

# 데이터마이닝 기법을 활용한 상수 이용현황 분석 및 단기 물 수요예측 방법 비교

## The Comparison Among Prediction Methods of Water Demand And Analysis of Data on Water Services Using Data Mining Techniques

안지훈<sup>1</sup> · 김진화<sup>2\*</sup>

서강대학교 경영전문대학원<sup>1</sup>

서강대학교 경영대학<sup>2</sup>

### 요 약

본 연구에서는 상수관망에 설치된 유량, 압력 센서를 통해 취득한 빅데이터에 대해 데이터마이닝 기법을 활용하여 해당 공급권역의 특성을 파악하고 그 정보에 기반하여 상수 공급에 있어서 유의할 점 등을 도출해보고자 하였다. 또한, 상수 사용에 대한 단기 수요예측을 수행하는데 있어서도 통계적 방법인 다중회귀분석과 데이터마이닝의 인공신경망 기법을 비교하여 좀 더 정확한 수요예측을 할 수 있는 모델을 제시해보고자 하였다. 데이터 수집과 테스트를 위하여 지자체 한 군의 소블록 지역을 대상으로 선정하였다. 해당 지역은 가정용 수요 외에도 관공서, 병원 등의 대형 업무용 수요도 일부 존재하고 있는 지역이다. 해당 지역의 센서를 통해 취득되는 연속 발생 데이터를 수집하였다. 이런 방식을 통해 취득된 데이터는 총 2,728건이며 이 중 2,632건은 예측모델을 생성하는데 96건은 예측모델의 예측력을 테스트 하는 데에 활용하였다. 이러한 테스트를 수행한 결과 상수 수요예측에 있어서 인공신경망이 다중회귀분석에 비교하여 더 좋은 예측율을 보였다.

- 중심어 : 수요예측, 상수도 공급예측, 데이터 마이닝, 인공신경망, 다중회귀분석

### Abstract

This study identifies major features in water supply and introduces important factors in water services based on the information from data mining analysis of water quantity and water pressure measured from sensors. It also suggests more accurate methods using multiple regression analysis and neural network in predicting short term prediction of water demand in water service. A small block of a county is selected for the data collection and tests. There is a water demand on business such as public offices and hospital too in this area. Real stream data from sensors in this area is collected. Among 2,728 data sets collected, 2,632 sets are used for modelling and 96 sets are used for testing. The shows that neural network is better than multiple regression analysis in their prediction performance.

- Keyword : Demand Prediction, Prediction of Water Supply, Data Mining, Neural Networks, Multiple Regression Analysis

## I. 서 론

### 2.1 주제 선정배경

상수도 시설은 도시생활을 영위하는 시민들의 기본적인 생활기반 마련을 위한 도시기반 시설로, 근래에 들어서 생활 및 경제 수준 향상으로 물에 대한 양적 및

질적 요구가 높아짐에 따라 그에 대한 효율적인 계획 및 관리가 요구되고 있다. 그러나, 안정적인 물 공급을 하기 위한 상수관망의 유지관리 활동에 있어 현재까지는 주로 비정상적인 이벤트가 발생한 이후에 이를 조치하는 사후 대응에 치우쳐 있고, 조속한 누수지점 검색 및 보수 등의 대응 행위를 하는데 있어 해당 담당자의 경험에 높은 의존도를 보이고 있다. 이에 계획되지

않은 사고발생에 대한 대처를 신속히 하고 상수도 관리조직 내 작업자의 원활한 업무순환 등을 위해 사전인지를 통한 조기대응의 요구가 높아지고 있다[3].

이에 상수도 시설 운영상태의 신속한 감지 및 안정적인 운영, 유지관리를 위해서 현장 시설물에 센서를 부착하여 유량, 압력 등의 실시간 상태정보를 원격 취득하고, 취득한 데이터를 상수관망의 이상현상 감지에 활용하려고 하는 움직임이 국내·외적으로 발생하고 있다. 이러한 센서를 통해 데이터는 몇 초 내지 몇 분간의 간격으로 취득되고 있으며, 이는 몇 년 전부터 부상하고 있는 빅데이터의 형태를 가지고 있다. 이 데이터의 분석을 통해 상수관망의 이상을 감지, 공급 구역의 특성 파악, 물 수요예측 등의 여러 작업을 수행할 수 있다. 예로 이스라엘의 TaKaDu, WhiteWater 등 해외에서는 현장에 설치된 센서 데이터를 축적하고 이 데이터에 대한 분석 작업을 통해 상수관망의 이상을 감지하는 솔루션을 판매하는 사업을 수행하고 있다. 또한, 국내에서도 이와 비슷한 환경의 소프트웨어를 개발하여 각 지자체에 판매하려는 모습을 보이고 있다. 이와 비슷하게 전력분야의 스마트 그리드 사업을 통해 각 가정에서 발생하는 전기 사용에 대해 스마트 플러그인, 스마트 미터 등을 통해 축적된 데이터를 기반으로 각 고객의 사용패턴을 분석하고 전기사용을 절감할 수 있도록 유도하는 정보를 서비스로 제공하는 활동들도 미국, 유럽 등의 해외에서는 많이 수행되고 있다[3, 4, 6].

