A review of artificial intelligence based demand forecasting techniques

Hyerin Jeong[a] · Changwon Lim[a,1]

[a]Department of Applied Statistics, Chung-Ang University

Abstract

In recent years, 'big data' has been generated in various fields. Many companies have tried to make profits by building a system capable of analyzing big data based on artificial intelligence (AI) techniques. By integrating AI technology, analyzing and utilizing vast amounts of data are increasingly valuable. In particular, demand forecasting with maximum accuracy is absolutely critical to national and business management in various fields such as finance, procurement, production and marketing. In this case, it is important to apply an appropriate model considering the demand pattern for each field. It is possible to analyze complex patterns of real data, which is also enlarged by a traditional time series model or regression model.
However, choosing the right model among the various models is difficult without prior knowledge. Many studies based on AI techniques such as machine learning and deep learning have proven that they can overcome these problems. In addition, demand forecasting through the analysis of stereotyped data and unstructured data of images or texts has shown high accuracy. In this paper, we introduce important areas where demand forecasts are relatively active. And we introduce machine learning and deep learning techniques considering the characteristics of each field.

Keywords: big data, artificial intelligence, demand forecasting, machine learning, deep learning

1.서론

인간의 삶에 예측이란 불가피하며 이를 바탕으로 대부분의 의사결정이 이루어진다. 또한 예측은 개인의 사사로운 결정 뿐 아니라 가구나 기업의 수 전력, 교통수단, 관광객, 기존 제품 및 신제품의 수요 등을 결정하는데 중요한 역할을 한다. 이 때 수요예측에서 발생되는 오차는 국가의 성장이나 기업의 매출에 아주 큰 영향을 미치기 때문에 이를 최소화해야하며 최대의 정확도를 갖췄을 때 잘 된 수요예측이라 할 수 있다.
전통적으로 쓰이고 있는 수요예측기법에는 시계열분석법과 회귀분석법이 있다. 시계열분석법은 일정한 시간 간격을 두고 기록된 데이터를 바탕으로 미래를 예측하며 대표적으로 Autoregressive(AR), Moving Average(MA), Autoregressive Integrated Moving Average(ARIMA), Exponential Smoothing(지수평활법) 등이 있다. 회귀분석법은 단일 또는 다수의 독립변수들과 종속변수 간의 관계를 찾기 위해 관계식을 추정하고, 추정된 관계식으로 독립변수들을 통해 미래를 예측하는 방법이다. 또 수요가 거의 없거나 전혀 없는 새로운 부품에 대한 보정된 예측을 하거나 간헐적인 수요패턴을 보이는 제품의 재고관리나 생산량에 대한 예측에는 베이지안 모델을 적용하기도 한다 (Bergman 등, 2017). 그러나 최근에는 인공지능 기반의 수요예측기법을 활발히 적용하고 있는 추세이다. 구글, 패션브랜드, 카드사 등 다양한 업체에서는 이미 인공신경망(Aritificial Neural Network)을 활용한 수요예측기법을 통해 다양한 사업 및 서비스를 제공하고 있다. 한 예로, 해외 패션 브랜드인 Zara에서는 빅데이터를 분석하여 수요예측 및 매장별 재고량을 파악하여 상품별 가격결정, 운송과정까지 실시간으로 파악할 수 있는 재고관리 시스템을 개발하여 공급망을

This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT (NRF-2017M3C4A7083281).
[1]Corresponding Author: Department of Applied Statistics, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea. E-mail: clim@cau.ac.kr

개선하고 있다 (Nenni 등, 2013). 또한 국내 기업의 한 예로 SK텔레콤은 SNS, 소셜데이터, 검색키워드를 바탕으로 고객, 사회 트렌드를 파악하여 자체적으로 개발한 빅데이터 기반의 서비스 모델을 기업이미지나 홍보에 적용하고 지속적으로 개발하고 있다. 또 주요 카드 업체들은 소비자의 행동패턴을 분석하여 마케팅이나 신규상품 개발 및 상품추천에 활용하고 있다. KB국민카드의 경우, 실시간 마케팅 시스템의 개발로 고객의 현 상황에 가장 최적인 조건과 혜택을 제공하고 있다. 이는 각 고객들에게 맞춤화된 최적의 서비스를 제공할 수 있으며 전략적인 마케팅 활동이 가능하다는 장점이 있다 (김정경, 2016).
또 다른 실제 사례로 김인중 등 (2017)은 T-커머스를 통한 시간별 매출 예측을 위해 Deep Neural Network(DNN)와 통계모델을 결합한 연구를 하였다. 연구에 쓰인 데이터는 실제로 T-커머스를 운영하는 ㈜더블유쇼핑의 2015년 7월부터 2016년 9월까지의 판매 기록이다. 입력 변수로 쓰인 요소는 정량화가 용이한 상품의 가격, 날짜, 요일, 휴일 여부, 날씨이다. 그러나 결과적으로 시간대별 트렌드 분석을 통한 매출 예측에는 최근 3개월간의 데이터로 학습한 경우가 더 낮은 Normalizaed Mean Absolute Error(NMAE)를 보였음을 알 수 있었다. 해당 연구에서 제안하는 매출 예측 과정의 특징은 T-커머스의 각 상품은 모든 시간대에 고르게 판매되지 않으므로 시간대별로 집계할 경우 데이터의 편중이 심하기 때문에 희박성의 문제를 고려한 것이다. 이를 해결하기 위해 가우시안 평활화를 진행한 통계적 모델과 SVD를 적용하였으며 DNN과 함께 사용해서 희박성 문제와 예측 성능 문제를 동시에 고려할 수 있음을 알 수 있었다.
이종민 등 (2017)은 텍스트마이닝을 활용하여 전세가격 예측에 관한 연구를 하였다. 전국을 대상으로 하는 국내 주요 언론사 16개의 뉴스 기사에서 '전세'라는 키워드가 포함되어있으며 중복뉴스를 제외한 총 42,474개를 한국언론진흥재단의 BIGKINDS로 수집하였다. 전세가격지수로는 KB국민은행에서 제공하는 전국 월간 전세가격지수를 사용하였다. 수집된 데이터의 기간은 2012년 1월부터 2016년 12월까지 총 60개월로 분석 기간과 동일하였다. LDA알고리즘을 이용해서 총 8개의 토픽으로 이루어진 토픽모형을 얻고 통계 프로그램 R의 'Sejong' Package로 명사만을 추출한 후 전체 기사에 대한 각 토픽의 비중과 해당 토픽에 포함된 상위 10개의 명사들을 알 수 있었다. 토픽의 비중이 전세가격 예측력을 향상시키는데 유의미하다는 것을 그랜저 인과관계 검정법의 결과를 통해 밝혔다. 그랜저 인과관계 검정법이란 시차분포모형을 이용해서 설명변수와 종속변수의 구분을 명확하게 하기위해 쓰이는 방법이다. 그리고 예측력 향상에 관한 분석을 위해 전세가격지수 데이터를 활용한 ARIMA 기본 모형과 이 기본모형과 토픽 비중을 추가한 토픽 포함 모형 간의 RMSE와 MAE를 비교하였다. 그 결과 '전세' 뉴스의 토픽 비중을 활용한 토픽 포함 모형으로 전세가격 예측을 했을 시 좀 더 정교한 예측이 가능하다는 것을 보였다.
인공지능기반 수요예측기법의 가장 큰 장점은 예측의 정확도와 계산량의 효율성이라 할 수 있다. 기존 수요예측기법에 비해 상대적으로 높은 정확도를 보이며 예측실행시간도 덜 소요되기 때문이다. 또 독립변수와 종속변수간의 관계를 나타내는 모델을 설정할 필요 없이 알고리즘을 구축할 수 있다. 회귀분석의 경우, 실제로 Parry 등 (2011)은 확률론적 신경망, Chu와 Cao (2011)는 동적 큐빅(Dynamic Cubic) 신경망 같은 새로운 기법이 기존 기법보다 더 나은 예측력을 지닌다는 것을 보였다.
본 논문에서는 수요예측의 대표적인 분야들과 각 분야의 특징에 맞는 인공지능기반 수요예측모델들을 소개하고자 한다. 기존 리뷰논문들과 차별을 둔 점은 첫째, 수요예측의 다양한 적용분야와 각 분야에 맞는 다양한 수요예측모델들을 인공지능 기반 기법을 중심으로 소개한 것이다. 둘째, 전통적인 시계열 모델, 회귀 모델이나 간단한 신경망 네트워크 구조의 모델이 아닌, 기존 모델과 인공지능 기반 기법을 결합시켜 더욱 심화된 하이브리드 모델들을 추가해서 정리하고 언급한 것이다. 셋째, 단순히 기존 연구의 나열이 아닌, 수식전개를 통한 네트워크 구조의 흐름을 정리하여 통계학적인 특징으로 접근한 것이다.
이어서 2장에서 수요예측모델의 다양한 적용분야를 소개한다. 그리고 3장에서는 머신러닝 기반의 수요예측 기법으로 Support Vector Regression(SVR), 부스팅(Boosting), 가우시안 프로세스(Gaussian Process, GP)를 소개한다. 4장에서는 딥러닝 기반의 수요예측 기법으로 Recurrent Neural Networks(RNN) 및 Long Short-Term Memory models(LSTM), Restricted Boltzmann Machine(RBM)들을 기반으로 변형 및 발전된 다양한 기법들을 소개하고 마지막으로 5장에서 수요예측모델 리뷰의 결론 및 논의점을 언급하고자 한다.

2.수요예측모델의 적용 분야

본 장에서는 수요예측모델이 적용되는 다양한 분야들을 소개하고자 한다. 분야별로 데이터가 갖는 특징이 다르기 때문에 이를 고려해서 알맞은 수요예측 모델을 적용하는 것이 중요하다. 수많은 분야들 중 특히 수

요예측이 중요한 도시용수 분야, 전력 및 에너지 자원 분야, 현금수요 분야, Ride-services 분야, 관광분야, 단기예측분야 그리고 예비부품 및 신제품 분야로 구분하였다.

2.1.도시 용수 분야(Urban Water Demand)

물은 사람들의 삶의 질을 보장하고 유지하기 때문에 보존해야하는 천연자원이다. 인구와 산업이 증가함에 따라 알맞은 물 공급은 도시개발에 있어 중요한 요소로 자리 잡고 있으므로 도시 용수 수요예측은 높은 신뢰도를 갖추고 효율적으로 이루어져야한다. 식수 데이터는 시간에 의존하므로 시계열 데이터로 분류할 수 있다. 예를 들어, 도시의 식수 데이터의 특징은 낮에는 수요량이 많지만 밤에는 수요량이 적다는 것이다 (Ji 등, 2014).
식수 데이터 분석에 주로 사용되어져 온 모델은 전통적인 시계열 모델인 ARIMA이었다 (Bretan 등, 2014). 그러나 Voitcu와 Wong (2006)에 의하면 이러한 전통적인 방법은 선형성이라는 가정을 만족해야한다는 한계를 주요 단점으로 꼽고 있다. 실제 데이터는 더욱 복잡하고 비선형적인 패턴을 보이기 때문이다. 이를 보완하기 위해 다양한 비선형 모델들을 적용할 수 있지만 해당 모델에 대한 지식이 없으면 분석하고자 하는 데이터에 알맞은 비선형 모델을 선택하는 것과 그 모델에 최적인 매개변수를 선택하는 것 또한 어렵다. 이러한 문제는 최근 이루어진 많은 연구결과들을 통해 ANN 모델로 극복할 수 있음을 알 수 있다. ANN 모델은 다중 변수들 사이에 존재하는 비선형 관계와 예측 과정을 노드에 연결된 가중치나 뉴런을 이용해 입출력 매핑을 수행하는 비선형 회귀모델의 한 형태라 할 수 있다. 여러 개의 은닉 층으로 구성되는 Multilayer Perceptron(MLP) 유형의 ANN 모델구조로 인해 이 모델을 적용한 예측은 노이즈(noise)가 많은 실제 시계열 환경에서 효율적으로 이루어진다 (Abrahart 등, 2012).
그러나 Tiwari와 Adamowski (2013, 2014, 2017)의 연구를 통해 Wavelet-bootstrap-neural network (WBNN) 모델이 더 뛰어난 성능을 보였음을 알 수 있다. 이들은 2013년에 캐나다의 Montreal과 2014년, 2017년에 캐나다의 Calgary의 물 수요 데이터를 이용했다. 2013년의 연구에서는 전체 데이터셋을 훈련 데이터셋, 검증 데이터셋, 검정 데이터셋의 3가지로 나누었다. 분리한 각 데이터셋으로 주별 예측과 월별 예측을 수행하였다. 연구에 사용한 모델은 Wavelet Neural Network(WNN), Bootstrap-based Neural Network(BNN), WBNN, ARIMA, ARIMAX이었으며 RMSE와 MAE를 기준으로 모델 간 성능 비교를 하였다. 결과적으로 WBNN 모델이 일별, 주별, 월별 예측에서 모두 우수한 성능을 보였음을 알 수 있었다. 추가로 부트스트랩 수행 시 표본의 수를 다르게 적용해볼 것을 제안하였다.
이어서 Bretan 등 (2017)의 연구에서는 브라질의 도시인 Franca에서 2012년 5월부터 2013년 12월까지 20분마다 측정한 물 소비 데이터를 사용하였다. 570일 동안 원래 데이터의 시간 당 수요율을 기준으로 모델의 성능을 측정하였다. 그 중 훈련 데이터셋은 400일, 검증 데이터셋은 140일, 검정 데이터셋은 30일간의 데이터로 분할하여 SVR 모델을 적용하여 연구를 진행하였다. 이 연구에서는 SVR 모델과 푸리에 급수(Fourier Series)를 기반으로 한 Adaptive Fourier Series(AFS) 기법을 결합한 하이브리드 모델을 제안하였다. 이 연구에서 제안한 혼합 모델의 성능은 RMSE와 MAE을 기준으로 측정했을 때, ANN의 성능보다 좋다는 결론을 얻었다.
Farias 등 (2018)의 연구에서는 바르셀로나의 식수 수요량 데이터를 사용하였으며 ANN을 기반으로 한 QMMP^{+}(Qualitative Multi-Model Predictor Plus) 예측모델을 제시하였다. MAE, RMSE, MAPE를 기준으로 Radial Basis Function(RBF)-ANN, ARIMA, Double Seasonal Holts Winter(DSHW) 모델들과 비교하여 측정한 결과, 해당 연구에서 제시한 모델의 성능이 가장 좋았음을 알 수 있었다.

2.2.전력 및 에너지 분야

최근 인구 증가와 이에 따른 경제 발전으로 인해 인간의 전력 및 에너지 소비량은 환경을 위협할 정도로 증가하였다. 건물의 전력 소비량 예측은 건물의 에너지 효율성과 시스템 결함을 감지 및 진단하는데 유용하므로 중요한 역할을 하고 있다 (Wang과 Srinivasan, 2017). 잘못된 예측에서 생기는 오류는 이익, 시장 점유율이나 주주가치에 중요한 영향을 미치기 때문에 예측의 정확성이 중요하지만 가격의 변동성과 비정상성으로 인해 예측이 어렵다는 것이 문제이다. 따라서 기존의 수요예측모델보다 더 정확한 예측이 가능한 모델을 고려해야한다 (Fan과 Chen, 2006). 최근 인공지능 기술의 발달로 빠른 속도와 최적의 솔루션으로 미래 전력 수요량을 예측할 수 있는 AI 기반의 전력 수요 예측 모델이 인기를 얻고 있다 (Wang과 Srinivasan, 2017).

