

## 급배수관망 누수에측을 위한 확률신경망

# Probabilistic Neural Network for Prediction of Leakage in Water Distribution Network

하성룡<sup>1\*</sup> · 류연희<sup>2</sup> · 박상영<sup>3</sup>

Ha, Sung-Ryong<sup>1\*</sup> · Ryu, Youn-Hee<sup>2</sup> · Park, Sang-Young<sup>3</sup>

1 충북대학교 도시공학과

2 충북대학교 도시공학과 대학원 졸업

3 한국수자원공사 수자원연구원

(2006년 6월 21일 논문 접수; 2006년 12월 6일 최종 수정논문 채택)

### Abstract

As an alternative measure to replace reactive stance with proactive one, a risk based management scheme has been commonly applied to enhance public satisfaction on water service by providing a higher creditable solution to handle a rehabilitation problem of pipe having high potential risk of leaks. This study intended to examine the feasibility of a simulation model to predict a recurrence probability of pipe leaks. As a branch of the data mining technique, probabilistic neural network (PNN) algorithm was applied to infer the extent of leaking recurrence probability of water network. PNN model could classify the leaking level of each unit segment of the pipe network. Pipe material, diameter, C value, road width, pressure, installation age as input variable and 5 classes by pipe leaking probability as output variable were built in PNN model. The study results indicated that it is important to pay higher attention to the pipe segment with the leak record. By increase the hydraulic pipe pressure to meet the required water demand from each node, simulation results indicated that about 6.9% of total number of pipe would additionally be classified into higher class of recurrence risk than present as the reference year. Consequently, it was convinced that the application of PNN model incorporated with a data base management system of pipe network to manage municipal water distribution network could make a promise to enhance the management efficiency by providing the essential knowledge for decision making rehabilitation of network.

**Key words:** data mining; pipe leaks; probabilistic neural network (PNN); water distribution network

**주제어:** 데이터마이닝, 관로 누수, 확률신경망, 상수관망

\*Corresponding author Tel: +82-43-261-2497, FAX: +82-43-261-2497, E-mail: simplet@chungbuk.ac.kr (Ha, S.R.)

## 1. 서론

누수의 주요 원인을 분석해 보면 관의 노후화와 아주 밀접한 관련이 있음을 알 수 있다. 국내의 경우 상수관망을 비롯한 수도시설물에 대한 관리가 이루어지기 시작한 것은 불과 몇 년 사이의 일이다. 그로 인해 관로의 특성, 관로 교체 여부 또는 누수관련 자료 등 과거 몇십년 전부터 누적된 관의 이력에 대한 자료는 구축이 미비한 실정이고 현재 구축되어 있는 자료도 정확한 정보를 가지고 있지 않아 신뢰성이 떨어진다. 외국의 경우와 비교해 볼 때 국내에선 급배수관망의 누수예측과 관련된 연구가 아직도 체계적으로 진행되지 못하고 있는 실정이다.

상수관망의 유지·관리 방법은 주로 관의 노후와 파손, 부식에 중점을 두고 연구가 이루어지고 있다. 미국의 경우 노후관로의 평가 및 교체를 위해 상수관로의 기능을 평가하는 인자를 설정하고 실험적 연구를 통해 관로의 노후도를 평가하고 관을 효과적으로 개량할 수 있는 프로그램을 개발하였다(Eisenbeis, 1997). 관로의 노후도 평가를 위하여 관의 파손율은 중요한 인자중의 하나이다. 관의 파손특성을 예측하는 수학적 모델은 파손발생률과 환경 및 매설년수의 관계 분석과 파손의 시공간적 군집분석(Clustering)에 의한 예측 등이 있으며 다양한 확률적 기법을 이용한 관로의 파손모델이 연구되어 왔다(Goulter et al., 1993). 지리정보시스템을 이용하여 관로 정보를 관재질, 환경, 관령 등의 여러 범주로 구분한 후 파손확률을 예측하는 기법 또한 연구의 한축을 이루고 있다(Malandain et al., 1998). 국내에서는 점수평가법(Numerical weighting system)을 이용하여 관중, 관경, C 값, 최대수압, 누수 및 파손 경력 등의 인자들을 고려한 수도관 노후화 평가모델을 제시한 바가 있다(최승일, 1995).

관로를 효율적으로 유지, 관리하고 수요량을 관리하기 위해서 누수 문제는 간과할 수 없다. 누수의 방지는 원수의 취수로부터 수용가의 관말까지 용수공급 시스템을 구성하는 모든 요소들을 복합적으로 고려하여야 한다. 종합적 검토를 위해서는 우선 각 구성 요소들의 개별적 분석이 이루어져야 하며 이어서 유지관리 측면의 유량관리 및 누수탐사분석이 요구되고

마지막으로 무수요인 제거를 위한 각종 예방적 방법이 필요하다(한국수자원공사, 1998). 국가지리정보체계(NGIS; National Geographic Information System) 사업의 일환으로 구축된 상수도시설물에 대한 데이터베이스는 단순히 편리한 장표 출력에만 그치는 등 그 활용 역시 미비하다.

이렇게 오늘날 정보시스템에 대한 모순 가운데 하나가 데이터 기술의 발전, 글로벌 정보네트워크의 확산으로 인해 데이터의 양은 크게 증가하였음에도 불구하고 실제 활용 가능한 정보는 그리 많지 않다는 것이다. 그러한 한계를 극복하고자 여러 분야에서 다각적인 노력이 진행되고 있는데 그 중 하나가 데이터 마이닝(Data Mining)이다. 데이터마이닝 기법은 의사결정, 예측, 예보 등 산업 전반에 걸쳐 여러 가지 다양한 상황에 적용될 수 있고 2000년대의 데이터베이스 응용기술로 주목받고 있으며 데이터 웨어하우스(Data Warehouse)와 연계하여 전문가 시스템 개발 지식으로 활용하고자 하는 연구가 한창 진행 중이다(Hadzilacos et al., 2000; Saegrove et al., 1999).

거대한 양의 상수도시설물 데이터베이스 혹은 자료로부터 데이터 마이닝의 개념을 도입하여 관로를 유지·관리하기 위한 확률적 접근에 활용할 수 있다(Hadzilacos et al., 2000). 과거 경험적 방법에 의존하던 관로의 유지·관리 방법에서 관의 파손, 누수 등을 예측함으로써 미리 대처 가능한 적극적인 관로의 유지·관리를 가능하게 할 것이며 지자체의 예산편성에 효과적인 제안을 할 수 있다(Saegrove et al., 1999). 본 연구에서는 상수도시설물에 대한 데이터베이스로부터 누수에 영향을 미칠 수 있는 인자들을 추출, 분석하고 이를 공간적 속성과 결합하여 장래 누수 발생 확률을 예측해 보고자 하였다.