## 2.2 주요 고려사항

본 작업에 사용될 지자체 D시의 중블록 데이터의 경우 관망의 이상이 발생한 시점의 데이터를 갖지 않고 정상적인 운영 상태에서의 데이터만을 보유하고 있기에 현장에서 취득되는 유량, 압력 정보를 기반으로 그 지역이 어떠한 곳이라는 정보를 도출하는 단계에 머무르고자 한다.

또한, 수요예측 작업을 수행하기 위해서는 우선적으로 데이터의 축적이 중요하다. 특히, 주기성, 계절성 등의 특성을 예측모델에 반영하기 위해 최소 1년의 사용 유량 데이터를 필요로 한다. 그러나, 예측 작업에 있어 사용된 지자체 Y군 내 소블록 지역의 경우 센서 설치가 최근에 이뤄졌기에 2014년 3월 말부터 5월 중순까지 54일간의 데이터를 예측모델 생성에 활용하고 그 이후 데이터를 예측모델 검증에 활용하기로 하였다. 또

한, 예측모델 생성에 사용되는 데이터 기간이 짧으므로 온도, 습도 등의 날씨 정보는 금번 예측 작업에서 제외하기로 하였다.

## II. 공급권역 특성 파악

### 2.1 데이터 취득 방식

지자체 D시의 상업, 업무용 시설이 다수 위치한 중심가 지역에 물을 공급하고 관리하는 중블록 지역을 대상으로 하였다. 2012년 1월부터 4월까지의 기간 중 취득된 데이터에 대해 10분 간격의 단위로 조정된 사용유량, 압력 데이터를 활용하고자 하며, 도출된 데이터 건수는 17,418건이다. 또한, 이와 관련한 외부 환경 변수로 월, 요일 등의 시간 관련한 변수와 동일 시간대의 온도, 습도 등의 날씨 관련 변수를 추가하였다. 날씨 자료의 경우에는 해당 중 블록 지역에 가장 인접한 기상청 관측소의 데이터를 추출하여 센서 데이터와 동일하게 10분 단위의 데이터로 변환하여 활용하였다.

### 2.2 분석변수 선정

본 작업의 Target 변수는 센서에서 취득한 물 사용량(ton, m<sup>3</sup>), 압력(kg/cm<sup>2</sup>)으로 하고, Input 변수는 일반적으로 도시생활에서 물 사용에 영향을 미치는 것으로 판단되는 시간, 날씨와 관련된 정보를 Input 변수로 선정하며 각각에 대한 내용은 다음과 같다.

#### 2.2.1 월(Month)

년간 물 사용량 추이를 봤을 때, 계절에 따라 차이가 발생함을 알 수 있다. 이에 계절성을 반영하기 위해 봄, 여름 등의 사계절 구분은 봄은 3~5월 등의 주관적인 요소가 존재하기에 보다 객관적인 계절성 반영을 위해 월을 추가한다.

#### 2.2.2 요일

주간 물 사용량 추이를 봤을 때, 요일에 따라 차이가 발생하기에 이를 반영하기 위해 요일을 추가한다.