이형로와 신현정 (2011)은 지역난방공사의 대규모로 생산된 전기나 열을 넓은 범위의 지역에 일괄적으로 공급하는 시스템인 열병합 발전소에 대한 지역난방 모형의 사업 허가신청 및 분석 시 월별 단위의 중장기 전력 및 열 수요예측의 필요성을 언급하고 있다. 이를 근거로 필요시마다 신속, 간편, 정확하게 수요예측을 수행할 수 있는 월별 전력 수요 예측 모델을 제안하였다. 연구에는 한국 에너지 경제 연구원에서 제공한 지역난방 공사의 월별 전기 에너지 데이터를 사용하였다. 해당 데이터는 시계열 데이터로 변수별로 모두 단위가 다르기 때문에 최대-최소 정규화(Min-Max normalization)을 통해 데이터의 범위를 일치시키고 분포를 유사하게 만들어 처리하였다. 이후 변수선택을 통해 예측에 중요한 요인이 되는 입력 변수들만 선별하였으며 사용된 예측모델은 SVR이다. 예측 성능에 대한 평가 기준으로 MAPE 값과 시프트인덱스(shiftindex)를 사용하였다. 시프트인덱스는 예측값과 이전 시점 간의 차이가 현시점 간의 차이보다 클 경우 1 이하의 값을 갖고 반대의 경우 1 이상의 값을 갖게 된다. 즉, 1 이상의 값을 갖게 될 경우 이전 시점의 값을 따라간다고 할 수 있기 때문에 모델이 시프트 되는 경향이 있다는 것을 알 수 있었다. 따라서 예측 값이 이전 시점의 값을 계속 따라가는 형태일 때 MAPE 값을 기반으로 하는 측정의 한계를 극복할 수 있음을 알 수 있다. 예측성능을 비교하기 위해 쓰인 모델들은 AR, ANN이었으며 SVR을 적용했을 때 MAPE 값이 가장 낮았으며 시프트인덱스 값도 1 보다 낮았으므로 시프트되는 경향도 낮다고 할 수 있었다. 추가로 SVR 모델에 Correlation 기준으로 변수선택을 수행한 SVR_{CORR}모델에는 기온, 증발량, 풍속, 기압 총 4개의 변수만이 사용되었으므로 더 우수함을 알 수 있었다.

같은 맥락으로 김항석과 신현정 (2013)은 중장기의 전기가격 예측 모델을 제안하였다. 연구에 사용된 데이터는 에너지 경제 연구원이 제공한 한국전력거래소의 2000년 1월부터 2008년 7월까지 월별 전기 가격이다. 제안하는 예측 모델은 준지도학습(Semi-Supervised Learning, SSL)과 ANN을 결합한 하이브리드 모델이다. SSL로 전기 생산과 연관된 변수들로 다음 달 전기가격의 상승과 하락을 분류하는 문제를 다루었으며, ANN로는 다음 달 전기가격을 예측하는 문제를 다루었다. 그리고 각각 두 모델로 얻은 결과들을 제안하는 하이브리드 공식을 통해 다음 달 전기가격의 최종 예측값을 도출한다. 제안하는 하이브리드 공식은 다음과 같다. 현 시점을 t라고 할 때 실제 전기가격은 y_t이고, 다음 시점인 $t+1$에서 각각 SSL과 ANN으로 예측한 전기가격은 $\hat{y}_{t+1}^{SSL}$, $\hat{y}_{t+1}^{ANN}$이다. SSL을 이용해서 분류 문제로 접근하기 때문에 $\hat{y}_{t+1}^{SSL}$은 -1 또는 1의 값을 갖는다. 이때 $t+1$시점에서의 전기가격은 $\hat{y}_{t+1}^{SSL}(\hat{y}_{t+1}^{NN}-y_t)\geq 0$을 만족하면 $\hat{y}_{t+1}^{ANN}$, 만족하지 못하면 y_t로 예측한다. 또한 국내 전기가격 형성에 원유가격, 환율과 같은 경제나 금융 관련 변수가 가장 큰 영향을 미친다는 현실을 고려하여 SSL의 결과값에 비중을 주도록 한 것이 특징이다.

길소현 등 (2016)은 시계열 데이터인 에너지 수요량 데이터를 이용한 수요예측에는 RNN의 확장형 알고리즘인 Elman Neural Network 방법을 적용하는 것이 더 낫다는 것을 Xia와 Li (2012), Anbazhagan과 Kumarappan (2012)의 연구를 근거로 밝혔다. 신동하와 김창복 (2016)의 연구를 통해 ANN을 적용하여 일주일치의 하계 전력수요량 데이터와 기상요소들을 입력패턴으로 했을 때 가장 좋은 예측 결과를 얻을 수 있었음을 알 수 있었다. 추가로 습도나 풍속 등의 기상요소를 활용해서 체감온도, 불쾌지수 예측 요소를 추가함으로써 더 좋은 예측 결과를 얻을 수 있었다.

안준영 등 (2017)은 2009년부터 2017년까지의 전력통계 정보시스템과 기상청의 데이터와 10개의 기상요소, 시간요소를 사용해서 신경망을 활용한 동계 전력수요예측에 관한 연구를 진행하였다. 예측률의 향상을 위해 주말을 제거한 데이터셋과 주말과 공휴일을 모두 제거한 평일 데이터셋을 사용한 것이 특징이다. 실험을 진행한 결과 일주일치의 전력데이터와 기상요소들을 입력패턴으로 하여 DNN을 적용한 예측 결과가 가장 좋았음을 알 수 있었다. 같은 맥락으로 박준호 등 (2017)은 2013년부터 2016까지의 전력통계 정보시스템, 기상청의 전력수요데이터와 기상데이터를 이용하였다. 기존의 데이터를 그대로 사용하지 않고 평일, 주말, 주말 및 공휴일의 전력 수요 패턴이 다르다는 것을 고려하여 별도의 데이터셋 세 가지를 구축하여 진행하였다. 이렇게 나눈 각 데이터셋과 적합한 딥 러닝 네트워크를 적용하면 전력 수요예측의 오차를 감소할 수 있음을 밝혔다. 주말 및 특수일 데이터는 시계열 데이터가 아니기 때문에 DNN을 적용하였으며 시계열 데이터인 전체 데이터셋에는 LSTM을 적용하였다. 전력 수요패턴은 일주일 간격으로 비슷하기 때문에 7개의 LSTM셀을 구성하였다. 그 결과 평일과 특수일을 구분하는 구분요소를 사용함으로써 상당한 예측률의 향상을 기대할 수 있음을 밝혔다.

박지현 등 (2017)은 호주의 National Electricity Market 데이터셋의 전처리를 거친 후에 얻은 데이터를 RNN의 입력 데이터로 사용하여 예측한 결과가 Single Layer Perceptron(SLP), MLP, CNN을 사용했을 때보다 월등히 좋았음을 밝혔다. 서한석과 신광섭(2018)은 MLP를 사용한 학습을 통한 범용적인 열수요예측 모델 개발 목적의 연구를 하였다. 연구에 쓰인 데이터는 일부 지역의 2012년 1월부터 2015년 1월 31일까지의 열 수요데이터와 해당 지역에서 같은 기간동안의 외기온도이다. 연구에서 고려한 변수는

외기온도, 소비자가 사회적 요인에 의해 행동 할 시 나타나게 되는 열 소비 패턴인 소비자의 사회적 행동 패턴, 같은 외기온도일지라도 열 수요는 다른 결과를 보이는 심리적 요인이다. 결과적으로 온도와 열수요는 거의 반비례 관계를 갖지만 일정 온도 이상부터 열 수요는 수렴 상태를 유지하였다. 그리고 일요일과 금요일을 비교했을 때, 비슷한 온도지만 일요일의 열 수요가 더 높다는 결과를 통해 주중과 공휴일에는 직장을 가지 않기 때문에 소비자의 사회적 행동 패턴에 따른 상이한 결과를 얻을 수 있음을 증명하였다. 목요일과 금요일 간 열 수요량 차이가 별로 없다는 결과를 통해 심리적 요인의 반영 효과를 증명하였다.
김태영과 조성배 (2018)는 효율적인 주거용 전력수요 예측을 위해 Particle Swarm Optimization(PSO) 기반 CNN-LSTM 신경망을 제안하였다. PSO알고리즘은 최적의 하이퍼파라미터를 갖춘 최적인 신경망을 검색하기 때문에 더 정확한 예측률을 달성할 수 있는 장점이 있다. 연구에 사용된 데이터는 프랑스의 한 가정집에서 실제로 수집한 1분 단위의 가정용 전력 소비 데이터이며 기간은 2006년 12월 16일부터 2010년 11월 26일까지이다. 최적의 CNN으로 전력 수요 예측에 영향을 미치는 변수들의 특징을 추출한 후 LSTM으로 전력 수요에 존재하는 불규칙한 추세를 학습한다. 기존에 전력 수요 예측에 쓰였던 모델들인 선형회귀, 의사 결정 나무, 랜덤 포레스트, MLP 뿐만 아니라 CNN-LSTM보다 낮은 MSE 값을 근거로 PSO 적용의 유용성을 증명하였다. 정호철 등 (2019)은 다양하게 존재하는 시계열 에너지 소비량에 대한 소비패턴을 추출하기 위해 1Dimension-Convolution and Bidirectional LSTM(1D-ConvLSTM)을 제안하였다. Bidirectional LSTM을 적용함으로써 과거와 미래의 정보를 모두 이용할 수 있으므로 예측 성능이 향상됨을 알 수 있었다.
또한 다수의 전력 회사와 고객들로 구성된 스마트 그리드 네트워크를 통한 전력 수요 예측을 위해 강화학습 및 게임 이론 기법을 적용하기도 한다. 예를 들어 Apostolopoulos 등 (2018)은 두 단계로 이루어진 Power Company Learning Selection and Demand Response Management(PC-DRM) 알고리즘을 바탕으로 한 전력수요 예측에 관한 연구를 진행하였다. Lu와 Hong (2019)은 ANN을 기반으로 강화학습을 적용한 스마트 그리드 시스템을 위해 새로운 '실시간 인센티브 기반 수요 알고리즘'을 제안하였다. 제안된 알고리즘은 ANN으로 에너지 수요를 예측한 후 강화학습을 적용해서 전기회사와 고객을 동시에 고려하여 최적의 이익을 만드는 방법이다.
또한 부스팅을 기반으로 한 전력 수요 예측 기법을 연구한 사례들도 있다. 예를 들어, Taieb과 Hyndman (2014)은 20개의 지리적 영역이 있는 미국 전력회사의 시간당 부하량을 백캐스트하고 예측하는데 머신러닝의 XGBoost 기법으로 접근하였다. Chen 등 (2018)은 가정의 연간 전력 소비량을 예측하기 위해 XGBoost와 ridge 회귀모델을 결합한 하이브리드 모델을 제안하였으며 선형회귀모델, XGBoost 모델, Feedforward Deep Neural Network(FDNN)와 RMSE를 기준으로 비교한 결과 제일 좋은 성능을 갖췄음을 보였다.
특히 다른 분야들과 달리 전력 및 에너지 분야는 예측 기간을 단기와 중장기로 나눠서 예측한 연구들이 있는 것이 특징이다. Singh 등 (2013)에 의하면 단기 수요예측의 기준은 1시간에서 1주까지이고, 중기 수요예측의 기준은 1주에서 1년까지이며 장기 수요예측의 기준은 1년 이상이라는 것을 알 수 있다. 그 중 전기 및 에너지 분야에서는 주로 단기 수요예측에 관한 연구가 대부분이었지만 장기 수요예측에 관한 연구도 진행된 사례가 있었다. 예를 들어, Chae 등 (2016)은 15분 단위로 건물의 일일 전기 사용량에 대한 부하 예측 연구를 하였으며 ANN 모델을 RBF 커널과 다항 커널을 적용한 GP 모델, SVR 모델들과 비교하였다. ANN 모델의 RMSE 값이 가장 낮은 값을 가져 좋았음을 알 수 있었다. Setiawan 등 (2009)은 매우 짧은 기간 동안의 전력 부하량을 5분마다 예측하였으며 SVR모델을 적용하였다. 반면 Ghods와 Kalantar (2011)는 전력 부하량의 장기 수요예측을 위해 ANN기법, Genetic algorithm(GA)와 SVM의 하이브리드 모형인 SVMG 모델을 적용한 연구를 진행하였다. Kaytez 등 (2015)은 이전에는 전기 에너지 소비 예측 연구에는 잘 사용되지 않았던 기법인 Least square-SVM(LS-SVM)모델을 1970년부터 2009년까지 측정된 터키의 전기 에너지 소비 데이터에 적용하여 연구를 진행하였다. LS-SVM모델은 민감도와 특이도를 기준으로 Multi linear regression(MLR), ANN과 비교했을 때 가장 우수한 성능을 보였다.

2.3.현금 수요 분야

오늘날 은행을 비롯하여 다양한 분야의 산업에서 국제화 경험은 매우 중요하게 되었다. 은행 산업이 글로벌 뱅킹으로서 성공할 수 있는 방법으로 현금 자동 지급기(Automated Teller Machine, ATM)의 도입이 한 예이다. 국제적인 경영 전략뿐만 아니라 인력의 효율성 문제도 효과적으로 해결할 수 있기 때문이다 (Jegede, 2014). ATM의 원활한 운영은 은행의 통화 관리를 최적화하여 통화 유통뿐만 아니라 은행의

운영 경비를 효과적으로 절감할 수 있다는 장점이 있다. 그러나 통화 관리를 최적화하기 위해 은행은 너무 많은 현금을 보유하고 있으면 안되기 때문에 정확한 ATM의 현금 수요예측은 중요하다. ATM의 효과적인 관리를 위한 ATM의 현금 수요 예측은 ATM의 현금 수요 데이터 특성에 적합한 수요예측 모델을 적용하는 것이다.
Simutis 등 (2008)은 ATM의 현금 수요 데이터의 특징으로 매월 초 많은 돈이 인출되는 경향과 소비가 잦은 쇼핑센터 근처에 있는 ATM의 현금 인출량은 금요일, 토요일에 가장 많은 경향이 있다는 점을 꼽았다. 해당 연구에서는 시뮬레이션된 데이터와 실제 ATM 15개에서 2년간 수집된 데이터가 사용되었다. 그리고 ANN 모델과 SVR 모델로 현금 수요를 예측하였다. 두 모델들을 사용하여 50일 후의 일일 현금 수요 예측의 MAPE를 비교하였고, 그 결과 SVR 모델보다 ANN이 더 좋은 결과를 보였음을 알 수 있었다.
Venkatesh 등 (2014)의 연구에서 사용한 데이터셋은 시계열 데이터의 예측에서 Computational Intelligence(CI)의 정확성을 평가하는 NN5 Competition에서 쓰인 데이터로 영국에서 111개의 무작위로 선택된 ATM에서 매일 인출된 현금 금액을 관찰한 결과이다. 일일마다 현금 인출 패턴이 비슷한 ATM 센터 별로 클러스터링한 후 각 클러스터 별로 네 가지의 다른 신경망 네트워크로 실험한 후 General regression Neural Network(GRNN) 모델의 성능이 가장 좋다는 것을 보였다.

