

선별적 샘플링을 이용한 국내 도시가스 일별 수요예측 절차 개발

이근철¹, 한정희^{2*}

¹건국대학교 경영대학, ²강원대학교 경영학과

Forecasting Daily Demand of Domestic City Gas with Selective Sampling

Geun-Cheol Lee¹, Jung-Hee Han^{2*}

¹College of Business Administration, Konkuk University,

²Department of Business Administration, Kangwon National University

요약 본 연구에서는 국내 도시가스 일일 수요 예측에 대한 문제를 다룬다. 정확한 일일 수요 예측은 안정적인 도시가스의 수급을 위해서 필수적인 사항으로 실제 가스 공급기관의 일상 업무에 해당한다. 본 연구에서는 수요예측 방법을 고안하기 위하여 일일 도시가스 수요 시계열에 대한 데이터 분석을 수행하였으며, 예측일 수요에 영향을 주는 주요한 요인으로 직전일 수요, 기온, 요일 등을 파악하였다. 본 연구에서는 이러한 요인들을 고려한 회귀 모형과 국내 도시가스 수요 특성에 맞는 선별적 샘플링 절차를 제안하였다. 제안 모형과 선별적 샘플링 절차로 구성된 예측 방법의 성능 검증을 위하여 실제 도시가스 수요에 대한 예측을 수행하였다. 문헌에 소개된 기존 방법과 예측 성능을 비교한 결과, 본 연구에서 제안한 방법의 평균절대백분율오차는 약 2.22%로서 개선 비율은 대략 7%에 해당한다.

Abstract In this study, we consider a problem of forecasting daily city gas demand of Korea. Forecasting daily gas demand is a daily routine for gas provider, and gas demand needs to be forecasted accurately in order to guarantee secure gas supply. In this study, we analyze the time series of city gas demand in several ways. Data analysis shows that primary factors affecting the city gas demand include the demand of previous day, temperature, day of week, and so on. Incorporating these factors, we developed a multiple linear regression model. Also, we devised a sampling procedure that selectively collects the past data considering the characteristics of the city gas demand. Test results on real data exhibit that the MAPE (Mean Absolute Percentage Error) obtained by the proposed method is about 2.22%, which amounts to 7% of the relative improvement ratio when compared with the existing method in the literature.

Keywords : City gas, Daily demand, Forecasting, Regression, Selective Sampling

1. 서론

천연가스는 석유, 석탄, 원자력과 더불어 국내 주요 에너지원으로 자리 잡고 있으며 기존 화석연료에 비하여 친환경적이고 사용이 용이하며 안전하다는 장점으로 그 활용도가 증가되는 추세이다. 1980년대 후반 천연가스

가 국내에 처음 도입된 이후 그 수요가 꾸준히 증가하여 2014년에는 전체 에너지원 중 17% 정도의 비중을 차지하는 것으로 나타났다. 이는 석유, 석탄에 이어 세 번째에 해당하는 사용량이다[1, 2, 3].

천연가스는 도시가스, 발전, 지역난방 등으로 그 용도가 나뉘고, 도시가스는 가정용, 산업용, 수송용 등으로

*Corresponding Author : Jung-Hee Han (Kangwon National University)

Tel: +82-33-250-6151 email: jhhan@kangwon.ac.kr

Received July 22, 2015

Revised (1st August 17, 2015, 2nd August 21, 2015)

Accepted October 8, 2015

Published October 31, 2015

다시 분류될 수 있다. 도시가스 사용량은 포화상태에 다른 가정용 도시가스의 보급으로 인해 그 증가세가 둔화되고 있으나, 산업용 및 수송용 수요는 계속 증가하는 추세이다. 국내 천연가스 수요에서 도시가스 수요의 비중은 2013년 기준 49%로 발전(43%)과 지역난방(6%)에 비해 여전히 가장 크다[4].

이와 같이 국가의 주요 에너지원으로 많은 국민이 사용하는 도시가스는 필수재적 성격을 갖고 있는 공공재의 특성이 있다[5]. 따라서, 도시가스를 사용하고 있는 공공의 이익을 극대화하기 위해서는 도시가스에 대한 안정적이고 효율적인 운영이 매우 중요하다고 볼 수 있다. 여러 가지 운영 활동 중, 도시가스는 일별 수요에 맞게 공급이 이루어져야 하기 때문에 안정정인 공급을 달성하기 위해서 정확한 수요의 예측이 중요하고 필수적이다. 수요 예측이 잘못되면, 가스 수급에 영향을 미치게 되거나, 공급 배관 내 압력 불균형을 초래하는 등의 심각한 문제가 발생하게 된다[6].

본 연구에서는 국내 도시가스의 단기 수요, 즉 일별 수요를 예측하는 방법을 제안한다. 국내에서 사용되는 에너지원 수요 예측에 대한 기존의 연구는 전력에 치중되어 있고, 도시가스에 대한 수요예측 연구는 그리 많지 않다. 경제성장률, 인구증가율 등의 거시 경제변수를 이용한 가스 수요 예측에 대한 연구가 수행되었으나, 이는 장기 수요에 대한 예측으로 가스 도입 및 시설 확장 계획의 기본 자료로 이용된다[2, 5, 7]. 본 연구에서 고려하는 단기 가스 수요 예측에 대한 연구와 유사한 연구는 Choi et al.[8]과 Park et al.[9]의 연구가 있다. Choi et al.[8]은 중단기 도시가스 수요 예측을 위한 시계열 모형을 제안하고 있으나, 본 연구와 동일하게 도시가스의 일별 국내 수요 총량을 예측하는 연구는 Park et al.[9]의 연구가 거의 유일하다. Park et al.[9]이 제안한 모형의 상세한 내용은 Jung[10]에서도 확인할 수 있다.

