

특별고압 수전설비 관리에 데이터 마이닝 기법을 적용한 파급고장 발생가능고객 예측시스템 구현 연구

논 문
58-12-25

A Study on Constructing the Prediction System Using Data Mining Techniques to Find Medium-Voltage Customers Causing Distribution Line Faults

배 성 환* · 김 자 희† · 임 한 승**
(Sung Hwan Bae · Ja-Hee Kim · Han-Seung Lim)

Abstract - Faults caused by medium-voltage customers have been increased and enlarged their portion in total distribution faults even though we have done many efforts. In the previous paper, we suggested the fault prediction model and fault prevention method for these distribution line faults. However we can't directly apply this prediction model in the field. Because we don't have an useful program to predict those customers causing distribution line faults. This paper presents the construction method of data warehouse in ERP system and the program to find customers who cause distribution line faults in medium-voltage customer's electric facility management applying data mining techniques. We expect that this data warehouse and prediction program can effectively reduce faults resulted from medium-voltage customer facility.

Key Words : Customer Faults, Data Mining, Distribution Fault, Fault Prevention, Facility Management System, Reliability, Medium-Voltage Customer Fault

1. 서 론

특별고압 고객에서 발생한 고장이 배전선로로 파급되어 정전이 발생하는 고장은 전체 배전고장이 감소하고 있음에도 불구하고 건수와 점유비율이 매년 증가하고 있다[1]. 현재 신배전정보시스템 내에 있는 특별고압 수전설비 관리시스템은 파급고장 분석방법으로 정전설비와 정전원인을 분석하여 제공하고 있으나 특별고압 파급고장을 감소시키는데 한계에 도달하였다. 그러므로 이러한 파급고장을 감소하고자 은행, 보험, 통신 분야에서 주로 사용하고 있는 데이터마이닝 기법을 적용하여 특별고압 파급고장 발생가능 고객을 예측하는 모델을 연구하여 제시하였다[2]. 그러나 현재까지 예측시스템으로 구현하지는 못하였다. 우리나라 배전분야 정보시스템은 1976년 주상변압기의 부하관리를 위한 '저압부하관리시스템'을 시작으로[3] 2005년 지리정보시스템(Geographic Information System)을 기반으로 한 '신배전정보시스템(New Distribution Information System)' 구축을 완료하여 모든 배전설비의 개체관리와 이력관리가 가능한 특징을 갖고 있다[4]. 특별고압으로 전력을 공급하는 고객의 업무처리와 고장관리에도 이 신배전정보시스템이 사용되고 있

다. 지금까지 국내에서 신배전정보시스템을 이용한 노후설비의 교체주기정립, 신뢰도 기반의 유지보수 등의 여러 가지 연구가 진행되었으며[5]-[9], 해외에서도 지리정보시스템과 배전설비의 개체 및 이력관리를 하는 배전자산관리시스템(Distribution Asset Management System)을 구축하고[10], 이를 통한 설비교체주기 설정, 유지보수자원관리, 신뢰도기반의 유지보수, 고객정보와 연계분석 등에 관한 여러 가지 연구가 진행되어왔다[11]-[16]. 그러나 국내외에서 데이터마이닝 기법을 적용한 특별고압수전설비의 파급고장 발생가능 고객을 예측하여 미리 보수함으로써 고장을 예방하도록 하는 시스템에 관한 연구사례는 찾아볼 수 없었다. 따라서 신배전정보시스템의 특별고압 수전설비 관리에 앞서 연구한 결과를 토대로 데이터마이닝 기법을 적용한 파급고장 발생가능고객을 예측하는 시스템을 최초로 구현하고 장기적으로 예측 정확도를 향상시키고자 데이터웨어하우스 구축방안을 제시하여 특별고압 파급고장을 효율적으로 감축하고자 한다.

2. 특별고압 파급고장 발생가능 고객 예측시스템 구현

2.1 구현 방향

특별고압 고객에서 발생하는 파급고장을 예방하기 위해서는 정확하게 고장발생 가능고객을 찾아내어 고장이 발생하기 전에 미리 보수하는 것이 중요하다. 이는 지금까지 연구되거나 구현된 적이 없는 새로운 시스템을 구현하는 것이므로 우선 예측에 필요한 변수들의 데이터를 확보하고 예측변수를 설정한다. 그 다음 검증된 알고리즘과 상용화된 도구를 사용하여 파급고장 발생가능 고객을 정확히 예측할 수 있는 모델을 만들고 이를 예측시스템으로 프로그래밍 하여 기존 특별고압 수전설비 관리시스템에 연계하여 사용할 수 있도록 구성한다.

* 시니어회원 : 국립서울산업대학교 IT정책전문대학원
산업정보시스템전공 박사과정,
한국전력공사 기술기획처장

† 교신저자, 정회원 : 국립서울산업대학교 IT정책전문대학원
교수·공학박사.

E-mail : jahee@snut.ac.kr

** 정 회원 : 한국전력공사 차장대리

접수일자 : 2009년 10월 28일

최종완료 : 2009년 11월 16일

2.2 데이터마이닝 의사결정나무 알고리즘

의사결정나무의 대표적인 알고리즘은 1980년 Kass가 개발한 CHAID, 1984년 Breiman et al.이 개발한 CART, 1993년 Quinlan이 개발한 C4.5, 1997년 Loh 와 Shih이 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) 알고리즘을 사용하였다. 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.

$$\text{Person의 카이제곱량은 } X^2 = \sum_{i,j} (f_{ij} - e_{ij})^2 / e_{ij} \quad (1)$$

여기서 f_{ij} 는 관측빈도수(Frequency), e_{ij} 는 분포의 동일성 또는 독립성의 가설 하에서 계산된 기대빈도수(Expected Frequency)를 말하며 다음과 같다.

$$e_{ij} = \frac{f_{it} \times f_{tj}}{f_{tt}}$$

f_{it} 는 $i \times j$ 범주 행열에서 i 행의 합계

f_{tj} 는 $i \times j$ 범주 행열에서 j 열의 합계

f_{tt} 는 $i \times j$ 범주 행열에서 i 행과 j 열의 총합계

2.3 데이터마이닝 의사결정나무 해석결과

2.3.1 해석용 데이터 수집 및 정제

특별고압고객의 파급고장 발생을 예측하고자 2005년부터 2007년까지 전국에서 발생한 특별고압고객의 파급고장이 발생한 경험 데이터와 자료검증을 위해 고장이 발생하지 않은 특별고압고객의 경험 데이터를 우리나라 배전을 담당하고 있는 K사의 데이터베이스에서 랜덤 샘플링으로 확보하고 필수항목이 누락된 것은 고객에게 전화로 확인하여 입력하는 등 데이터를 정제하여 표 1과 같이 7,034개의 해석용 데이터를 확보하였다.

표 1 데이터마이닝 해석용 데이터

Table 1 Data for Analyzing as Data Mining

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

2.3.2 데이터마이닝 분석 변수 설정

수집한 데이터의 분석을 위해 목표변수와 입력변수를 설정하고 데이터 입력을 위해 변수 이름을 정한 다음 데이터의 분류를 기호화하고 변수의 속성을 표 2와 같이 정의하였다.