## 2. 연구 방법

### 2.1. 누수예측 체계의 구성

본 연구에서 제안하는 누수예측 체계는 크게 4단계로 구성되어 있다. 1단계에서는 관련 기초조사를 한 후, 2단계에서는 확률적 신경망(Probabilistic Neural Network, PNN) 모델을 구성한다. 그리고 3단계에서는 구축된 모델을 검증하고 4 단계에서는 검증된 모델에 시나리오를 구성하여 적용하였다(Fig. 1).

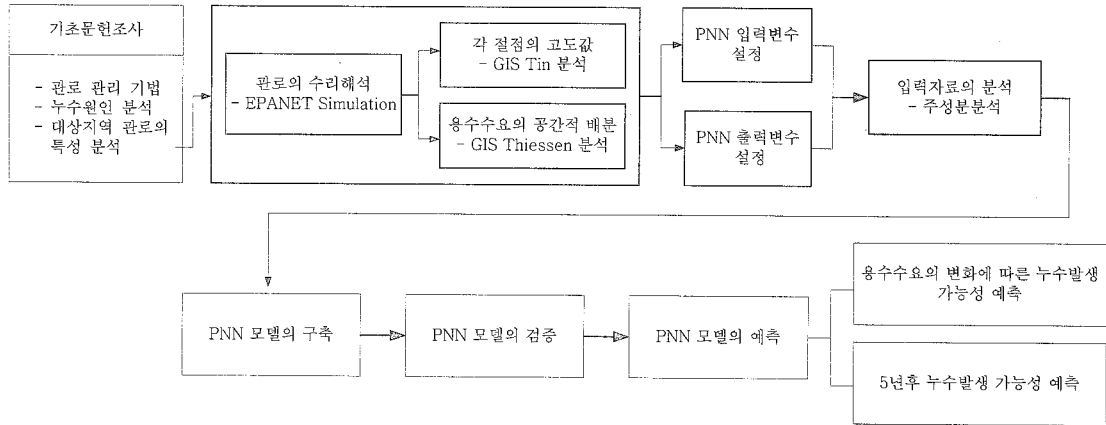


Fig. 1. Diagram of the study flow.

1단계에서 관로의 관리 기법과 누수 발생원인, 대상지역 관로의 특성 분석 등 기초조사를 실시하였다. 대상지역 관로의 특성을 파악하기 위해 관망 레이어에 대해 EPANET을 이용하여 관로해석을 시행한다. EPANET의 입력 자료는 구축된 데이터베이스로부터 GIS 분석을 통해 만들어질 수 있다. GIS의 TIN 분석을 통해 각 절점의 고도값을 구하며 절점의 수요량을 산정하기 위해서는 Thiessen 분석을 시행하여 공간적 배분을 한다. 그 다음 기초조사에서 얻어진 이룬을 바탕으로 누수 발생 가능성을 예측하고자 하는 PNN 모델의 입력변수와 출력변수를 설정한다. 설정된 입·출력변수를 데이터베이스로부터 추출, 분석한 후 주성분분석을 실시하여 모델에 맞는 적절한 형태로 가공, 변환한다. 이렇게 구성된 모델의 검증이 끝나면 설정된 시나리오를 적용하여 연구 대상 관로시스템의 누수 가능성을 예측하게 된다.

## 2.2. 주성분 분석

주성분분석(PCA; Principal Component Analysis)이란 여러 변수가 서로 상관관계를 가지고 있을 때 이들을 설명할 수 있는 구조를 찾아내는 방법이다. 요인분석에서와 마찬가지로 요인이라 할 수 있는 변수들을 찾는다. 각 변수들은 종속변수와 독립변수로 분류되지 않는다. 즉, 이 분석에서는 서로 상관관계를 가지고 있는 변수들을 이용해서 상관관계를 갖지 않는 새로운 변수들을 찾아내는데 목적이 있다. 이러한 변수들을 주성분이라고 한다. 각 주성분분석은 원래 변수들의 선형결합으로 구성된다. 요인분석에서와 마찬가지로

설명력이 가장 높은 식이 주성분식 1로 추출이 되며, 주성분식의 변화가 높아짐에 따라 점차적으로 설명력이 낮아진다. 주성분분석은 요인분석과 마찬가지로 설명력이 높은 서로 상관관계가 없는 주성분분석을 몇 개 찾아서 분석을 하는데 사용할 수도 있다.

## 2.3. 확률신경망

확률신경망은 신경망 구조로 구성된 베이저안(Bayesian) 분류기라 할 수 있다. 이 구조는 입력층(Input layer), 학습층(Pattern layer), 합산층(Summation layer), 출력층(Output layer)의 4개의 layer로 구성되어 있다. PNN의 실행은 검증하기 위한 입력벡터를 학습층에 전달하면 학습층에서 각각의 뉴런을 나타내는 학습자료와 입력벡터와의 거리를 측정 계산하여 측정된 거리를 뉴런의 활성화함수에 종속시킨다. 여기서 단순히 상이한 다변량 정규분포들의 평균으로서 확률분포함수의 추정을 나타내고, 각 클래스에 관한 확률밀도함수를 계산하기 위해 식(1)과 같은 가우스 함수를 사용한다.

$$g(x) = \frac{1}{n} \sum_{i=1}^n \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (1)$$

여기서,  $n$ 은 클래스 내에 있는 자료들의 개수이고,  $x_i$ 는 클래스 내에 있는  $i$ 번째 자료를 나타내며,  $x$ 는 입력이고,  $\sigma$ 는 너비 파라메타를 나타낸다.

계산된 확률밀도함수 값은 합산층에서 각 클래스

Table 1. Number of pipe leak occurrence corresponding to the diameter, pressure and installation age

구분		A동	B동	C동	D동	E동	구성비(%)
매설년수	0-10년	23	39	11	7	7	54.72
	11-20년	7	7	2	5	3	15.09
	21년 이상	3	4	4	0	0	6.92
	결측데이터	2	25	3	6	1	23.27
관경(mm)	75	4	2	0	0	0	3.77
	80	0	4	1	12	2	11.95
	100	13	38	7	2	7	42.14
	150	7	21	5	0	0	20.75
	200	8	5	6	0	1	12.58
	250	1	2	0	0	0	1.89
	300	2	3	1	4	1	6.92
수압(kg/cm <sup>2</sup> )	0-2.5	4	0	0	2	2	5.13
	2.5-5.0	17	17	6	8	4	33.33
	5.0-7.5	8	32	13	4	4	39.10
	7.5-10	6	26	1	2	0	22.44

의 활성함수 값을 평가한 후, 출력층에서 그 활성함수 값이 최대치인 클래스에 입력벡터를 분류한다. PNN은 빠르게 훈련될 수 있고 견고한 이론을 바탕으로 하고 있으며, 신뢰구간을 생성할 수 있는 장점이 있다.