천연가스에 대한 소비 증가는 국내뿐 아니라 전 세계적인 추세로써 관련 수요예측 연구도 많이 수행되어 왔다[11]. Soldo[11]의 문헌조사 연구에서 천연 가스 수요 예측 연구의 역사, 분류, 주요 설명 변수 등을 확인할 수 있다. 몇몇 국가의 천연가스 단기(일일) 수요 예측에 관한 최근 연구를 몇 가지 소개하면 다음과 같다. Azadeh et al.[12]은 적응 네트워크 기반 퍼지 추론 시스템 방법을 제안하여 이란의 일일 가스 수요를 예측하였다. Taspinar et al.[13]은 터키의 일일 천연가스 수요 예측을

위해 다양한 예측 방법을 적용하고 그중 계절형 ARIMA 방법이 가장 좋은 성능을 보이는 것을 확인하였다. 가장 최근에 Zhu et al.[14]은 SVR(Support Vector Regression)에 기반한 예측 방법을 영국의 일일 천연가스 수요 예측에 적용하여 기존 인공신경망과 ARMA모형에 비해 향상된 예측 결과를 얻었다. 특히, 이 연구에서는 FNF(False Neighbours Filtering)라는 잘못된 과거 자료(False Neighbours)를 샘플링에서 제외하는 절차를 포함하여 예측력 개선에 큰 효과를 얻었다.

본 논문에서는 국내 도시가스 단기 수요에 대해 고려하며, 도시가스 일일 수요 특성에 적합한 모형을 제시하고, 모형의 성능을 극대화할 수 있는 샘플링 절차를 통한 예측 방법을 제안한다. 그리고, 제안 방법의 성능을 검증하기 위해서 기존 방법과의 비교 실험 등을 수행한다. 본 논문의 나머지 부분은 다음과 같이 구성되어 있다. 다음 장에서는 국내 도시가스 수요가 어떠한 요인들과 상관관계가 있는지 분석하고, 그 분석을 바탕으로 3장에서는 수요 예측 방법을 제안한다. 4장에서는 기존에 알려진 수요 예측 방법과의 비교 실험을 비롯한 성능 실험과 그 결과를 소개한다. 마지막 장에서는 논문의 결론과 추후 연구 방향에 대해 언급한다.

2. 도시가스 수요 분석

국내 도시가스의 용도는 2013년을 기준으로 가정용과 산업용에 77%가 치중되어 있기 때문에[4], 그 수요의 변동은 일반 국민의 생활 패턴으로부터 많은 영향을 받는다. 예를 들어, 온도가 낮아지는 겨울철에는 난방을 위한 수요가 증가하고, 주말이나 공휴일 등에는 많은 기업이 휴무에 들어가므로 수요가 감소하는 모습을 볼 수 있다. Fig. 1은 2011년부터 2013년까지의 연도별 일일 도시가스 수요를 나타낸 그래프이다 (단, y축 값인 일일 수요는 3년 중 최대 일일 수요값으로 정규화되어 있다. x축 값은 1월1일부터 12월31일까지가 1부터 순서대로 번호 매겨져 있다). 3년동안의 수요 자료를 사용하였지만, 연도별 수요 패턴이 유사하게 나타나기 때문에 연도별 그래프를 중첩하여 나타내었다. 중첩된 그래프를 통해 연중 공통적으로 나타나는 도시가스 수요의 패턴을 어느 정도 파악할 수 있다.

Fig. 1을 통해서 도시가스 수요에 영향을 주는 몇 가

지 주요한 요인을 유추할 수 있다. 첫째, 온도에 따른 변화가 두드러진다. 겨울철에 수요가 가장 많으며, 여름철에 수요가 가장 적다. 온도가 낮아지는 가을철엔 수요가 차츰 증가하고, 온도가 상승하는 봄철엔 수요가 점점 감소한다. 즉, 계절의 온도 변화에 따라 연중 수요도 완만하게 변화하는 것을 볼 수 있다. 둘째로, 연간 수요 패턴 속에 작은 수요 패턴이 존재하는데, 이는 요일 패턴으로 볼 수 있다. 즉, 주말에 수요가 줄었다가 주중에 다시 회복하는 패턴이 반복되고 있다. 온도 요인의 영향이 적은 여름철에 이러한 패턴을 온전히 확인할 수 있는데, 조금 더 자세히 살펴보기 위해서 Fig. 2에 2013년 여름철 몇 주간의 수요를 표시하였다.

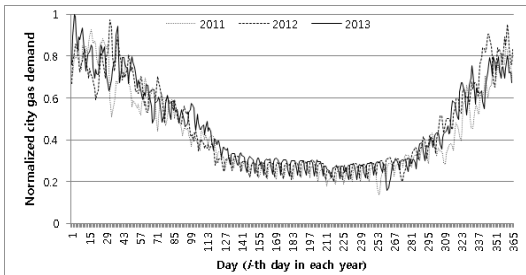


Fig. 1. City gas demand throughout the years

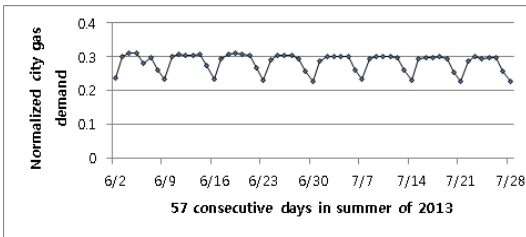


Fig. 2. Demand pattern in summer season

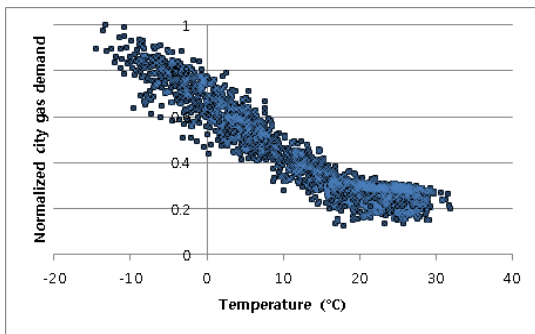


Fig. 3. Scatter chart between city gas demand (y-axis) and temperature (x-axis)