#### 2.2.3 휴일

앞에서 언급한 요일 항목 외에 토요일, 일요일 외의 국경일, 명절 등을 포함한 휴일과 평일 사이에 물 사용

〈표 1〉 수집된 데이터의 형식

시간	이벤트	월	요일	휴일	시간대	유량	압력	강수여부	강수_6H	강수_12H	기온	습도
12-01-01 01:00	0	1	7	1	1	10	3.80	0	-	-	2.30	50.00
12-01-01 01:10	0	1	7	1	1	10	3.70	0	-	-	2.36	50.00
12-01-01 01:20	0	1	7	1	1	11	3.70	0	-	-	2.36	50.00
12-01-01 01:30	0	1	7	1	1	9	3.80	0	-	-	2.36	49.90
12-01-01 01:40	0	1	7	1	1	12	3.70	0	-	-	2.34	50.00
12-01-01 01:50	0	1	7	1	1	10	3.70	0	-	-	2.35	50.20
12-01-01 02:00	0	1	7	1	2	9	3.70	0	-	-	2.39	50.20
12-01-01 02:10	0	1	7	1	2	10	3.80	0	-	-	2.38	50.90
12-01-01 02:20	0	1	7	1	2	11	3.70	0	-	-	2.40	50.80
12-01-01 02:30	0	1	7	1	2	9	3.70	0	-	-	2.34	51.00
12-01-01 02:40	0	1	7	1	2	11	3.70	0	-	-	2.32	51.80
12-01-01 02:50	0	1	7	1	2	9	3.70	0	-	-	2.13	52.80
12-01-01 03:00	0	1	7	1	3	9	3.70	0	-	-	1.90	54.20
12-01-01 03:10	0	1	7	1	3	9	3.70	0	-	-	1.97	54.10
12-01-01 03:20	0	1	7	1	3	9	3.70	0	-	-	2.00	54.00
12-01-01 03:30	0	1	7	1	3	9	3.70	0	-	-	1.88	54.10
12-01-01 03:40	0	1	7	1	3	10	3.70	0	-	-	1.68	55.00
12-01-01 03:50	0	1	7	1	3	8	3.70	0	-	-	1.72	55.00
12-01-01 04:00	0	1	7	1	4	8	3.80	0	-	-	1.64	55.60
12-01-01 04:10	0	1	7	1	4	9	3.80	0	-	-	1.63	56.00
12-01-01 04:20	0	1	7	1	4	8	3.80	0	-	-	1.83	55.00
12-01-01 04:30	0	1	7	1	4	7	3.80	0	-	-	1.97	54.10
12-01-01 04:40	0	1	7	1	4	9	3.80	0	-	-	1.87	55.00
12-01-01 04:50	0	1	7	1	4	17	3.70	0	-	-	2.05	53.90
12-01-01 05:00	0	1	7	1	5	28	3.70	0	-	-	2.00	54.10
12-01-01 05:10	0	1	7	1	5	36	3.70	0	-	-	1.84	55.00
12-01-01 05:20	0	1	7	1	5	28	3.60	0	-	-	1.95	54.50
12-01-01 05:30	0	1	7	1	5	30	3.70	0	-	-	2.09	54.00
12-01-01 05:40	0	1	7	1	5	17	3.70	0	-	-	2.19	53.40
12-01-01 05:50	0	1	7	1	5	6	3.80	0	-	-	2.18	53.70
12-01-01 06:00	0	1	7	1	6	9	3.70	0	-	-	2.19	53.70
12-01-01 06:10	0	1	7	1	6	8	3.80	0	-	-	2.25	53.70
12-01-01 06:20	0	1	7	1	6	7	3.80	0	-	-	2.27	54.00
12-01-01 06:30	0	1	7	1	6	7	3.80	0	-	-	2.27	54.00
12-01-01 06:40	0	1	7	1	6	9	3.80	0	-	-	2.20	54.40
12-01-01 06:50	0	1	7	1	6	8	3.80	0	-	-	2.27	54.10
12-01-01 07:00	0	1	7	1	7	8	3.80	0	-	-	2.38	54.00
12-01-01 07:10	0	1	7	1	7	9	3.90	0	-	-	2.38	54.00
12-01-01 07:20	0	1	7	1	7	17	3.80	0	-	-	2.30	54.00
12-01-01 07:30	0	1	7	1	7	28	3.80	0	-	-	2.34	54.30
12-01-01 07:40	0	1	7	1	7	25	3.70	0	-	-	2.21	56.00
12-01-01 07:50	0	1	7	1	7	10	3.80	0	-	-	2.27	55.40
12-01-01 08:00	0	1	7	1	8	9	3.80	0	-	-	2.34	54.50
12-01-01 08:10	0	1	7	1	8	8	3.80	0	-	-	2.30	54.80
12-01-01 08:20	0	1	7	1	8	9	3.80	0	-	-	2.37	54.50
12-01-01 08:30	0	1	7	1	8	9	3.70	0	-	-	2.39	54.60
12-01-01 08:40	0	1	7	1	8	8	3.70	0	-	-	2.47	54.20
12-01-01 08:50	0	1	7	1	8	9	3.70	0	-	-	2.50	54.20
12-01-01 09:00	0	1	7	1	9	10	3.70	0	-	-	2.61	54.20
12-01-01 09:10	0	1	7	1	9	6	3.70	0	-	-	2.74	54.00

량에 대한 차이가 발생하는 것을 반영하기 위해 휴일 여부 항목을 추가한다.

2.2.4 시간대

1일간 사용현황을 시간단위로 분석하였을 때, 시간대별로 물 사용에 있어 차이가 발생하기에 1시간 단위로 하는 시간대 항목을 추가한다.

2.2.5 온도

물 사용이 계절과 연관성이 있음을 앞에서 언급하였

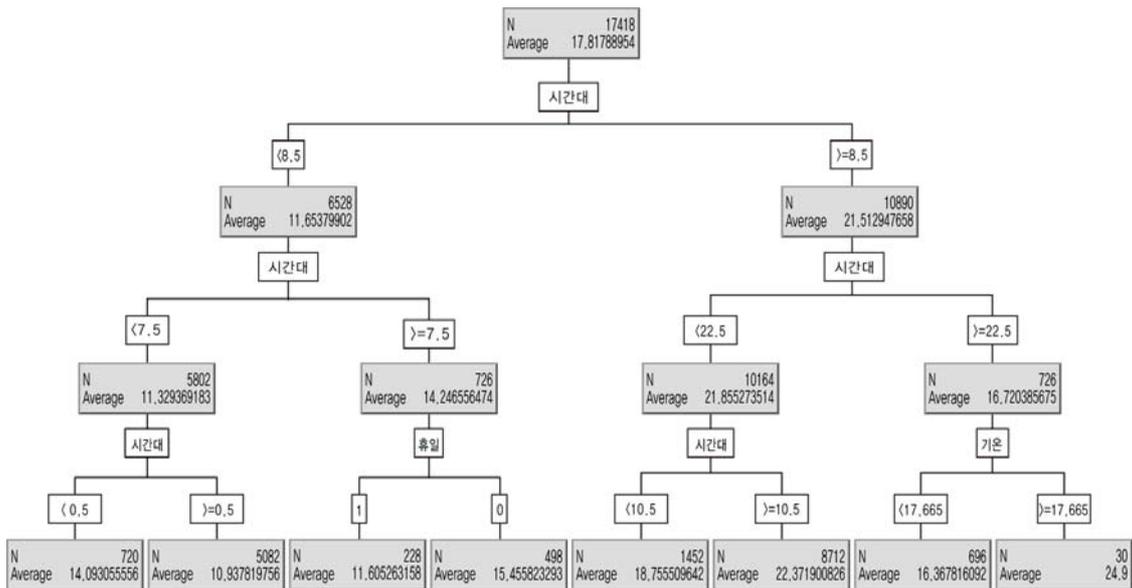
으며, 계절은 온도와 밀접한 연관관계가 있기에 온도를 Input 변수로 추가한다.