표 2 특별고압고객 파급고장 예측용 변수 정의

Table 2 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) 수전설비위치	LOCATION	Nominal
(9) 고장횟수	OUTAGES	Interval
(10) 안전관리자구분	ELECTRICIAN	Nominal
(11) 선로종류	WIRETYPE	Nominal
(12) 선로규격	WIRESIZE	Nominal
(13) 선로거리	WIREDST	Interval
(14) 개폐기종류	SWTYPE	Nominal
(15) 개폐기사용연수	SWUY	Interval
(16) 개폐기제작회사	SWCOM	Nominal
(17) 차단용량	BCAPACITY	Ordinal
(18) 차단기제작회사	BCOM	Nominal
(19) 차단기사용연수	BUY	Interval
(20) 피뢰기제작회사	LACOM	Nominal
(21) 피뢰기사용연수	LAUY	Interval
(22) MOF 제작회사	MOFCOM	Nominal
(23) MOF사용연수	MOFUY	Interval
(24) MOF전류강도	MOFCS	Ordinal
(25) 변압기용량	TRCAPACITY	Ordinal
(26) 변압기제작회사	TRCOM	Nominal
(27) 변압기사용연수	TRUY	Interval
(28) 고장여부	FAULT	Binary 목표변수

2.3.3 의사결정나무 해석결과

의사결정나무의 상용분석도구인 Enterprise Miner 9.1을 사용하여 표 1의 데이터를 인터랙티브 의사결정나무로 해석하였다. 그 결과 그림 1과 같이 첫 단계에서 고장횟수(OUTAGES)를 기준으로 나무가 분리되었다. 즉 한번이상 고장이 발생한 경험이 있는 고객의 수전설비에서 다시 고장이 발생하는 확률이 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 노후되었어도 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 고객의 경우는 차단기사용연수(BUY)를 기준으로 분리되었다. 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 전기사용연수가 2년 이상인 경우는 피뢰기제작회사(LACOM)를 기준으로 동일한 사유에 의해 분리되었다.

또한 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

차식마디로 사업소, 전기사용연수, 차단기제작회사(BCOM)를 기준으로 추가 분리되어 피뢰기 설비관리를 잘하지 못하는 사업소와 전기사용연수가 3년 이상 된 사업소의 차단기 제작회사가 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 피뢰기제작회사가 대규모 등인 경우, 고장횟수 1이상이고 사업소가 3970 등인 경우는 그림 1과 같이 추가 분리되었으나 의미가 없어 가지치기를 하였다.

인터랙티브 의사결정나무는 그림 2와 같이 분석(Training), 검증(Validation) 모두 Leaf = 27에서 수렴하였고 오분류율은 'Training' 12.82%, 'Validation' 14.85%로 나타나서 모형은 정확도가 85%이상으로 좋은 것으로 판단된다.

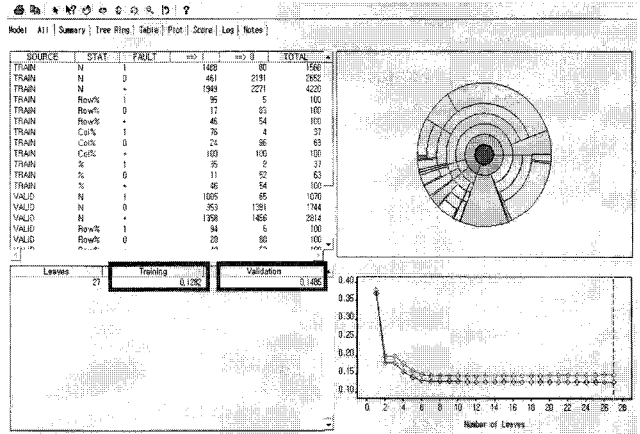


그림 2 인터랙티브 의사결정나무 결과
Fig. 2 Result of Interactive Decision Tree

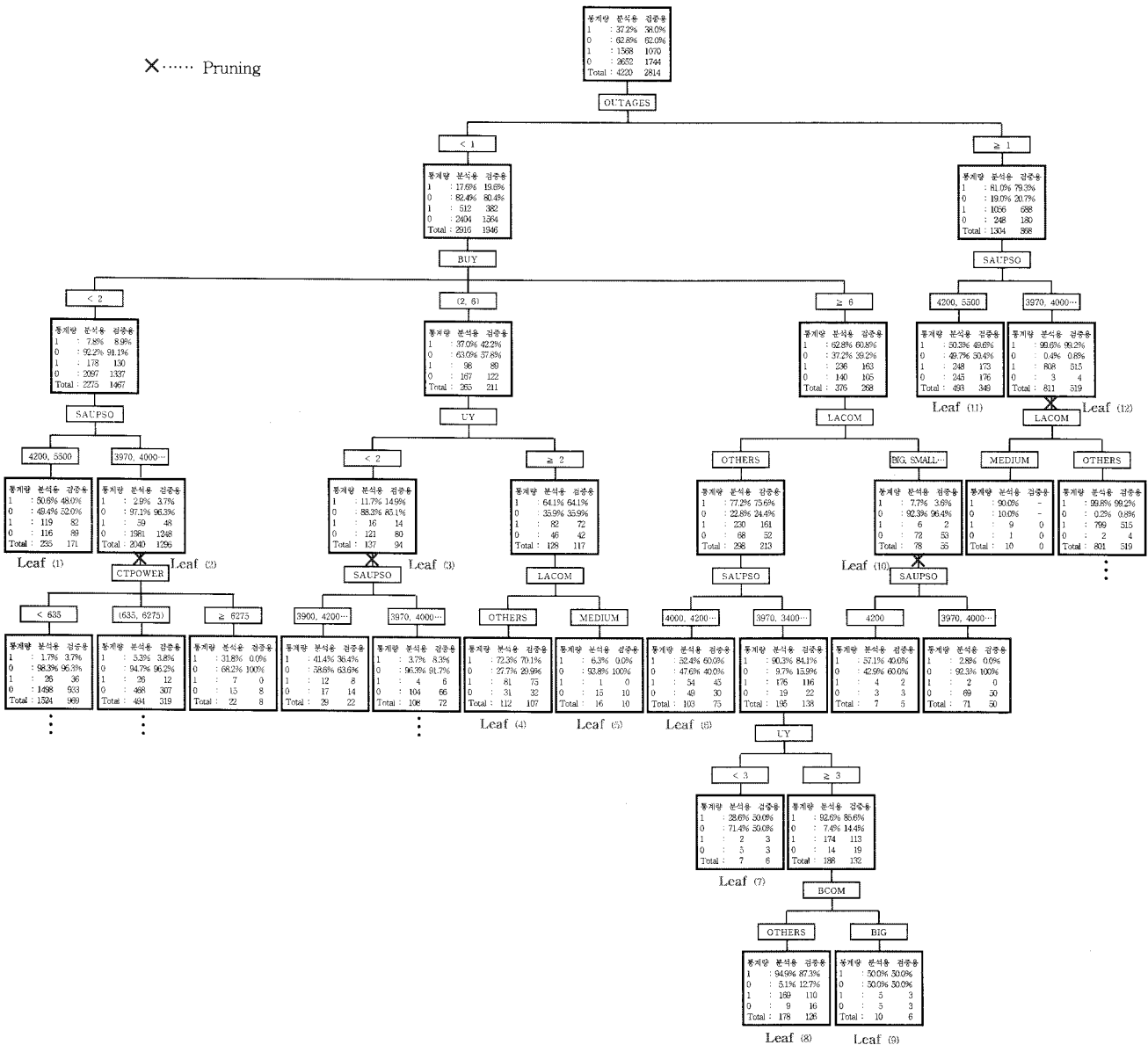


그림 1 특별고압 파급고장 발생유형 인터랙티브 의사결정나무 및 가지치기
Fig. 1 Interactive Decision Tree and Branch Pruning of Medium-Voltage Customer Fault Pattern

