사결정나무에서는 차단기사용연수(BUY), 피뢰기사용연수(LAU), 계기용변성기사용연수(MOFUY), 변압기사용연수(TRUY)가 모두 동일한 카이제곱분류계수 $-\log(P)=115$ 로 나타나 분리규칙 적용을 위한 대상의 선택이 필요하였다. 결국 표 1에서 나타낸 것과 같이 분석 대상기간(2005~2007)에서는 차단기의 고장이 제일 많이 발생하고 있으므로 차단기사용연수(BUY)를 선택하여 분리하였다. 차단기는 사용기간이 6년 이상인 고객에서 고장이 62.8%로 발생률이 가장 높게 나타났으며 2~5년이 37%, 2년 미만에서 7.8%인 것으로 나타났다. 세번째 단계에서 차단기사용연수 2~5년이 전기사용연수(UY) 2년을 기준으로 분리되어 2년 이상 경과고객에서 주로 고장이 발생함을 알 수 있다. 차단기사용연수 6년 이상은 피뢰기제작회사 규모를 기준으로 분리되어 주로 기타인 소규모의 이름 없는 회사 제품이 품질이 낮아 고장이 주로 발생하고 있는 것으로 나타나고 있다. 네 번째 단계에서 일반트리는 전기사용연수가 2년 이상인 경우는 피뢰기제작회사(LACOM)를 기준으로 동일한 사유에 의해 분리되었다.

인터랙티브 의사결정나무도 분리된 트리의 기본구조는 일반의사결정나무와 동일하나 진행단계별로 분리규칙을 직접 보고 선택하여 트리를 세분화하고 가지치기를 할 수 있어 보다 정교한 트리를 만들 수 있었다. 그 결과 일반 트리에서는 볼 수 없었던 고장횟수 1이상과 차단기사용연수 2년 미만이 사업소를 기준으로 분리되어 설비를 잘 관리하고 있는 사업소와 그렇지 못한 사업소가 구별되었다. 그리고 차단기사용연수 6년 이상이고 피뢰기제작회사가 기타(OTHERS)인 경우가 자식마다로 사업소, 전기사용연수, 차단기제작회사(BCOM)를 기준으로 추가 분리되어 피뢰기 설비관리를 잘하지 못하는 사업소와 전기사용연수가 3년 이상된 사업소의 차단기제작회사가 기타로서 소규모이고 이름 없는 회사제품인 경우 고장이 많이 발생하고 있는 것이 추가로 파악되었다. 한편 차단기사용연수 2년 미만이고 사업소가 3970 등인 경우, 차단기사용연수 2~5년이고 전기사용연수 2년 미만인 경우, 차단기사용연수 6년 이상이고 피뢰기 제작회사가 대규모 등인 경우, 고장횟수 1이상이고 사업소가 3970 등인 경우는 인터랙티브 의사결정나무로 그림 4와 같이 추가 분리되었으나 의미가 없어 가지치기를 하였다.

일반 의사결정나무는 분석(Training)에서 Leaf가 26개 였으나 검증(Validation)에서 가지치기를 하여 Leaf=7이 되었으며, 오분류율은 'Training' 12.99%, 'Validation' 14.57%로 나타났다.

인터랙티브 의사결정나무는 그림 5와 같이 분석(Training), 검증(Validation) 모두 Leaf = 27에서 수렴하였고 오분류율은 'Training' 12.82%, 'Validation' 14.85%로 비슷한 결과를 보였다.