### 3. 자료의 수집

#### 3.1. 연구 대상 지역

본 연구는 일찍이 상수도 전산화 사업이 이루어진 경기도 S시를 대상으로 그 중 구시가지에 해당되는 S구, J구 일대 5개동의 배수관에 대해 실행하였다.

수도관의 형태별 누수 발생 현황을 분석해 보면, 누수는 주로 수용가와 직접 연결되는 급수관에서 대부분 발생하고 있으나, 우리나라는 아직 세부 급수관까지의 자료구축이 미비한 실정이어서 관련 자료의 획득이 어려운 관계로 배수관을 대상으로 자료를 수집하였다. 본 연구지역의 배수관은 관의 재질과 관경을 기준으로 구분되어 1,734개 단위관로(segment)로 구성되어 있다. 배수관은 강관, 주철관, 닥타일주철관, 3중 벽구조 내층격수도관 4가지의 관종으로 구성되어 있는데 이중 주철관과 닥타일주철관이 96% 이상으로 대부분을 차지하고 있다.

#### 3.1.1. 누수 발생 이력

누수관의 이력에 대한 자료는 1993년부터 1999년까지 7년간의 자료이다. 기간 중 누수가 발생했던 경력을 가진 관의 segment는 105개이며 여기서 2번 이상의 중복된 누수 경력을 가진 관의 segment는 51개가 있다. 누수관은 대상지역 전역에 걸쳐 분포하고 있다. 누수발생 자료를 분석해본 결과, 관종별로는 배수관의 대부분을 차지하고 있는 주철관에서의 누수 비중(64.15%)이 상대적으로 높았고, 닥타일관(27.04%), 강관(8.81%) 순이었다.

#### 3.1.2. 매설년수

매설년수별 누수발생 현황을 살펴보면 10년 이하의 관에서 누수발생이 많은 것으로 나타났는데 이는 누수관의 이력 자료가 구축되기 시작한 것이 최근 10년 사이의 일이며 따라서, 과거 몇 십년 전 관의 이력에 대해서 누적된 자료가 아니기 때문으로 분석된다. 또한, 누수의 원인인 부실공사로 인한 결과이거나 누수가 발생한 기록이 있는 관은 또 다시 누수가 발생할 확률이 높기 때문인 것으로 판단된다(Table 1, Fig. 2).

#### 3.1.3. 관경별 누수발생이력

관경별 누수발생 자료에서는 관경이 100mm인 것이 다른 관경보다는 비교적 누수발생 확률이 높은 것

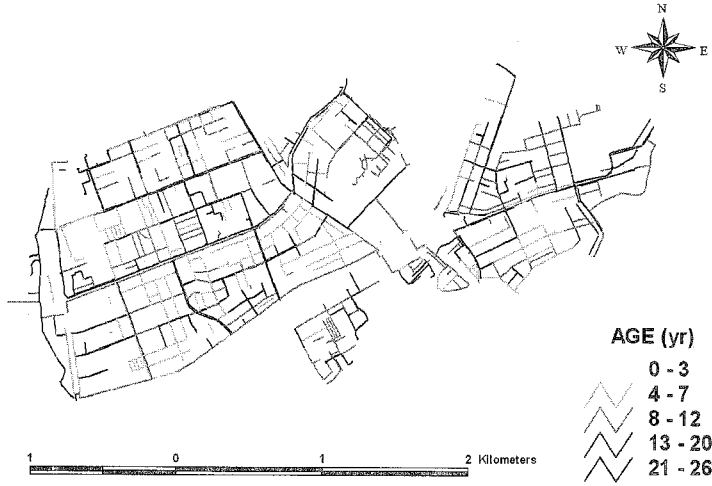


Fig. 2. Pipe age distribution of the study water network.

으로 나타났다. 또한, 전체적으로 볼 때, 관경이 80mm 이상 200mm 이하인 것에서 주로 누수가 발생하는 것으로 조사되었는데, 이와 같은 결과는 분석대상 관로의 관경이 동일 범위에 대부분 분포하기 때문인 것으로 판단된다(Table 1).

### 3.1.4. 수압분포

수압은 자료를 4단계로 범위를 나누어 분석해 보았다. 그 결과, Table 1의 내용을 보면 7.5-10kg/cm<sup>2</sup>의 범위에서보다는 5.0-7.5kg/cm<sup>2</sup>의 범위에서 누수가 많이 발생하는 것을 알 수 있다. 일반적으로 수압이 높으면 높을수록 누수발생 가능성도 높아지는 것으로 알려져 있으나, 대상지역의 특성과 분석 대상을 배수관망으로 제한한 것에 이와 같은 결과가 나타난 것으로 보인다.

## 4. EPANET 모형을 이용한 수리해석

본 연구 대상지의 상수도는 B정수장 계통과 수도권 광역상수도인 S정수장 계통으로 구분되어 있으며 총 16개의 배수지에서 물을 공급하고 있다. 대상지는 배수관망이 복잡하며, 본관과 지관이 중복되어 있고 대용량의 급수체계로 되어 있다. 따라서 모델의 시뮬레이션을 위해서는 기존 관망을 단순화하는 과정이 필요하여 급수관로의 분기점이 있는 배수관망의 절점을 주요 절점으로 선정하였다.

### 4.1. 대상 공간의 Terrain 분석 및 고도정보 추출

EPANET의 입력 자료로 각 절점의 고도값을 산출하기 위해 대상 지역의 지형을 분석하였다. 이를 위해 대상지역 등고자료로부터 TIN을 형성하여 고도값을 추출하였다. 본 연구 대상지는 인구가 밀집된 시가지 지역의 일부로 각 배수지가 위치한 지역을 제외하고는 대부분 고도가 100m 이하로써 비교적 평탄한 지형으로 이루어져 있다.

### 4.2. 용수 수요량의 공간적 배분

모델의 입력자료 중 각 절점의 수요량을 산정하기 위해 대상지역 관망 레이아웃에 대해 GIS 분석을 시행하였다. 대상지역의 관망에 1,073개의 절점을 설정하고, GIS 분석기법 중 Thiessen 분석을 이용하여 설정된 각 절점별로 영역을 할당한 후 그 영역 안에 포함

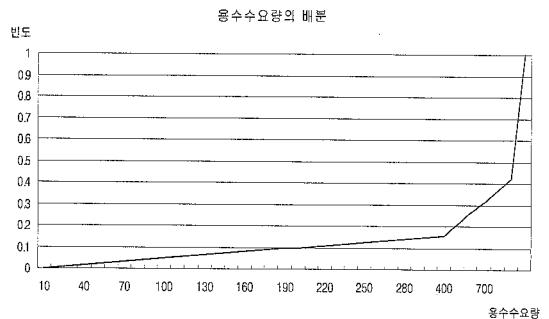


Fig. 3. Frequency plot of the junction demand.