2.4 특별고압 파급고장 발생가능 고객 예측 모델

표 3에서 보는 바와 같이 오분류율이 5%미만인 리프(Leaf)를 기준으로 그림 1의 데이터마이닝 의사결정나무 해석 결과를 이용하여 파급고장 발생가능 고객 분류예측 논리식을 분류 정확도가 높은 순서로 표 4와 같이 만들었다.

표 3 Leaf Node별 오분류율

Table 3 Misclassification Rate by Each Leaf Node

Leaf	Training						
	고장 발생률[%]	1	0	소계	오분류율		
					오분류수	전체수	오분류율[%]
(12)	99.6	808	3	811	3	811	0.4
(8)	94.9	169	9	178	12	989	1.2
(4)	72.3	81	31	112	43	1,101	3.9
(6)	52.4	54	49	103	92	1,204	7.6
(1)	50.6	119	116	235	208	1,439	14.5
(11)	50.3	248	245	493	453	1,932	23.4
(9)	50.0	5	5	10	458	1,942	23.6
(7)	28.6	2	5	7	463	1,949	23.8
(3)	11.7	16	121	137	584	2,086	28.0
(10)	7.7	6	72	78	656	2,164	30.3
(5)	6.3	1	15	16	671	2,180	30.8
(2)	2.9	59	1,981	2,040	2,652	4,220	62.8

즉, 표 4의 논리조건은 의사결정나무에서 해당 리프가 나오기까지의 단계별 분류과정을 논리조건식으로 표현한 것이며 구체적 의미는 다음과 같다. Leaf(12)는 사업소가 4200 또는 5500이 아닌 대부분의 사업소에서 한번 고장이 발생한 고객에서 또다시 고장이 발생할 확률이 가장 높다는 것을 의미한다. 다음으로 Leaf(8)은 고장이 한 번도 발생하지 않은 고객에서는 차단기사용연수가 6년 이상이고 피뢰기제작회사가 소규모이고 이름이 없는 기타인 경우에서 사업소가 3970 등인 경우로 전기공급 후 3년 이상 경과하고 차단기제작회사가 소규모이고 이름이 없는 기타인 경우에서 고장이 많이 발생한다는 것을 의미한다. 또한 Leaf(4)는 고장이 한 번도 발생하지 않은 고객으로 차단기 사용연수가 2~5년 이고 전기사용연수가 2년 이상이며 피뢰기제작회사가 소규모로 이름이 없는 기타인 경우에서 고장이 발생하고 있음을 나타낸다.

표 4 특별고압 파급고장 발생가능 고객 분류 논리식

Table 4 Logic Equation Classifying Customers Who Cause Medium-voltage Customer Fault

Leaf	논리 조건
(12)	고장횟수 ≥ 1 \cap 사업소 $\neq 4200, 5500$
(8)	고장횟수 < 1 \cap 차단기 사용연수 ≥ 6 \cap 피뢰기제작사 = 기타 \cap 사업소 = 3970 등 \cap 전기사용연수 ≥ 3 \cap 차단기제작회사 = 기타
(4)	고장횟수 < 1 \cap 차단기 사용연수 = 2~5년 \cap 전기사용연수 ≥ 2 \cap 피뢰기제작회사 = 기타

이 논리식을 근거로 하여 특별고압 파급고장 발생가능 고객을 예측할 수 있는 모델을 그림 3과 같이 고안하였다. 이 모델에서 보면 첫째, 고장이 발생한 실적이 있는 고객 중 사업소 4200과 5500을 제외한 나머지 사업소의 고객은 파급고장 발생 가능성이 높으므로 최우선적으로 점검해야 한다. 둘째, 고장발생 실적이 없는 고객 중 차단기의 사용연수가 2~5년이고 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 3970, 3900, 3400, 4500, 4600, 5000, 5200, 5400이고 전기사용기간이 3년 이상 및 차단기 제작회사가 기타인 고객은 파급고장 발생 가능성이 높으므로 우선적으로 점검해야 한다는 결론을 얻었다.

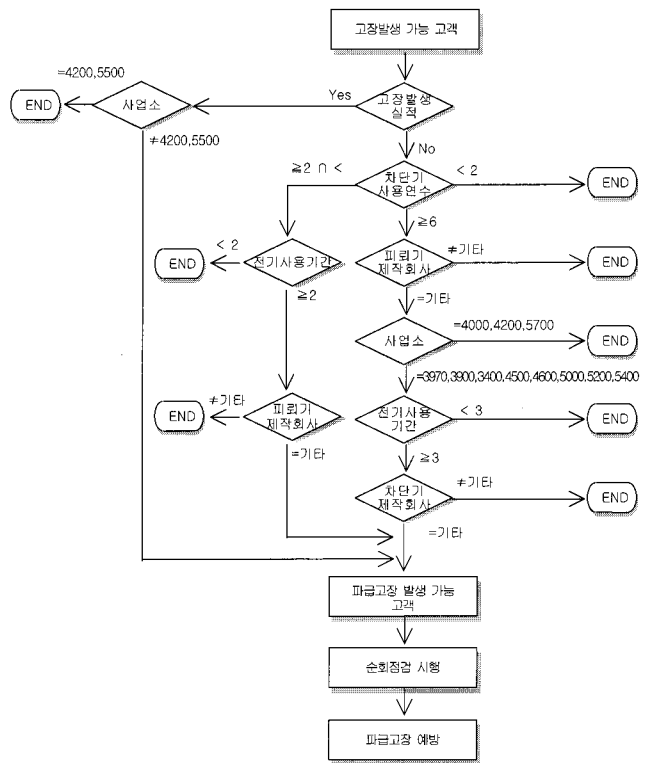


그림 3 특별고압 파급고장 발생가능 고객 예측모델
Fig. 3 Prediction Model of Customers Who Occur Medium-Voltage Customer Fault

2.5 특별고압 파급고장 발생가능 고객 예측시스템 구현

2.5.1 예측시스템 구성

신고객정보시스템(NCIS)의 고객계약 데이터베이스와 신배전정보시스템(NDIS)의 수전설비 및 파급고장 데이터베이스에서 파급고장 발생가능 고객 예측에 필요한 표 4의 예측변수들을 추출하여 그림 3의 예측모델을 실현할 수 있도록 그림 4와 같이 예측시스템을 구성한다. 이렇게 하여 업무시스템에서 필요한 예측변수의 데이터를 직접 고객예측시스템으로 입력하고 결과를 해석하도록 예측시스템을 고안하였다.