여러 수요 변동 요인과 도시가스 수요량간의 관계를 조금 더 자세히 살펴보자. 먼저, 온도와 도시가스 수요간의 관계를 자세히 파악하기 위하여, Fig. 3에 2009년부터 2013년까지 총 1826일분의 도시가스 수요량과 서울 지역 해당 일 평균온도간의 산포도를 표시하였다(서울지역이 도시가스 수요가 가장 많이 발생함). 여기서도 Fig. 1과 마찬가지로 도시가스 수요량은 정규화된 값을 사용한다. 특정 지역에 대한 일 평균 온도값은 기상청 국가기후데이터센터(<http://sts.kma.go.kr/>)에서 구할 수 있다.

Fig. 3에도 보듯이 온도와 도시가스 수요량과는 전체적으로 음의 상관관계가 있음을 알 수 있다(동일 자료에 대한 상관계수값은 -0.94이다.). 다만, 온도가 약 20도 이상인 경우에는 음의 상관관계가 많이 희석되고 있기 때문에, 계절을 구분하여 도시가스 수요를 예측할 필요가 있다는 것을 알 수 있다.

도시가스 수요에 영향을 주는 추가적인 요인을 자세히 파악하기 위하여 일일 수요와 해당일의 1일전 수요, 2일전 수요, ..., 7일전 수요간의 상관계수 값을 구하고, 그 결과를 Table 1에 정리하였다. 마찬가지로 2009년부터 2013년까지의 도시가스 일일 총 수요 자료를 이용하였다. 예를 들어, 특정일(d day)과 직전일(d-1 day) 수요의 상관계수 값은 2009년 1월 2일부터 2013년 12월 31일까지의 자료와 2009년 1월 1일부터 2013년 12월 30일까지의 자료의 상관계수 값을 의미한다.

Table 1. Correlation coefficient values with demands of the latest days

d-1 day	d-2 day	d-3 day	d-4 day	d-5 day	d-6 day	d-7 day
0.983	0.959	0.944	0.935	0.931	0.936	0.942

Table 1을 보면 도시가스 수요가 온도 보다도 직전일 수요와 더 높은 상관관계를 갖고 있다는 것을 확인할 수 있다. 직전일 뿐만 아니라 2일전, 3일전 수요와도 상당히 높은 상관계수 값을 보이는데, 이는 자기회귀성이 도시가스 수요에 영향을 미치는 또 하나의 요인이라는 것을 설명한다. 직전 일들과의 상관계수 값은 해당일과 멀어질수록 작아지다가 7일전 수요와는 3일전 수요만큼의 높은 상관계수 값을 보이고 있다. 7일전은 해당일과 같은 요일이라는 점에서 요일 요인이 도시가스 수요에 영향을 주는 또 하나의 요인임을 다시 확인할 수 있다.

중장기 수요 예측에는 가격변수와 소득변수를 고려하기도 하지만 본 연구에서 고려하는 일일 수요 예측에 대해서는 일별 소득 및 일별 도시가스 공급 가격 데이터가 존재하지 않기 때문에 해당 변수들은 고려하지 않는다.

3. 제안 방법

이번 장에서는 2장에서 파악된 도시가스 수요 변동 요인을 바탕으로 새로운 회귀 모형과, 도시가스 수요 특성에 맞는 선별적 샘플링 절차를 제안한다. 최종적으로 제안된 모형과 샘플링 절차로 구성된 도시가스 일일 수요 예측 절차를 제시한다.

3.1 제안 모형

본 연구에서는 앞서 데이터 분석을 통해 파악된 온도 요인, 요일 요인, 자기회귀성 등을 독립변수로 하는 다중 회귀모형을 다음과 같이 제안하고자 한다.

$$y_d = \beta_1 y_{d-1} + \beta_2 y_{d-2} + \beta_3 y_{d-7} + \beta_4 t_d + \beta_5 g_d + \beta_6 w_d^{\text{월}} + \beta_7 w_d^{\text{화}} + \beta_8 w_d^{\text{수}} + \beta_9 w_d^{\text{목}} + \beta_{10} w_d^{\text{금}} + \beta_{11} w_d^{\text{토}} + \beta_0 + \epsilon \quad (1)$$

모형에서 사용되는 기호에 대한 설명은 다음과 같다.

- y_d d일의 도시가스 총 수요량
- t_d d일의 외기온도
- g_d d일과 직전일의 온도차, 즉 $t_d - t_{d-1}$
- $w_d^\delta = 1$, 만약 d일이 δ 요일이면; 그렇지 않으면 0
- β_i i번째 독립변수의 회귀계수, $i=1, 2, \dots, 11$
- β_0 회귀모형의 상수
- ϵ 회귀모형의 오차

모형에 대한 부연 설명을 하자면 다음과 같다. 종속변수, y_d 는 특정 일자(d일)의 국내 도시가스 수요의 총량으로, d일이 예측일이라면 d-1일까지의 일별 총 수요량은 알고 있다고 가정한다. 모형에서는 총 11개의 독립변수를 이용하여 종속변수를 설명하고자 한다. 독립변수들은 앞서 데이터 분석에서 확인된 요인들 이외에 추가적인 요인까지 포함하고 있다. 먼저, 데이터 분석에서도 확인

되었듯이 도시가스 일일 수요는 자기회귀성이 매우 크게 나타나기 때문에 직전일의 수요(y_{d-1})뿐 아니라, 직전일 전일의 수요(y_{d-2})도 독립변수에 포함시킨다. y_{d-1} 만 모형에 포함시킬 경우 직전일 수요의 이상 변동에 대해서 예측값이 지나치게 예민하게 반응할 가능성이 있으므로 y_{d-2} 를 y_{d-1} 과 함께 자기회귀요소로 고려함으로써(즉, 직전일 수요에 대한 민감도를 다소 둔화시킴으로써) 수요 예측의 강건성(robustness)을 확보하고자 한다. 또한 예측일과 동일한 요일이면서 가장 최근 수요량인 y_{d-7} 을 독립변수로 고려하여, 자기회귀성과 요일 요인을 동시에 반영하고자 한다.