2.4.Ride-services 분야

자전거 공유(Bike-sharing) 시스템은 친환경적인 도시에서 지속 가능한 교통수단 이용 시스템이다 (Xu 등, 2018). 우리나라에도 일명 '따릉이'가 있듯이 전 세계적으로 수많은 도시에서도 널리 이용되어지고 있다. 중국에서도 이 시스템이 주요 도시를 시작으로 중소 도시로 빠르게 확산되어지고 있는 추세이다. 자전거는 짧은 거리를 이동하는데 있어 편의성뿐만 아니라 비용적인 면에서도 효율적이고 시간절약도 가능하다는 면에서 선도적인 분야라 할 수 있다 (Ai 등, 2019).
2017년 초 중국에 다량으로 배치된 자전거를 공유하여 탈 수 있는 시스템이 도입되었는데 이 시스템은 기존과 달리 자전거를 반납하는 구역이 따로 정해져 있지 않고 사용자가 임의의 장소에 주차하는 것이다. 정해진 장소에 주차를 하진 않지만 자전거에 내장된 GPS 추적 모듈을 통해 이용자들이 스마트폰의 앱을 이용하여 근처에 있는 자전거를 찾을 수 있다. 주변에 있는 자전거를 찾을 때 자전거 본체에 인쇄된 QR 코드를 스캔하여 자전거의 지능형 잠금 장치를 해제할 수 있다 (Xu 등, 2018). 이 시스템은 자율성과 편리성 덕분에 출퇴근 시간에도 많은 직장인들이 이용하게 되면서 도시 교통 혼잡 현상의 해결책 중 하나가 되었다. 그러나 여전히 기존의 자전거 공유 시스템과 같이 공간적, 시간적 수요의 변동 문제를 겪고 있다. 이러한 문제는 비효율적인 자전거 위치조정과 높은 운영비용의 문제로 이어질 가능성이 있음을 우려할 수 있다 (Pal과 Zhang, 2017). 따라서 문제 해결을 위해 공유되는 자전거의 수요를 공간적, 시간적 두 영역 모두에서 예측해야 하고, 이를 통해 공유되는 자전거의 균형을 재조정해야한다 (Liu, 등 2018).
Xu 등 (2018)의 연구에서는 단기간 수요예측을 위해 자전거 공유 운영회사의 데이터를 사용하였다. 해당 연구에 쓰인 데이터는 파이썬(Python)으로 개발된 멀티 프로세싱 크롤러(Multi-processing crawler)를 통해 한 달 동안 수집되었다. 웹 크롤러는 식별된 API를 기반으로 자전거 공유 모바일 앱에서 보낸 자전거 찾는 요청을 실행한다. 그 후 서버는 자전거의 위치 정보에 대한 데이터를 특정 영역 내에서 반환한다. 이런 방식으로 개발된 크롤러는 해당 지역에서 나눠진 그리드 내에서 자전거를 공유하는 위치 정보를 지속적으로 요청하고 반환된 정보를 기록하면서 데이터를 수집한다. 수집된 데이터를 LSTM 모델에 적용하여 분석하였고 기존의 시계열모델보다 성능이 뛰어났음을 증명하였다. 또 다른 사례로 Ai 등 (2018)의 연구에 따르면 Conv-LSTM 모델을 적용한 결과가 LSTM 모델을 적용했을 때보다 더 좋은 성능을 나타냈음을 보였다.
또한 택시 수요예측은 스마트시티에서 지능형 교통 시스템 구현에 있어 중요한 요소이다. 택시 호출 서비스의 인기가 높아짐에 따라 우리나라에서도 '카카오택시'와 같은 모바일 앱 서비스도 등장하였으며 대규모의 택시 수요 데이터의 수집이 가능해졌다. 택시 호출량 증가에 따른 혼선에 현명하게 대처하기 위해서는 택시 수요 예측의 정확성을 높여야한다. 그 이유는 첫째, 택시 수요 예측이 정확할수록 현 시점에 남아 있는 택시 수를 제대로 파악할 수 있기 때문이다. 둘째, 택시 호출시 배차 지연으로 인한 시간 간격을 줄일 수 있기 때문이다. 셋째, 승객을 태우지 않을 채로 거리에 있는 택시들을 다시 배치하여 에너지 낭비나 교통 혼잡을 해결할 수 있기 때문이다 (Yao 등, 2018).
Yao 등 (2018)의 택시 수요 예측에 대한 연구에서는 공간과 시간의 문제를 동시에 고려하여 DMVST-Net(Deep Multi-View Spatial-Temporal Network)을 제안하였다. 이 네트워크는 공간 뷰에서 지역적

(Local) CNN으로 지역 공간 상관관계를 모델링하고, 시간 뷰에서 LSTM 모델로 가까운 시점에서 미래의 수요 간 상관관계를 모델링하며 의미론적 뷰에서 유사한 패턴을 갖는 지역 간 상관관계까지 모델링하였으며 단일 뷰만 사용했을 때보다 더 좋은 성능을 나타냈음을 보였다.
Wang 등 (2017)은 온라인 카헤일링(Online car-hailing) 시스템을 이용한 수요예측 연구를 하였다. 온라인 카헤일링 시스템은 택시 호출 서비스와 같은 맥락으로 사용자가 있는 가까운 장소에 위치한 운전자가 픽업 요청을 받아들이면서 이루어지고 모바일 앱으로도 실행가능하다. 이 연구에 쓰인 모델은 end-to-end를 사용하는 발전된 DeepSD(Advanced Deep Supply-Demand)라는 딥러닝 모델이며 MAE와 RMSE를 기준으로 비교했을 때, 기존의 모델보다 더 좋은 성능을 나타냈음을 보였다.

2.5.관광 분야

과거에 비해 국내 및 국외로의 관광 빈도가 급격히 늘어나고 있다. 특히 모바일, 소셜 네트워크의 활성화로 관광객들이 목적지에 대한 다양하고 많은 정보들을 얻기 시작하면서 빅데이터의 수집과 분석이 용이해졌다. 잘된 수요예측을 통한 성공적인 관광산업은 해당 지역을 특색 있게 만들 수 있으며 이로 인한 경제적 파급효과도 기대해 볼 수 있다. 즉, 개인에게 적합한 정보를 기반으로 하고 있는 전략적인 광고가 관광산업의 핵심이라 할 수 있다. Song과 Liu (2017)은 관광 빅데이터 활용의 핵심목표는 통계모델의 구축을 통한 서비스 제공 뿐 아니라 관광객들의 요구를 예측하고 문제를 해결할 수 있는 사전 예방 차원임을 강조하고 있다.
관광산업에서 피드백은 관광객들의 선호도를 파악할 수 있어 높은 성장률을 달성하고, 고객 요구를 더 잘 충족하는 전략을 수립하는 데 있어 중요하다 (Chen 등, 2014). 요즘은 관광산업에서 고객과 관광사업자 사이의 피드백 과정을 구축하기 위해 주요방법으로 기계학습을 사용한다. 기계학습은 기존의 피드백 마케팅보다 개인에게 집중적이며 더욱 빠르다는 장점이 있다. 기계학습으로 훈련된 수요예측모델로 관광 서비스를 경험한 사용자를 식별하고 성공률을 측정함으로써 피드백 시스템을 수정할 수 있다. 그 후 해당 정보를 사용하여 관광 기업은 자사의 서비스 비효율성, 고객 불만에 솔루션을 제공할 수도 있다 (Bajari 등, 2015).
박수지 등 (2017)은 관광지식정보시스템에서 제공하는 원래의 데이터를 활용하는 기존의 수요예측 방법으로는 무료관광지에 대한 과대 집계 논란으로 인해 특정 관광지에 대한 측정만 가능하다는 한계점을 근거로 하고 있다. 관광분야에서 온라인 검색 엔진을 통한 관광객의 정보탐색행동과 관광지 선택행동에 연관관계가 존재할 것이라 판단하고 텍스트 마이닝을 통한 새로운 방식의 수요예측을 실시하였다. 텍스트 마이닝을 통해 얻은 키워드 검색 추이가 관광 수요예측에 온라인 검색 엔진이 뒷받침될 수 있음을 보였을 뿐만 아니라 관광지를 방문하기 전에 검색했던 정보를 기반으로 데이터를 추출하기 때문에 잠재적인 관광객까지 고려한다는 점에서 단기 예측에 있어 훨씬 뛰어나다고 할 수 있음을 알 수 있다.
최석재 등 (2016)은 온라인 뉴스 기사 제목에 형태소 분석을 수행하여 행사 개최 예정인 특색 지역의 참가 인원을 예측하는 성과 예측에 대한 연구를 하였다. 해당 연구를 관광분야에 포함한 이유는 주로 특정 지역에 존재하는 특징을 활용한 이벤트성 축제는 관광지역에서 개최하기 때문이다. 연구에서 뉴스 기사 제목을 활용한 이유는 기사의 본문보다 더욱 명확하고 핵심적인 내용은 제목을 통해 한 눈에 파악이 가능하고 모바일 환경에서는 제목의 영향이 더욱 크기 때문이다. 특히 뉴스 기사는 리뷰와 다르게 행사 개최 전에도 접할 수 있기 때문에 사전예측을 위한 연구에 활용 가능하다. 연구에 쓰인 데이터는 포털 사이트에서 '축제'로 검색했을 때 상위 5에 등장하는 고양국제꽃박람회, 담양대나무축제, 보령머드축제, 부여서동연꽃축제, 포항불빛축제의 성과 자료는 정부3.0 대한민국정보공개 사이트를 통해 수집하였다. 그리고 각 축제에 관한 기사는 네이버의 기사검색으로 수집해서 제목만 추출하였다. 수집된 기사 중 여러 명의 모든 전문가들의 긍부정성에 대한 견해가 만장일치한 기사만 채택하였다. 이후 형태소 분석에 쓰인 한글 형태소 분석기는 해당 연구소에서 자체적으로 개발한 'RHINO 2.0'이다. 해당 연구가 갖는 특이점은 뉴스라는 데이터의 성격을 고려했을 때 축제 시작 전 다량의 기사 수 증가현상을 참가 인원 미달에 대한 우려가 있음으로 보이기 때문에 예상 참가 인원 수의 하락 변수로 판단한 것이다. 그리고 사회적으로 긍부정적인 현상과 개최 직전 축제에 대한 기사 및 검색량의 증가 현상은 예상 참가 인원수와 비례한다고 판단한 것이다. 제안 방법을 평가한 결과 70.6%의 예측률을 달성했음을 알 수 있었다.
관광분야의 수요예측을 위해 인공지능 기반의 수요예측 모델로는 머신러닝 기법의 SVR이 주로 쓰인다. SVR은 오류를 최소화하는 대신 일반화 오류의 상한을 최소화하기 위해 구조적 위험 최소화 원칙을 적용한다 (Hong 등, 2011). 특히 이 모델은 Pai 등 (2005, 2006)에 의하면 관광객 수요예측에 성공적으로

적용되었음을 알 수 있다. Wu 등 (2012)은 일반적인 Gaussian Process Regression(GPR) 모델의 변형 모델인 희박한(Sparse) GPR 모델을 제안하였으며 MAE, MAPE, MSE를 기준으로 ARMA모델보다 뛰어나다는 성능을 증명하였다. 또한 Wang (2004)은 관광객 수요 예측을 주제로 마코프 체인과 Grey-Markov(GM) 모델을 결합한 MCGM(1,1)의 효율성을 증명하였으며 Sun 등 (2016)은 외국인 관광객 도착 수요 예측을 위해 Cuckoo 탐색 알고리즘을 사용하여 MCGM(1,1)의 변형을 제안하기도 하였다. Hu (2017)는 대만과 중국 국가 관광청에서 수집한 데이터로 외국인 관광 예측을 위해 소프트 컴퓨팅기술을 접목시킨 새로운 GM 모델인 Soft Computing-MCGM(SC-MCGM(1,1))을 제안하였다. SC-MCGM(1,1) 모델은 GM(1,1)을 사용해서 수집된 데이터의 추세를 식별한 후 마코프 체인을 적용하여 잔차를 수정하기 때문에 예측 정확도를 보다 향상시킬 수 있음을 알 수 있었다.

2.6.단기 예측 분야

단기 예측 분야는 제품의 출시기간 동안 예측한 수요량과 최대한 비슷한 수요를 달성하는 것이 중요하다. 우리가 일상에서 쉽게 접할 수 있는 단기 예측 분야로는 영화와 패션을 예로 들 수 있다. 따라서 본 논문에서는 영화와 패션분야의 단기 예측에 관한 연구들을 소개하고자 한다.
문화상품의 특징을 지닌 영화는 수익의 생명주기가 다른 상품들에 비해 비교적 짧기 때문에 영화의 수익과 직결되는 영화의 흥행을 예측하는 것이 필수적이기 때문에 영화 흥행 예측에 관한 연구는 다양하게 이루어지고 있다. (임준엽과 황병연, 2014). 이 때 영화 흥행 예측을 위한 변수들로는 영화 내적인 요소와 영화 외적인 요소로 구분하여 진행된다. Liu (2006)와 Chintagunta 등 (2010)의 연구를 통해 영화 외적인 요소인 영화에 대한 구전효과가 흥행에 영향이 있음을 알 수 있었다. 더불어 Liu (2006)는 사용자들에 의한 영화에 대한 구전활동은 영화개봉 직전과 영화개봉 1주일 후의 기간 동안에 가장 활발했고 영화에 대한 기대도는 개봉 후 1주일을 기준으로 점점 낮아짐을 밝혔다. 이를 근거로 임준엽과 황병연 (2014)이 영화 흥행 예측 연구에 사용한 데이터는 네이버에서 제공하는 네이버 영화의 설문자료를 2013년 11월 27일에 수집된 포털 자료와 트윗 데이터이다. 연구에 쓰인 예측 방법은 나이브 베이지안 분류(Naive Bayes' Rule)이며 결과적으로 포털자료와 트윗 데이터를 사용해 함께 분석했을 때의 정확도가 가장 높았음을 알 수 있었다.
같은 맥락으로 이오준 등 (2014)은 2013년 12월 18일에 개봉했던 영화 '변호인'의 트윗데이터를 이용해 관객들이 영화의 어떤 속성들을 선호하는지 살펴보고 이를 통해 영화 흥행에 크게 영향을 미치는 요인을 밝히고자 하였다. 영화 '변호인'을 선택한 이유는 정치적 소재로 인하여 개봉 전부터 트위터를 통한 담론이 활발했기 때문에 관객들이 해당 영화를 무엇에 근거하여 선택하였는지 분석하여 영화 흥행 요인을 연구하는데 적합하다고 판단했기 때문이다. 수집한 트윗 데이터로 트윗의 내용을 대표하는 토픽 키워드를 추출하여 관객들의 관심을 반영하는 영화의 속성들이 무엇인지 파악하였다. 분석을 위해 '한나눔' 형태소 분석기를 이용하여 고빈도 명사만을 추출하였다. 다음으로 토큰화 과정을 거친 뒤 단어들의 빈도수를 계산하여 토픽을 차례로 얻을 수 있었다. 명사만 고려한 이유는 텍스트에서 가장 높은 빈도를 보이는 품사이기 때문에 중요한 정보를 전달할 수 있으며 문장에서 명사만으로도 내용의 핵심을 파악하는 것이 가능하기 때문이다. 추가로 영화 개봉 후 시간이 지남에 따라 각 토픽들이 트윗에 노출되는 비율의 변화에 대해서도 언급하고 있다.
이상훈 등 (2015)은 텍스트마이닝과 로지스틱 회귀분석으로 영화 흥행 여부에 대한 예측모형을 개발하는 연구를 하였다. 연구에 쓰인 데이터는 포털 사이트 'D'사의 2013년부터 2014년까지 47편의 영화에 대한 55,028개의 리뷰와 영화 평점, 영화진흥위원회에서 얻을 수 있는 스크린 수이다. 본래 데이터 셋에 있는 모든 영화에서 공통적으로 빈도 수 100 이상을 갖는 용어 중 상위 10개를 추출하였다. 이때 텍스트마이닝을 통해 리뷰에 대한 긍정 부정을 최대한 정확하게 분류하기 위해 특이값 분해(Singular value decomposition, SVD)를 수행한다. 이를 통해 차원을 축소함으로써 이 많은 희박한 행렬을 얻게 되는 문제를 방지할 수 있기 때문이다. 로지스틱 회귀분석을 위해 SVD를 통해 얻어진 총 10개의 고유벡터들과 평점평균, 스크린 수까지 모두 12개의 설명변수였으며 영화의 흥행여부가 종속변수이었다. 로지스틱 회귀분석 수행 시 단계별 변수선택법을 실시하였으며 SVD로 얻은 고유벡터들 중 첫 번째와 세 번째 벡터와 평점평균 설명변수들이 모두 유의했다. 총 47개의 영화 중 관객 30만명을 기준으로 흥행한 34개의 영화와 흥행에 실패한 13개의 영화에 대해 95.74의 확률로 정확하게 분류했음을 알 수 있었다. 당시 미개봉 영화였던 '기술자들'에 대해 흥행이라고 예측했으며 실제로 흥행한 결과도 확인할 수 있었다. 하지만 어디에서 얼마나 많은 양의 텍스트 데이터를 수집하는지에 따라 분석결과가 달라지는 한계 또한 알 수 있었다.