종합 분석결과를 고장이 한번이상 발생한 고객에서 다시 고장이 발생하는 확률이 81%로 매우 높다. 이 중에서 강원본

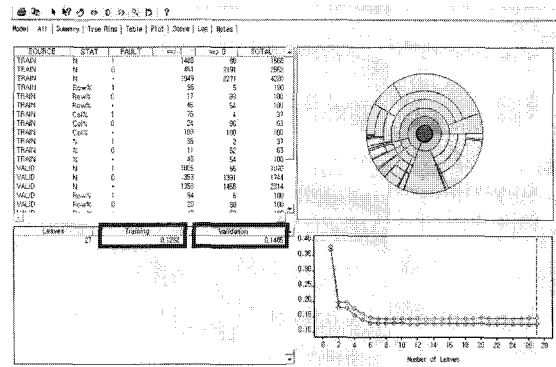


그림 5 인터랙티브 의사결정나무 결과
Fig. 5 Result of Interactive Decision Tree

부(4200)와 부산본부(5500)는 한번이상 고장이 발생한 특별고압고객의 수전설비 관리를 잘 하여 다른 사업소에 비해 50% 고장이 발생하고 있다. 반면 강원본부와 부산본부 이외의 사업소는 한번이상 고장이 발생한 고객에서 99.6% 다시 고장이 발생하여 상대적으로 수전설비의 관리를 잘 하고 있지 못함이 Leaf⑫에서 나타나고 있다. 다음으로 고장이 한 번도 발생하지 않은 고객에서는 차단기사용연수(BUY)가 6년 이상이고 피뢰기제작회사(LACOM)가 소규모이고 이름이 없는 기타(OTHERS)인 경우에서 사업소가 서울(3970), 경기북부(3400), 경기(3900), 충남(4600), 전남(5000), 대구경북(5200), 충북(4500)인 경우로 전기공급 후 3년 이상 경과하고 차단기제작회사(BCOM)가 소규모이고 이름이 없는 기타(OTHERS)인 경우에서 고장이 많이 발생하고 있음을 Leaf⑧에서 알 수 있다. 그리고 고장이 한 번도 발생하지 않은 고객에서 차단기사용연수(BUY)가 2~5년이고 전기 공급 후 2년 이상으로 피뢰기제작회사(LACOM)가 소규모이고 이름이 없는 기타(OTHERS)인 경우에도 고장이 많이 발생하고 있음을 Leaf④에서 알 수 있다. 또한 고장이 한 번도 발생하지 않은 고객에서 차단기사용연수(BUY)가 2년 미만으로 사업소가 강원(4200), 부산(5500)인 경우에도 고장이 많이 발생하고 있음을 Leaf①의 분석을 통해 알 수 있다.

4.4.2 회귀분석 결과

SAS사의 Enterprise Miner 9.1의 회귀분석(Regression) 해석결과 'Stepwise' 9 단계에서 수렴되었으며 단계별 선택 변수와 카이제곱량은 다음 그림 6에서 보는 바와 같다.

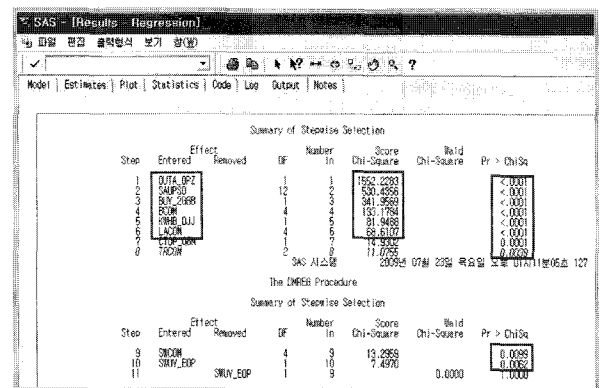


그림 6 회귀분석 단계별 진행결과
Fig. 6 Result of Stepwise Selection in Regression

회귀분석에서 연관성이 높은 변수는 유의확률($Pr > ChiSq$)로 판단하는데 0.05 보다 작으면 입력변수들이 목표변수를 설명하는데 의미가 있다고 보며 0.0001보다 작으면 통계적으로 극히 유의하다고 판단할 수 있다. 그림 6에서 유의확률이 0.05 이하이고 카이제곱량(Chi-Square)이 커서 의미가 있는 입력변수는 고장횟수(OUTAGES), 사업소(SAUPSO), 차단기 사용연수(BUY), 차단기 제작회사(BCOM), 전력사용량 배수(KWHBAESU), 피뢰기제작회사(LACOM)인 것으로 나타났다.

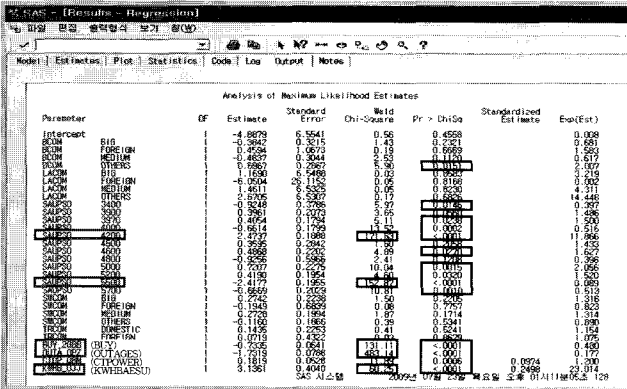


그림 7 회귀분석 변수 항목별 최대 가능성 평가 결과
Fig. 7 Analysis of Maximum Likelihood Estimates by Regression Variables

그림 7에서 회귀계수에 대한 유의성 검증을 위해 유의확률($Pr > ChiSq$)이 0.05 이하이고 월트카이제곱량(Wald Chi-Square)이 커서 의미가 있는 입력변수는 고장횟수(OUTAGES), 사업소(SAUPSO), 차단기사용연수(BUY), 전력사용량배수(KWHB)인 것으로 판명되었다. 회귀분석에서 목표변수인 고장발생 여부의 실제값과 분류결과에 의한 예측값의 결과는 그림 8과 같이 나타났다. 즉 실제 '0'의 값 2,652개를 사용하여 '0'으로 맞게 분류한 예측값이 2,424개 이고 '1'로 잘못 분류한 예측값이 228개인 것으로 나타났으며 실제 '1'의 값 1,568개를 사