온도 요인을 반영하기 위하여 예측일의 기온(t_d)을 고려한다. 온도 자료를 수집할 때, d일이 예측일이라면 실제 온도를 알 수 없기 때문에 기상예보를 활용하여야 한다. 이 후 성능실험에서는 과거기간에 대해 소급하여 예측을 실행하므로 일 평균 온도의 실측치 자료를 사용한다. 그리고, 추가적인 온도 요인으로 직전일과의 온도차를 고려한다. 앞서 데이터 분석에서는 다루지 않았지만, 겨울철 온도 급강하 또는 일시적 기온 급상승에 대한 가수요 및 관성 효과를 예상하여 온도차 독립변수, g_d 를 모형에 추가하였다.

요일 요인에 대해서는 요일마다 이진변수를 도입하여 d일이 해당 요일이면 1값을, 그렇지 않으면 0값을 갖도록 하였다. 요일은 일곱 종류이지만 변수는 여섯 개만 사용하여도 충분한 설명이 가능하기 때문에 일요일을 제외한 요일별 이진변수를 추가하였다. 모든 이진 변수가 0인 경우가 예측일이 일요일인 경우이다. 제안 모형에서는 독립변수 y_{d-7} 과 더불어 요일별 이진변수를 통해 요일 요인을 설명한다.

본 연구에서는 제안 모형의 설명력에 대한 객관적인 검증을 위하여 독립변수 일부만을 포함하는 단축 모형들과 결정계수(R2) 값을 비교하고자 한다. 비교 대상으로 사용된 단축모형은 다음과 같다.

$$y_d = \beta_1 y_{d-1} + \beta_2 t_d + \beta_3 w_d^{\text{월}} + \beta_4 w_d^{\text{화}} + \beta_5 w_d^{\text{수}} + \beta_6 w_d^{\text{목}} + \beta_7 w_d^{\text{금}} + \beta_8 w_d^{\text{토}} + \beta_0 + \epsilon \quad (2)$$

$$y_d = \beta_1 y_{d-1} + \beta_0 + \epsilon \quad (3)$$

$$y_d = \beta_1 t_d + \beta_0 + \epsilon \quad (4)$$

Table 2에는 2011년부터 2013년까지 1096일치 도시

가스 일일 수요 자료에 대해 제안 모형과 위에 소개된 단축 모형들을 적용한 경우의 결정계수 값이 정리되어 있다. 표에 보듯이 제안 모형(식 1)은 0.99이상의 아주 높은 결정계수 값을 보인다. 첫 번째 단축 모형(식 2)의 결정계수 값은 제안 모형보다 근소하게 작은 값이다. 독립변수 y_{d-2} , y_{d-7} , g_d 가 제안 모형의 결정계수 값을 작게나마 증가시켜주는 역할을 하고 있음을 확인할 수 있다. 결정계수 값의 근소한 차이가 예측력의 유사함을 의미하지는 않는다. 두 모형(식 1과 2)간 예측 성능의 차이는 4장에서 비교한다. 그리고, 직전일 요인만 고려한 모형(식 3)과 온도 요인만을 고려한 모형(식 4)의 결정계수 값은 상대적으로 매우 작은 값을 갖는다.

Table 2. Determinant coefficient values of the models

Models	Equation	R^2
The proposed model	(1)	0.993
	(2)	0.989
The truncated models	(3)	0.965
	(4)	0.912

Table 3에는 본 절에서 소개된 모든 예측 모형의 회귀계수 추정 결과가 정리되어 있다. 표에서 보듯이 제안 모형에 포함된 모든 회귀계수의 p-value가 0.001미만의 유의한 수준으로 나타났고, 세 가지 단축 모형의 대부분의 변수에 대해서도 유의한 수준으로 나타났다. 따라서 본 연구에서 고려하는 변수들이 수요 예측에 의미 있는 요인이라는 점을 확인할 수 있다.

Table 3. Interferences on regression coefficients

(a) Regression coefficients of model (1)				
(1)	Coefficient values	Standard error	t statistics	p-value
β_0	4501.96	726.33	6.20	<0.001
β_1	1.18	0.02	49.31	<0.001
β_2	-0.30	0.02	-14.37	<0.001
β_3	0.04	0.01	4.33	<0.001
β_4	-180.46	20.28	-8.90	<0.001
β_5	-708.61	31.76	-22.31	<0.001
β_6	8885.54	244.37	36.36	<0.001
β_7	1424.99	335.19	4.25	<0.001
β_8	1845.69	265.56	6.95	<0.001
β_9	2591.79	254.90	10.17	<0.001
β_{10}	1799.75	256.41	7.02	<0.001
β_{11}	-1403.74	243.81	-5.76	<0.001

(b) Regression coefficients of model (2)				
(2)	Coefficient values	Standard error	t statistics	p-value
β_0	13638.67	777.51	17.54	<0.001
β_1	0.78	0.01	77.70	<0.001
β_2	-475.62	21.08	-22.57	<0.001
β_3	9048.80	288.15	31.40	<0.001
β_4	5189.04	288.01	18.02	<0.001
β_5	4275.46	289.26	14.78	<0.001
β_6	4554.57	289.31	15.74	<0.001
β_7	3827.60	289.95	13.20	<0.001
β_8	128.14	289.75	0.44	0.658