Jeon 과 Son (2016)은 영화 흥행의 가장 직접적인 기준인 관객 수의 예측에 관한 연구를 하였다. 한국에서 2012년 1월 1일 이후 개봉하여 2015년 12월 31일까지 상영된 영화들 중 총 관객 수가 50만 명 이상을 기준으로 276개의 영화만을 선별하여 데이터로 사용하였다. 총 관객 수의 기준을 넘지 못한 영화들은 포털의 평점과 평가자 수, 블로그 수, 뉴스 수와 같이 총 관객 수에 영향을 미치는 온라인 구전 변수(Online-Word-Of-Mouth, OWOM)의 효과를 파악하기 힘들기 때문에 OWOM 변수를 수집할 수 없어 제거하였다. 예측에 사용된 모델들은 의사결정나무, MLP 유형의 ANN, 다항 로짓 모형, SVM이다. 예측은 개봉 전, 개봉 일, 개봉 1주 후, 개봉 2주 후의 4가지 시점에서 예측하였으며 각 시점에서 관측 가능한 변수들만 예측변수로 사용하였다. 결론적으로 MLP 유형의 ANN으로 예측한 결과의 정확도가 가장 높았지만 OWOM 변수는 해당 영화의 상영기간이 거의 끝나갈 때 쯤 수집이 가능하므로 해당 연구에서 설정한 시점에서는 사용할 수 없었다. 그러나 최종적으로 이 변수를 MLP 유형의 ANN으로 예측한 결과 2015년 상영 영화에 대한 총 관객 수 예측의 정확도가 개선되었음을 알 수 있었다.
남수현과 권오병 (2017)은 텍스트마이닝과 딥러닝을 활용하여 영화의 시나리오 분석을 통해 사전 흥행을 예측한 연구를 진행하였다. '필름메이커스', '오발탄' 등의 시나리오 공유 사이트에서 수집한 총 200개의 영화 시나리오와 영화진흥위원회의 영화 데이터베이스로부터 각 영화의 흥행 결과와 외적 요인 정보를 수집하였다. 동일 단어를 형태소에 따라 구분하기 위해 해당 연구소에서 자체적으로 개발한 'RHINO 2.5.1' 형태소 분석기를 사용하였으며 영화 시나리오에 존재하는 많은 단어들 간의 관계를 파악하기 위해 형태소 분석이 된 데이터로 워드 임베딩(Word embedding)을 학습하며 학습에는 Word2vec을 활용하였다. 그리고 예측을 위해 LSTM과 DNN 두가지 모델을 사용하였다. 먼저 LSTM의 입력 값은 영화 시나리오이며 워드 임베딩을 활용하여 형태소 분석을 진행하였다. 그리고 누적 관객수 500만 이하, 500만 이상, 1000만 이상으로 세 개의 범주로 나누어 분류한 결과 값을 DNN의 입력 값으로 사용하였다. 추가로 개봉 날짜, 장르, 상영 시간, 연령 제한, 감독, 주연 배우, 시나리오 작가, 제작사, 배급사를 입력 값으로 사용하여 영화 흥행 예측에 관한 연구를 진행한 바 있음을 알 수 있었다.
Kim과 Lim (2018)은 '베리 굿 걸', '제보자', 미개봉 상태였던 '드라큘라'에 대한 영화진흥원에서 제공하는 일별 박스오피스와 댓글 정보의 댓글 수와 평점 평균을 데이터로 사용하였다. GLS(Generalized Least Square) 모형과 Bass모형을 결합한 하이브리드 모델로 영화의 특정 일차 누적 관객 수 예측에 관한 연구를 하였다. 하이브리드 모델을 제안한 이유는 GLS 모델로 얻은 예측 값을 Bass 모델로 보정하여 선형 예측기법과 비선형예측기법을 접목시키기 위함이다. 변수 선택 방법 중 하나인 All subset regression으로 얻은 변수로 회귀식을 결정하였으며 모형의 예측 정확도 기준은 5중 교차 검증 방법으로 계산된 RMSE로 하였다. 미개봉 영화를 제외하고는 과거 데이터가 있는 영화에 대해 하이브리드 모델을 적용했을 때 예측력이 향상했음을 알 수 있었다.
패션 산업이 다른 산업과 차이를 띄는 가장 큰 특징은 제품의 다양성으로 인한 짧은 수명 주기, 변동이 심한 고객의 수요, 긴 공급 과정이라는 것이다 (Sen, 2008). 대부분의 패션 아이템 판매량은 계절성이라는 특징을 강하게 띄고 있다. 자주 변화하는 패션 시장의 환경에서 고객의 요구와 패션 판매에 영향을 미치는 다양한 설명 변수들은 판매량의 불규칙성이나 무작위성을 증가시키며 이러한 특성은 판매 예측의 복잡성을 증가시킨다. 따라서 패션산업의 판매 데이터 특성을 고려하여 수요를 예측하기에 적합한 예측 모델을 개발할 필요가 있다 (Wong과 Guo, 2010). Frank 등 (2004)은 여성의 캐주얼 의류 판매량 수요를 예측하기 위해 다변량 퍼지 논리 모델(Multivariate Fuzzy Logic Model)을 제안하였다. Huang 등 (2004)은 FDNN의 느린 학습속도를 개선하고자 무작위로 입력가중치를 선택하고 Single Hidden Layer FNN(SLFN)의 출력 가중치를 결정하는 Extreme Learning Machine(ELM) 알고리즘을 개발하였으며, Sun 등 (2008)은 실제 홍콩의 패션 소매업체의 데이터로 수요에 의미 있는 설명변수들 간의 관계를 분석하기 위해 ELM 알고리즘을 적용했다. 이들은 패션산업 데이터의 특징인 높은 수요의 불확실성과 강한 계절성에 적합한 ELM 알고리즘을 장착한 신경망모델을 구축하였다. 이들은 이 모델에 판매량과 이에 미치는 색상, 크기, 가격 등 다양한 설명변수들 간의 관계를 조사하고자 하였다. Thomassey와 Happiette (2007)은 신경망 네트워크 기반의 모델을 개발하였다. 이 모델은 자체 조직 지도 NN 기반 시스템(Self-organizing Map NN Based System) 클러스터링 절차와 확률론적 신경망기반 의사결정 나무 기술로 이용 가능한 데이터를 추출하고 분석하여 새로운 의류 품목의 판매량을 예측하는 신경망을 기반으로 한다.
그러나 최근에는 많은 연구자들이 신경망 네트워크 모델이 일부 전통 모델보다 더 나은 예측 성능을 갖췄다는 것을 실제 데이터셋을 적용한 성공적인 실험 결과로 증명하였다. 예를 들어, Ansuj 등 (1996)은 신경망 네트워크 모델의 MAE값과 잔차의 분산이 ARIMA 모델보다 작음을 보였고, Chu와 Zhang (2003)가 판매 데이터를 이용한 연구들이 있다.

Wong 과 Guo (2010)의 연구에서 실제 패션 판매 데이터로 실험한 결과가 있다. 이 연구에서 쓰인 데이터는 중국의 한 지역에서 1999년 1월부터 2006년 12월까지 중저가 패션 제품을 판매하는 가장 큰 패션 소매 회사 중 하나에서 수집된 것이다. 이 데이터셋에는 각 도시의 월별 판매량과 각 도시의 다른 품목 범주가 포함되어있다. 최근 2년간의 데이터를 테스트 셋으로 사용하여 예측 모델의 정확도를 비교하고 평가하였다. 그리고 이 연구의 특징은 최초로 기존의 NN기반 예측모델의 단점을 극복하기 위해 Hybrid Intelligent(HI) 모델을 제안하여 패션 산업에 판매량 예측 문제 해결을 시도한 것이다. 이 모델의 예측자(forcaster)에서 NN의 최적의 가중치를 얻기 위해 Harmony Search(HS) 알고리즘과 ELM 알고리즘이 통합된 것이 특징이다. HI 예측자에 휴리스틱한 fine-tuning을 가하여 NN의 일반화 성능을 향상시켰다. 그 결과 RMSE와 MAPE를 기반으로 성능을 측정했을 때, 기존의 시계열 모델인 ARIMA와 AR보다 뛰어나다는 것을 증명함으로써 HI 모델은 패션 산업의 불규칙한 수요패턴에 적합한 모델이라 할 수 있음을 보였다.

2.7.예비부품 및 신제품 분야

예비부품(Spare parts) 및 신제품의 공통적인 특징은 데이터가 부족하거나 없다는 점이다. 먼저 예비부품의 수요예측에 관하여 언급하고자 한다. Hua 등 (2006)은 예비부품에 대한 재고관리는 기업의 운영관리에 있어 중요하기 때문에 절충의 필요성을 강조한다. 예비부품의 재고량이 많아지게 되면 그만큼 많은 양의 자본이 묶이게 되고 반면에 예비부품의 재고량이 적어지게 되면 소비자의 요구에 제때 대응하지 못해 고객 서비스 부실이나 고비용이 소모되는 현상이 초래되기 때문이다. 간헐적인 수요패턴을 갖는 예비부품의 구매단가는 주로 높은 편이며 이 때문에 재고 관리 이론에서 보류 비용과 재주문 비용 사이에 잘 알려진 균형을 악화 시킨다 (Lolli 등, 2017). 이러한 이유들로 예비부품을 제공하는 서비스 공급망에 AI를 사용하여 수요 예측의 정확도를 향상시키는데 관심을 가져야한다 (Amirkolaii 등, 2017).
예비부품의 수요예측에 관하여 현재까지 이루어진 연구에 쓰여진 모델들로는 Croston, 지수 평활화, Exponentially Weighted Moving Average(EWMA), Trend Adjusted Exponential Smoothing, Weighted Moving Average, Double EWMA 등의 기본적인 시계열모델 및 변형모델이 있다. Croston (1972)와 Bier (1984)의 예비부품 수요예측에 관한 연구에서 위와 같은 많은 전통 모델들 중 각각 Croston 방법, EWMA 모델의 성능이 가장 좋았지만 MAPE 값이 80%를 넘는 좋지 못한 결과를 얻었다. 또 제품의 주문일시와 인도일시 사이에 경과된 시간을 뜻하는 리드타임(Lead-Time Demand, LTD)의 정확한 분포를 파악하는 것 역시 재고 관리 시스템 실행에 가장 우선시되어야 하는 조건이지만 간헐성으로 인해 LTD 추정도 어려운 것이 문제이다. Hua 등 (2006)에서는 이러한 LTD 추정을 목표로 제안하는 모델은 로지스틱 회귀 모델과 SVR 모델을 결합한 Logistic Regression SVM(LRSVM)이라는 하이브리드 모델이며 처음으로 0이 아닌 값의 수요를 부트스트랩 방법으로 샘플링하여 LTD를 추정한다. 중국의 한 석유화학 기업의 30가지 예비부품에 대한 실제 데이터셋에 지수 평활화, Croston, 마코프 부트스트랩, SVM 모델, LRSVM 모델을 적용한 실험 결과 제안한 모델의 성능이 가장 뛰어났음을 보였다.
Jung 등 (2017)은 부트스트랩 방법이 간헐적인 수요예측 연구에 적용 가능한 방법 중 하나이지만 LTD 수요에 존재하는 기존의 종속 구조는 고려되지 않기 때문에 부정확한 예측이 이루어질 가능성이 있음을 우려하고 있다. LTD에 존재하는 종속성이란 0과 0이외의 값들이 연속적으로 존재함에 따른 자기 상관 구조이다. 그래서 LTD가 독립적이라고 단정지을 수 없고 LTD에서 기존의 1차 자기 상관 구조를 고려해야 한다. 따라서 이 연구에서는 예측 정확도를 향상시키기 위해 간헐적인 수요의 고유한 특성을 고려한 새로운 부트스트랩 방법으로 수정된 마코프 부트스트랩(Modified markov bootstrap)을 제안한다.
윤현민과 김수환 (2017)은 해군 수리부속의 수요예측 정확도 향상을 통해 군수운영에 발전이 되고자 다양한 머신러닝 기반 기법을 적용한 연구를 하였다. 연구에 쓰인 데이터는 DDH-O 구축함 O척의 정비실적 데이터이며 연구의 목적을 명확히 하고자 수리부속별, 연도별(2009년–2016년)로 재정리하였으며, 수리부속 품목수는 총 29,083개였다. 수리부속 소요수량예측을 목적으로 총 10회 측정한 MSE의 평균을 최종 MSE로 하여 예측력을 측정하였다. 연구에 사용된 모델은 회귀나무, 랜덤포레스트, ANN, 선형회귀이다. 최종 선택된 모델은 랜덤포레스트로 약 90%의 정확도를 얻었으며 기존 시계열 모델로 얻은 40–50% 정도의 정확도를 극복했음을 보였다. 그러나 해당 데이터로 측정한 결과를 전체 수리부속에 적용하기에 한계점이 있음 또한 알 수 있었다.
다음으로 판매 전 단계에 있는 신제품 수요예측에 관하여 언급하고자 한다. 제품의 종류가 다양해지면서 수명은 짧아지는 추세로 인해 새로운 제품의 수요예측은 더욱 중요해졌다. 그러나 신제품은 과거 데이

터가 없다는 특징 때문에 수요예측에 어려움을 겪고 있다. Lee 등 (2014)은 신제품은 시간이 지나면서 유사한 제품의 확산 패턴을 띄게 된다는 가정 하에 신제품 수요예측에 대한 매개변수 추정치를 얻을 수 있지만 매개변수 선택에 있어 객관성이 결여된다는 한계점이 있음을 언급하였다. 이 연구는 신제품에 대한 수요예측에 알맞은 Bass 확산 모델과 다중 선형회귀 또는 머신러닝 기반의 회귀 알고리즘인 Gaussian process 회귀 등을 활용한 새로운 방법을 제안하였다. 해당 방법은 사람의 주관이 배제된 채로 오로지 제품의 속성에만 기반 할 수 있다는 장점이 있다. 이들은 RMSE와 MAE를 기준으로 성능을 평가하였으며 높은 예측 정확도를 가지고 있음을 보였다.

3.머신러닝 기반 수요예측 기법

3.1.SVM 기반 기법

SVM은 Vapnik 등 (1992)의 연구에서 처음 제시되었으며 분류 및 회귀에 사용되는 지도학습 방법 중 하나로 예측의 정확도를 극대화할 뿐 아니라 과적합(over-fitting)을 방지한다는 장점이 있다. 이 때 확장된 특성 공간(Feature space)에서 각 데이터들의 거리를 계산하는 커널(Kernel) SVM 기법이 효율적이다. Boser 등 (1992)이 제안한 커널 SVM은 원래 훈련 데이터셋을 현재 차원에서 분류가 불가능할 때 비선형 매핑(Mappping)을 통해 고차원 데이터셋으로 변환시킨다. 그 후 최대 마진(margin)을 만드는 최적의 의사결정경계(Decision Boundary)인 하이퍼플레인(Hyperplane)으로 선형 분류를 가능하게 한다.
선형 학습 기계(Linear learning machine) 특징을 갖는 SVM 알고리즘의 원리를 바탕으로 회귀문제로 접근하여 예측하는 기법은 SVR이라 한다 (Msiza 등, 2008). 비선형 회귀문제에서는 비선형 함수를 통해 원 데이터의 입력 공간 χ를 커널 함수를 이용해 새로운 고차원 특성 공간 F로 매핑시킨다. 그리고 이 공간에서 다시 선형 회귀를 수행한다 (Chen과 Wang, 2007). 따라서 기존 데이터셋에서의 비선형 회귀문제는 입력 공간 χ가 아닌 특성 공간 F에서 모델 복잡성이 적은 가장 평평한 함수(flattest function)를 찾는 문제로 변환된다.
다음과 같이 입력 벡터인 x_i와 출력 값인 d_i의 순서쌍 n개로 이루어진 원래의 데이터셋 D가 주어졌다고 하자.

$$D = \{(x_i, d_i)\}_{i=1}^{n}$$

SVR의 목적은 회귀 함수인 $y = f(x)$를 이용했을 때, 출력 값들의 집합인 $\{d_i\}$로 올바르게 예측되는지 확인하는 것이다 (Chen과 Wang, 2007). 회귀 함수는 매핑 함수를 통해 아래와 같이 다시 표현된다.

$$y = f(x) = \omega\phi(x) + b$$

$$\phi : R^n \to F,\ \omega \in F,$$

여기에서 $\phi(x)$는 입력공간 χ로 부터 비선형매핑이 된 고차원의 특성 공간 즉 F와 같고 ω와 b는 모두 훈련 데이터셋인 D로 추정가능한 모수이다. 과적합을 방지하고 일반화 능력을 향상시키기 위해 ω와 b는 각각 아래의 risk function인 $R(C)$를 최소화 하는 값으로 추정되며 식 (3.1)과 같이 표현된다.