2.2.6 습도

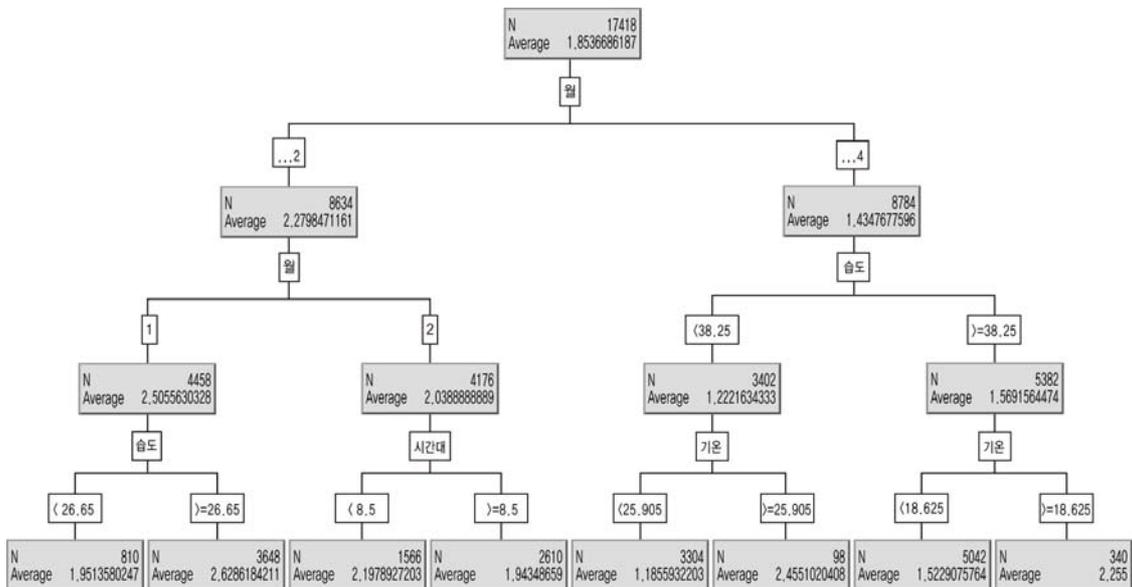
우리나라의 경우 고온 다습한 여름철에 물 사용량이 많은 것으로 미뤄 온도 외 습도의 경우에도 물 사용에 영향을 미치기에 추가한다.

2.2.7 강수여부와 강수량

물 사용에 있어 비가 내리는 경우에도 증가하는 현



<그림 1> Decision Tree 적용 결과-Target 변수 : 사용량



<그림 2> Decision Tree 적용 결과-Target 변수 : 압력

상을 보이는 것으로 파악되어 Input 변수로 추가한다. 강수여부는 해당 시점에 강수량의 존재 여부를, 강수\_6H의 경우 해당 시점 이전 6시간부터 누적된 강수량을, 강수\_12H의 경우 이전 12시간부터 누적된 강수량을 의미한다.

### 2.2.8 물 사용량

압력의 경우 단독으로 발생하기보다 물 사용량에 따라 변화가 있기에 Target 변수를 압력으로 하는 경우 물 사용량을 Input 변수로 추가한다.

## 2.3 적용 방법

주요 고려사항에서 언급한 바와 같이 현재 관망 운영상황에 대한 분석작업을 수행하기에 어떤 환경 하에서 물 사용, 압력이 어떤 모습인지를 우선 파악하고자 한다. 그런 다음 향후 그와 비슷한 상황 하에서 비슷한 수치를 보이는지를 알아보려고 한다. 그러하기에 현재 운영상황이 어떠한 규칙을 갖고 있는지 파악해 보고자 Decision Tree 기법을 활용한다. 분석 도구로는 SAS Enterprise Miner를 활용한다[5, 8, 11, 12, 14, 15].

이 작업에 활용된 데이터 파일의 형식은 앞의 <표 1>과 같다.

## 2.4 결과 및 분석

### 2.4.1 Decision Tree를 활용한 결과

사용량, 압력을 각각 Target 변수로 하여 Decision Tree 기법을 적용한 결과는 앞의 <그림 1>, <그림 2>와 같다.