(c) Regression coefficients of model (3)				
(3)	Coefficient values	Standard error	t statistics	p-value
β_0	922.0	323.2	2.9	0.004
β_1	0.98	0.0	174.6	<0.001

(d) Regression coefficients of model (4)				
(4)	Coefficient values	Standard error	t statistics	p-value
β_0	77288.0	315.1	245.3	<0.001
β_1	-2025.3	18.9	-107.1	<0.001

3.2 선별적 샘플링 절차

회귀 모형을 이용하여 특정일에 대한 예측을 하기 위해서는 모형에 포함되어 있는 독립변수 값과 회귀계수 값을 미리 알아야 한다. 독립변수 값은 주어진 자료로부터 쉽게 정리할 수 있지만, 회귀계수 값은 샘플링을 통해 수집한 과거자료를 갖고 최소자승법을 이용해 구해야 한다. 이때, 샘플링에서 어떠한 과거 자료를 얼마나 수집하느냐에 따라 회귀계수 값이 달라질 수 있고, 모형의 예측력에도 영향을 줄 수 있다. 본 연구에서는 회귀 모형의 예측력을 향상시킬 수 있는 과거 자료만을 선별적으로 샘플링하는 절차를 제안한다.

일반적으로 회귀모형의 계수 값을 구하기 위해서는 단순히 가장 최근 자료 N개를 샘플링한다. 본 연구에서는 최근 과거 데이터를 무조건 샘플링하지 않고, 최근 과거 데이터 중에서 다음 두 가지 기준에 적합한 것들만 선별적으로 샘플링한다. 첫 번째 기준은 예측일과 되도록 유사한 과거 데이터만 샘플링에 포함시키는 것이고, 두 번째 기준은 이상치가 포함된 과거 데이터는 샘플링에서 제외시키는 것이다. 본 연구에서는 예측일과 과거 데이터의 유사도 기준을 동일한 계절에 속해있는지 여부로 판단하였다. 단, 여기서 계절은, 도시가스 일일 수요 패턴에 근거하여 다음과 같이 정의된 계절을 의미한다: 계절1 = {3월, 4월, 10월, 11월}, 계절2 = {5월, 6월, 7월, 8월, 9월}, 계절3 = {1월, 2월, 12월}. 일반적인 계절

정보와는 다른 이 계절을 이후에는 도시가스 계절이라고 지칭하겠다.

두 번째 샘플링 기준으로, 본 연구에서는 과거 데이터가 특수한 날인 경우, 즉, 공휴일, 명절 등인 경우는 해당 데이터를 샘플링에서 제외하였다. 이러한 특수일에는 신정, 삼일절, 노동절, 어린이날, 석가탄신일, 현충일, 광복절, 개천절, 한글날, 성탄절, 설날연휴, 추석연휴가 포함된다. 과거 데이터의 종속변수 값이 특수일에 해당되는 경우뿐 아니라, 독립변수 값이 특수일에 해당되는 경우도 샘플링에서 제외한다. 특수일 수요는 정상적인 값이 아니라고 판단되어, 샘플링에 포함할 경우 오히려 예측력을 떨어뜨릴 수 있다고 우려되기 때문에 샘플링에서 제외시킨다. 위 두 가지 기준을 이용하여 특정일(d일)의 예측을 위한 샘플링 절차를 다음에 정리한다. 절차에서 d는 예측일을 N은 샘플링 개수를 각각 의미한다.

선별적 샘플링 절차(d, N)

- Step 0. $x:=d-1, n:=0$
- Step 1. 만약 x일이 d일과 동일한 도시가스 계절에 속한다면, Step 2로 이동; 그렇지 않으면 $x:=x-1$, Step 1로 이동.
- Step 2. 만약 x일, x-1일, x-2일, 또는 x-7일이 특수일에 속하면, $x:=x-1$, Step 2로 이동; 그렇지 않으면, x일의 데이터를 샘플링에 포함시키고 Step 3으로 이동.
- Step 3. $n:=n+1$. 만약 $n=N$ 이면, 종료; 그렇지 않으면, Step 1로 이동.

3.3 도시가스 수요 예측 절차

이 절에서는 앞에서 소개한 회귀모형과 선별적 샘플링 절차로 구성된 도시가스 수요 예측 절차에 대해 설명한다.

특정일에 대한 예측을 위해서는 먼저 앞 절에서 소개한 선별적 샘플링 절차를 이용하여 샘플링을 완료한 후, 샘플링된 데이터 셋을 갖고 최소자승법을 적용하면 특정일에 대한 회귀모형의 회귀계수 값을 모두 구할 수 있고, 이렇게 구해진 회귀계수 값을 반영한 회귀모형을 이용하면 특정일에 대한 예측치까지 구할 수 있다. 이와 같은 절차를 여러 날에 대해 반복 적용하면 특정 기간에 대한 일일 수요 예측이 가능하다. 다만, 설날과 추석과 같은 명절 연휴 기간은 극히 이례적인 수요가 발생하는 날이

기 때문에 제안하는 일반적인 회귀모형만으로는 정확한 예측이 어렵다. 기존 에너지 수요예측 관련 연구에서는 명절이나 공휴일만을 대상으로 한 별도로 방법을 제안하기도 하였다 [9, 15, 16].

본 연구는 공휴일이나 명절과 같은 특수일의 수요 예측에 초점을 맞추기보다는 연중 일일 수요를 전반적으로 정확하게 예측할 수 있는 방법을 제안하는 것에 초점을 맞추고 있기 때문에 별도의 명절 수요 예측 방안을 제시하지는 않는다. 대신, 간단한 가중치 적용을 통해 명절 수요 변동에 대응한다. 구체적인 가중치 적용 방법은 다음과 같다. 제안 모형과 샘플링 절차를 통해 구해진 예측치에 대해 해당 예측일이 설날이라면, 0.85를 곱해주고, 설날전후 연휴일이라면 0.9를 곱해준다. 그리고 해당일이 추석이라면 0.8를 곱해주고, 추석전후 연휴일이라면 0.85를 곱해준다. 이와 같은 명절 보정을 포함한 제안 예측 방법을 특정 기간에 대해 적용하는 절차를 아래와 같이 정리할 수 있다.