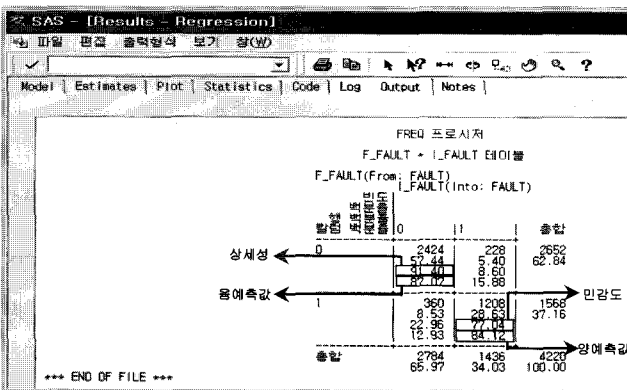


그림 8 회귀분석 모델의 목표변수 실제값과 예측결과
Fig. 8 Result of True and Forecast Value of Objective Variable, 'Fault', in Regression Model

용하여 '0'으로 잘못 분류한 예측값이 360개 이고 '1'로 맞게 분류한 예측값이 1,208개인 것으로 나타났다. 모형에서 사용한 총 데이터는 4,220개이다. 모형 평가에 사용되는 상세성, 민감도, 양예측값과 음예측값은 '1'에 가까울수록 좋다. 그림 8에서 상세성은 0.914, 민감도는 0.77, 양예측값은 0.841, 음예측값은 0.87로 민감도를 제외하고는 모두가 우수한 것으로 나타났으며 민감도도 양호한 편이다. 또한 오분류율은 낮을수록, 정확도는 높을수록 좋은 모형이다. 그림 8에서 오분류율은 $(228 + 360) / 4,220 = 0.13934$, 정확도는 $(2,424 + 1,208) / 4,220 = 0.86066$ 으로 오분류율 13.9%, 정확도 86.1%로 모형은 좋은 것으로 판단된다.

4.4.3 인공신경망 분석결과

SAS사의 Enterprise Miner 9.1의 인공신경망(Neural Networks)을 이용하여 학습 방식(Training Technique)으로 임의(Default), 레벤버그-마르퀴르트(Levenberg-Marquardt), 결합기울기(Conjugate-Gradient), 준뉴튼(Quasi-Newton), 역전파(Standard Back-propagation)를 적용시켜 해석하여 보았다. 인공신경망 해석결과는 학습방식별로 비슷하였으나 ROC 커브를 기준으로 볼 때 준뉴튼과 임의, 결합기울기, 레벤버그-마르퀴르트, 역전파 순으로 해석결과가 좋은 것으로 그림 9와 같이 나타났다.

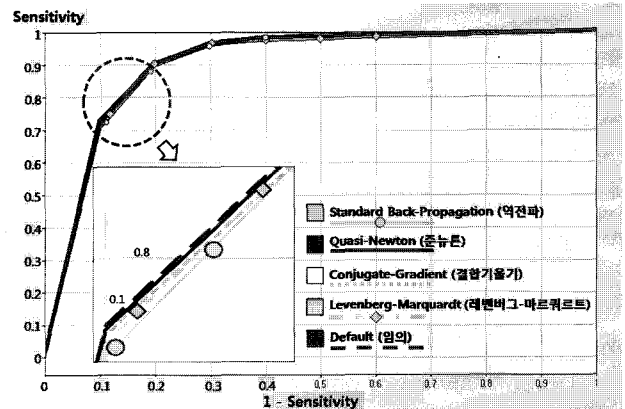


그림 9 인공신경망 학습방식별 해석결과
Fig. 9 Neural Network Result by Training Technique

4.4.4 분석결과 종합

앞서 살펴본 의사결정나무(Decision Tree), 인터랙티브 의사결정나무(Interactive Decision Tree), 회귀분석(Regression), 인공신경망(Neural Networks)을 해석한 SAS Enterprise Miner 9.1은 그림 10과 같이 구성하였다. 해석 틀의 종합구성은 'Input Data Source'에 데이터를 여러 가지로 분석해 볼 수 있는 'Multiplot'과 데이터 분석용(Train) 및 검증용(Validation)을 설정하는 'Data Partition'을 연결한다. 그 다음 'Data Partition'에 변수의 결측값을 설정·변경할 수 있는 'Replacement'를 연결한 다음 데이터 마이닝 기법별 도구를 연결하고 이를 종합평가할 수 있는 'Assessment'를 연결하여 구성한다. 'Regression'은 연속형(Interval) 변수를 범주형으로 변환할 수 있는 'Transform Variables'를 하나 더 연결한다.