$$R(C) = \frac{1}{2}\|\omega\|^2 + C\frac{1}{n}\sum_{i=1}^{n} L_\epsilon(y_i, d_i) \tag{3.1}$$

첫번째 항인 $\frac{1}{2}\|\omega\|^2$은 유클리디안 정규항이며 두번째 항인 $C\frac{1}{n}\sum_{i=1}^{n} L_\epsilon(y_i, \hat{y_i})$은 empirical error이자 훈련 데이터셋 D의 오차이다. Vapnik (1995)의 선형 ϵ-intensive 손실함수(loss function)인 $L_\epsilon(y, d)$는 다음과 같다.

$$L_\epsilon(y, d) = \max(0, |y - d| - \epsilon)$$

또한 정규화 된 상수항 C는 오차가 발생했을 때, 경험적 손실과 정규항 사이의 조절(trade-off)을 결정함으로써 패널티(penalty)를 계산한다. C의 값이 증가하면 정규항에 대한 경험적 손실의 유의성이 증가하고 패널티는 ϵ보다 클 때에만 수용가능하다. ϵ-intensive 손실 함수는 추정을 안정화하는데 기여하므로 노이즈를 감소시킨다. Figure 3.1과 같이 ϵ는 훈련 데이터셋 D의 근사 정확도를 의미하는 튜브(tube)의 크기와 같다고 할 수 있다. ω와 b를 추정하기에 앞서, 식 (3.1)은 여유 변수(slack variable)인 ξ_i와 ξ_i^*로

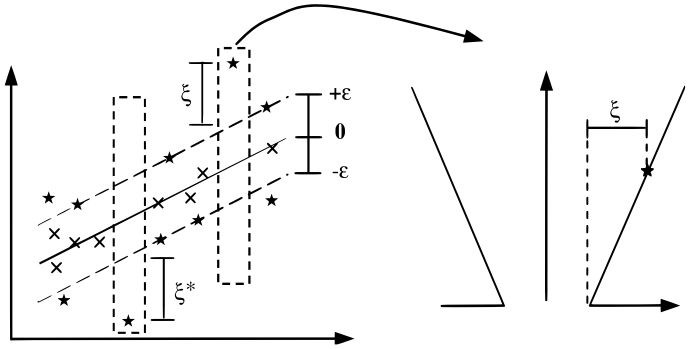


Figure 3.1. Set soft margin loss for linear SVR (Schölkopf and Smola, 2002).

$i = 1, 2, \ldots, N$에 대하여 식 (3.2)과 같이 다시 표현되고 이 변수들은 각각 오차 범위의 상한과 하한을 넘어선 범위에 존재하며 0이 아닌 값이라 가정한다. 이 때, ξ_i와 ξ_i^*를 최소화함으로써 훈련 데이터셋 D 오차를 최소화하고 $f(x)$ 모델의 복잡성을 줄이기 위해 $\frac{1}{2}\|\omega\|^2$는 작아진다.

$$R(\omega, \xi^*) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{3.2}$$

이 때 제약식은 아래와 같다.

$$d_i - \omega\phi(x_i) - b_i \leq \epsilon + \xi_i^*, \quad \xi_i^* \geq 0$$
$$-d_i + \omega\phi(x_i) + b_i \leq \epsilon + \xi_i, \quad \xi_i \geq 0$$

다음은 라그랑지안 승수 $\alpha_i, \alpha_i^*(\alpha_i\alpha_i^* = 0, \alpha_i, \alpha_i^* \geq 0, i = 1, \ldots, N)$를 사용함으로써 제약을 없애는 과정을 나타내는 수식이다.

$$\begin{aligned} L = &\frac{1}{2}\|\omega\|^2 + C\frac{1}{n}\sum_{i=1}^{n}(\xi_i + \xi_i^*) - \sum_{i=1}^{n}\alpha_i(\epsilon + \xi_i - d_i + \omega\phi(x_i) - b_i) \\ &- \sum_{i=1}^{n}\alpha_i^*(\epsilon + \xi_i^* + d_i - \omega\phi(x_i) + b_i) - \sum_{i=1}^{n}(\eta_i\xi i + \eta_i^*\xi i^*) \end{aligned} \tag{3.3}$$

식 (3.3)에서 η_i, η_i^*는 듀얼 변수(dual variable)들로 $\eta_i, \eta_i^* \geq 0$ 의 조건을 만족해야한다. 라그랑지안 승수법의 등식인 L은 ω와 b에 대해 최소화 되어져야하므로 최적의 안장점인 $(\omega, b, \alpha_i, \alpha_i^*)$을 찾아야한다.
이제 식 (3.3)을 ω와 b로 각각 편미분함으로써 다음과 같은 식을 얻을 수 있다.

$$\frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^{n}\alpha_i - \alpha_i^*\phi x_i = 0 \tag{3.4}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n}\alpha_i - \alpha_i^* = 0 \tag{3.5}$$

위의 식 (3.4)와 식 (3.5)로부터 ω값은 다음과 같이 구할 수 있다.

$$\omega = \sum_{i=1}^{n}\alpha_i - \alpha_i^*\phi x_i$$

이제 SVR 모델의 함수식은 커널 함수인 $\mathbf{K}(x_i,x_j)$로 다시 표현되며 특성 공간에서의 내적을 나타내는 기호 $\circ$ 를 이용한 해당 식은 다음과 같다.

$$\mathbf{K}(x_i, x_j) = \phi(x_i) \circ \phi(x_i)$$

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\mathbf{K}(x_i, x_j) + b$$

이 때 커널 함수로는 Mercer의 조건인 $\mathbf{K}(x_i,x_j) = \phi(x_i) \circ \phi(x_i)$을 만족하는 어떤 함수든지 사용될 수 있다. 가장 널리 쓰이는 커널 함수는 가우시안(Gaussian) 커널로도 불려지는 Radial basis function(RBF) 커널이다. 확장된 차원에서 데이터 간의 거리는 RBF 커널에 의해 다음과 같이 구할 수 있다.

$$\mathbf{K}(x_i, x_j) = \exp\left(-\frac{1}{2}\|x_i - x_j\|^2/\sigma^2\right)$$

커널을 이용해 SVR 모델로 수요예측을 했을 시, 정확도는 SVR 모델의 하이퍼파라미터(hyperparameter)인 C, ϵ과 커널의 매개변수 σ에 의해 결정되기 때문에 모델의 하이퍼파라미터 최적화도 중요한 요소이다. Peng 등 (2004)은 GA 값을 사용함으로써 최적의 하이퍼파라미터를 선정하여 구성된 SVR 모델인 GA-SVR모델을 제안하였다. Chen과 Wang (2007)는 Real-value genetic algorithm(RGA)를 채택함으로써 모델의 모든 하이퍼파라미터를 동시에 추정하지 못하는 문제를 해결하였고 이를 통해 동시에 여러 개의 SVR 모델들의 하이퍼파라미터 최적화를 가능하게 하고 예측의 효율성을 높였다. 일반적으로 모델에 있는 모든 하이퍼파라미터 선택과정은 여러 개의 세트로 다양한 SVR 모델을 구축하고 검증 데이터셋으로 유효성을 판단하는 시행착오(trial-and-error) 절차를 거치기 때문에 많은 시간이 소비된다는 단점이 있다. 그러나 GA-SVR 모델은 채택된 RGA의 진화과정을 거쳐 최적화된 하이퍼파라미터 값으로 최적의 SVR 모델을 구축한다. 즉 SVR 모델이 다양한 데이터셋에 적합될 수 있도록 RGA의 채택을 통해 효율적으로 최적의 하이퍼파라미터를 찾는 것이다. 이 모델의 순서 및 구조는 Figure 3.2와 같다.
Holland (1992)가 제시한 GA 알고리즘 구현 방식은 랜덤한 염색체 집단으로 구조를 평가하고 문제에 대한 더 나은 솔루션을 나타내는 염색체가 빈약한 염색체보다 복제될 수 있는 기회를 부여받음으로써 번식 기회를 받는 것이다. SVR 모델의 하이퍼파라미터인 C, σ^2, ϵ는 $G = \{p_1, p_2, p_3\}$인 염색체를 생성하기 위해 코딩되어 표현된다. 여기서 p_1, p_2, p_3는 각각 C, σ^2, ϵ을 나타낸다. 그리고 적합성을 정의하는 단계에서 k-배 교차검증을 시행하여 훈련 데이터셋에 대한 과적합을 방지한다. Duan 등 (2001)의 연구를 통해 k 값을 5로 설정하여 시행한 k-배 교차검증은 적은 훈련 데이터셋에서도 매개변수의 추정치를 구할 때 효율적인 계산량으로 신뢰성 있는 값을 구할 수 있음을 알 수 있다. 그리고 가장 마지막 하위집합의 MAPE에 의해 선정된 매개변수 세트의 성능이 평가된다. 같은 맥락으로 Yuan 등 (2002)이 제시한 Chaotic Genetic Algorithm(CGA)은 전역(global)의 최적해를 찾는데 유용하며 Hong 등 (2011)의 연구에서 SVRCGA 모델을 제안하였고 MAPE, MAE, RMSE 값 모두 SVR 모델보다 낮아 더 나은 성능을 보였다.
또한 도시용수 분야에서의 수요예측 연구 중 Brentan 등 (2017)이 제안한 혼합모형 SVR+AFS 모델이 있다. 외생 변수와 주기적인 수요 현상 간의 비선형 상관관계는 오류의 특정한 임계값을 유지하면서 예측하기 어렵기 때문에 매 시간마다의 도시용수 수요 패턴은 단순한 현상으로 설명되기 어렵다. 이 때, 푸리에 급수(Fourier series)는 주기적인 신호나 함수를 간단한 진동 함수의 합으로 확장하는데 적합하기 때문에 일일, 주간, 계절에 따라 다르게 나타나는 수요패턴을 기반으로 수요예측 하는데 적용할 수 있다. 제안한 모델은 SVR 모델에서 얻은 예측결과 편차에 푸리에 급수 층을 모델링하여 SVR 모델로 해결할 수 없는 시계열 주기의 최대 수요 피크를 더욱 잘 포착한다. 이와 같이 푸리에 급수를 적용한 하이브리드 모델을 통해 SVR 모델의 결과를 향상시킬 수 있다. 즉 관측된 수요와 SVR 모델에 의해 예측된 수요 간 편차는 매 시간마다 계산되는데 푸리에 급수 계수를 첨가하여 새로운 시계열 모델을 사용하면서 업데이트되며 모델의 오류도 모델링 가능하다. 수요를 예측하고자 하는 기간에 동일한 간격의 t값을 사용하면 0부터 2π까지 범위를 갖는 정규화 된 시점 t_i에서 다음과 같은 식으로 표현된다.

$$f^*(t_i) = a_0 + \sum_{j=1}^{M}(a_j \cos(jt_i) + b_j \sin(jt_i)) \tag{3.6}$$

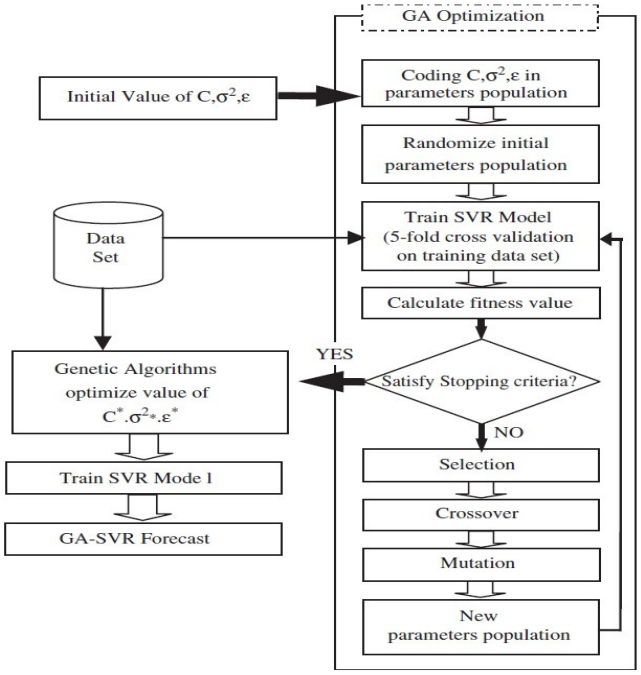


Figure 3.2. Structure of GA-SVR model (Chen and Wang, 2007).

여기에서 t_i는 $\frac{2\pi i}{N}$이다. 함수 f는 시점t_i에서 일반적인 함수 f에 대한 푸리에 급수 근사 값이며 시점 t_i의 i는 평가 간격에서 지점의 위치이고 M은 푸리에 다항식의 길이이며 a_0, a_j, b_j는 모두 조정 가능한 푸리에 계수를 나타낸다. 함수 f로 얻은 실제 값과 식 (3.6)을 통해 함수 f^*로 얻은 근사 값의 차이를 이용해서 다음과 같은 제곱편차값 s를 얻을 수 있다.

$$s = \sum_{i=0}^{N-1} \{f(t_i) - f^*(t_i)\}^2 \tag{3.7}$$

여기에서 N는 수요예측에 관심 있는 기간의 총 시간 간격 수를 나타낸다. 식 (3.7)에서 구한 s에 최소제곱법을 적용하고 직교 조건을 고려하면 푸리에 계수 a_0, a_j, b_j는 다음과 같이 표현된다.

$$a_0 = \frac{\sum_{i=0}^{N-1} f(t_i)}{N},\ a_j = 2\frac{\sum_{i=0}^{N-1} f(t_i)\cos(jt_i)}{N},\ b_j = 2\frac{\sum_{i=0}^{N-1} f(t_i)\sin(jt_i)}{N}$$

이러한 모수들이 결정되고 나면 식 (3.6)은 함수 f를 추정해서 측정된 데이터를 조정하는 데 사용된다.
검증 전에 초기 SVR 모델의 매개변수들 중 C와 ϵ은 잘 훈련되고 보정되어야 하며 각 매개변수들은 다음과 같은 조건을 따른다. 첫 번째 매개변수 C는 모델 복잡성과 ϵ보다 큰 편차가 허용되는 정도 사이에 절충을 제공하며 C의 값에 따라 회귀 모델의 견고성을 파악할 수 있다. 다음 매개변수 ϵ은 Figure 3.1의 튜브의 크기를 조절하기 때문에 서포트 백터의 수를 결정한다. 너무 큰 값은 더 적은 서포트 벡터 수를 결정하기 때문에 더 많은 데이터 포인트가 ϵ튜브에 맞게 되고, 덜 복잡한 회귀 함수가 나타나게 된다. 두 매개변수의 경계 범위는 $1 \leq C \leq 1500$이고 $0.05 \leq \epsilon \leq 0.9$이다.
Figure 3.3는 SVR-AFS의 전체적인 온라인(on-line) 학습 과정을 나타내는 그림이다. AFS를 결합함으로써 SVR 모델의 재생을 위한 최적의 주기 삽입이 가능한 점이 특이다. 모델의 업데이트는 예측의 특정한 결과 값 이후부터 진행된다. 관측된 수요량과 SVR 모델에 의해 예측된 수요량 간의 편차는 매 시간마다

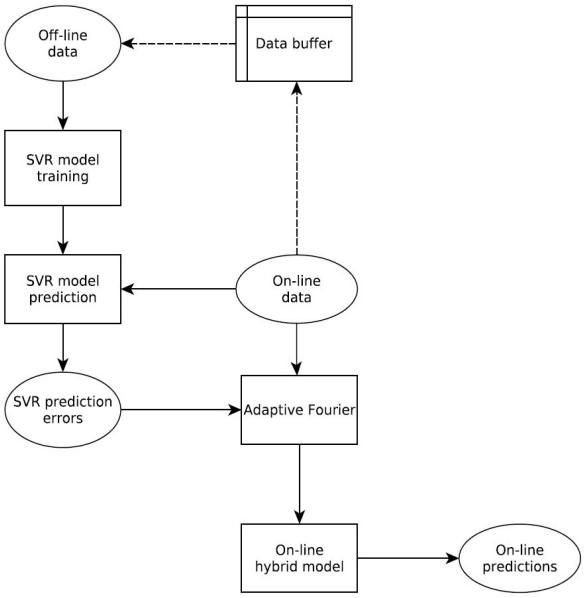


Figure 3.3. Flow diagram of SVR-AFS (Brentan *et al.*, 2017).