도시가스 수요 예측 절차(N, 시작일, 종료일)

- Step 0. $x:=\text{시작일}$
- Step 1. 선별적 샘플링 절차(x, N)를 적용하여 데이터 셋을 준비한다.
- Step 2. 준비된 데이터 셋에 대해 최소자승법을 적용하여 제안 모형(식 1)의 회귀계수 값을 구한다.
- Step 3. 구해진 회귀계수 값과 독립변수 값을 제안 모형에 대입하여, x일의 예측치를 구한다.
- Step 4. 만약 x일이 명절 연휴에 속하면, 미리 지정된 가중치를 곱하여 예측치로 사용한다.
- Step 5. 만약 $x=\text{종료일}$ 이면, 종료; 그렇지 않으면, $x:=x+1$, Step 1로 이동.

4. 계산 실험

이번 장에서는 제안된 수요예측 방법의 성능을 평가하기 위해서 먼저 기존에 알려진 수요예측 방법과 동일한 기간에 대한 예측을 통해 성능 비교를 해보고, 최근 기간에 대한 추가적인 실험을 수행한다. 본 연구에서는 특정일에 대한 수요예측 성능을 측정하기 위하여 실측치 대비 절대백분율오차, 즉 APE(Absolute Percentage Error)값을 사용한다.

4.1 벤치마킹 방법

이 절에서는 벤치마킹 방법으로 국내 도시가스 일일 수요를 예측한 Park et al.[9]의 방법을 소개한다.

Park et al.[9]에서는 두 가지 회귀 모형의 가중평균 모형을 제안하고 있다. 첫 번째 모형은 자기회귀 모형으로 직전일 수요만을 독립변수로 사용하고 있으며, 두 번째 모형은 온도 모형으로 예측일의 기온만을 독립변수로 사용하는데, 기온의 2차항과 3차항까지 고려하고 있다. 위 두 모형으로부터 나온 예측치의 가중평균을 통해 최종 수요를 예측한다. Park et al.[9] 방법의 또 다른 특징은 요일별로 개별 모형을 사용한다는 점이다. 즉, 요일별로 각각 7개의 자기회귀 모형과 7개의 온도 모형을 사용한다. 따라서, 특정일의 예측을 위한 샘플링 시 예측일과 같은 요일에 해당하는 과거 자료만을 이용한다.

한편, 본 연구에서 제안한 방법은 2일전, 7일전 수요, 직전일과의 기온차, 2진(binary) 요일 변수를 하나의 모형에 포함시키고, 기온에 대한 2, 3차항은 고려하지 않는다는 차이가 있다. 그리고, 벤치마킹 방법과는 달리 요일 구분 없이 계절을 고려하여 데이터를 선별적으로 샘플링한다는 점 또한 차이점으로 볼 수 있다.

4.2 성능 비교 실험

벤치마킹 방법과의 성능 비교를 위하여 Park et al.[9]에서 예측에 사용한 2011년 11월 1일부터 2012년 10월 31일까지의 도시가스 일일 총량에 대한 예측을 제안 방법으로 수행하였다. Table 4에 벤치마킹 방법의 결과 및 제안 방법의 결과가 정리되어 있다. 벤치마킹 방법의 예측 결과와 더불어 지수평활법(ES, Exponential Smoothing)과 ARIMA모형의 결과도 Park et al.[9]에서 참조되었다. Table 4에는 ‘단순 샘플링 방법(Simple sampling)’의 예측 결과도 정리되어 있는데, 이 방법은 제안 모형을 사용하지만 선별적 샘플링 절차는 사용하지 않고 최근 100개의 자료 전체를 회귀계수 추정에 사용한다.

Table 4에서 보듯이 제안 방법이 여러 APE 측면에서 가장 우수한 성능을 보이고 있다. 먼저, MAPE(mean APE)값, 즉, 예측 기간 366일의 APE값의 평균이 가장 작은 값(2.22%)을 보여 전체적으로 가장 좋은 예측 성능을 보이고 있다. 또한 25%, 50%, 75%, 90% 백분위수에서도 가장 낮은 값을 보여 넓은 범위에서 가장 좋은 성능을 보이고 있음을 알 수 있다. 비록, 최대 오차가 Park

et al.[9]보다 큰 값을 보이나, 이는 추석 명절에 발생한 오차로 공휴일 모형이 별도로 존재하는 기존 연구와 비교한다면 열악한 성능은 아니라고 볼 수 있다.

Table 4. Comparative test results in terms of APEs (Unit: %)

	ES	ARIMA	Park et al.[9]	Simple sampling	The proposed method
mean (MAPE)	7.39	3.95	2.40	2.97	2.22
25 percentile	2.40	1.10	0.85	0.92	0.64
median	5.62	2.58	1.90	2.15	1.59
75 percentile	11.07	5.10	3.36	3.95	2.97
90 percentile	16.30	8.74	4.52	6.16	4.49
max	31.88	40.27	14.33	24.53	17.75

Table 4의 결과에서 주목할 것은 선별적 샘플링 절차의 적용 유무에 따른 성능 차이이다. 동일한 회귀모형을 사용함에도 불구하고 ‘단순 샘플링 방법’은 제안 방법에 비해 성능이 크게 떨어지고 벤치마킹 방법에 비해서도 높은 오차율을 보이고 있다. 회귀모형을 이용한 수요 예측시 수요 특성에 적합한 샘플링 절차를 고안하여 사용하는 것이 반드시 필요하겠다.