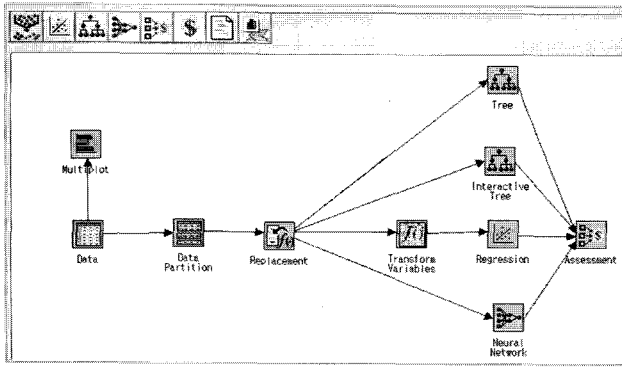


그림 10 데이터 마이닝 종합 구성도
Fig. 10 Data Mining Configuration Diagram

이렇게 구성한 Enterprise Miner를 해석한 결과 리포트 차트는 그림 11과 같다. 리포트 차트는 반응률이 높을수록 결과가 좋다는 것을 나타내는데 그림 11에서 인터랙티브 의사결정나무(Interactive Decision Tree : 그림 11의 Tree-2)가 제일 좋은 것으로 나타났다. 이는 표 4에서 보는바와 같이 입력변수가 대부분 명목형(Nominal)이고 연속형(Interval)과 순서형(Organal)의 경우도 데이터의 범위가 넓지 않은 특성과 일반 의사결정나무보다 유의미한 변수가 3단계 더 분리되어 세분화 및 정밀도가 향상됨으로 인해 나타난 결과로 판단된다. 다음으로 인공신경망(Neural Networks), 회귀분석(Regression), 의사결정나무(Decision Tree)의 순서로 해석결과가 좋음을 알 수 있다.

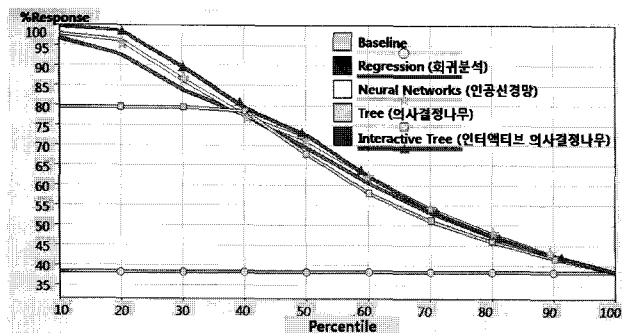


그림 11 데이터 마이닝 기법별 리프트차트 비교
Fig. 11 Lift Chart Comparison by Data Mining Method

모형의 정확도를 나타내는 ROC커브는 'Sensitivity(민감도)'와 '1-Specificity(상세성)'이 1에 가까울수록 좋다는 것을 나타낸다. 즉, 곡선 왼쪽으로 갈수록 분류 결정값이 높고 곡선 오른쪽으로 갈수록 분류 결정값이 낮음을 나타내는데 곡선 아래의 면적이 클수록 모형의 정확도가 높다. 그림 10의 구성도에 따라 데이터 마이닝 기법별 해석결과를 비교하면 그림 12와 같다. 이 그림에서 보면 리프트 차트와 같이 인터랙티브 의사결정나무(Interactive Decision Tree : 그림 12의 Tree-2)가 제일 모형의 정확도가 높다. 다음으로 인공신경망(Neural Networks), 회귀분석(Regression), 의사결정나무(Decision Tree)순으로 정확도가 나타났다. 인터랙티브 의사결정나무의 해석결과가 제일 좋은 것은 앞서 리프트차트에서 설명한 바와 같이 해석한 자료에 명목형 입력변수가 많고 연속형과 순서형 데이터의 범위가 넓지 않은 특성과 일반

의사결정나무보다 유의미한 변수가 3단계 더 분리되어 세분화 및 정밀도가 향상됨으로 인해 나타난 결과로 판단된다.

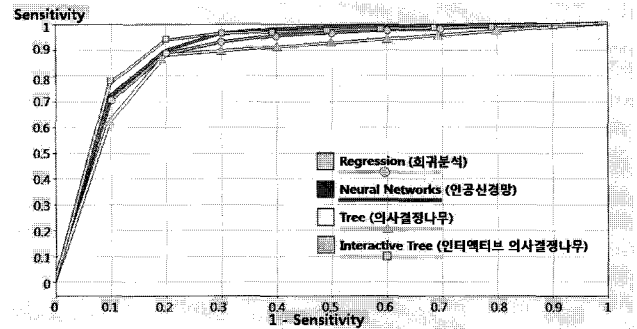


그림 12 데이터 마이닝 기법별 ROC 커브 비교
Fig. 12 ROC Curve Comparison by Data Mining Method

5. 특별고압 파급고장 발생가능고객 예측모델 설정

5.1 특별고압 파급고장 발생가능 고객 분류 논리식

그림 4에서 Leaf Node를 고장 발생률이 큰 순서로 정리하면 ⑫, ⑧, ④, ⑥, ①, ⑪, ⑨, ⑦, ③, ⑩, ⑤, ②로 나타나고 이를 토대로 게인차트를 그리면 그림 13과 같다.