### 2.4.2 분석 내용

<그림 1>에서 보면, 우선 사용량에 제일 영향을 많이 미치는 변수는 시간대로 오전 8:30 이전과 이후로 분리가 되어 하루 중 사용량이 오전 8:30 이후에 몰려 있음을 알 수 있다. 그리고, 오전 8:30 이후에는 역시 오후 10:30 이전과 이후로 분할되어 오후 10:30 이전에 주로 많이 사용됨을 알 수 있다. 이는 본 작업대상 지역의 특성이 업무, 상업용 밀집 지역임을 고려한다면 직장인들의 근무 시간과 맞물려 집중적으로 물 사용이 발생되고 있음을 알 수 있다. 그리고, 오후 10:30 이후 시간이면서 기온이 높은 시기에 물 사용이 많이 발생

하는 것으로 미루어 기온이 올라 사람들의 야간활동이 늘어나면서 해당 지역 내 호프집 등과 같은 음식점 등의 야간 시간대 영업을 활발해져 물 사용이 늘어나는 것으로 판단된다.

압력을 Target으로 하는 분석결과를 보기 전에 우선 압력의 경우 물 사용이 늘어나면 압력이 저하하고 사용이 줄어들면 증가하는 모습을 띠는 음의 상관관계를 갖고 있다. <그림 2>를 보면 1월과 2월 대비 3월, 4월에 더 낮은 압력분포를 보이고 있음을 알 수 있다. 이는 겨울에서 봄으로 넘어가는 시기에 물 사용량이 늘어나 상대적으로 더 낮은 압력을 가지게 되는 것으로 판단할 수 있다. 이러한 점을 고려하여 해당 중블록에 상수를 공급하는 배수지에서는 해당 지역의 증가한 물 사용으로 압력저하 등에 따른 원활하지 못한 물 사용의 시민 불편을 방지하기 위해 겨울이 지나 봄이 도래하면서 펌프 가동 등의 작업을 통해 물 공급 압력을 높여줄 수 있게끔 시설을 준비하게끔 운영계획을 수립할 수 있다.

## III. 상수 사용에 대한 단기 수요예측

### 3.1 데이터 취득 방식

지자체 Y군의 소블록 지역을 대상으로 하였다. 해당 지역은 가정용 수요처 외 관공서, 병원 등의 대형 업무용 수요처도 일부 존재하고 있는 지역이다. 해당 지역의 센서를 통해 1~3초 간격으로 취득되는 데이터에 대해 15분 간격의 데이터로 1차 조정을 수행하였다. 이런 방식을 통해 취득된 데이터는 총 2,728건이며 이 중 2,632건은 예측모델을 생성하는데 96건은 예측모델에 의한 예측값과 실제값 비교에 활용하였다.

15분 간격의 사용유량 데이터 생성 후 이전 45분, 30분, 15분 데이터와 가중치를 적용한 이동평균값을 현 시점의 사용유량 값으로 정하는 2차 조정작업을 수행하였다. 그 수식은 아래와 같다.

$$F_{t0} = F_{t0} \times 0.3 + F_{t-1} \times 0.3 + F_{t-2} \times 0.2 + F_{t-3} \times 0.2$$

$$F_{t0} : \text{현 시점}, F_{t-1} : 15\text{분전 시점},$$

$$F_{t-2} : 30\text{분전 시점}, F_{t-3} : 45\text{분전 시점}$$

여기에서 이동평균을 적용한 이유는 취득한 센서 데이터 중 센서이상, 통신불량 등 예기치 못한 이상치 데이터가 예측작업에 영향을 미치는 것을 감소하기 위해

서이며, 또한 원래 발생한 15분 단위의 사용 추이와 비교하여 그 패턴이 크게 차이를 벗어나지 않는 수준의 추이를 갖도록 하기 위해 가중치를 부여하였다.

### 3.2 분석변수 선정

본 작업의 Target 변수는 앞으로 사용될 물 사용량 (ton, m<sup>3</sup>)이며, Input 변수는 앞의 고려사항에서 언급한 바와 같이 짧은 학습 데이터 기간으로 날씨에 관련된 변수는 배제하였다. 추가적으로 현재 시점의 상수 사용 패턴과 과거 시점의 사용 패턴간의 일치 관계를 분석한 결과 동일 시간대 2주일 전, 1주일 전의 사용 패턴과 각각 피어슨 상관계수를 산출한 결과 평균 0.5 이상의

상관관계가 있음을 보이기에 Input 변수로 예측하는 시점에서 90분전부터 15분전까지의 사용량 데이터 6개, 1주일 전 예측 시간대 데이터를 포함한 동일 시간대 사용량 데이터 7개와 요일, 휴일 등의 시간과 관련한 데이터를 Input 변수로 선정하며 각각에 대한 내용은 다음과 같다.

#### 3.2.1 월(Month)

년간 물 사용량 추이를 봤을 때, 계절에 따라 차이가 발생함을 알 수 있다. 이에 계절성을 반영하기 위해 봄, 여름 등의 사계절 구분은 봄은 3~5월 등의 주관적인 요소가 존재하기에 보다 객관적인 계절성 반영을 위해 월을 추가한다.