4.3 추가 실험

제안 예측 방법의 성능을 자세히 파악하기 위하여 최근 데이터에 대하여 예측 절차를 적용하고 다른 방법들과 비교해 보았다. 사용한 데이터는 2011년 1월 1일부터 2013년 12월 31일까지 1096일치 자료이며, 지수평활법, ‘단순 샘플링 방법’, 3장에서 소개된 ‘모형 (2)를 이용한 방법’, 그리고 제안 예측 방법을 이용하였다. ‘모형 (2)를 이용한 방법’은 단축 모형을 사용하지만, 본 연구에서 제안하는 선별적 샘플링 절차를 이용한 방법이다. 지수평활법의 지수평활 모수는 사전 실험을 통해 0.9로 지정하였다. 각 예측 방법에 대해 연도별 MAPE값이 Table 5에 정리되어 있다.

Table 5. MAPEs of the tested methods without holiday adjustment (Unit: %)

	ES	Simple sampling	The truncated model (2)	The proposed method
2011	7.07	2.97	2.44	2.26
2012	7.08	2.98	2.39	2.26
2013	7.28	2.98	2.43	2.27
Overall	7.14	2.98	2.42	2.26

Table 5에서 보듯이 제안 방법은 모든 연도에 걸쳐 가장 우수한 성능을 보여주고 있다. Table 4에서와 마찬가지로 단순 샘플링 방법과 제안 방법의 예측 결과로부터 선별적 샘플링 절차가 예측 성능에 미치는 영향을 가늠할 수 있다. 동일한 회귀 모형을 사용하여도 선별적 샘플링 절차를 적용하지 않으면 오차율이 약 0.7% 증가한다. 또한 Table 5에서는 단축 모형(식 2)과 제안 모형의 성능 차이도 확인할 수 있다. 앞서 Table 2의 결정계수 값 비교에서는 제안 모형이 0.993, 단축 모형이 0.989로 작은 차이를 보였지만, 실제 예측 성능에서는 제안 모형이 단축 모형에 비해 일관성 있게 우수한 성능을 보이고 있다. 제안 모형에 포함된 독립변수 y_{d-2} , y_{d-7} , y_d 가 의미 있는 역할을 수행하는 것으로 파악할 수 있다.

Table 5에서 ‘모형 (2)를 이용한 방법’과 단순 샘플링 방법을 비교하면, ‘모형 (2)를 이용한 방법’이 더 우수한 성능을 보인다. 단순 샘플링 방법은 제안 모형을 사용하되 선별적 샘플링 절차를 사용하지 않고, ‘모형 (2)를 이용한 방법’은 선별적 샘플링 절차를 사용하되 제안 모형을 사용하지 않는다는 점에서, 선별적 샘플링 절차가 제안 회귀 모형보다 예측 성능 향상에 더 큰 공헌을 하고 있다고 볼 수 있으며, 또한 회귀모형과 샘플링 절차의 적절한 조합을 찾는 것이 중요함을 알 수 있다.

주목 할 것은 Table 5의 결과는 설날 및 추석 연휴 예측에 아무런 보정을 적용하지 않고 각 방법들을 적용한 결과라는 것이다. 앞 장에서 소개된 간단한 명절 보정 절차를 적용한다면 더 개선된 결과를 얻을 수 있다. 실험에 사용된 네 가지 방법을 같은 기간에 대해 명절 보정 절차만 추가하여 예측한 결과를 Table 6에 정리하였다. 이 결과를 Table 5의 결과와 비교하면 모든 경우에 대해 오차율이 감소한 것을 볼 수 있다. 1년 중 단 6일에 대한 보정으로 적지 않은 개선을 이루었다는 점에서 명절 보정 절차의 중요성을 확인할 수 있다. 그리고 명절 보정 절차 적용 여부에 상관없이 제안 방법이 항상 가장 낮은 오차율을 보이는 점에서, 본 연구에서 제안한 예측방법의 우수성을 다시 한 번 확인할 수 있다. 실제로 2011년부터 2013년까지 각각 세 번의 설날과 추석 당일의 수요를 일주일 전 수요와 비교하면, 설날의 경우 2%~42%, 추석의 경우 20%~47% 가량 도시가스 사용량이 감소함을 확인할 수 있다.

Table 6. MAPEs of the tested methods with holiday adjustment (Unit: %)

	ES	Simple sampling	The truncated model (2)	The proposed method
2011	6.93	2.85	2.28	2.17
2012	7.12	2.92	2.28	2.19
2013	7.13	2.89	2.27	2.12
Overall	7.06	2.89	2.28	2.16

5. 결론

본 연구에서는 국내 도시가스 일일 수요의 총량을 예측하는 방법을 제안하였다. 정확한 수요 예측은 안정적인 도시가스 공급을 위한 필수 의사결정이다. 본 연구에서는 데이터 분석을 통해 도시가스 수요 특성에 맞는 회귀 모형과, 선별적 샘플링 절차로 구성된 예측 방법을 제안하였다. 비교실험을 통하여 기존의 도시가스 일일 수요 예측 방법보다 제안 방법의 성능이 우수함을 보였고, 최근 실제 자료에 대한 추가적인 실험을 통해 그 성능을 검증하였다. 실험을 통해 우수한 예측 성능이 제안된 회귀 모형보다 선별적 샘플링 절차에 의한 것이 더 크다는 것을 확인할 수 있었다.