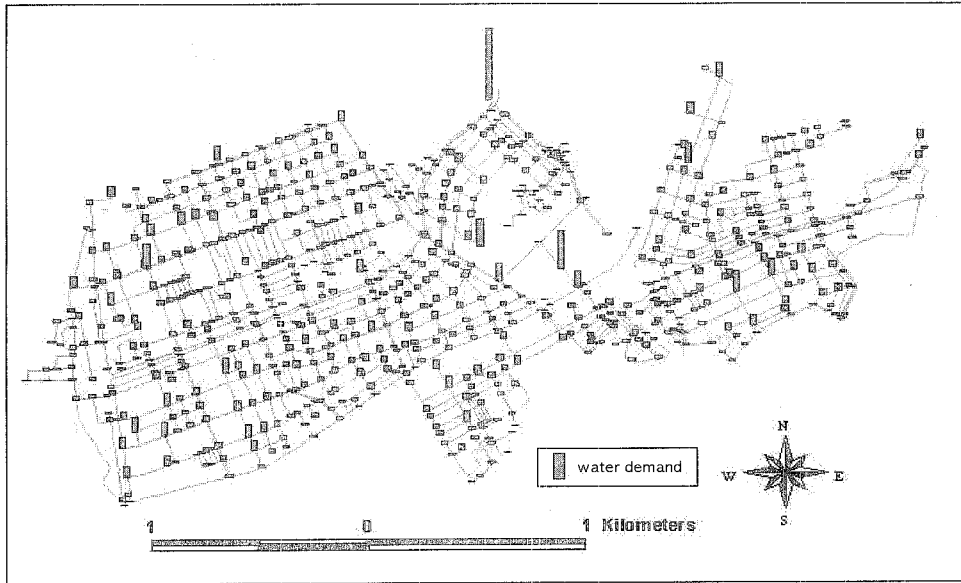


Fig. 4. Spatial distribution of water demand.

된 급수전의 평균 사용량을 집계하여 절점의 수요량으로 산정하였다.

Fig. 3는 할당된 영역의 월평균 사용량을 EPANET의 형식에 맞게 일평균 사용량으로 계산한 후, 배분된 용수 수요량을 그래프로 나타내었다. 배분된 용수 수요량은  $10\text{m}^3/\text{day}$  이하인 것부터  $1,800\text{m}^3/\text{day}$  이상인 것까지 분포하고 있다. Fig. 4은 배분된 용수수요량 자료를 S-PLUS를 이용하여 대상지역의 각 절점별 용수 수요량을 공간상에 출력한 결과이다.

#### 4.3. EPANET을 이용한 수리해석 결과

모델의 시뮬레이션을 통하여 대상지역 관망의 유량, 수압, 속도, 수도손실 등을 계산할 수 있었다. 단, 여기서 시간에 따른 유량의 변동은 고려하지 않았다. 대상지역의 관로 해석 결과를 살펴보면, 수압의 경우 대부분의 관이  $5\text{--}10\text{kg}/\text{cm}^2$  사이에 많이 분포하고 있으며, 절점의 유량은  $200\text{m}^3$  이하에 대부분 분포하고 있다.

### 5. 주성분분석 결과

본 연구에서 관로 누수의 주된 영향인자로 선정된 6개의 성질 즉, 관경, 관중, C값, 도로폭, 매설년수, 수압의 선형결합으로 이루어진 누수발생의 주성분을

도출하기 위하여 주성분 분석을 실시하고 관로 누수 발생 확률 계산을 위한 PNN 모델의 출력 변수를 산출하기 위한 기초 정보로 사용하였다.

#### 5.1. 변수간의 상관성 분석

먼저, 6개의 주된 영향인자(변수)들간의 상관관계 행렬을 산출하였다. 예를 들면, C값과 관중의 상관관계는 0.78로 강한 정(+)의 상관관계를 보이는 반면에 C값과 수압은 0.07로 대단히 약한 상관관계 가졌다. 상관관계행렬의 행렬값(determinant)은 0.135이었다. 결과로부터, 6개 주요 영향인자 상호간의 종속 정도에 관한 정보를 얻을 수 있고, 이 통계량을 이용하여 변수쌍들 간의 상관관계가 다른 변수에 의해 설명되는 정도를 분석하였다.

#### 5.2. 요인수 결정

요인수를 결정하기 위한 측도로는 Kaiser-Meyer-Olkin(KMO) 측도와 Bartlett's 측도를 사용하였다. KMO 측도는 변수쌍들 간의 상관관계가 다른 변수에 의해 잘 설명되는 정도를 나타내는 것이므로 이 측도의 값이 적으면 요인분석을 위한 변수들의 선정이 좋지 못함을 나타낸다. 일반적으로 측도의 값이 0.50 미만이면 받아들일 수 없는 것으로 판단하지만 그대로 사용하기도 한다(정충영 등, 1992). 한편 Bartlett's

**Table 2.** Communalities by altering numbers of principal component

Major factor	Communalities						
	Initial	6 Component	5 Component	4 Component	3 Component	2 Component	1 Component
관경	1.000	.998	.998	.946	.743	.673	3.368E-02
관중	1.000	.973	.973	.748	.724	.708	.708
C값	1.000	.931	.931	.929	.928	.927	.916
도로폭	1.000	1.000	1.000	.966	.691	.633	5.870E-02
수압	1.000	1.000	1.000	.969	.913	9.865E-04	2.557E-04
매설년수	1.000	.982	.982	.782	.640	.640	.632

**Table 3.** Total variance explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.349	39.143	39.143	2.349	39.143	39.143	2.344	39.072	39.072
2	1.234	20.564	59.707	1.234	20.564	59.707	1.234	20.565	59.637
3	1.057	17.624	77.332	1.057	17.624	77.332	1.062	17.694	77.332
4	.701	11.684	89.015						
5	.543	9.053	98.069						
6	.116	1.931	100.000						

Test of Sphericity(BTS)는 요인분석 모형의 적합성 여부를 나타내는 척도로, 본 검토에서는 위의 6개 변수에 대한 분석결과의 검정치인 근사카이제곱의 값이 3461.915이고 자유도가 15, 유의확률이 .000이므로 요인분석의 사용이 적합하며 공통요인이 존재한다고 결론을 내릴 수 있다.