〈표 2〉 분석에 적용된 데이터 파일 형식

MM	WD	OFF	HH	M	W06	W05	W04	W03	W02	W01	W00	T06	T05	T04	T03	T02	T01	OUTPUT
5	4	0	0	0	12,526	11,702	11,128	10,691	10,165	9,334	8,398	13,381	13,130	12,338	11,295	10,223	9,185	8,315
5	4	0	0	15	11,702	11,128	10,691	10,165	9,334	8,398	7,712	13,130	12,338	11,295	10,223	9,185	8,315	7,739
5	4	0	0	30	11,128	10,691	10,165	9,334	8,398	7,712	7,347	12,338	11,295	10,223	9,185	8,315	7,739	7,379
5	4	0	0	45	10,691	10,165	9,334	8,398	7,712	7,347	6,919	11,295	10,223	9,185	8,315	7,739	7,379	6,868
5	4	0	1	0	10,168	9,334	8,398	7,712	7,347	6,919	6,620	10,223	9,185	8,315	7,739	7,379	8,868	6,441
5	4	0	1	15	9,334	8,398	7,712	7,347	6,919	6,620	6,510	9,185	8,315	7,739	7,379	6,868	6,441	6,356
5	4	0	1	30	8,398	7,712	7,347	6,919	6,620	6,510	6,197	8,615	7,739	7,379	6,868	6,441	6,356	6,022
5	4	0	1	45	7,712	7,347	6,919	6,620	6,510	6,197	6,001	7,739	7,379	6,868	6,441	6,356	6,022	5,704
5	4	0	2	0	7,347	6,919	6,620	6,510	6,197	6,001	5,848	7,379	6,868	6,441	6,356	6,022	5,704	5,420
5	4	0	2	15	6,919	6,620	6,510	6,197	6,001	5,848	5,608	6,868	6,441	6,356	6,022	5,704	5,420	5,137
5	4	0	2	30	6,620	6,510	6,197	6,001	5,848	5,608	5,390	6,441	6,356	6,022	5,704	5,420	5,137	5,097
5	4	0	2	45	6,510	6,197	6,001	5,848	5,608	5,390	5,328	6,356	6,022	5,704	5,420	5,137	5,097	5,173
5	4	0	3	0	6,197	6,007	5,848	5,608	5,390	5,328	5,751	6,022	5,704	5,420	5,137	5,097	5,173	5,260
5	4	0	3	15	6,001	5,848	5,608	5,390	5,328	5,751	7,256	5,704	5,420	5,137	5,097	5,173	5,260	5,255
5	4	0	3	30	5,848	5,608	5,390	5,328	5,751	7,256	7,137	5,420	5,137	5,097	5,173	5,260	5,255	5,116
5	4	0	3	45	5,608	5,390	5,328	5,751	7,256	7,137	6,606	5,137	5,097	5,173	5,260	5,255	5,116	4,959
5	4	0	4	0	5,390	5,328	5,751	7,256	7,137	6,606	6,334	5,097	5,173	5,260	5,255	5,116	4,956	4,899
5	4	0	4	15	5,328	5,751	7,256	7,137	6,606	6,334	5,376	5,173	5,260	5,255	5,116	4,956	4,899	4,936
5	4	0	4	30	5,751	7,256	7,137	6,606	6,334	5,376	5,368	5,260	5,255	5,116	4,956	4,899	4,936	4,992
5	4	0	4	45	7,256	7,137	6,606	6,334	5,376	5,368	5,568	5,255	5,116	4,959	4,899	4,936	4,992	5,263
5	4	0	5	0	7,137	6,606	6,334	5,376	5,368	5,568	5,712	5,116	4,956	4,899	4,936	4,992	5,263	6,353
5	4	0	5	15	6,606	6,334	5,376	5,368	5,568	5,712	6,102	4,956	4,899	4,936	4,992	5,263	6,353	7,865
5	4	0	5	30	6,334	5,376	5,368	5,568	5,712	6,102	6,773	4,899	4,936	4,992	5,263	6,353	7,865	8,273
5	4	0	5	45	6,376	5,368	5,568	5,712	6,102	6,773	8,529	4,936	4,992	5,263	6,353	7,865	8,273	8,426
5	4	0	6	0	5,368	5,568	5,712	6,102	6,773	8,529	10,663	4,992	5,263	6,353	7,865	8,273	8,426	8,592
5	4	0	6	15	5,568	5,712	6,102	6,773	8,529	10,663	12,661	5,263	6,353	7,865	8,273	8,426	8,592	8,771
5	4	0	6	30	5,712	6,102	6,773	8,529	10,663	12,661	14,507	6,353	7,865	8,273	8,426	8,592	8,771	10,373
5	4	0	6	45	6,102	6,773	8,529	10,663	12,661	14,507	15,535	7,865	8,273	8,426	8,582	8,771	10,373	11,887
5	4	0	7	0	6,773	8,529	10,663	12,661	14,507	15,535	17,014	8,273	8,426	8,592	8,771	10,373	11,887	13,387
5	4	0	7	15	8,529	10,663	12,661	14,507	15,535	17,014	18,329	8,426	8,592	8,771	10,373	11,887	13,387	14,871
5	4	0	7	30	10,663	12,661	14,507	15,535	17,014	18,329	19,926	8,592	8,771	10,373	11,887	13,387	14,871	15,989

3.2.2 요일

주간 물 사용량 추이를 봤을 때, 요일에 따라 차이가 발생하기에 이를 반영하기 위해 요일을 추가한다.

3.2.3 휴일

앞에서 언급한 요일 항목 외에 토요일, 일요일 외의 국경일, 명절 등을 포함한 휴일과 평일 사이에 물 사용량에 대한 차이가 발생하는 것을 반영하기 위해 휴일 여부 항목을 추가한다.