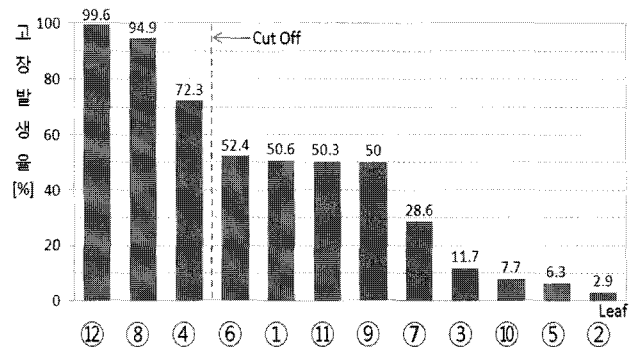


그림 13 게인차트
Fig. 13 Gain Chart

특별고압 파급고장 발생가능 고객 분류 논리식을 만들기 위해 오분류율이 5%이하로 적은 Leaf를 선택하고자 그림 13 게인차트에서 Leaf ④와 ⑥ 사이를 분류결정값(Cut Off)으로 설정하고 Leaf ⑫, ⑧, ④를 선택하여 해당 Leaf Node가 결정된 의사결정나무를 사용하여 논리식을 표 7과 같이 작성하였다.

표 7 특별고압 파급고장 발생가능 고객 분류 논리식

Table 7 Logic Equation Classifying Customers Who Cause Medium-voltage Customer Fault

Leaf	논리 조건
⑫	고장횟수 ≥ 1 ∩ 사업소 ≠ 4200, 5500
⑧	고장횟수 < 1 ∩ 차단기 사용연수 ≥ 6 ∩ 피뢰기제작사 = 기타 ∩ 사업소 = 3970 등 ∩ 전기사용연수 ≥ 3 ∩ 차단기 제작회사 = 기타
④	고장횟수 < 1 ∩ 차단기 사용연수 = 2~5년 ∩ 전기사용연수 ≥ 2 ∩ 피뢰기 제작회사 = 기타

이 논리식은 앞에서 분석한 회귀분석 결과, 그림 6에서 카이제곱량이 크고 '유의확률($Pr > ChiSq < 0.0001$)'이하로 유의미한 입력변수는 고장횟수(OUTAGES), 사업소(SAUPSO), 차단기사용연수(BUY), 차단기제작회사(BCOM), 전력사용량배수(KWHB), 피뢰기제작회사(LACOM)였다. 그림 7에서 왈드카이제곱량이 크고 '유의확률($Pr > ChiSq < 0.0001$)'로 유의미한 입력 변수가 고장횟수(OUTA), 사업소 4200(강원)과 5500(부산), 차단기사용연수(BUY), 전력사용량배수(KWHB)인 것으로 나타났다. 따라서 의사결정나무의 분리기준 변수는 회귀분석의 유의미한 변수와 유사함을 알 수 있다.

5.2 특별고압 파급고장 발생가능고객 예측 모델

현재 특별고압 파급고장 예방을 위해서는 사업소 단위로 3년을 주기로 1회씩 그림 14와 같이 임의로 특별고압고객을 지역별로 선정할 후 순회하여 정기점검을 시행하고 있다.

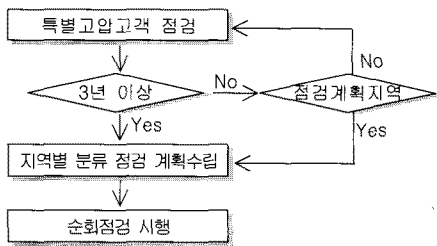


그림 14 특별고압 파급고장 예방을 위한 순회점검 시행절차
Fig. 14 Work Flow of Patrol Test to Prevent Medium-Voltage Customer Fault

그러나 이러한 점검 방법은 특별고압 파급고장을 효과적으로 예방하지 못하고 있다. 이에 의사결정나무를 통해 얻은 결과를 토대로 그림 15와 같이 파급고장 발생가능 고객

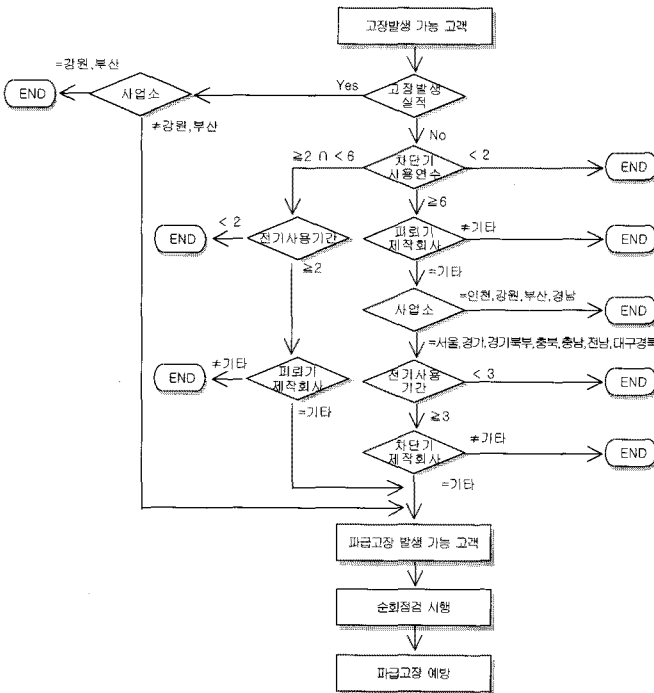


그림 15 특별고압 파급고장 발생가능 고객 예측 모델
Fig. 15 Prediction Model of Customers Who Cause Medium-Voltage Customer Fault