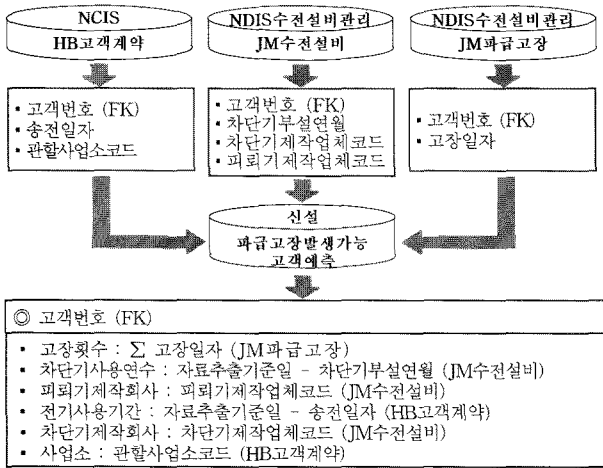


그림 4 파급고장 발생가능 고객예측 시스템 구성도
 Fig. 4 Configuration of Prediction System to Find Customers Who Cause Medium-Voltage Customer Faults

2.5.2 필요사항 및 조건

특별고압고객 파급고장 발생가능고객 예측시스템 구현을 위해서는 표 4에서 언급된 필수항목들의 자료가 반드시 입력관리되어야 한다. 따라서 도면검토등록, 수전설비 순회시험점검, 파급고장 입력 시에 차단기설치연월, 차단기제작회사, 피뢰기제작회사의 입력이 필수적이다. 각각의 화면에서 입력한 데이터는 최근 입력된 데이터를 기준 데이터로 한다. 또한 입력데이터 정확도를 향상시키기 위하여 차단기와 피뢰기 설치연월은 제조연월 이후이어야 하는 것과 같은 입력데이터 상관체크 및 검증기능이 필요하다. 그리고 전체 입력항목을 필수 항목으로 관리하여 데이터 누락을 방지하고 고객의 폐지, 휴지 등의 사유로 인한 변경데이터에 대한 이력관리를 하여 고객사항과 설비데이터간의 연결 포인트를 지속적으로 관리할 수 있도록 예측시스템을 만든다.

2.5.3 시스템 구축 기본 방향

기존 특별고압 수전설비 관리시스템 화면에 차단기와 피뢰기 사항을 입력란에 추가하여 데이터를 입력받는다. 차단기와 피뢰기의 입력된 정보는 이력관리를 하도록 한다. 표 4의 필수 예측변수를 그림 4와 같이 추출하여 그림 3의 예측모델을 예측시스템으로 구현한 다음 특별고압 파급고장 발생가능 고객 명세서를 출력하여 사업소에서 해당 고객에 대한 고객설비 고장예방 점검을 할 수 있도록 한다. 또한 최종적인 특별고압 파급고장 발생가능 예측 고객 명세는 화면으로 조회하고 사용자가 현장에서 활용할 수 있도록 엑셀 형태의 출력 기능을 추가로 만들도록 한다.

2.5.4 입력 설계

현재 구성된 입력화면에서 특별고압 파급고장 발생가능고객 예측모델 구현을 위해 그림 3에서 도출된 예측변수들의 필수입력 항목으로 수전설비 최초등록 시 입력에서 누락된 피뢰기의 입력란을 보호장치에 추가하여 반드시 입력하도록 신규입력설계 한다.

다음은 전력사용량의 정상계량 여부를 확인하기 위하여 3년에 1회 실시하는 순회점검 후 결과입력 시, 특별고압 파급고장 발생가능 여부를 예측하는 순회점검 결과 등록화면에

필수입력항목으로 입력에서 누락된 차단기, 피뢰기의 제조연월과 설치연월 및 제조사 입력란을 추가하여 입력하도록 신규입력설계 한다. 그리고 특별고압 파급고장 발생 시 현장 조사를 실시하여 결과를 입력하는 파급고장 데이터베이스의 수전설비 파급고장 점검결과 등록화면에 파급고장 발생가능 고객 예측변수로 필수입력항목이나 누락된 차단기, 피뢰기의 제조연월, 제조사 입력란을 추가하여 입력하도록 신규입력설계 한다.

2.5.5 출력 설계

출력화면은 지금까지 입력한 데이터를 이용하여 특별고압 파급고장 발생가능 고객을 예측하고 사업소별로 활용할 수 있도록 파급고장 데이터베이스에 고객별로 주소, 상호, 계약 전력, 인입주 번호, 공급일자, 차단기의 설치연월과 제작회사, 피뢰기의 설치연월과 제작회사, 고장횟수를 신규 출력 설계하여 특별고압 고객의 수전설비 점검에 활용한다. 특별고압 파급고장 발생가능고객의 현장을 방문하여 적외선 열화상 탐지장치 등으로 특별고압 수전설비의 차단기와 피뢰기 등을 정밀 점검하고 문제점이 있는 설비를 사전 교체함으로써 파급고장을 미리 예방할 수 있게 한다.

2.5.6 특별고압 파급고장 발생가능 고객 예측시스템 프로그램명

지금까지의 예측 시스템 구성, 입력설계, 출력설계를 바탕으로 특별고압 파급고장 발생가능 고객을 예측하여 해당 고객리스트를 출력해주는 시스템을 그림 5와 같이 프로그래밍 하였다. 본 프로그램은 기존의 특별고압 수전설비 관리시스템 프로그램에 예측시스템을 구축해야하는 관제로 새로운 예측시스템을 기존시스템에 연계하는 부분은 불가피하게 기존 시스템을 개발한 K사의 협력회사 도움을 받아 연결하였음을 밝혀둔다.

```

[특별고압 파급고장 발생가능 고객 분류 예측 논리식 쿼리]
SELECT A.CUSTNO
      ,A.USER_NM
      ,A.JUSO
      ,A.TRADENM
      ,A.CNTR_PWR
      ,B.SVCPOLE_NO
      ,A.ONSVC_YMD
      ,A.BRK_ATT_YM
      ,A.BRK_MANUFR_CD
      ,A.LA_ATT_YM
      ,A.LA_MANUFR_CD
      ,A.GOJANG
INTO :H-CUSTNO
WHERE ONSVC_YMD < TO_CHAR(ADD_MONTHS(TO_DATE
(H-STATI-YMD,'YYYYMMDD'),-24),'YYYYMMDD')
AND NOT EXISTS (SELECT CUSTNO
FROM JMT0090
WHERE CUSTNO = C.CUSTNO)
AND C.CUSTNO = B.CUSTNO
AND C.CREATE_SEQNO = (SELECT MAX(CREATE_SEQNO)
FROM JMT0200
WHERE CUSTNO = C.CUSTNO)
AND C.LA_MANUFR_CD = '90'
AND D.CUSTNO = B.CUSTNO
AND D.CREATE_SEQNO = (SELECT MAX(CREATE_SEQNO)
FROM JMT0210
WHERE CUSTNO = D.CUSTNO)
AND D.BRK_ATT_YM BETWEEN
TO_CHAR(ADD_MONTHS(TO_DATE
(H-STATI-YMD,'YYYYMMDD'),-72),'YYYYMMDD')
AND
TO_CHAR(ADD_MONTHS(TO_DATE
(H-STATI-YMD,'YYYYMMDD'),-24),'YYYYMMDD')
AND B.JURIS.OFFICECD = :H-OFFICECD
) A
..JLT0440 B
WHERE A.CUSTNO > :H-N-CUSTNO
AND A.CUSTNO = B.CUSTNO(+)
ORDER BY A.CUSTNO
    
```

그림 5 특별고압 파급고장 발생가능 고객 예측 프로그램명
 Fig. 5 Prediction System Program to Find Customers Who Cause Medium-Voltage Customer Faults

이렇게 구현한 예측시스템으로 여러 곳의 데이터베이스에 산재된 파급고장 예측변수들의 데이터를 예측시스템으로 모아 C사업본부에서 특별고압 파급고장 발생가능 고객을 찾아 실제 예측한 결과를 그림 6과 같이 출력하였다.