3.2.4 시간대

1일간 사용현황을 시간단위로 분석하였을 때, 시간 대별로 물 사용에 있어 차이가 발생하기에 1시간 단위로 하는 시간대 항목을 추가한다.

3.2.5 과거 사용량

예측할 사용량에 영향을 미칠 90분 전부터의 사용량 데이터와 사용추이에 영향을 미칠 1주일 전의 동 시간대 사용량 데이터를 Input 변수로 추가한다.

3.3 적용 방법

수요예측 작업을 위해 통계기법으로 많이 활용되는 회귀분석과 데이터마이닝 기법을 활용한 방법으로 인공신경망 기법을 활용하여 비교해보기로 한다. 인공신경망 기법을 활용한 이유로는 기존 수요예측과 관련하여 발표된 논문, 학회지 자료 등을 통해 분석 도구로는 EXCEL에 3rd 제품으로 설치하여 사용 가능한 XL Miner를 활용하였다[1, 2, 7, 9, 10, 16, 17].

이 작업에 활용된 데이터 파일의 형식은 <표 2>와 같다.

3.4 다중회귀분석과 인공신경망 방법의 비교

예측모델은 현 시점에서 향후 6시간 동안의 15분 단위의 데이터를 예측하는 것이며, 현 시점에서 30분 이후 데이터 예측 시 Input 데이터로 활용되는 직전 데이터는 15분 이후 예측작업에서 산출된 예측치를 Input 변수로 활용하는 방식을 사용한다. 동일한 데이터를 사용하여 두 가지 방법으로 예측모델을 생성한 결과 아래 그림과 같이 Validation Data 결과는 다중회귀분석이 인공신경망 보다 좋은 성능을 보이고 있다.

인공신경망을 통한 작업 결과는 다음과 같다:

Training Data scoring-Summary Report

Total sum of squared errors	RMS Error	Average Error
474070827.5	507.451711	25.0356051

Validation Data scoring-Summary Report

Total sum of squared errors	RMS Error	Average Error
256266919.4	569.19101	-93.3208155

다중회귀분석을 통한 작업 결과는 아래와 같다:

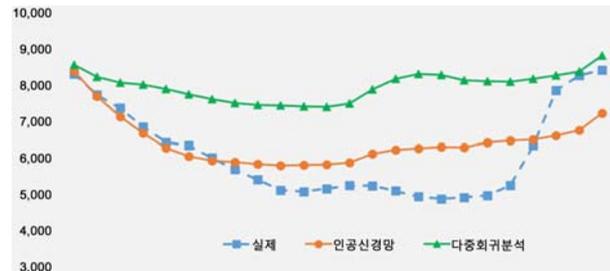
Training Data scoring-Summary Report

Total sum of squared errors	RMS Error	Average Error
555820727.4	549.465558	0.0010453

Validation Data scoring-Summary Report

Total sum of squared errors	RMS Error	Average Error
238649450.7	549.27713	-59.723289

그러나, 각 모델 기반으로 6시간 동안의 데이터를 예측한 결과 다음 <그림 3>과 같이 인공신경망이 다중회귀분석보다 더 좋은 예측율을 보이고 있음을 알 수 있다.



<그림 3> 인공신경망, 다중회귀분석, 실제값의 비교 그래프

IV. 결론 및 향후 계획

Decision Tree 규칙에 따라 사용량은 시간대에 따라 영향을 받는다는 것을 알 수 있었으나, 이는 해당 지역의 특성에 따라 변동이 있을 것이라는 판단이 들며, 또한 1~4월 데이터를 활용함에 따라 온도, 습도 등의 날씨 영향도가 미미한 것으로 판단이 된다. 어쨌든 해당 지역의 경우 원활한 물 공급을 위해서는 직장인들의 주 근무 시간대에 배수장의 운영시설에 문제가 발생하지 않도록 좀 더 시설관리에 주의를 기울일 필요가 있을 것으로 보인다.

〈표 3〉 〈그림 3〉 인공신경망, 다중회귀분석, 실제값의 비교

시간	실제	인공신경망		다중회귀분석	
		신경망	오차율	회귀	오차율
14-05-15 00:00	8,315	8,394	1.0%	8,571	3.1%
14-05-15 00:15	7,739	7,703	0.5%	8,240	6.5%
14-05-15 00:30	7,379	7,140	3.2%	8,083	9.5%
14-05-15 00:45	6,868	6,698	2.5%	8,029	16.9%
14-05-15 01:00	6,441	6,289	2.4%	7,905	22.7%
14-05-15 01:15	6,356	6,065	4.6%	7,759	22.1%
14-05-15 01:30	6,022	5,945	1.3%	7,628	26.7%
14-05-15 01:45	5,704	5,900	3.4%	7,513	31.7%
14-05-15 02:00	5,420	5,846	7.9%	7,461	37.7%
14-05-15 02:15	5,137	5,817	13.2%	7,451	45.0%
14-05-15 02:30	5,097	5,821	14.2%	7,431	45.8%
14-05-15 02:45	5,173	5,841	12.9%	7,417	43.4%
14-05-15 03:00	5,260	5,897	12.1%	7,510	42.8%
14-05-15 03:15	5,255	6,135	16.7%	7,887	50.1%
14-05-15 03:30	5,116	6,240	22.0%	8,191	60.1%
14-05-15 03:45	4,956	6,284	26.8%	8,321	67.9%
14-05-15 04:00	4,899	6,319	29.0%	8,292	69.3%
14-05-15 04:15	4,936	6,300	27.6%	8,138	64.9%
14-05-15 04:30	4,992	6,437	28.9%	8,112	62.5%
14-05-15 04:45	5,263	6,495	23.4%	8,104	54.0%
14-05-15 05:00	6,353	6,525	2.7%	8,178	28.7%
14-05-15 05:15	7,865	6,621	15.8%	8,278	5.3%
14-05-15 05:30	8,273	6,766	18.2%	8,383	1.3%
14-05-15 05:45	8,426	7,241	14.1%	8,825	4.7%