을 예측할 수 있는 모델을 새롭게 구축하였다. 이 모델에서 보면 첫째, 고장이 발생한 실적이 있는 고객 중 강원본부와 부산본부를 제외한 나머지 사업소의 고객은 파급고장 발생 가능성이 높으므로 최우선적으로 점검해야 한다. 둘째, 고장 발생 실적이 없는 고객 중 차단기의 사용연수가 2~5년이요 전기사용기간이 2년 이상 경과하였으며 피뢰기 제작회사가 기타로서 소규모 이름 없는 회사제품을 사용하고 있는 고객은 파급고장 발생 가능성이 높으므로 순회점검을 시행해야 한다. 셋째, 고장발생 실적이 없는 고객 중 차단기의 사용연수가 6년 이상, 피뢰기 제작회사가 기타이며 사업소가 서울, 경기, 경기북부, 충북, 충남, 전남, 대구경북이고 전기사용기간이 3년 이상 및 차단기 제작회사가 기타인 고객은 파급고장 발생 가능성이 높으므로 우선적으로 점검해야 한다는 결론을 얻었다.

6. 특별고압 파급고장 예방대책을 통한 배전계통 신뢰도 향상방안

앞에서 살펴본 바와 같이 배전계통의 전력공급 신뢰도는 배전 전문가들의 꾸준한 노력의 결실로 지속적으로 향상되고 있다. 그러나 특별고압 파급고장은 매년 지속적으로 증가하고 있어 이에 대한 효과적인 대응방안이 필요하다. 그러므로 본 논문에서는 데이터 마이닝 기법을 적용하여 고장 발생 가능 고객을 예측하고 이들 고객을 대상으로 순회점검을 집중적으로 실시하고자 한다. 이를 통해서 특별고압 파급고장을 효과적으로 줄인다면 배전계통의 신뢰도는 크게 향상시킬 수 있을 것으로 판단된다. 이에 배전계통에서 데이터 마이닝 기법을 적용하여 특별고압 고객의 파급고장을 예방하기위한 대책을 다음과 같이 제시하고자 한다.

첫째, 지금까지 사용한 일차원 및 이차원 통계분석에서 벗어나 은행, 보험, 백화점 등이 데이터 마이닝 기법을 적용해 매출을 올릴 수 있는 특정 고객을 찾아내어 타겟 마케팅하는 기법을 전력분야에 도입해야 한다. 배전계통에서도 본 논문에서 제시한 데이터 마이닝 예측모델을 적용하여 특별고압 파급고장을 유발시킬 가능성이 높은 고객을 찾아내어 중점적으로 순회점검 함으로써 효과적으로 고장을 예방하도록 한다.

둘째, 데이터 마이닝 예측 모델을 적용하기 위하여 데이터 웨어하우스를 구축한다. 현재 특별고압 파급고장은 단순한 통계분석을 위한 기본 자료만을 입력하고 있다. 그러나 정확한 특별고압 파급고장 발생가능 고객을 예측하기 위해서는 본 논문의 분석결과 중요 변수로 판단되는 고장발생 이력, 차단기 사용기간, 피뢰기 제작회사, 전기사용기간, 차단기 제작회사 등을 비롯한 특별고압 파급고장 관련 변수가 될 수 있는 요소들에 대한 데이터 웨어하우스 구축이 필요하다.

셋째, 특별고압고객 관련 정보는 한국전력의 회사 정보가 아닌 관계로 전산 시스템에 입력항목이 설정되어 있음에도 많은 항목들이 누락된 채로 공란으로 관리되고 있다. 따라서 신규 전력공급 및 순회점검 시 필요한 데이터는 현장 조사하여 입력하고 관리가 필요하지 않은 항목들은 삭제하여 불필요한 전산시스템의 메모리 사용에 따른 비용을 절감하도록 데이터베이스 정비 및 입력데이터의 수정을 전사적으로 시행할 필요가 있다.

넷째, 잘 정비된 데이터와 데이터 웨어하우스를 이용하여 효과적으로 파급고장 발생 가능고객을 분석하기 위한 SAS, SPSS, IBM 등의 적정 데이터 마이닝 분석 툴을 도입한다.