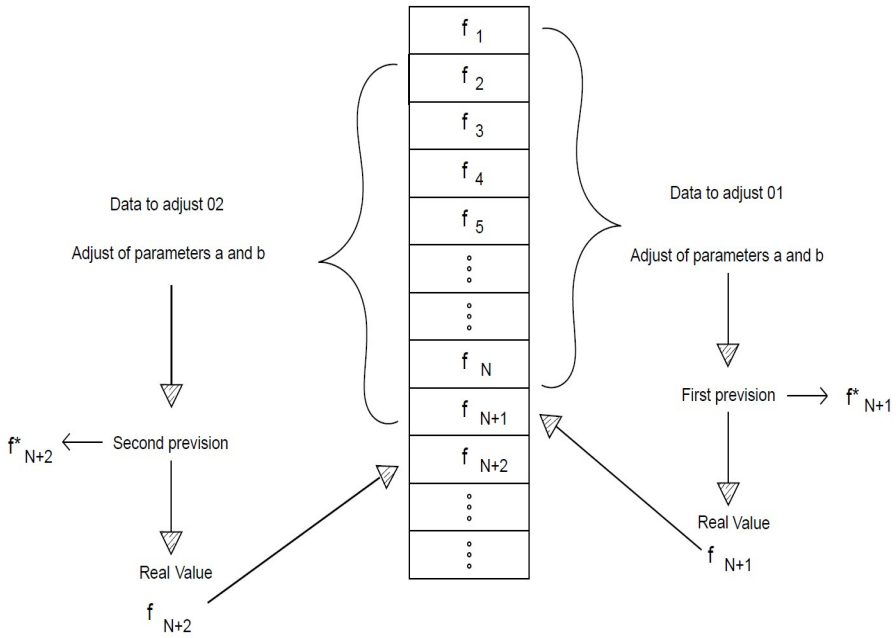


Figure 3.4. Update process of AFS (Brentan *et al.*, 2017).

계산되고 푸리에 급수의 계수는 새로운 시계열을 사용하여 업데이트되며 모델의 온라인 오류를 모델링 한다. 최종적으로 SVR-AFS 모델로 예측된 수요량 y는 SVR 모델에 의해 예측된 물 수요량인 y_{SVR}과

AFS모델에 의해 추정된 편차 f_{AFS}의 합으로 나타낸다.
Figure 3.4는 AFS의 업데이트 과정을 나타내는 그림이다. 예측된 보정은 전체 모델에 온라인 학습(online learning)을 통해 통합하고 예측 프로세스에 마지막으로 측정된 물 수요량은 다시 새로운 변수로 추가되어 데이터를 업데이트한다. 적응형 푸리에급수(Adaptive Forier series)의 결합으로 가장 최근의 물 수요 정도를 추가함으로써 실시간 예측으로 수요를 업데이트 할 수 있는 것이다. 여기서 온라인 학습이란 $t = 1, 2, \ldots$에 대해 순차적이고 연속적으로 실행되며 $\mathbf{x}_t \in \chi$인 질문을 받고, $\mathbf{p}_t \in \mathbf{P}$라고 답변을 예측하면 실제 정답인 $\mathbf{y}_t \in \mathbf{Y}$와의 차이를 손실함수 $\mathbf{L}(p_t, y_t)$로 나타낸다. 이 때 반드시 $\mathbf{P}$와$\mathbf{Y}$가 같을 필요는 없으며 $\mathbf{P}$가 더 큰 집합일 때 예측이 더 편리하다는 점이 있다 (Shalev-Shwartz, 2012).

3.2.부스팅(Boosting) 기반 기법

Schapire (1990)이 약한 학습 알고리즘을 임의의 높은 정확도를 얻는 알고리즘으로 바꾸는 방법을 제안했고 Freund (1995)이 부스팅을 위한 새로운 알고리즘을 제안하였다. Friedman (2001)의 그래디언트 부스팅(Gradient Boosting)은 함수 공간에서 수치적 최적화의 관점에서 부스팅을 함수 추정을 위한 방법이라 뜻한다.
Regue와 Recker (2014)는 bike sharing 운영의 비효율을 해결하고자 4가지 모델을 동시에 고려하여 연구했다. 4가지 모델은 각각 스테이션 레벨에서의 수요예측모델, 스테이션 재고모델, 재배포 요구모델, 차량경로 지정모델이다. 그 중 수요예측 모델을 그래디언트 부스팅 기반 기법으로 구축하였다. 그래디언트 부스팅의 목적은 회귀분석과 마찬가지로 데이터셋 $D = \{(x_i, d_i)\}_{i=1}^{n}$가 주어졌을 때, 손실함수 $L(F(x), d)$의 기댓값을 최소화하는 모델 $F(x)$를 찾는 것이다.

$$F^*(x) = \arg\min_{F(x)} E_{d,x}[L(F(x), d)] \tag{3.8}$$

식 (3.8)의 $F^*(x)$는 Friedman (2002)에서 가법 확장(additive expansion) 형식인 $\widehat{F}(x) = \sum_{z=1}^{Z} \beta_z h(x_s; q_z)$에 의해 추정되는 모델임을 알 수 있다. 여기서 base learner인 함수 $h(x; q)$는 의사 결정 나무와 같은 입력 벡터 x와 모수 벡터 q의 간단한 함수로 선택된다. x_s는 랜덤하게 추출된 x의 하부표본(subsample)이고 q_z은 z번째 의사 결정 나무의 모수들이다. 그리고 확장 계수들의 집합인 $\{\beta_z\}_{z=0}^{Z}$과 모수들의 집합 $\{q_z\}_{z=0}^{Z}$들은 초기함수(initial guess)인 $F_0(x)$을 시작으로 forward stagewise 과정을 따르면서 훈련 데이터셋에 적합된다.

$$(\beta_z, q_z) = \arg\min_{\beta, q} \sum_{i=1}^{n} L(F_{z-1}(x_i) + \beta h(x_i; q), d_i) \tag{3.9}$$

이어서 식 (3.9)로 적합된 β_z와 q_z으로 모델의 업데이트 과정은 식 (3.10)과 같이 나타낼 수 있다.

$$F_z(x) \leftarrow F_{z-1}(x) + \nu_1 \beta_z h(x_s; q_z) \tag{3.10}$$

여기에서 ν_1는 정규화 모수로서 과적합을 방지하고, 주로 작은 수일수록 더 좋은 추정값을 얻을 수 있다.
Taieb과 Hyndman (2014)는 Kaggle Global Energy Forecasting Competition 2012에서 그래디언트 부스팅 방법을 적용해서 시간대별, 요일별, 공휴일의 유무에 따라 변화하는 전력 수요예측에 관한 연구를 진행하였으며 우수한 성과를 보였다. 이전에 있었던 Ramanathan 등 (1997)의 연구에서는 하루 중 매시간마다 각각 회귀모델을 학습시켰는데, 제안한 모델은 여기서 착안되었다. 시간대별로 생성한 모델들은 미래의 시간대에 따른 예측 값을 추출하기 위해 재귀적으로 사용된다. 즉, 다음날의 예측 값을 생성하면 그 값은 그 다음날의 예측 값을 생성하기 위한 입력 값이 되는 것이다. 그리고 t시간에 대한 모델을 추정하기 위해 이전시간인 $t-1$시간의 관측 데이터와 다음시간인 $t+1$시간의 관측데이터를 추가하여 사용한다. 이 모델로 연도별, 요일별, 일별 시간대, 공휴일별로 수요변경이 가능하다. 각 회귀모델은 전력 수요를 예측하기 위해 패널티 항이 적용된 회귀 스플라인이 있는 비모수적 가법 모델(Nonparametric additive model)을 사용한다. 이 모델은 비선형관계를 가지며 k일의 t시점에 대한 수요는 다음과 같이 나타낸다.

$$y_{k,t} = c_t(k) + f_t(\mathbf{y}_{k,t}) + g_t(s_{k,t}) + \epsilon_{k,t} \tag{3.11}$$

여기에서 $y_{k,t}$는 위에서 언급한 k일의 t시점에 대한 수요를 뜻하고, $c_t(k)$는 calendar effect에 대한 모델이다. $\mathbf{y}_{k,t}$는 k일 t시점 이전의 과거수요를 나타내는 벡터들이고 $f_t(\mathbf{y}_{k,t})$는 최근 수요변수의 효과를 나타내는 모델로 lagged demand 효과를 포함한다. $s_{k,t}$는 k일의 t시간 및 그 이전에서의 최근 온도 변수의 벡터이고 $g_t(s_{k,t})$는 총 전기 수요량에 대한 최근 기온의 영향을 나타내는데 건물의 열 관성을 고려했을 때 현재 온도 뿐만 아니라 lagged temperature도 함께 고려해 온도 효과까지 포함한 모델이다. 그리고 $\epsilon_{k,t}$는 k일의 t시점에서의 오차를 뜻한다. 이와 같은 모델을 설정함으로써 모델 내에서 수요 시계열의 계열 상관성(serial correlation)을 포착 할 수 있다. 이어서 비모수적 가법 모델의 구성 요소별 그래디언트 부스팅(component-wise gradient boosting)은 적합 과정 중 변수 선택 과정을 포함해서 각 모델평가에 사용된다 (Friedman 등, 2000). 부스팅의 그래디언트 하강 뷰(gradient descent view)는 더욱 일반적인 최적화 뷰를 이용해서 부스트를 연결한다. 그래디언트 부스팅은 함수 공간에서 수치적 최적화의 관점으로 부스팅이 함수 추정을 위한 방법이라 하고 있으며 설명변수의 변환과 특이점에 대해 크게 민감하지 않다는 이점이 있음을 알 수 있다. 또한 이전 모델에서 부스팅 과정을 계속 할 수 있기 때문에 새로운 데이터를 수집하더라도 전체 모델을 다시 학습할 필요가 없으므로 계산량이 적어 효율적이라 할 수 있다 (Friedman, 2001).

Taieb과 Hyndman (2014)는 구성 요소 별 그래디언트 부스팅과 penalised regression splines(P-spline)을 사용해서 분석하였는데, 구성 요소 별 그래디언트 부스팅은 모델 적합 과정 중에 자동 변수 선택을 추가한다는 이점이 있고 P-spline으로 수요량의 매끄러운(smooth) 추정량을 구할 수 있다.

식 (3.11)은 $y_{k,t} = F_t(\mathbf{x}_k) + \epsilon_{k,t}$로 다시 표현할 수 있으며, 여기에서 $\mathbf{x}_k = [k, \mathbf{y}_{k,t}, s_{k,t}]$이다. 그리고 표본 데이터셋 D에 기반하여 손실함수의 기댓값을 최소화하는 함수 F_t를 찾게 된다. $z(z = 0, \ldots, Z)$번째 단계에서 F_t의 추정량인 $\widehat{F}_t^{(z)}(\mathbf{x}_k)$는 $z = 0$일 때에는 시간 k의 평균 수요량을 나타내는 $\bar{y}_k$와 같고 마찬가지로 stagewise 과정을 따르면서 모델이 업데이트된다. 이 연구에서는 추정하려는 모수를 θ로 정의하였으며 업데이트 과정에서 z단계에서의 약한 학습자(learner) 추정값 $h_z(\mathbf{x}_k; \hat{\theta}_z)$에 곱해지는 정규화모수 ν_1의 범위는 0부터 1까지다. $h_z(\mathbf{x}_k; \hat{\theta}_z)$를 $\hat{h}_z(\mathbf{x}_k)$라고 하자. 이 때 잔차는 손실함수를 오차제곱합으로 설정하였을 때의 negative gradient와 같다. $\widehat{F}_t^{(z-1)}$이 주어졌을 때, 각각 추가항인 $\hat{h}_z(\mathbf{x}_k)$은 다음과 같이 negative gradient $u_{k,t}^z$을 계산함으로써 얻어진다.

$$\begin{aligned} u_{k,t}^z &= -\frac{1}{2}\frac{\partial(y_{k,t} - F(\mathbf{x}_k))^2}{\partial F(\mathbf{x}_k)}\Bigg|_{F(\mathbf{x})=F_t^{(z-1)}(\mathbf{x})} \\ &= y_{k,t} - F_t^{(z-1)}(\mathbf{x}_k) \end{aligned} \tag{3.12}$$

식 (3.12)을 통해서 모수의 추정값 $\hat{\theta}_z = \arg\min_{\theta\in\Theta} \sum_{k=1}^{K} [u_{k,t}^z - h_z(\mathbf{x}_k; \theta)]^2$을 얻을 수 있다. 즉, $\hat{h}_z(\mathbf{x}_k)$는 이전 모델인 $F_t^{(z-1)}(\mathbf{x}_k)$로부터 잔차를 가장 잘 예측할 수 있게 선택된다. 이로써 최종목표로 하는 해인 $\widehat{F}_t(\mathbf{x}_k)$는 z단계의 추가 구성요소인 $\nu_1\hat{h}_z(\mathbf{x}_k)$에 의해 계속해서 개선되며 식 (3.13)과 같다. 또한 하이퍼파라미터인 Z은 구성요소의 개수에 한계를 두어 과적합을 방지한다.

$$\widehat{F}_t(\mathbf{x}_k) = \widehat{F}_t^{(Z)}(\mathbf{x}_k) = \hat{h}_0(\mathbf{x}_k) + \sum_{z=1}^{Z} \nu_1 \hat{h}_z(\mathbf{x}_k) \tag{3.13}$$

이 때, 식 (3.13)는 약한 학습자가 모든 예측변수를 동시에 사용해서 추정하고 있는 과정을 나타는데 Bühlmann과 Yu (2003)는 새로운 구성 요소 별 큐빅 스무딩 스플라인을 사용한 L2부스팅이 고차원의 예측변수가 있는 경우에 효과적임을 입증하였다. 핵심은 한 번에 변수 하나로 약한 학습자를 사용한 후 적합성에 큰 기여를 하는 학습자를 선택하는 것이다. 추정 값 $\hat{\theta}_z$는 component-wise 그래디언트 부스팅을 적용해서 다음과 같이 다시 표현된다.

$$\hat{\theta}_z^{(j)} = \arg\min_{\theta\in\Theta} \sum_{k=1}^{K} [u_{k,t}^z - h_z(x_{jk}; \theta)]^2$$

x_{jk}는 $j = 1, \ldots, J$일 때 $\mathbf{x}_k$의 j번째 변수이다.

$$j_z = \arg\min_{j\in\{1,\ldots,J\}} \sum_{k=1}^{K} [u_{k,t}^z - h_z(x_{jk}; \hat{\theta}_z^{(j)})]^2$$

그리고 z번째 단계에서 $\hat{\theta}_z^{(j_z)}$을 사용해서 최종 해는 다음과 같이 표현할 수 있다.

$$\widehat{F}_t(\mathbf{x}_k) = \hat{h}_0(x_{j_0 k}) + \sum_{z=1}^{Z} \nu_1 \hat{h}_z(x_{j_z k})$$

여기에서 $\hat{h}_z$은 j_z번째 예측변수의 함수이다. 결론적으로 그래디언트 부스팅이 효과적인 예측 모델이라는 것과 적합 과정에서 자동 변수선택이 가능한 component-wise 그래디언트 부스팅의 이점을 증명하였다.
Mayrink와 Hippert (2016)은 단기 전기 수요량을 예측하기 위해 그래디언트 부스팅과 지수 평활의 DSHW을 결합한 모델을 제시하였다. 제안한 결합모델은 초기 예측값을 계산한 후 각 단계에서 그래디언트 부스팅 알고리즘을 반복적으로 사용해서 이전 단계의 오차를 맞춰가는 것이다. 예측 변수 집합 x_t로부터 오차e_t를 예측하기 위한 오차 적합 과정은 식 (3.14)와 같다.