한편, 각 변수의 공통성(Communality)은 추출된 요인에 의해 설명되는 비율을 나타낸다. 예를 들어, Table 2에서 3가지 요인만으로 각 주요 변수의 설명력, 즉 공통분산치를 살펴보면, C값과 같은 경우에는 공통성이 0.928로 이들 3가지 요인으로 92%가 설명되고 있다는 것을 알 수 있다. 공통성이 낮은 변수는 요인분석에서 제외하는 것이 좋다. 일반적으로 0.4이하이면 낮다고 판정한다. 공통성은 추출된 각 요인의 제곱의 합으로 표시된다. 예를 들어, 3개의 요인으로 추출되었을 경우, 다음 식(2)와 같이 계산된다.

$$\text{Communality} = (\text{Factor } 1)^2 + (\text{Factor } 2)^2 + (\text{Factor } 3)^2 \quad (2)$$

### 5.3. 요인 부하량 산출

고유치는 그 요인이 설명하는 분산의 양을 나타내므로 이 값이 큰 요인이 중요한 요인이 된다. Table 3

의 결과로부터, 고유치가 1이상인 3가지 성분(요인)을 추출하였다. 이 3가지 요인이 전체의 77.3%를 설명하고 있다. 물론, 요인1이 가장 높은 설명을 하고 있다. 요인을 더 많이 추출하였다면 그 설명력은 77.3%보다 높아질 것이지만 요인의 수가 많아진다는 단점이 있다. 따라서 요인은 가능한 한 적으면서 설명력을 높이는 것이 중요한 과제가 된다.

### 5.4. 주성분식의 도출

요인점수는 케이스별로 이 행렬에 제시된 계수에 의해 계산된다. 즉, 누수의 주된 6개 영향인자, C값, 관중, 매설년수, 관경, 도로폭, 수압을 이용하여 새로이 도출된 3개의 주성분식 Z<sub>1</sub>, Z<sub>2</sub>, Z<sub>3</sub>는 행렬요소로부터 다음과 같이 정의된다.

$$\begin{aligned} Z_1 &= (.074_{X_1}) + (.364_{X_2}) + (.409_{X_3}) + (.081_{X_4}) + (-.044_{X_5}) + (.336_{X_6}) \\ Z_2 &= (.654_{X_1}) + (-.002_{X_2}) + (.080_{X_3}) + (-.608_{X_4}) + (.006_{X_5}) + (-.077_{X_6}) \quad (3) \\ Z_3 &= (.236_{X_1}) + (-.099_{X_2}) + (.001_{X_3}) + (.252_{X_4}) + (.903_{X_5}) + (.026_{X_6}) \end{aligned}$$

실제 각 케이스에 대한 요인점수는 위의 식에 각 변수의 값을 대입하여 계산한다.

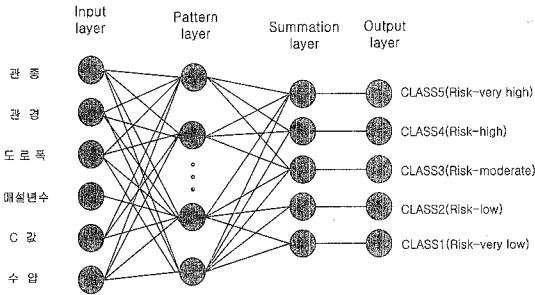


Fig. 5. PNN model structure for prediction of leaking rate.

## 6. 확률 신경망의 적용

누수발생 확률을 예측하기 위해 PNN을 이용한 모델을 구성하였다. 모델의 입력변수는 누수영향 인자로서 관중, 관경, 도로폭, 매설년수, 유속계수(C값), 수압을 설정하였고 모델의 출력변수는 누수가 발생할 확률로 이는 누수발생 위험도 등급에 따라 5개의 클래스로 분류된다. 누수발생 위험도가 가장 낮을 경우 CLASS-1로, 또한 누수발생 위험도가 가장 높을 경우 CLASS-5로 분류하게 된다. Fig. 5는 PNN 모델의 구조를 나타낸 것이다.

### 6.1. 모델 입력변수

모델의 입력은 관중, 관경, 도로, 매설년수, 유속계수, 도로폭, 수압의 6개 인자를 0부터 1사이의 값으로 가중치를 주어 구성하였다.

#### 6.1.1. 관중

입력변수 중 관중은 대상지역 누수발생 경력 관의 관중 분석결과를 토대로 0과 1 사이에서 누수발생 가능성에 따라 가중치를 주어 입력하였다. 3중 벽구조 내충격 수도관은 0, 강관은 0.3, 닥타일주철관은 0.6을, 누수발생 확률이 높은 것으로 나타난 주철관은 1의 값을 입력하였다.

#### 6.1.2. 관경

관경 역시 대상지역의 과거 누수발생 경력 관의 관경 분석 결과를 토대로 누수발생 빈도에 따라 0부터 1사이의 값을 주어 입력하였다. 75mm 관은 0, 250mm는 0.16, 300mm는 0.32를, 80mm는 0.49,

200mm는 0.83, 그리고 100mm의 관은 1의 값을 입력하였다.

#### 6.1.3. 도로폭

대상지역의 도로는 국지도로, 집산도로, 소로, 골목길로 나뉘어져 있다. 이러한 도로의 구분에 따라 각 도로는 다른 비중의 교통량을 지니며, 이는 지하에 매설된 관로에 외부하중으로 작용하게 된다. 따라서, 도로폭이 넓은 것에 높은 점수를 주어 1, 0.66, 0.33, 0으로 변환하여 입력하였다.

#### 6.1.4. 매설년수

매설년수 자료는 연속형 자료이므로 식(4)를 이용하여 0부터 1 사이의 값으로 정규화 하였다.

$$Y_i = \frac{1}{X_{\max} - X_{\min}} (X_i - X_{\min}) \quad (4)$$

여기서,  $Y_i$ 는 정규화된 자료값,  $X_i$ 는 실제 자료값,  $X_{\min}$ 은 실제 자료의 최소값,  $X_{\max}$ 는 실제 자료의 최대값이다.

#### 6.1.5. 유속계수(C)

일반적으로 매설된 관로의 유속계수(C)값은 관 내면의 조도와 관로 중의 굴곡, 분기부 등의 개소 및 통수년수, 라이닝여부와 관경에 따라 다르다. 따라서, 이는 국내·외 시설기준과 문헌을 참고로 하여 설정하였다. 주철관의 예를 들면 Table 4와 같다.