고객번호	고객명	주소	상호	계약전력	인입주번호	공급일자	차단기 설치연월	차단기 설치
0602031998	간척동	대전광역시 서구 도	1000	17562601	19970715	200410		30
060304827	박대용	대전광역시 서구 둔	3	17671793	20030719			
060309707	계룡대근무지점	충청남도 계룡시 문	40	19503318				
0602096127	박상식	대전광역시 서구 도	3	16679704	19980331			
061196076	이관균	충청남도 계룡시 문	8	11559421	20040213			
0614089713	서한우	대전광역시 서구 둔	5	17574361	20000522			
0614082593	정민환	대전광역시 서구 둔	5		1991123			
0617204531	김연애	대전광역시 서구 둔	15		19941118			
0617385825	삼부도건	대전광역시 서구 둔	74		20020611			
0618018688	공정숙	대전광역시 서구 둔	3	19504077				
0619844036	채종갑	대전광역시 서구 둔	3	19504332	19940505			
0622012590	김태민(삼정제회)	대전광역시 서구 둔	510	17639333	19951216			
0622013667	리순덕	대전광역시 서구 둔	360	17638361	20000327			
0622014641	권리단	대전광역시 서구 둔	950	17638282	20030702			
062203722	허공빈씨	대전광역시 서구 둔	300	17500381	19970712			
062203947	한봉교	대전광역시 서구 둔	400	18508341	19980613			
0622036168	김영남	대전광역시 서구 둔	150	18508791	19911111			

그림 6 C사업본부 파급고장 발생가능 고객출력 샘플
Fig. 6 Sample of Customer List Causing Medium-Voltage Customer Faults at C-Division

3. 데이터웨어하우스 구축방안

3.1 구축 방향

특별고압 파급고장 발생가능 고객 예측시스템은 예측에 필요한 변수들의 데이터를 장기적으로 확보하여 예측의 정확도를 높이는 것과 동시에 향후 기술발전에 따라 새로운 시스템의 출현에 대비하고 미래에 지속적으로 활용 가능하도록 하기위해 데이터웨어하우스의 구축이 필요하다. 데이터웨어하우스에서 고장발생 가능고객을 예측하는데 필요한 메타데이터를 저장해 놓고 고장발생 가능 고객의 분류예측 논리식을 적용한 예측모델을 시스템으로 구현하여 해당고객을 예측해낸다면 효과적으로 특별고압 수전설비의 파급고장을 감축할 수 있을 것이다. 이에 따라 현행 신고객정보시스템과 신배전정보시스템에서 데이터마이닝 기법을 적용한 변수들의 데이터를 추출하여 메타데이터로 그림 7과 같이 데이터웨어하우스에 저장하도록 고안한다. 그 다음 데이터웨어하우스에 저장된 변수들의 데이터 중에서 특별고압 파급고장 발생가능 고객을 예측하는데 필요한 변수의 데이터를 불러내어 그림 7과 같이 앞서 구현한 예측시스템으로 해당고객을 예측해 내도록 시스템을 구성한다.

이렇게 예측시스템을 구현하면 IT기술의 발전으로 시스템이 업그레이드되거나 변경되더라도 데이터는 지속적으로 활용할 수 있는 이점이 있다. 그러나 Enterprise Resource Planning (ERP)시스템으로 구성된 K사의 데이터웨어하우스에 특별고압 파급고장 발생가능 고객 예측변수들을 본 논문에서 데이터웨어하우스로 직접 구축하기에는 패키지 프로그램을 수정해야 하므로 많은 재정적 부담이 수반되어 현실적으로 불가능하다. 따라서 ERP시스템에 데이터웨어하우스를 구축하는 것은 장기적으로 추진하도록 구축방안만을 제시하고자 한다.

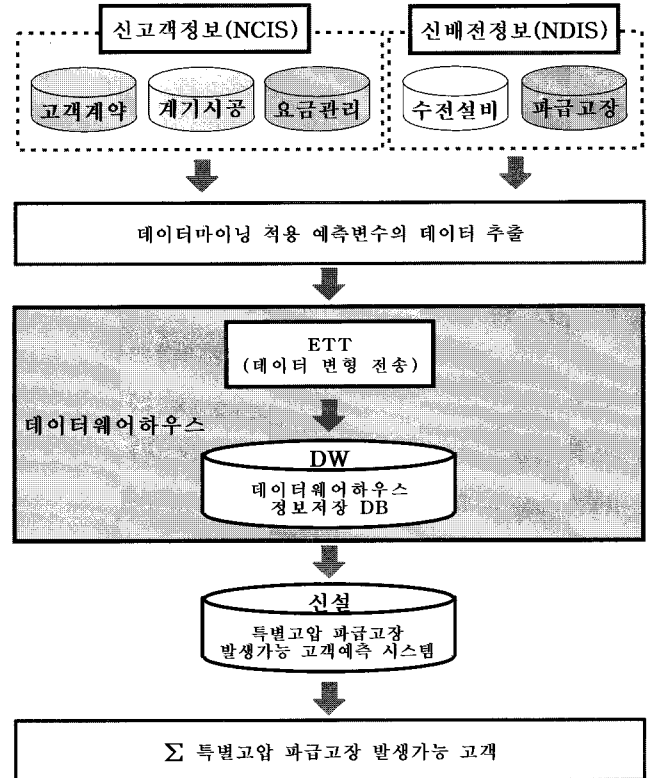


그림 7 데이터웨어하우스와 고객예측 시스템 구성도
Fig. 7 Configuration Concept of Prediction System and Data Warehouse

3.2 데이터마이닝 적용 예측변수 추출

데이터마이닝 기법을 적용하여 특별고압 파급고장 예측에 사용된 표 2의 변수를 신고객정보시스템(NCIS)과 신배전정보시스템(NDIS)에서 찾아내어 해당 데이터베이스와 예측변수들의 객체관계도를 그림 8과 같이 분석하여 직접 작성하였다.