$$\begin{aligned}
\hat{e}_t^{(0)} &= f_0(\mathbf{x}_t) = \bar{e} \\
\hat{e}_t^{(1)} &= f_0(\mathbf{x}_t) + \eta f_1(\mathbf{x}_t) = \hat{e}_t^{(0)} + \eta f_1(\mathbf{x}_t) \\
\hat{e}_t^{(2)} &= f_0(\mathbf{x}_t) + \eta f_1(\mathbf{x}_t) + \eta f_2(\mathbf{x}_t) = \hat{e}_t^{(1)} + \eta f_2(\mathbf{x}_t) \\
&\vdots \\
\hat{e}_t^{(M)} &= \sum_{m=0}^{M} f_m(\mathbf{x}_t) = \hat{e}_t^{(M-1)} + \eta f_M(\mathbf{x}_t)
\end{aligned} \tag{3.14}$$

다음으로 반복 횟수를 나타내는 모수 m은 1부터 M까지고 훈련 데이터의 하위집합 $\mathbf{x}_{i1}^{\hat{n}}$에 대해 z번째 트리의 잔차 r_{iz}은 i가 1부터 $\hat{n}$까지 일 때 $y_i - f_{z-1}(\mathbf{x}_i)$로 정의된다. 목적함수를 최적화하는 모수들 θ_z은 식 (3.15)과 같이 손실함수와 정규화함수의 합으로 표현된다.

$$\begin{aligned}
Obj(\theta_z) &= L(\theta_z) + \Omega(\theta_z)) \\
L(\theta_z) &= \sum_t (e_t - \hat{e}_{tz})^2 = \sum_t (e_t - f_z(\mathbf{x}_t \mid \theta_z))^2 \\
f_z(\mathbf{x} \mid \theta_z) &= f_z(\mathbf{x} \mid \{b_{ez}, R_{ez}\}_1^E) = \sum_{e=1}^{E} (b_{ez} 1(\mathbf{x} \in R_{ez}))
\end{aligned} \tag{3.15}$$

여기에서 $\{R_{ez}\}_1^E$ 은 z번째 트리의 터미널 노드에 의해 정의되는 입력 공간과 분리된 영역이다. E는 터미널 노드의 수이며 b_{ez}은 해당 영역 R_{ez}에서 트레이닝 인스턴스(training instance)의 결과값의 평균이다. 그리고 정규화함수는 다음과 같다.

$$\Omega(\theta_z) = \gamma_1 E + \frac{1}{2}\lambda_1 \sum_{e=1}^{E} w_e^2$$

여기에서 γ_1는 잎 노드의 정규항이며 너무 많은 터미널 노드가 있는 트리에 패널티를 준다. λ_1는 과적합을 방지하고 출력값의 분산을 줄이는 L2 정규항이고 w_e는 잎 스코어이다. 이렇게 정의된 목적함수를 최소화하는 b_{ez}와 $\mathbf{R}_{ez}$으로 이루어진 E개의 순서쌍은 $\hat{\theta}_z$이 된다.

$$\hat{\theta}_z = \{b_{ez}, R_{ez}\}_1^E = \arg\min_{b_e, R_e} Obj(\hat{\theta} = \{b_{ez}, R_{ez}\}_1^E)$$

$f_z(\mathbf{x}) = f_{z-1}(\mathbf{x}) + \eta \rho_z f_z(\mathbf{z} \mid \theta_z)$에 의해 업데이트 되고 마지막으로 추정의 목표인 $\hat{f}(x) = f_Z(x)$를 얻게 된다. 이를 통해 지수평활법으로 예측한 결과의 잔차에 그래디언트 부스팅 기법을 적용함으로써 예측 정확도를 향상시킬 수 있음을 보였다.
Chen 등 (2018)은 앙상블 학습 기법을 이용하여 연간 가계 전력 소비량을 예측하기 위해 두 가지 모델인 XGBoosting과 능형회귀(ridge regression)을 결합한 새로운 모델을 제안하였다. 첫 번째 모델은 여러

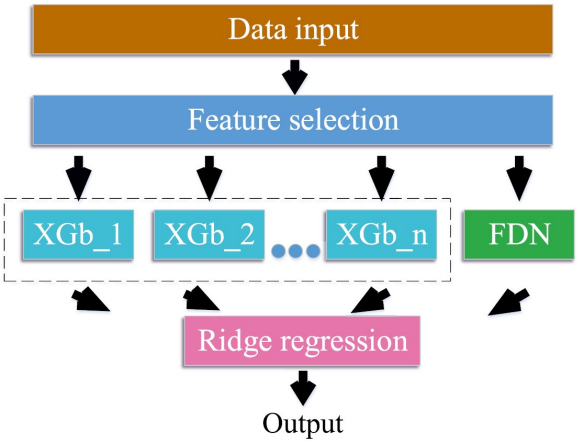


Figure 3.5. Hybrid model structure of XGBoost and ridge regression (Chen *et al.*, 2018).

개의 다른 XGBoost 트리 모델을 결합한 XGBoost 포레스트 모델이고 두 번째 모델은 FDNN이다. 이 때 두 모델을 결합하기 위해 ridge regression를 도입하였다. 단일 XGBoost 트리 모델의 가장 적합한 하이퍼파라미터들은 검증 데이터셋에서 그리드 서치(grid search)에 의해 결정된다. Figrue 3.5는 제안하고 있는 모델구조를 나타낸 그림이다. 여러 개의 XGBoost모델과 FDNN모델이 있을 때 개별 모델들은 입력 데이터를 받고 예측 결과를 독립적으로 나타낸다. 그 후 각 모델의 예측 결과들은 능형 회귀의 입력으로 들어가고 최종적인 예측 결과를 얻을 수 있게 된다. 제안된 앙상블 모델은 RMSE를 기준으로 비교했을 때 단일 모델보다 좋은 성능을 가졌음을 알 수 있었다.

3.3.가우시안 프로세스(Gaussian Proess) 기반 방법

Wu 등 (2012)은 Gaussian Process Regression(GPR)을 Mercer의 조건을 만족하는 커널을 사용하며 비모수 모델링과 확률론적 추론을 위해 베이지안 프레임을 기반으로 한 모델로 정의한다. 이 연구에서는 관광 분야의 수요예측을 위해서 희박한 GPR 모델을 제안하였다. 이 모델은 계산의 편리성과 향상된 일반화 성능을 갖추고 있으며 ARMA 모델, Wu (2009)의 v-SVM 모델, Wu (2010)의 g-SVM 모델보다 예측성능이 우수함을 보였다. 일반적인 GP 모델의 주요 장점은 예측의 불확실성까지 포함한 전체적인 확률적 예측분포를 구할 수 있다는 것과 베이지안 커널의 모수를 학습할 수 있다는 것이다. 하지만 naive 구현을 위해 계산량의 증가가 문제 되기 때문에 이 한계를 극복하기 위해 희박한 근사치를 제안하고 있다. 일반적인 GPR 모델에서 훈련 데이터셋 D가 $d_i = f_{x_i} + \epsilon_i$인 (noisy)프로세스로부터 생성된 것이라 하자. 여기에서 x_i는 $\mathbf{R}_d$공간에 있는 입력 벡터이고 y_i는 $\mathbf{R}$공간에 있는 출력 값이므로 스칼라 값을 갖는다. 그리고 노이즈 항(noise term)인ϵ_i은 $N(0, \sigma^2)$를 따른다. 모든 입력벡터를 합친 $d * d$크기의 행렬 $\mathbf{X}$가 평균이 0인 가우시안 분포를 따르며 입력 값으로 주어지면 출력 값 y에 대한 결합 확률 분포는 다음 식과 같다.

$$p(y \mid \mathbf{X}) = N(0, \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 I) \tag{3.16}$$

여기에서 $\mathbf{K}(\mathbf{X}, \mathbf{X})$는 입력 벡터 간의 거리를 측정한 커널 원소 $\mathbf{K}_{ij}(x_i, x_j)$를 가지며 커널 행렬이라고도 불린다. $\sigma^2 I$항은 가우시안 잡음을 발생시키며 ϵ_i항과 같은 역할을 한다. 이 때 커널 $\mathbf{K}_{ij}$는 다음 식과 같다.

$$\mathbf{K}_{ij}(x_i, x_j) = k(x_i, x_j) = \nu_2^2 \exp(-\frac{(x_i - x_j)}{2\lambda_2^2}) \tag{3.17}$$

여기에서 ν_2^2은 사전분포의 분산을 통제하고 λ_2^2는 x_i와 x_j가 얼마나 떨어져 있어야 하는지 결정한다. 이를 바탕으로 훈련 데이터셋 D와 검정 데이터셋으로 점 X^*가 주어졌을 때 GPR 모델의 목적은 신뢰구간을

통해 예측의 출력 값인 f^*를 찾는 것이다. 이 때 $p(y \mid f)$는 독립적 우도함수이고 $p(f, f^*)$는 결합 사전 분포로 가우시안 분포를 따르며 다음과 같다.

$$p(f, f^*) = N\left(0, \begin{bmatrix} \mathbf{K}_{ff} & \mathbf{K}_{f^*f} \\ \mathbf{K}_{ff^*} & \mathbf{K}_{f^*f^*} \end{bmatrix}\right)$$

여기에서 f와 f^*는 $\mathbf{K}(\mathbf{X}, \mathbf{X})$의 원소들의 첨자이며 베이지안 추론법에 의해 사후 예측 분포는 다음 식과 같다.

$$p(f^* \mid y) = \int p(f, f^* \mid y)\, df = \int \frac{p(y \mid f)p(f, f^*)}{p(y)}\, df = N(\mu, \Sigma) \tag{3.18}$$

여기서 μ와 Σ는 각각 다음과 같다.

$$\mu = \mathbf{K}(\mathbf{X}^*, \mathbf{X})[\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 I]^{-1} y$$

$$\Sigma = \mathbf{K}(\mathbf{X}^*, \mathbf{X}) - \mathbf{K}(\mathbf{X}^*, \mathbf{X})[\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 I]^{-1}\mathbf{K}(\mathbf{X}, \mathbf{X}^*)$$

GPR 모델의 하이퍼파라미터로 이루어졌고, 커널 함수의 모수인 $\theta = [\lambda_2, \nu_2, \sigma_2]$는 d차원의 입력 벡터가 주어졌을 때 훈련 데이터 출력 값의 최대 우도 추정법으로 학습되고 식 (3.19)와 같이 표현된다.

$$\log(p(y \mid \mathbf{X}, \theta)) = -\frac{1}{2} y^T (\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 I)^{-1} y - \frac{1}{2} \log \left|\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 I\right| - \frac{n}{2} \log 2\pi \tag{3.19}$$

$$\theta_{max} = \arg\max_{\theta} \log(p(y \mid \mathbf{X}, \theta))$$

그러나 위에서 언급했듯이 GPR 모델의 naive 구현을 위해서는 많은 연산이 필요하고 구체적으로는 훈련 데이터셋 n개에 대해 n^3번의 연산이 필요하다는 문제점이 있다. 이를 극복하고자 Wu 등 (2012)은 다음과 같은 희박한 근사 절차를 제안하였다.
우선 m개의 잠재변수이자 유도변수 세트 $\mu = [\mu_1, \mu_2, \ldots, \mu_m]$ 를 추가하여 식 (3.20)과 같이 표현한다.

$$p(f^*, f) = \int p(f, f^*, \mu) = \int p(f, f^* \mid \mu) p(\mu)\, d\mu \tag{3.20}$$

여기에서 $p(\mu)$는 $N(0, \mathbf{K}_{\mu,\mu})$를 따른다. μ가 주어졌을 때 f^*와 f가 조건부 독립이라 가정함으로써 결합 사전분포를 다음과 같이 근사시킬 수 있다.

$$p(f, f^*) \approx q(f, f^*) = \int q(f^* \mid \mu q(f \mid \mu) p(\mu))\, d\mu \tag{3.21}$$

식 (3.18)의 두 개의 근사 유도 조건인 와 에 대한 다른 추가 가정은 다음과 같다.

$$p(f \mid \mu) = N(\mathbf{K}_{f,\mu}\mathbf{K}_{\mu,\mu}^{-1}\mu, \mathbf{K}_{ff} - \mathbf{Q}_{ff})$$

$$p(f^* \mid \mu) = N(\mathbf{K}_{f^*,\mu}\mathbf{K}_{\mu,\mu}^{-1}\mu, \mathbf{K}_{f^*f^*} - \mathbf{Q}_{f^*f^*})$$

여기에서 $\mathbf{Q}_{a,b} = \mathbf{K}_{a,\mu}\mathbf{K}_{\mu,mu}^{-1}\mu\mathbf{K}_{\mu,b}$ 이다. ARMA 모델, v-SVM 모델, g-SVM 모델과 성능비교를 한 결과, 희박한 GPR 모델이 MAE, MAPE, MSE를 기준에서 모두 더 뛰어났음을 보였다 (Wu 등, 2012).

4.딥러닝 기반 수요예측 기법

4.1.RBM 기반 방법

Hinton과 Salakhutdinov (2006)이 제안한 RBM은 에너지 함수를 기반으로 한 딥러닝의 비지도 학습 방법 중 하나이다. 이 때 가시 층(visible layer)과 은닉 층에 포함된 모든 노드들끼리의 연결은 없으며 Figure 4.1과 같이 각 층끼리의 연결만 존재하기 때문에 신경망과 같은 학습 원리를 가진다. 가시 층은

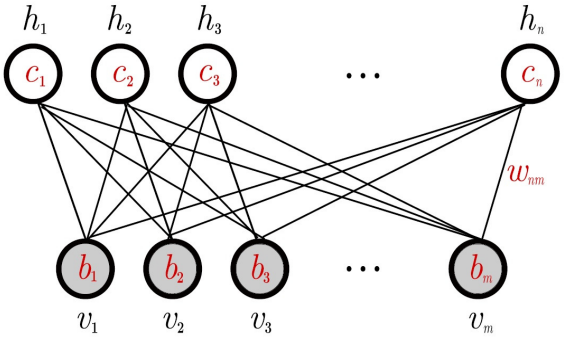


Figure 4.1. Undirected graph of Restricted Boltzmann Machine (Fischer and Igel, 2012).