#### 6.1.6. 수압

대상지역 상수관망에 대해 EPANET으로 관로해석을 시행한 후 구해진 수압을 기존 문헌을 참고하여

Table 4. Willams-Hazen coefficient C

재 료	C
보통 콜타르칠한 주철관, 신품	130-135
보통 콜타르칠한 주철관, 통수 5년	120
보통 콜타르칠한 주철관, 통수 10년	110
보통 콜타르칠한 주철관, 통수 15년	105
보통 콜타르칠한 주철관, 통수 20년	95
보통 콜타르칠한 주철관, 통수 30년	85
보통 콜타르칠한 주철관, 통수 40년	80
보통 주철관, 고(very old)	100



Table 5. Example of factor score

관로_ID	FAC1	FAC2	FAC3
67	0.11991	-1.79230	0.27340
260	0.11991	-1.79230	0.27340
270	-0.57925	-0.67763	0.30620
272	-0.37268	-1.16607	0.17084
278	0.11991	-1.79230	0.27340
295	-0.61303	-1.09708	-2.80275
337	-0.61303	-1.09708	-2.80275
407	-0.45742	-2.46942	0.56675
416	-0.58422	-0.11400	1.07561
419	1.04178	0.71129	0.59741
420	2.12908	0.65256	0.64646

Table 6. Example of pipe leaking rate at each group

Group	소속 관로 수(a)	누수발생관 수(b)	7년간 누수 발생건수	누수발생확률 P=(b/a)*100
G15	15	1	1	6.67
G16	4	2	2	50.00
G17	1	0	0	0.00
G18	11	6	6	54.55
G19	14	4	5	28.57
G20	4	0	0	0.00
G21	2	0	0	0.00
G22	7	1	2	14.29

0부터 1 사이의 값으로 가중치를 주어 입력하였다. 수압이 높을수록 높은 점수를 부여하여 3kg/cm<sup>2</sup> 이하는 0, 3-5kg/cm<sup>2</sup>은 0.5, 5kg/cm<sup>2</sup> 이상은 1의 값을 설정하였다.

6.2. 누수발생 확률 산정

입력자료의 주성분분석을 시행하여 계산된 요인점수에 의해 입력자료를 그룹화시켰다. 1,734개 각 관의 segment별 요인점수를 Table 5에 나타내었다.

여기서 3가지 요인의 점수가 같을 경우 하나의 그룹으로 간주하였고 이에 따라 입력자료는 282개의 그룹으로 나뉘어졌다. 1,734개의 자료는 282개의 그룹으로 같은 요인점수에 의해 나뉘어졌으며, 여기에는 7년간의 누수자료에 의해 누수 경력을 가진 관의 자료가 포함되어 있다. Table 6은 각 그룹별 누수발생 확률에 대한 내용을 나타낸 것이다. 이로써 지난 7년 동안 누수 발생 경력이 있는 105개 관의 분포 그룹을 알 수 있다. 이 중 두 번 이상의 누수 경력이 있는 자료까지 포함된 7년간 누수발생 건수는 156건이다.

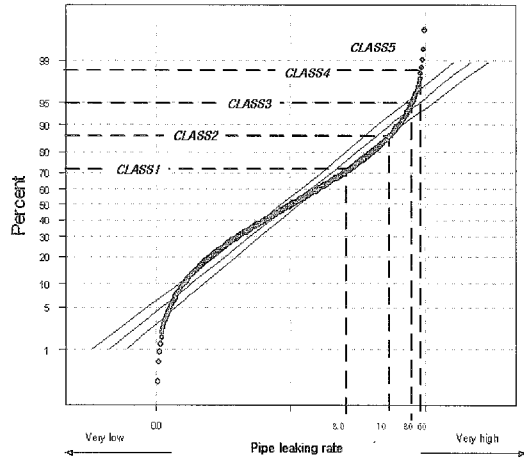


Fig. 6. Composition of class by pipe leaking rate.

따라서, 나뉘어진 각 그룹별로 그룹의 모집단수, 즉 그룹의 소속관수에 대한 누수발생관 수의 비로 누수 발생 확률을 구할 수 있다.

누수발생 확률이 10% 이하인 자료는 전체 1,734개의 관로 중 1,497개로 86%를 차지하고 있으며, 50% 이하인 것은 97%를 나타내고 있다. 배수관에서의 누수발생 확률은 그 분포가 비교적 낮은 편이며 이는 누수가 대부분 급수관에서 많이 발생하고 있기 때문이다.

각 그룹의 확률분포를 누적확률밀도함수로 표현하여 분석해 보았다. Fig. 6의 곡선의 평균( $\mu$ )은 0.5, 표준편차( $\sigma$ )는 0.288165이다. 이 때, 평균  $\mu$ 로부터 각 방향으로 표준편차만큼 떨어진  $\mu - \sigma$ 와  $\mu + \sigma$ 의 구간은 0.683의 확률을 가지고 있었으며,  $\mu - 2\sigma$ 로부터  $\mu + 2\sigma$ 까지의 구간은 0.954의 확률을 가지고 있었다. 그러므로 이 곡선은 정규분포를 따르고 있음을 알 수 있다.

이 누수발생 확률의 확률분포와 자료의 특성을 고려하여 PNN의 최종 출력값으로 나오는 클래스를 결정하였다.

위의 Fig. 6에서 보는 바와 같이, 최종 출력값이 되는 클래스는 5개로 나누었다. 0-3% 사이의 매우 낮은 누수발생 확률을 가진 자료는 CLASS-1로, 그 다음 3-10% 사이의 누수발생 확률을 가진 자료는 CLASS-2로, 10-30% 사이의 누수발생 확률을 가지게 되면 CLASS-3으로, 30-60% 사이의 확률은 CLASS-4로, 그리고 마지막으로 60% 이상의 매우 높

Table 7. Classification accuracy of the PNN model

Classification class True class	1	2	3	4	5	Total	Accuracy(%)
1	471	3	9	3	10	496	94.96
2	0	103	0	0	0	103	100
3	0	0	61	0	0	61	100
4	6	0	0	14	0	20	70.00
5	5	0	0	0	9	14	64.29

은 누수발생 확률을 가지는 자료는 CLASS-5로 분류된다.

### 7. 모델 검증 및 시나리오 분석결과

#### 7.1. 모델 검증

대상지역 1,734개 관의 자료 중 학습자료로 60%를 사용하고 검증자료로 40%를 사용하여 모델을 검증해 보았다. 학습 과정에서 가우스 활성화 함수의  $\sigma$ 값인 너비파라메타는 0.04로 고정되었다. 1,040개의 자료에 대해 입력변수 값과 출력변수 값을 모두 모델에 학습시키고 난 후, 나머지 694개의 자료에 대해서는 입력변수 값만을 넣어 이미 알고 있던 출력변수 값과 모델을 거쳐 분류된 출력변수 값과 비교하여 검증하였다. 모형의 예측 정확도는 94.81%이었다.

검증 결과 Table 7을 보면, CLASS-1, 2, 3은 분류 정확도가 꽤 높은 반면, CLASS-4, 5는 이에 비해 정확도가 다소 떨어지는 것으로 나타났다.