구체적으로 그림 8에서 살펴보면 신고객정보시스템(NCIS)에서 추출할 수 있는 예측변수들은 다음과 같다. 고객계약 데이터베이스에서는 계약종별(4), 계약전력(5), 사용연수(7)를 계산할 수 있는 송전일자이고, 계기시공 데이터베이스에서는 MOF제작회사(22), MOF사용연수(23)를 알 수 있는 MOF제조연월, MOF 전류장도(24)이다. 요금관리 데이터베이스에서는 사용량배수(6)를 산출할 수 있는 전력사용량을 추출할 수 있다. 그리고 신배전정보시스템(NDIS)에서 추출할 수 있는 예측변수들은 다음과 같다. 수전설비 데이터베이스에서는 선로종류(11)인 고압인입선종류, 선로규격(12)인 고압인입선규격, 선로거리(13)인 고압인입선거리, 변압기용량(25), 변압기제작회사(26), 변압기사용연수(27)를 계산할 수 있는 변압기설치일자, 수전설비위치(8), 개폐기종류(14), 개폐기사용연수(15)를 계산할 수 있는 개폐기제작연도, 개폐기제작회사(16), 안전관리자구분(10), 차단기차단용량(17), 차단기제작회사(18), 차단기사용연수(19)를 계산할 수 있는 차단기부설연월이다. 파급고장 데이터베이스에서는 고장횟수(9)를 계산할 수 있는 고장일자를 추출할 수 있다.

그림 8 객체관계도에 있는 예측 변수들의 메타데이터를 추출하여 필요한 형태로 변형하여 정보저장소에 보관한 다음 필요한 때에 가져다가 가공하여 활용할 수 있는 데이터웨어하우스를 구성하도록 한다.

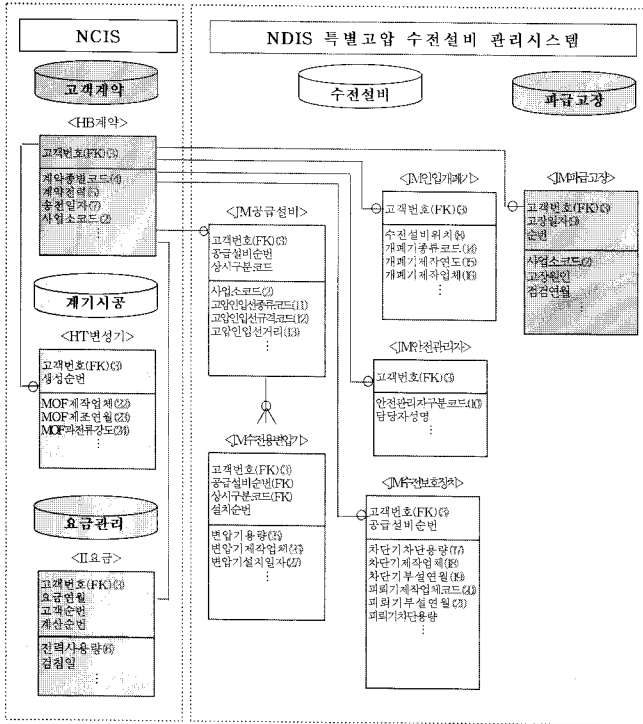


그림 8 파급고장 발생가능 고객예측 변수의 객체관계도
 Fig. 8 Entity Relationship Diagram of Variables Predicting Customers Who Cause Medium-Voltage Customer Faults

3.3 데이터웨어하우스 구축방안

데이터웨어하우스의 구축은 많은 시간과 노력이 소요되는 작업으로 구체적인 구축 방법은 그림 9와 같다. 데이터웨어하우스는 일반업무 데이터베이스, 데이터 추출·변형·전송, 정보저장소 및 정보활용의 4가지 부분으로 구성된다.

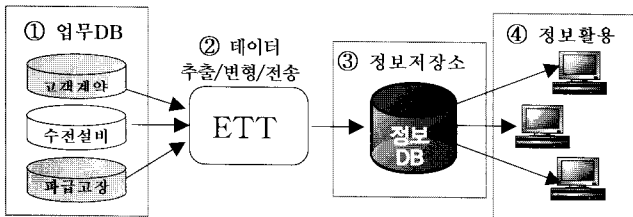


그림 9 데이터웨어하우스의 구성
 Fig. 9 Organization of Data warehouse

그림 9의 업무 데이터베이스는 고객계약, 요금관리, 계기시공, 수전설비, 파급고장처럼 다양한 업무영역의 데이터베이스들을 의미하며, 각각의 시스템에서 개별적으로 관리되고 있다. 여기서 ETT(Extraction, Transformation, Transportation의 약자이며 ETL(Extract, Transform, Load)이라고도 함)은 일반 업무데이터베이스를 데이터웨어하우스용 정보데이터베이스로 변환하는 자동화 툴이다[17]-[18].

K사는 2007년 경영정보를 실시간으로 제공하기 위해 BW (Business Information Ware house)라는 데이터웨어하우스를 구축하였다. BW에서 정보데이터베이스는 Infocube라고 하는 Star Schema 구조를 가진 다차원구조의 읽기전용 데이터베이스를

갖고 있다[19].

BW에는 ETT기능을 내장하고 있어 BW자체에서 필요한 대부분의 데이터를 직접 취득하고 있으나, 신고객정보시스템과 신배전정보시스템은 ERP가 구축되지 않아 그림 10과 같이 EAI(Enterprise Application Integration)시스템을 통해 해당시스템에서 필요한 예측변수들의 데이터를 추출하여 ERP의 BW로 보내도록 만든다.