다섯째, 배전시스템의 특별고압고객 수전설비 관리 업무에 데이터 마이닝 기법을 도입하여 적용한다. 또한 상용화된 툴을 사용할 수 있는 데이터 마이닝 전문 인력을 양성한다.

이상의 제시방안을 기준으로 특별고압 파급고장 예방을 위한 순회점검의 개선된 절차를 그림 16과 같이 작성하였다. 향후 이러한 개선된 절차에 따라 특별고압고객의 순회점검을 시행하면 보다 효과적으로 특별고압 파급고장을 예방할 수 있을 것으로 기대된다.

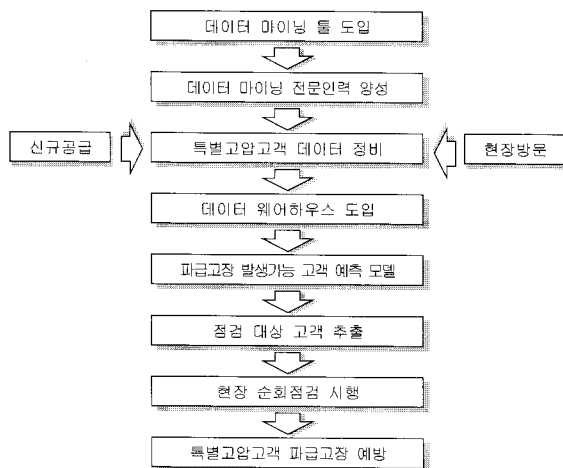


그림 16 특별고압 파급고장 예방을 위한 순회점검 절차 개선
 Fig. 16 Improved Work Flow of Patrol Test to Prevent Medium-Voltage Customer Fault

7. 결 론

본 연구에서는 SAS Enterprise Miner를 이용하여 배전시스템의 특별고압고객에서 발생하는 전기설비의 고장이 연계된 배전선로로 파급되어 건전한 고객이 정전되는 것을 예방하는 연구를 하였다. 특별고압 파급고장 발생 가능 고객을 은행 대출시 신용 불량고객을 탐지해 내는 것과 같이 데이터 마이닝 기법을 이용하여 예측할 수 있도록 하였다. 이렇게 예측된 고장유발 가능 고객을 효율적으로 집중 예방 보수함으로써 특별고압 파급고장을 예방하여 건전한 고객들이 특별고압 파급고장으로 인해 정전피해를 보는 일이 없도록 신뢰도 향상대책을 제시하였다. 구체적으로 과거 3년간 특별고압 파급고장 발생고객의 경험데이터와 건전고객의 경험데이터를 추출해 모형을 만들었다. 그 다음 특별고압 파급고장 유발 가능 고객을 의사결정나무, 회귀분석 및 인공신경망을 이용하여 예측하고 이들의 결과를 상호 비교·분석하였다. 특히 지금까지 단순히 설비별 고장원인별 이원분석법으로 통계를 분석하던 것에서 새로운 데이터 마이닝 기법을 이용하여 다수의 입력변수를 통해 특별고압고객의 고장발생 가능 유무를 예측한 것은 최초로 시도한 것으로 매우 의미가 크다. 이 분석결과를 토대로 고장발생 가능고객에게 집중적으로 고장예방 대책을 시행하여 효율적으로 정전을 줄이고 전력공급 신뢰도를 향상시킬 수 있게 규명한 것은 커

다란 성과라 할 수 있다.

이번 연구에서는 특별고압고객의 수전설비 관련 데이터가 단순 참고용으로 실제 업무활용을 위해 체계적으로 관리하고 있지 않음으로서 많은 항목이 공란으로 되어 있었다. 따라서 이러한 데이터의 정제에 많은 시간이 소요되고 사용할 수 있는 데이터가 적은 문제점이 있었으며 일부 항목의 경우는 시간적 제약으로 자료 파악이 불가능하여 분석에서 불가피하게 제외할 수밖에 없는 한계점이 있었다. 그리고 데이터 편중에 따라 불가피하게 고장발생 실적 데이터를 오버샘플링하여 고장원인을 분석할 수밖에 없었다. 앞으로 보다 더 많은 자료를 시간적 여유를 갖고 수집하고 많은 항목을 체계적으로 분석하도록 함은 물론, 오버샘플링에 대한 문제점도 심층적으로 검토하여 특별고압고객의 파급고장 발생가능 고객을 더 정밀하게 예측할 수 있는 연구가 필요할 것으로 기대된다.