Fig. 7는 위의 검증 결과를 그래프로 나타낸 것이다. 참값(true value) 레이블이 있는 축은 원래 그 자료가 소속되어 있는 클래스를 나타내는 축이며, 분류

클래스(classification class) 레이블이 있는 축은 학습을 통해 분류된 클래스를 나타내는 축이다. 대각선 축을 따라서 나타나는 막대그래프의 높이는 정 분류된 검증치를 나타낸다.

#### 7.2. 시나리오 1

검증된 PNN 모델을 두 가지 시나리오를 구성하여 적용해 보았다. 첫 번째 시나리오는 현 시점에서 5년 후를 기점으로 매설년수와 C값의 변화에 따른 누수 발생 확률을 예측해보았다. 5년이라는 시간의 변화에 따라 입력변수 중 매설년수와 C값이 영향을 받는다. 따라서, 이는 매설년수와 C값의 변화에 따른 예측이라 할 수 있다. PNN 모델로 분류해본 결과, 전체 1,734개의 자료 중 0-3% 사이의 매우 낮은 누수발생 확률을 가지는 CLASS-1로는 1,431개의 자료가 분류되었고, 3-10% 사이의 누수발생 확률을 가지는 CLASS-2로는 124개의 자료가, 그리고 10-30% 사이의 확률로 가지는 CLASS-3으로는 65개의 자료가 분류되었다. 또한, 30-60% 사이의 다소 높은 누수발생 확률 범위를 가지는 CLASS-4로 42개의 자료가 분류되었으며, 60% 이상의 높은 누수발생 확률을 가지는 CLASS-5로는 72개의 자료가 분류되었다(Table 8).

예측 결과 5년 후 14.59%에 해당되는 관들이 지금

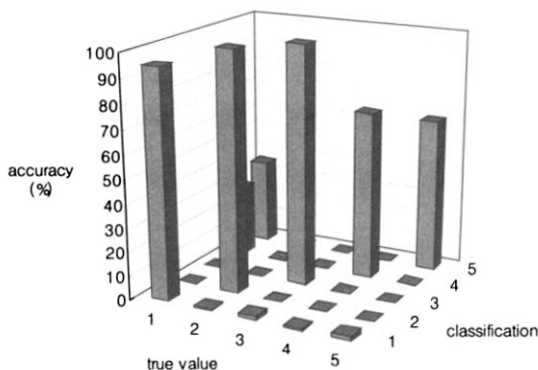


Fig. 7. Classification accuracy of the PNN model.

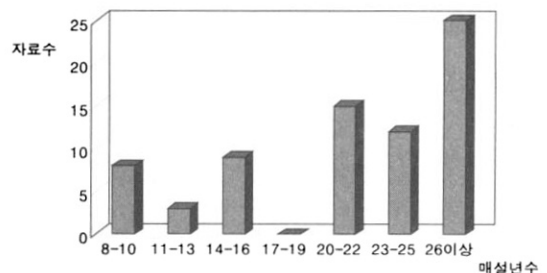


Fig. 8. Distribution of pipe age in CLASS-5 of scenario 1.

**Table 8.** Comparison of pipe leaking risk between present class and prediction class by the scenario

CLASS	Risk	소 속 관로수	구성비(%)	분류된 관로수(구성비%)		
				시나리오 I	시나리오 II	
1	I	very low	1,241	71.56	1,431(82.53)	1,333(76.87)
2	II	low	257	14.82	124(7.15)	180(10.38)
3	III	moderate	151	8.71	65(3.75)	139(8.02)
4	IV	high	49	2.83	42(2.42)	42(2.42)
5	V	very high	36	2.08	72(4.15)	40(2.31)

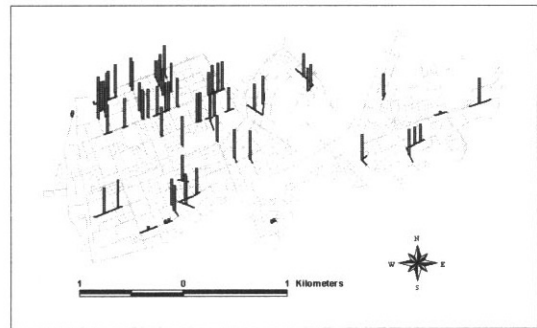
보다 상태가 악화되어 누수가 발생할 위험도가 높아지는 것으로 나타났다. 또한, 전체의 24.62%에 해당되는 관로가 5년 후에는 누수발생 위험도가 낮아지는 것으로 나타났는데, 검증결과에서 보다시피, CLASS-4와 CLASS-5의 오분류율이 다른 클래스에 비해 다소 높다는 점에서 그 원인을 찾을 수 있겠으며, 이는 근원적으로 모델의 구축에 사용된 관로 자료의 정확성 문제에 기인한 것으로 판단된다.

시나리오 I에서 누수가 발생할 위험도가 높은 것으로 분류된 CLASS-5의 자료를 분석해보았다. CLASS-5에 분류된 72개의 자료 중 관경은 100mm인 것이 36개, 150mm인 것은 27개로 두 가지의 경우가 대부분을 차지하는 것으로 나타났고, 관종의 경우에는 주철관이 63개로 87% 이상 차지하고 있었다. 매설년수는 대부분 20년 이상 된 관로가 많은 것을 알 수 있다(Fig. 8).

C 값은 매설년수와 관중에 의해 영향을 받는다. C 값은 매설년수의 증가로 인해 100 이하인 것이 50개로 70% 정도 분포하고 있었다. Fig. 9는 CLASS-5의 C값 분포를 나타낸 것이다. 막대 그래프의 높이가 높은 자료는 대부분 C값이 100이하인 것이다. 도로의 경우, 소로이면서 수압은 5kg/cm<sup>2</sup> 이상인 자료가 높은 비율을 차지하고 있는 것으로 나타났다. CLASS-5로 분류된 자료의 6가지 변수를 분석해본 결과, 대상 지역 관로의 특성 분석에서, 누수 발생 경력을 가진 관로의 성향과 비슷한 것으로 나타났다.

7.3. 시나리오 II

두 번째 시나리오에서는 용수 수요량 증가에 따른 수압 변화시 누수 발생 확률을 예측해보았다. 현재의 용수 수요량 69,801.4톤에서 공급량을 기준으로 하여 118,966.8톤으로 1.7배 증가시켰을 경우 수압변화에 따른 누수 발생 확률을 예측해보았다.