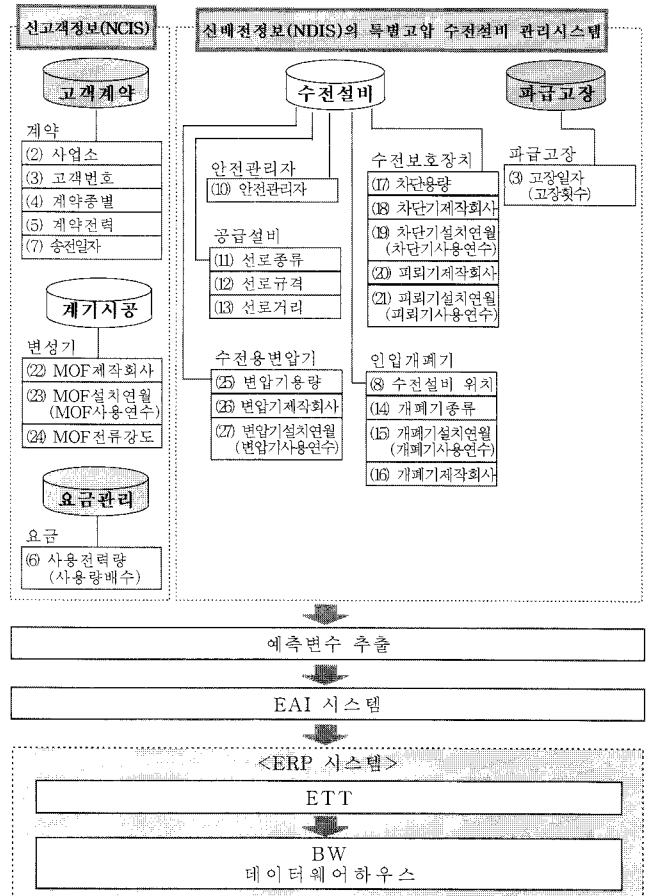


그림 10 ERP 데이터웨어하우스 구축 구성도
 Fig. 10 System Configuration of ERP Data Warehouse

ERP BW에 데이터를 저장하기 위해서는 인포큐브를 생성하고, 업무 데이터베이스의 위치를 나타내기 위해 인포소스를 지정한 다음 업무 데이터베이스에서 BW로 데이터를 전송하기 위해서 전송구조와 전송규칙을 지정한다. 그림 11에서 오른쪽이 특별고압 파급고장 발생가능 고객예측 데이터베이스의 데이터필드를 나타내고 왼쪽이 BW에 저장될 정보를 나타낸다. 특별고압 파급고장 발생가능 고객예측 데이터베이스가 BW의 요건에 맞게 구성되어 있다면 단순히 오른쪽에서 왼쪽으로 전송하기만 하면 된다. 그러나 요건이 맞지 않는 경우에는 별도의 변환작업을 거쳐야 하고 필요한 경우에는 원하는 필드만 전송하도록 변환조건을 지정해야 한다. 다음으로 인포큐브가 생성되어지고 전송규칙이 정해지면 언제 어떻게 데이터를 전송받을 것인지를 지정하면 데이터웨어하우스 구축이 완성된다.

이렇게 데이터웨어하우스를 구축함으로써 향후 IT기술발전 에 따른 시스템 변경에도 예측변수들의 데이터를 지속적으로 활용 가능하고 장기적으로 예측변수들의 데이터를 확보하여 예측시스템의 예측정확도를 향상시킬 수 있는 이점이 있다.

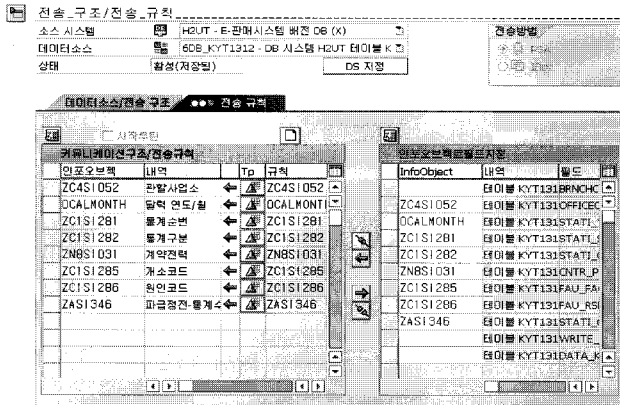


그림 11 전송구조 및 전송규칙 지정
Fig. 11 Define Transmission Rules and Structure

4. 의사결정나무 알고리즘을 적용한 예측시스템 적용효과

특별고압 파급고장 예방을 위해서 기존의 특별고압 수전설비 관리시스템에서는 고장설비와 고장원인의 이원분석 방법을 사용하여 고장예방 활동을 해 왔으나 실효성을 거두지 못하였다. 매년 고장건수는 증가하고 고장점유율의 비중도 늘어나고 있으나 특별한 예방대책은 없는 실정이다. 금융, 보험, 통신 분야에서 데이터마이닝 기법을 사용하여 고객의 기본정보를 토대로 신용불량 고객을 색출하고 대출 등의 업무에 반영하여 회사의 손실을 줄이고 있는 것에 착안하여 본 연구에서는 데이터마이닝의 의사결정나무 알고리즘을 전기분야에 적용하였다. 수많은 특별고압 고객 중에서 특별고압 파급고장 발생가능 고객을 예측하고 해당고객의 수전설비를 예방점검하여 불량 설비를 개선할 수 있게 함으로써 특별고압 파급고장을 효율적으로 줄일 수 있게 된 것은 지금까지 시도된 적이 전혀 없는 새로운 연구로 커다란 성과라 할 수 있다.

이번에 연구한 데이터마이닝 기법의 의사결정나무 알고리즘을 적용한 특별고압 파급고장 발생가능 고객 예측시스템의 구체적인 개발 효과는 다음과 같다.

첫째, 전통기술 분야인 전력분야에 새로운 IT와 통계기법을 융합하여 전력분야에서 가장 중요한 전기품질을 스마트하게 향상하도록 하였다. 기존의 일차원 통계분석과 이차원 통계분석을 근거로 전체 특별고압고객의 수전설비를 정기적으로 순회점검하는 업무처리방식에서 벗어나 표 2에서 보는 바와 같이 데이터마이닝 기법을 사용하여 28개 예측변수를 동시에 사용해서 특별고압 파급고장 발생가능 고객을 예측 하였다. 그리고 본 연구에서 개발한 예측시스템으로 해당 고객만을 직접 찾아내어 집중 점검할 수 있게 함으로써 효율적으로 특별고압 파급고장을 예방할 수 있다.

둘째, 특별고압고객 순회점검에 소요되는 비용과 인력을 절감할 수 있다. 현재 3년에 1회씩 전체 특별고압고객의 수전설비를 정기적으로 순회점검하고 있으나 본 연구에서 개발한 예측시스템을 활용하여 파급고장 발생가능 고객만을 찾아내어 점검하고 매년 점검을 완료한 후 설비를 개수한 고객을 제외해 나가면 특별고압고객 순회점검에 소요되는 비용과 인력을 효율적으로 줄일 수 있을 것이다.

셋째, 특별고압 파급고장으로 인한 건전설비 고객의 정전피해를 예방할 수 있다. 고도로 IT화된 사회로 발달하고 있는 우리나라에서 이제 고장으로 인한 정전피해는 막대하다.

무작정 열심히 고장예방을 위해 특별고압 수전설비를 순회점검하고 노력하는 것 보다는 체계적이며 같이 파급고장 발생가능 고객만을 찾아내어 수전설비를 보수한다면 효율적으로 특별고압 수전설비의 파급정전 고장을 예방하여 동일한 배전선로에 연결되어 있는 많은 건전한 고객들의 정전피해를 효과적으로 예방할 수 있을 것이다.

넷째, 전력분야의 다른 부분에도 응용하여 유용한 결과를 얻을 수 있다. 수만 가지의 기자재가 직렬로 연결되어 발전소에서부터 고객까지 전송되는 전기는 이제 과거와 같이 단순히 고장설비별 원인별 이차원적인 고장분석방법으로는 고장을 더 이상 줄일 수 없는 한계에 도달하였다. 다양한 자연환경에 노출되어 운전되는 송배전설비와 고객 수전설비의 고장감소를 위한 분석방법은 이제 IT기술과 통계기법을 접목하여 보다 차원 높은 분석이 이루어져야만 유용한 결과를 얻을 수 있을 것이다. 따라서 특정 기자재와 특정 고장에 대해서도 본 연구에서 수행한 것과 같이 데이터마이닝 기법과 IT기술을 적용하여 해석하면 효과적인 해석결과를 얻을 수 있다고 본다. 또한 전기요금의 미수금 고객을 분석하는 데에도 데이터마이닝 기법을 사용하면 어떤 유형의 고객이 요금을 많이 체납하거나 자주 체납하고 있는가 하는 것을 알 수 있으므로 효과적인 대책을 수립하여 전기요금 수금율을 향상시키는 데에도 기여할 수 있을 것으로 사료된다.