본 연구는 여러 가지 측면에서 확장 연구가 가능하다. 먼저, 실험 결과에도 보듯이 1년 중 최대오차는 주로 설날이나 추석과 같은 명절에 발생하는 것을 알 수 있다. 본 연구에서는 간단한 명절 보정 절차를 이용하였지만, 조금 더 정교한 명절 기간의 예측치 보정 절차를 개발한다면 예측력 향상에 큰 도움이 될 것이라 기대된다. 두 번째로, 샘플링 절차에 관한 추가 연구가 가능하다. 본 연구에서는 동일한 회귀 모형을 사용한다 하더라도 샘플링을 어떻게 하느냐에 따라 그 성능에 차이가 발생함을 확인할 수 있다. 샘플링의 크기를 얼마나 가져갈 것인지, 어떠한 유사일을 샘플링에 포함할 것인지 등에 대한 더 정교한 방법이 고안될 필요성이 있다.

References

- [1] Y. S. Chang, B. H. Kang, "Survey Analysis on Domestic Utilization of Natural Gas", *The 2013 Spring Conference Proceedings of the Society of Air-conditioning and Refrigerating Engineers of Korea*, 352-355, 2013.
- [2] J.-J. Her, H.-J. Lim, "An Analysis of Growth Factors on

- the City-gas Industry by Input-output Structural Decomposition Analysis", *Journal of Energy Engineering*, 21(2), 158-167, 2012.
DOI: <http://dx.doi.org/10.5855/ENERGY.2012.21.2.158>
- [3] KEEI Quarterly Energy Outlook, Korea Energy Economics Institute, 16(3), 2014.
- [4] KESIS, <http://www.ksesis.net/> (accessed July 2015)
- [5] S.-J. Lee, S.-S. Euh, S.-H. Yoo, "Estimation of City Gas Demand Function Using Time Series Data", *Journal of Energy Engineering*, 22(4), 370-375, 2013.
DOI: <http://dx.doi.org/10.5855/ENERGY.2013.22.4.370>
- [6] H.-Y. Oh, "Forecasting of the Short-Term Demand for the Natural Gas Using Time Series Analysis and Artificial Neural Networks", Master Thesis, Graduate School of Management, KAIST, 1997.
- [7] J.-S. Kim, C.-S. Yang, J.-G. Park, "An Empirical Study on the Consumption Function of Korean Natural Gas for City Gas", *Journal of Energy Engineering*, 20(4), 318-329, 2011.
DOI: <http://dx.doi.org/10.5855/ENERGY.2011.20.4.318>
- [8] B. Choi, H. Kang, K.-Y. Lee, S. T. Han, "A Development of Time-series Model for City Gas Demand Forecasting", *Korean Journal of Applied Statistics*, 22(5), 1019-1032, 2009.
DOI: <http://dx.doi.org/10.5351/KJAS.2009.22.5.1019>
- [9] J. S. Park, Y. B. Kim, C. W. Jung, "Short-Term Forecasting of City Gas Daily Demand", *Journal of the Korean Institute of Industrial Engineers*, 39(4), 247-252, 2013.
DOI: <http://dx.doi.org/10.7232/JKIIIE.2013.39.4.247>
- [10] C. W. Jung, "A Study on City Gas Demand Forecasting Based on Daily Characteristics", Master Thesis, Department of Industrial Engineering, Sungkyunkwan University, 2013.
- [11] B. Soldo, "Forecasting Natural Gas Consumption", *Applied Energy*, 92, 26-37, 2012.
DOI: <http://dx.doi.org/10.1016/j.apenergy.2011.11.003>
- [12] A. Azadeh, S.M. Asadzadeh, A. Ghanbari, "An Adaptive Network-based Fuzzy Inference System for Short-term Natural Gas Demand Estimation: Uncertain and Complex Environments", *Energy Policy*, 38, 1529-1536, 2010.
DOI: <http://dx.doi.org/10.1016/j.enpol.2009.11.036>
- [13] F. Taspinar, N. Celebi, N. Tutkun, "Forecasting of Daily Natural Gas Consumption on Regional Basis in Turkey Using Various Computational Methods", *Energy and Building*, 56, 23-31, 2013.
DOI: <http://dx.doi.org/10.1016/j.enbuild.2012.10.023>
- [14] L. Zhu, M. S. Li, Q. H. Wu, L. Jiang, "Short-term Natural Gas Demand Prediction Based on Support Vector Regression with False Neighbors Filtered", *Energy*, 80, 428-436, 2015.
DOI: <http://dx.doi.org/10.1016/j.energy.2014.11.083>
- [15] O.-S. Kwon, K.-B. Song, Development of Short-Term Load Forecasting Method by Analysis of Load Characteristics during Chuseok Holiday", *The Transactions of The Korean Institute of Electrical Engineers*, 60(12), 2215-2220, 2011.
DOI: <http://dx.doi.org/10.5370/KIEE.2011.60.12.2215>
- [16] K.-B. Song, J.-H. Lim, "Short-Term Load Forecasting for the Consecutive Holidays Considering Businesses' Operation Rates of Industries", *The Transactions of The Korean Institute of Electrical Engineers*, 62(12), 1657-1660, 2013.
DOI: <http://dx.doi.org/10.5370/KIEE.2013.62.12.1657>

이 근 철(Geun-Cheol Lee)

[정회원]



- 1997년 2월 : 한국과학기술원 산업공학과 (공학석사)
- 2003년 8월 : 한국과학기술원 산업공학과 (공학박사)
- 2003년 11월 ~ 2006년 2월 : UC Davis 박사후연구원
- 2006년 3월 ~ 현재 : 건국대학교 경영대학 교수

<관심분야>

제조시스템 운영 및 설계, 에너지 수요예측

한 정 희(Jung-Hee Han)

[정회원]



- 1996년 2월 : 고려대학교 산업공학과 (공학석사)
- 1999년 8월 : 고려대학교 산업공학과 (공학박사)
- 2000년 3월 ~ 2004년 2월 : LG전자 이동통신연구소 선임연구원
- 2004년 3월 ~ 현재 : 강원대학교 경영학과 교수

<관심분야>

최적화, 응용통계