**Fig. 9.** Distribution of C value at CLASS-5.

Table 8에서 보는 바와 같이, PNN 모델로 분류해본 결과, 전체 1,734개의 자료 중 0-3% 사이의 매우 낮은 누수발생 확률을 가진 CLASS-1로는 1,333개의 자료가 분류되었고, 3-10% 사이의 누수발생 확률을 가진 CLASS-2로는 180개의 자료가 분류되었다. 그리고 10-30% 사이의 확률로 분류되는 CLASS-3으로는 139개의 자료가, 30-60% 사이의 다소 높은 누수 발생 범위를 가지는 CLASS-4로 42개의 자료가 분류되었다. 또한, 60% 이상의 높은 누수발생 확률을 가진 CLASS-5로는 40개의 자료가 분류되었다. Table 8의 내용을 분석해 보면, 용수 수요량이 증가되어 수압에 변화가 생겼을 경우, 6.86%에 해당되는 관들이 현재보다 상태가 악화되어 누수가 발생할 위험도가 높아지는 것으로 나타났다. 또한, 전체 자료 중, 11.25%에 해당되는 관로가 용수 수요량 증가에 따른 수압의 변화시 누수발생 위험도가 낮아지는 것으로 나타났다. 이러한 결과 역시, 시나리오 I에서 살펴본 바와 같이, 검증시 발생하는 오분류율과 PNN 모델의 구축에 사용된 관로 자료의 정확성 문제에 기인한 것으로 판단된다.

두 번째 시나리오에 의해 예측해 본 결과에서, 누수발생 위험도가 가장 높은 CLASS-5의 자료로 분류

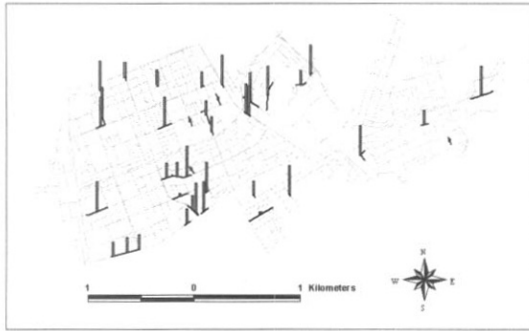


Fig. 10. Distribution of pressure at CLASS-5.

된 40개의 자료를 분석해 보았다. 관경의 경우 역시 100mm의 관이 20개로 가장 많은 부분을 차지하고 있었고, 관종은 주철관인 것이 28개, 그리고 닥타일 주철관인 것이 11개가 분류되었다. C값과 매설년수의 경우에는 특별한 분포를 보이기도는 자료의 성질이 산재되어 있는 것으로 나타났고, 도로는 소로가 26개, 수압은 3-5kg/cm<sup>2</sup>인 것이 18개, 5kg/cm<sup>2</sup> 이상인 것이 16개 차지하고 있었다. Fig. 10는 CLASS-5의 수압분포를 대상지역의 공간상에 나타낸 것이다. 그림에서 검은 선 위로 작은 높이로 나타난 것은 수압이 3kg/cm<sup>2</sup> 이하인 것이고, 가장 높은 막대그래프로 나타난 것은 5kg/cm<sup>2</sup> 이상인 것이다.

## 결 론

본 연구에서는 관로의 유지·관리를 위해 미리 대처가능한 적극적인 방법으로써 확률적 접근을 시도하였다. 이를 위해 데이터 마이닝의 개념을 도입하여 상수도시설물 데이터베이스로부터 누수발생 가능성을 예측하기 위한 목적에 필요한 전략적 정보를 추출하고 변형한 후, 데이터 마이닝의 기법 중, 인공신경망의 일종으로써 확률의 개념이 적용된 PNN을 이용하여 분석하였다. 지금까지의 연구를 통해서 얻어진 결론을 요약하면 다음과 같다.

첫째, 누수발생 가능성을 예측하기 위한 PNN 모델에서 입력변수는 누수 영향 인자로서 관경, 관종, C값, 도로폭, 수압, 매설년수의 6가지를 설정하였고, 출력변수는 누수발생 확률로 그 위험도에 따라 범위를 나누어 CLASS-1부터 CLASS-5까지 설정하였다. PNN 모델을 전체 1,734개의 관로 자료 중, 학습자료

로 60%, 검증자료로 40%를 사용하여 시뮬레이션 해본 결과, 총 94.81%의 정확성을 보여 이를 구성모델로 확정하였다.

둘째, 시나리오 I에서, 5년 후 C값과 매설년수의 변화에 따른 누수발생 확률을 예측해본 결과, 전체 관로의 14.6%가 현재와 비교해 볼 때, 누수가 발생할 위험도가 높아지는 것으로 분석되었다. 또한, 누수발생 위험도가 가장 높은 CLASS-5로는 전체 관로 중 4.2%가 분류되었다. CLASS-5의 관로 속성이 과거 누수가 발생한 이력을 가진 관의 성질과 유사한 것으로 분석되어 이에 대한 집중적인 관리가 필요할 것으로 판단된다.

셋째, 시나리오 II에서, 용수수요량의 증가에 따라 수압이 변화되었을 경우의 누수발생 확률을 예측해본 결과, 전체 관로의 6.9%에 해당하는 관들이 현 상태보다 누수가 발생할 위험도가 높아지는 것으로 분석되었다.

본 연구는 구축·관리 단계에 머물러 있던 상수도 시설물 데이터베이스로부터 데이터 마이닝의 개념을 도입하여 누수 영향 인자들을 추출, 분석하고 누수 예측을 위한 PNN 모델을 구성함으로써 데이터베이스의 활용가치를 높일 수 있었다.

## 사 사

이 논문은 2006학년도 충북대학교 학술연구지원사업의 연구비지원에 의하여 연구되었으며, 연구비 지원기관에 감사드립니다(This work was supported by the research grant of the chungbuk national university in 2006).

## 참고문헌

1. 정충영, 최이규 (1992) SPSSWIN을 이용한 통계분석, 가역경영사.
2. 최승일 (1995) 수도관 노후화 평가모델 개발, 수도, 22(3), pp. 34-55.
3. 한국수자원공사 (1998) 환경기술연구개발사업계획서.
4. Eisenbeis, P. (1997) Estimating the aging of a water mains network with the aid of a record of past failures, Proceedings of the 10th european junior scientist workshop.
5. Goulter, I. et al. (1993) Predicting water-main breakage rates,

- Journal of water resources planning and management*, 119(4), pp. 419-436.
6. Hadzilacos T., D. Kalles, N. Preston, P. Melbourne, L. Camarinopoulos, M. Eimermacher, V. Kallidromitis, S. Frondistou-Yannas and S. Saegrov (2000) UtilNets: a water mains rehabilitation decision-support system, *Computers, Environment and Urban Systems*, 24(3), pp. 215-232.
  7. Malandain, J. et al. (1998) Organizing a decision support system for infrastructure maintenance: application to water supply systems, First international conference for decision making in civil engineering.
  8. Saegrov S. et al. (1999) Rehabilitation of water networks survey of research needs and on-going efforts, *Urban water*, 1(1), pp.15-22.