다섯째, 데이터마이닝 기법을 적용한 특별고압 파급고장 발생가능 고객 예측시스템을 사용하여 특별고압 수전설비의 파급고장을 감소하는 사례는 정보화가 세계적으로 확산된 현대에서 해외의 다른 전력회사에서도 사용할 수 있다. 구체적으로 우리나라와 유사한 일본은 전력공급체제에 고압수전설비 체제를 시행하고 있고 고압고객이 50만호 정도, 파급고장이 10%정도 발생하고 있으므로 일본의 전력회사 데이터를 넣어 해석하면 해당 전력회사에서 파급고장 발생가능 고객을 예측해낼 수 있어 K사가 얻은 것과 똑같은 효과를 얻을 수 있다. 이외에도 전력공급에서 고압수전설비 체제를 사용하고 있는 동남아시아 등 다른 국가에서도 본 연구결과를 유용하게 활용할 수 있다.

5. 결 론

전체 배전고장이 매년 감소하고 있음에도 특별고압고객에서 발생하는 파급고장은 건수와 점유비율이 증가하고 있다. 기존의 고장설비와 고장원인의 이원분석 방법으로는 고장을 감소시키는데 한계에 도달하였다. 본 연구에서는 데이터마이닝 기법을 적용하여 파급고장 발생가능 고객을 예측하여 효율적으로 고장을 감소시킬 수 있도록 예측시스템을 실제 구현하고 데이터웨어하우스 구축방안을 제시하였다. 앞으로 배전계통의 특별고압 파급고장을 발생시킬 가능성이 높은 고객을 예측 시스템으로 추출하여 현장점검을 실시하고 불량설비를 개수함으로써 특별고압 파급고장을 예방할 수 있게 되었다. 구체적으로 특별고압 파급고장 발생가능 고객 예측용 입력·출력 설계 및 시스템 설계를 하고 프로그램을 작성하였다. 이를 통해 여러 곳으로 분산된 필요 데이터를 연계하여 가져오고 필수항목으로 입력되지 않고 있는 피뢰기, 차단기 등 관련 데이터는 입력항목으로 신규 설계하여 체계적으로 입력하고 관리하도록 개선하였다. 또한 본 연구에서 구현한 프로그램으로 데이터를 분석하여 특별고압 파급고장 발생가능 고객을 C사업본부에서 실제로 예측하였다. 본 예측 시스템에서 예측한 고객리스트를 기준으로 파급고

장 예방점검을 실시하여 전체 배전고장의 많은 점유율을 차지하고 있는 특별고압 고객의 파급고장을 효율적으로 줄일 수 있게 된 것은 지금까지 전혀 연구된 적이 없는 새로운 연구결과로 커다란 성과라 할 수 있다. 이번 연구에서는 경제적인 사유로 특별고압 파급고장 발생가능 고객예측 시스템만 구현하고 데이터웨어하우스는 ERP 시스템에 구축 방안을 제시하였다. 앞으로 재정적 투입을 통해 본 연구의 데이터웨어하우스 구축방안이 향후 추진될 판매분야 차세대 시스템 구축 시 프로그램으로 구현되기를 기대한다. 또한 본 예측시스템으로 예측한 결과를 1~3년간 현장에서 시행한 다음 그 결과를 수집하고 분석하여 본 연구에서 개발한 예측시스템의 시행 효과에 대한 연구가 후속 연구로 이어지기를 바란다.

참 고 문 헌

[1] 한국전력공사, “고객파급 정전 분석 및 예방대책”, 2000-2008

[2] 배성환, 김자희, 홍정식, 임한승 “데이터마이닝 기법을 이용한 특별고압 파급고장 발생가능 고객 예측모델 구축 및 신뢰도 향상방안에 관한 연구”, 대한전기학회지 제58권 제10호, pp. 1869-1880, 2009

[3] 대한전기학회(김재철 외 16인), “최신 배전시스템 공학”, 북스힐, pp.415-420, 2006

[4] 박영성, 박삼진, 오재신, “GIS를 이용한 신배전정보시스템(NDIS) 구축”, 대한전기학회 하계학술대회논문집, pp.18-20, 2007

[5] 황우현, 김자희 “3D GIS를 활용한 지중배전설비 시공·관리시스템 연구개발”, 대한전기학회 하계학술대회논문집, pp.36-37, 2007

[6] 허재행, 정하섭, 김유창, 손진만, 박종근, “배전계통의 다양한 유지보수기법을 적용한 신뢰도기반 최적 유지보수모델”, 대한전기학회 하계학술대회논문집, pp.50-51, 2006

[7] 박찬업, 권병국, 양원영, 이승철, “고장파급 시나리오에 기초한 광역정전 해석기법연구”, 대한전기학회논문지 제57권 2호, pp.155-163, 2008

[8] 배명석, “GIS를 이용한 배전설비 고장예측 알고리즘의 구현”, 한국공간정보시스템학회 춘계학술대회논문집, pp.89-94, 2002

[9] 이재식, 이영주, “사례기반 추론을 이용한 설비고장시기 예측”, 한국데이터베이스학회 춘계공동학술대회논문집, pp.315-322, 1999

[10] Tor, O, Shahidehpour, M, “Power Distribution Asset Management”, IEEE Power Engineering Society General Meeting, pp.7, 2006

[11] Berry C, Hirsch P, Tuel W.G, “Data Base Model for Distribution Facilities”, IEEE Transaction on Power Apparatus and Systems, vol. PAS-101 pp.363-370, 1982

[12] Ying Xiao, Shiming Tian, Qincheng Yuan, Xinyu Zhang, “An Approach to the Integration of GIS and DMS”, POWERCON'98 International Conference on Power System Technology, vol.1 pp.297-301, 1998

[13] Whei-Min Lin, Ming-Tong Tsay, Su-Wei Wu, “Application of Geographic Information System to Distribution Information Support”, IEEE Transactions on Power Systems, vol.11 pp.190-197, 1996

[14] Jae-Ju Song, Bong-Jae Yi, Jin-Ho Shin, Ju-Young Kang, Jung-II Lee, Seon-Ku Cho, “A Development of the Automatic Digitizing System for Power Distribution Facility Images”, 4th Annual ACIS International Conference on Computer and

Information Science, pp.105-110, 2005

[15] Nagura S, Masumoto T, Endo K, Wakasa F, Watanabe S, Ikeda K, “Development of Mapping System for Distribution Facility Management”, CIRED 10th International Conference on Electricity Distribution, vol.6 pp.520-524, 1989

[16] Reder W, Flaten D, “Reliability Centered Maintenance for Distribution Underground Systems”, IEEE Power Engineering Society Summer Meeting, vol.1 pp.551-556, 2000

[17] 홍릉과학출판사, W. H. Inmon, 유영일 번역 “데이터웨어하우스 구축 방법론”, 1998

[18] 이경아, 안기홍, “데이터웨어하우스 시스템 설계 및 구현”, 한밭대학교 정보통신전문대학원 논문집 제6권 제1호, pp. 47-52, 2008

[19] 한국전력공사, “PowerPia 관리자 매뉴얼”, 2007

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 방문

연구원
 2005년 ~ 현재 국립서울산업대학교 IT정책전문대학원 산업정보시스템공학 조교수
 2007년 대한전기학회 정회원
 Tel : 02-970-6867
 E-mail : jahee@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