상수 수요예측에 있어서는 인공신경망이 타 기법 대비 좋은 예측율을 보이고 있다. 하루 24시간을 해당 예측모델을 통해 작업한 결과 평균 12.7%의 오차율을 보이고 있으나, 사용량의 변동폭이 큰 경우에는 그 예측율이 현저히 떨어져 예측값이 실제값 대비 최대 약 40% 수준의 오차율을 보이는 경우도 발생하였다. 실제, 수요예측을 수행하는 기존 논문 중에는 예측율을 보이기 위해 변동이 큰 지역에 대해 예측작업 수행을 제외하는 경우도 보여지고 있다. 이에 대한 해결을 위해서는 변동폭이 커지는 시점의 예측방식에 대한 추가 검토가 필요한 상황으로, 인공신경망만을 이용하기 보다 다른 기법과 혼합한 모델을 연구해 보려 한다.

### 참 고 문 헌

[1] 삼성경제연구소, “효과적 수요 예측 방법과 사례”, SERI 이슈 페이퍼, 2012.  
 [2] 송수섭, 이의훈, “인공신경망을 이용한 소비자 선택 예측에 관한 연구”, 한국경영과학회, 제26권, 제4호,

pp.55-70, 2001.  
 [3] 주충남, 구자용, 유명진, “인공신경망을 이용한 단기 물수요 예측”, 대한환경공학회, 추계학술연구발표회 논문집(I), 1999.  
 [4] 최국렬, 데이터마이닝 이론과 실습 : 보건의료데이터 중심, 청구문화사, 2001.  
 [5] 최기선, 유철, 진력민, 유성근, 전명근, “AR모델과 MLP를 이용한 단기 물 수요 예측 알고리즘 개발”, 한국지능시스템학회논문지, 제19권, 제5호, pp.713-719, 2009.  
 [6] 최세일, “의주식시장 기술적 지표 분석”, 진리탐구, 1999.  
 [7] 추휘석, 민지경, 이인호, “다수의 인공신경망 모형을 통한 기업데이터의 분류 및 부도예측에 관한 연구”, 연세경영연구, 제41권, 제2호, pp.513-539, 2004.  
 [8] 한상만, 박승배, 정남호, “인공신경망과 로짓모형을 이용한 내구재의 구매의도 예측에 관한 비교연구”, 아시아마케팅저널, 제15권, 제3호, pp.71-92, 2000.  
 [9] 허명희, 이용구, 데이터마이닝 모델링과 사례, 한나래, 2008.  
 [10] 허 준, 최병주, 정성원, “클레멘타인을 이용한 데이터마이닝”, 1판, SPSS아카데미, 2001.  
 [11] Berry, M. and Linoff, G., *Data Mining Techniques*, Wiley, New York, 1997.  
 [12] Domingos, P. and G. Hulten, *Mining High-Speed Data Streams*, KDD, Boston, MA, USA ACM Press, 2000.  
 [13] Han, J. and M. Kamber, *Data Mining-Concepts and Techniques*, Morgan Kaufmann, 2001.  
 [14] Han, J., J. Pei, and Y. Yin, “Mining frequent patterns without candidate generation”, In *Proceedings of 2000 ACM-SIGMOD International Conference of Management of Data*, 2000.  
 [15] Hulten, G., L. Spencer, and P. Domingos, *Mining Time-Changing Data Streams*, KDD 01 San Francisco, C.A. USA, 2001.  
 [16] Pendharkar, P.C., “A Threshold-varying Artificial Neural Network approach for Classification and its Application to Bankruptcy Prediction Problem”, *Computers and Operations Research*, Vol.32, No.10, pp.2561-2582, 2005.  
 [17] West, D., “Neural Network Credit Scoring Models”, *Computers and Operations Research*, Vol.27, No.11, pp.1131-1152, 2000.

## 저 자 소 개



**안 지 훈(Jihoon Ahn)**

- 2014년: 서강대학교 경영전문대학원 졸업 (석사)
- 현재 : (주)삼천리 에너지종합연구소 근무
- 관심분야 : 에너지 및 자원 컨설팅



**김 진 화(Jinhwa Kim)**

- 2001년 : 위스컨신메디슨 대학 경영학 (박사)
- 1998년~2003년 : 오클라호마 주립대 경영대학 (교수)
- 2003년~현재 : 서강대학교 경영대학 (교수)
- 관심분야 : 데이터마이닝, 미래예측