참 고 문 헌

- [1] 대한전기학회(김재철 외 16인), "최신 배전시스템 공학", 북스힐, pp. 2-3, 2006
- [2] 한국전력공사, "배전설비 정전분석 및 예방대책", 2000 ~ 2008
- [3] 한국전력공사, "고객파급 정전분석 및 예방대책", 2000 ~ 2008
- [4] L. Xu. and M. Y. Chow, "A Classification Approach for Power Distribution Systems Fault Cause Identification.", IEEE Trans. on Power Systems, Vol. 21, No. 1, pp. 53-60, 2006
- [5] H. Manjari Dola, "Data mining for Distribution System Fault Classification", Power Symposium, Proc. of the 37th Annual North American, IEEE, pp. 457-462, 2005.
- [6] L. Xu., M. Y. Chow. and L. S. Taylor, "Analysis of Tree-caused Faults in Power Distribution System.", in Proc. 35th North American Power Symp., 2003
- [7] M. Y. Chow. and L. S. Taylor, "Analysis and Prevention of Animal-caused Faults in Power Distribution Systems.", IEEE Trans. Power Del., Vol. 10, No. 2, pp. 995-1001, 1995
- [8] M. Y. Chow., S. O. Yee. and L. S. Taylor, "Recognition Animal-caused Faults in Power Distribution Systems using Artificial Neural Networks.", IEEE Trans. Power Del., Vol. 8, No. 3, pp. 1268-1274, 1993
- [9] O. Dag. and C. Ucak, "Fault Classification for Power Distribution Systems via A Combined Wavelet-neural Approach.", in Proc. Int. Conf. Power System Technology, Vol. 2, pp. 1309-1314, 2004
- [10] K. L. Butler. and J. A. Momoh, "A Neural Network Based Approach for Fault Diagnosis in Distribution Networks.", in Proc. IEEE Power Engineering Society Winter Meeting, Vol. 2, pp. 1275-1278, 2000
- [11] J. T. Peng., C. F. Chien. and T. L. B. Tseng,

"Rough Set Theory for Data Mining for Fault Diagnosis on Distribution Feeder.", Proc. Int. Elect. Eng. ,Gener., Transm., Distrib., Vol. 151, No. 6, pp. 689-697, 2004

- [12] 황우현, 김자희, 장완성, 홍정식, 한득수, "데이터 마이닝 기법을 이용한 주상변압기 고장유형 분석 및 복구 예측 모델 구축에 관한 연구", 대한전기학회지 제57권 제 9호, pp. 1507-1515, 2008
- [13] 대한전기협회, "내선규정", pp. 399, 2006
- [14] 한국전력공사, "배전설비 통계", 2000 ~ 2008
- [15] 장남식, 홍성완, 장재호, "데이터 마이닝", 대청미디어, 2000
- [16] 최중후, 한상태, 강현철, 김은석, 김미경, 이성건, "Answer Tree 3.0을 이용한 데이터 마이닝 예측 및 활용", SPSS 아카데미, 2002
- [17] 배화수, 조대현, 석경하, 김병수, 최국렬, 이종언, 노세원, 이승철, 손용희, "SAS Enterprise Miner를 이용한 데이터 마이닝", 교우사, 2005
- [18] S. Walsh, W. Potts, D. Wielenga, J. Amrhein, K. Brown, I. Krammer, B. Lucas. and P. Terrill, "Applying Data Mining Techniques Using Enterprise Miner", SAS Institute Inc., 2004

저 자 소 개



배 성 환 (裴 成 煥)

1959년 8월 16일생
 1980년 건국대학교 전기공학 졸업
 1990년 미국 Union Graduate College
 전기 및 컴퓨터 공학과 졸업 (석사)
 2007년 ~ 현재 국립서울산업대학교
 IT정책전문대학원 산업정보시스템전공

박사과정 재학중
 1979년 ~ 현재 한국전력공사 기술기획처장
 1996년 발송배전기술사, 1998년 전기안전기술사
 1981년 ~ 현재 대한전기학회 종신회원
 Tel : 02-3456-3700
 E-mail : baeshwa@kepco.co.kr



김 자 희 (金 子 姬)

1973년 7월 27일생
 1995년 2월 한국과학기술원 전산학과 졸업
 1997년 2월 한국과학기술원 전산학과 졸업
 (석사)
 2003년 한국과학기술원 산업공학과 졸업
 (공학박사)

2004년 비엔나 대학 경영정보학과 방문연구원
 2005년 ~ 현재 국립서울산업대학교 IT정책전문대학원
 산업정보시스템공학 조교수
 2007년 대한전기학회 정회원
 Tel : 02-970-6867
 E-mail : jahee@snut.ac.kr



홍 정 식 (洪 涇 植)

1960년 3월 16일생.
 1982년 서울대학교 산업공학과 졸업
 1985년 서울대학교 대학원졸업(석사)
 1988년 서울대학교 대학원졸업(공학박사)
 1985~1987 서울대학교 조교
 1989 ~ 현재 국립서울산업대학교

산업정보시스템공학과 교수
 Tel : 02-970-6474
 E-mail : hong@snut.ac.kr



임 한 승 (林 漢 承)

1975년 3월 10일생.
 1998년 연세대학교 전기학과 졸업
 2000년 연세대학교 대학원졸업(석사)
 2005 ~ 현재 한국전력공사 차장대리
 Tel : 02-3456-3735
 E-mail : hslim75@kepco.co.kr