데이터를 나타내고 은닉 층은 임의의 복잡성으로 표현될 수 있는 분포 클래스를 확대하여 학습 용량을 증가시킨다 (Taylor 등, 2011). RBM에서 모든 노드들은 또는 의 값을 가지며 입력받은 데이터를 은닉 층으로 전달할지 확률에 따라 결정하기 때문에 종의 확률과정 네트워크라 볼 수 있다.
특정한 구성을 가질 확률로 가시 층과 은닉 층의 결합 확률 분포 는 에너지 함수를 사용하여 다음과 같이 표현된다.

$$p(\mathbf{v}, \mathbf{h}) = \frac{e^{-E(\mathbf{v},\mathbf{h})}}{Z} \tag{4.1}$$

여기에서 $\mathbf{v} = \{v_j\}$은 가시 층 노드들의 집합이고 v_j는 j번째 가시 층의 노드이다. $\mathbf{h} = \{h_i\}$는 은닉 층 노드들의 집합이고 h_i는 i번째 은닉 층의 노드이다. 또한 $Z = \sum_{v,h} e^{-E(v,h)}$는 모든 가능한 가시 층과 은닉 층의 노드들의 합으로 기댓값을 구함으로써 정규화된 계수이며 에너지 함수는 아래와 같이 표현된다.

$$E(\mathbf{v}, \mathbf{v}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} h_i v_j - \sum_{j=1}^{m} b_j v_j - \sum_{i=1}^{n} c_i h_i \tag{4.2}$$

여기에서 w_{ij}는 가중치, b_j와c_i는 각 층의 편차를 나타내며 w_{ij}, b_j, c_i는 RBM에서 학습되는 값들이다. 이 때 해당 상태의 확률은 다음과 같다.

$$p(V_j = 1 \mid \mathbf{h}) = \sigma(\sum_{i=1}^{n} w_{ij} h_i + b_j), p(H_i = 1 \mid \mathbf{v}) = \sigma(\sum_{j=1}^{m} w_{ij} v_j + c_i) \tag{4.3}$$

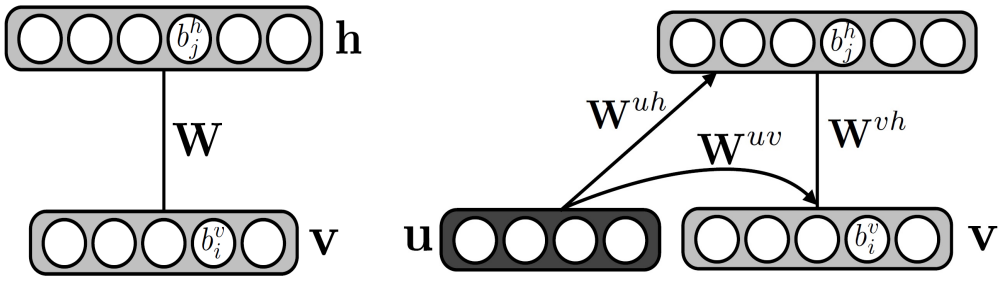
여기에서 V_j는 m개의 가시 층 노드로 이루어진 벡터V의 성분이고 H_i는 n개의 은닉 층 노드로 이루어진 벡터H의 성분이다. 랜덤 변수인 벡터H 와 벡터V는 $(\mathbf{v}, \mathbf{h}) \in \{0,1\}^{m+n}$의 값을 가지고, $\sigma(x) = \frac{1}{1+e^{-x}}$ 이다. 이렇게 구해진 확률의 최대 로그우도 값을 계산해서 모수 w_{ij}, b_j, c_i들은 그 값을 최대한 크게 하는 값들로 적합되기 위해 에너지 함수의 기울기는 각 모수에 대해 계산되어지고 식 (4.4)로 나타낼 수 있다.

$$\frac{\partial}{\partial w_{ij}} E(v, h \mid \theta) = v_i h_j, \quad \frac{\partial}{\partial b_j} E(v, h \mid \theta) = v_j, \quad \frac{\partial}{\partial c_{ji}} E(v, h \mid \theta) = h_i \tag{4.4}$$

하지만 보통 RBM에서는 최대우도함수를 그대로 적용해서 계산하는 것은 단순하지 않으며 다음과 같이 정의되는 Hinton (2002)의 Constrastive Divergence(CD) 알고리즘으로 학습된다.

$$CD_n = D_{KL}(p_0(x) \parallel p_\infty(x)) - D_{KL}(p_n(x) \parallel p_\infty(x)) \tag{4.5}$$

여기에서 $p_n(\cdot)$은 n번 실행된 마코프 체인의 분포이다. CD에서는 가시 층의 유닛들이 주어졌을 때 은닉 층의 유닛들을 모두 갱신하고 은닉 층의 유닛들이 주어지면 가시 층의 유닛들을 모두 갱신한다. 깁스 샘플링을 통해 은닉층의 노드를 구할 수 있으며 반복적인 실행을 통해 네트워크의 형태를 갖춘 채로 학습이



(a) Structure of RBM (b) Structure of CRBM

Figure 4.2. The general structure of (a) RBM and (b) CRBM (Mnih *et al.*, 2011).

가능하게 된다. 이렇게 얻어진 CD값의 왼쪽 항은 positive phase로 주어진 데이터에 대한 기댓값이고 negative phase는 번의 깁스 샘플링으로 재구성하여 얻어진 분포의 기댓값이다. 즉 CD값은 주어진 데이터의 분포와 재구성하여 얻은 두 분포의 유사성을 측정하는 방법인 Kullback Leibler Divergence(KLD)으로 구할 수 있다.

Mocanu 등 (2014)은 기상 조건, 냉난방 성능과 환경, 존재하는 사람의 수 등 많은 영향 요인과 에너지 행동의 복잡성으로 인해 빌딩의 정확한 에너지 소비의 수요량 예측 어려움을 해결하고자 Conditional Restricted Boltzmann Machine(CRBM)을 제안하였다. Mnih 등 (2011)로 CRBM 모델이 협업 필터링, 분류 및 행동 캡처 데이터 모델링을 포함하여 광범위한 문제에 적용 가능한 확률모델이라는 것을 알 수 있다. CRBM은 일반적인 RBM과 다르게 조건부 모델에도 적용가능하며 Figure 4.2는 RBM의 구조와 CRBM의 구조를 비교한 그림이다.

CRBM의 에너지 함수는 를 모델링하기 위해 RBM을 사용하고 u를 사용해서 RBM의 편차나 가중치를 동적으로 결정함으로써 분포$p(v \mid u)$를 모델링하며 이를 나타내는 식은 다음과 같다.

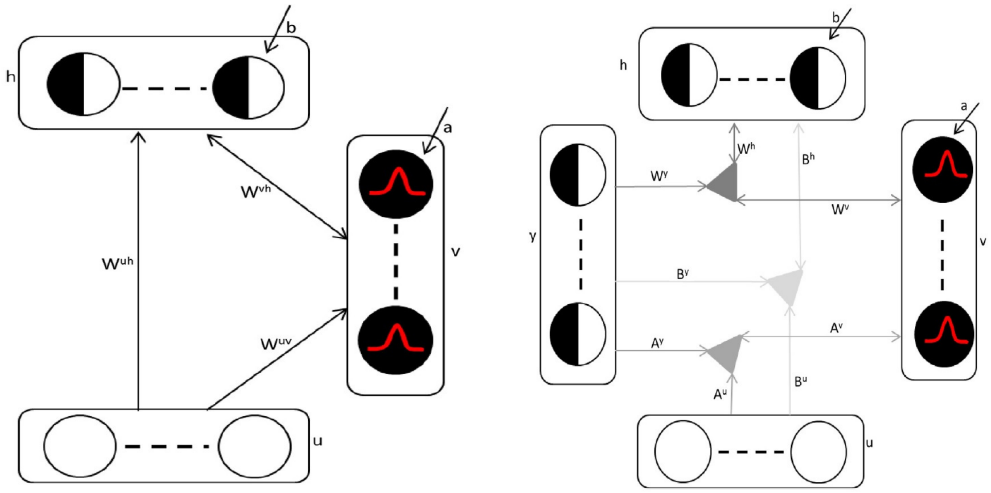
$$E(v, h, u; W) = -v^T W^{vhh} - v^T bv - u^T W^{uvv} - u^T W^{uhh} - h^T bh \tag{4.6}$$

$\mathbf{u}$는 길이가 n_u인 실수값을 갖는 벡터이며 n_u는 히스토리(history) 뉴런의 가장 마지막 뉴런이고 입력값이 된다. $\mathbf{v}$는 길이가 n_v이고 가시층의 모든 유닛인 v_i를 모아놓은 실수값을 갖는 벡터로 n_v는 가시 뉴런의 가장 마지막 뉴런이고 출력값이 된다. $\mathbf{h}$는 길이가 n_h이고 은닉 층의 모든 유닛인 h_j를 모아놓은 실수값을 갖는 벡터로 n_h는 은닉 뉴런의 가장 마지막 뉴런이다. 가중치 행렬 W^{vh}, W^{uv},W^{uh}은 각각 벡터 $\mathbf{v}$와 벡터 $\mathbf{h}$를 연결하는 노드, 벡터 $\mathbf{u}$와 벡터 $\mathbf{v}$를 연결하는 노드이고 벡터 $\mathbf{u}$와 벡터 $\mathbf{h}$를 연결하는 노드이다. 이 때 인공신경망과의 차이점은 W^{uv}와 W^{uh}은 단방향을 갖지만 W^{vh}는 양방향을 갖는다는 것이다. b^h는 은닉 뉴런의 편차이고 b^h는 가시 뉴런의 편차이다. Mocanu 등 (2014)에 의하면 CRBM에서의 확률적인 추론이란 두 가지 조건부 확률 분포인 $p(\mathbf{h} \mid \mathbf{v}, \mathbf{u})$와 $p(\mathbf{v} \mid \mathbf{h}, \mathbf{u})$를 정의하는 것임을 알 수 있다. 같은 층의 뉴런들끼리는 연결이 존재하지 않으므로 병렬적으로 추론이 가능하며 다음과 같은 분포를 얻을 수 있다. 뉴런들끼리는 연결이 존재하지 않으므로 병렬적으로 추론이 가능하며 다음과 같은 분포를 얻을 수 있다.

$$p(h = 1 \mid \mathbf{u}, \mathbf{v}) = \sigma(\mathbf{u}^T W^{uh} + \mathbf{v}^T W^{vh} + \mathbf{b}^h)$$
$$p(v \mid \mathbf{h}, \mathbf{u}) = N(W^{uv^T}\mathbf{u} + W^{uv}\mathbf{h} + \mathbf{b}^v, \sigma^2)$$

근사방법인 CD를 이용해 CRBM을 학습하는 과정에서 우도함수를 최대화하는 매개변수를 적합하고 입력 데이터 분포와 근사 모델의 확률분포 간의 차이인 KLD 값을 최소화한다. 각 가중치 행렬들과 편차들은 에너지함수에 대해 편미분되어 얻은 값을 더하고 학습률에 원래의 데이터에 대한 CD를 뜻하는 $\langle\rangle_{data}$와 재구성된 데이터에 대한 CD를 뜻하는 $\langle\rangle_{recon}$의 차이인 CD 값을 곱하는 과정으로 다음과 같이 업데이트된다.

$$\mathbf{W}^{uh}_{\tau+1} = \mathbf{W}^{uh}_{\tau} + \alpha(\langle \mathbf{u}\mathbf{h}^T \rangle_{data} - \langle \mathbf{u}\mathbf{h}^T \rangle_{recon}) \tag{4.7}$$



(a) Structure of CRBM (b) Structure of CRBM

Figure 4.3. The general structure of (a)CRBM and (b)FCRBM (Mocanu *et al.*, 2016a).

$$\mathbf{W}_{\tau+1}^{uv} = \mathbf{W}_{\tau}^{uv} + \alpha(\langle \mathbf{u}\mathbf{v}^T \rangle_{data} - \langle \mathbf{u}\mathbf{v}^T \rangle_{recon}) \tag{4.8}$$

$$\mathbf{W}_{\tau+1}^{vh} = \mathbf{W}_{\tau}^{vh} + \alpha(\langle \mathbf{v}\mathbf{h}^T \rangle_{data} - \langle \mathbf{v}\mathbf{h}^T \rangle_{recon}) \tag{4.9}$$

$$\mathbf{b}_{\tau+1}^{v} = \mathbf{b}_{\tau}^{v} + \alpha(\langle \mathbf{v} \rangle_{data} - \langle \mathbf{v} \rangle_{recon}) \tag{4.10}$$

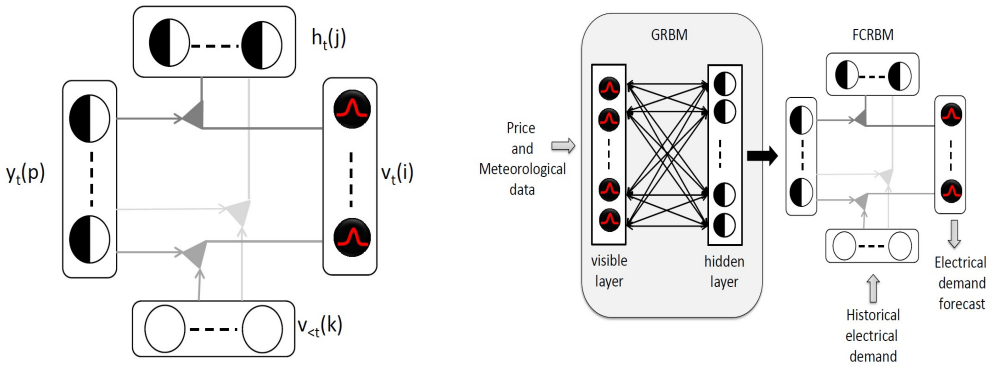
$$\mathbf{b}_{\tau+1}^{h} = \mathbf{b}_{\tau}^{h} + \alpha(\langle \mathbf{h} \rangle_{data} - \langle \mathbf{h} \rangle_{recon}) \tag{4.11}$$

여기에서 τ는 반복(iteration)이고 α는 학습률(learning rate)이다. Mocanu 등 (2016a)은 이전 연구에서 CRBM 모델의 성능을 증명했으나 해당 연구에서는 Factored Conditions Restricted Boltzmann Machine(FCRBM)을 적용하고 더 나은 결과를 증명하였다. Figure 4.3은 CRBM의 구조와 FCRBM을 비교한 그림이다.

이어서 같은 해에 Mocanu 등 (2016b)은 Factored Conditions Restricted Boltzmann Machine(FCRBM)을 제안하여 전기 에너지 수요의 적은 집계량에서도 높은 예측력을 달성했음을 보였다. FCRBM기법은 RBM기법을 기반으로 한 모델로 RBM의 에너지 함수와 다른 형태를 가진다. Figure 4.4(a)는 FCRBM의 구조그림이다. $v_{<t}$는 조건부 입력 층, h는 은닉 층, y는 레이블 층이고 v는 가시 층을 나타낸다. ◐는 이분형 노드, ◯와 나머지 기호는 실수 값을 나타낸다. 기존의 RBM모델의 에너지 함수와 달리 제안하는 모델의 에너지함수는 다음과 같다.

$$E(v_t, h_t \mid v_{<t}, y_t) = E_I E_{III} \tag{4.12}$$

$$E_I = \frac{1}{2}\sum_{i=1}^{n_1}(v_{i,t} - \hat{a}_{i,t})^2 - \sum_{j=1}^{n_2}\hat{b}_{j,t}h_{j,t}, \tag{4.13}$$



(a) Structure of FCRBM (b) Structure of Gaussian RBM(GRBM)

Figure 4.4. The general structure of (a)FCRBM and (b)GRBM as input to rhe FCRBM (Mocanu *et al.*, 2016b).

$$E_{III} = -\sum_{f=1}^{F}\left[\sum_{i=1}^{n_1} W_{if}^{v} v_{i,t} \sum_{j=1}^{n_2} W_{jf}^{h} h_{j,t} \sum_{p=1}^{n_3} W_{pf}^{y} y_{p,t}\right]$$
$$= -\sum_{f=1}^{F}\left[\sum_{i=1}^{n_1}\left[\sum_{j=1}^{n_2}\left[\sum_{p=1}^{n_3} W_{if}^{v} W_{jf}^{h} W_{pf}^{y} v_{i,t} h_{j,t} y_{p,t}\right]\right]\right] \tag{4.14}$$

여기에서 F, n_1,n_2 ,n_3는 각각 인자의 총 개수, 가시 층, 은닉 층, 라벨 층의 노드 개수를 나타낸다. 여기서 $\hat{a}_{i,t}$와 $\hat{b}_{j,t}$는 식 (4.15)과 같다.

$$\hat{a}_{i,t} = a_i + \sum_m A_{i,m}^{v} \sum_k A_{k,m}^{v<t} v_{k,<t} \sum_p A_{p,m}^{y} y_{p,t},$$
$$\hat{b}_{j,t} = b_j + \sum_n B_{j,n}^{h} \sum_k B_{k,n}^{v<t} v_{k,<t} \sum_p B_{p,n}^{y} y_{p,t} \tag{4.15}$$

$A_{i,m}^{v}$,$A_{k,m}^{v<t}$,$A_{p,m}^{y}$,$B_{j,n}^{h}$,$B_{k,n}^{v<t}$,$B_{p,n}^{y}$은 각 층들의 동적 편향 값(dynamic biases)이다. 이를 통해 FCRBM의 확률은 j번째 은닉 층의 노드과 i번째 가시 층의 노드로부터 추론가능하며 다음과 같이 표현된다.

$$p(h_{j,t} = 1 \mid v_t, v_{<t}, y_t) = \sigma(\hat{b}_{j,t} + h_{j,t}^{*})$$
$$p(v_{i,t} \mid h_t, v_{<t}, y_t) = N(\hat{a}_{i,t} + v_{i,j}^{*}, \sigma_i^2)$$

FCRBM을 이용한 모수들의 학습 및 업데이트는 식 (4.12)에서 정의한 에너지함수를 이 모델에서 정의하는 $\mathbf{W}_{if}^{v}, \mathbf{W}_{jh}^{h}, \mathbf{W}_{pf}^{y}, \mathbf{A}_{i,m}^{v}, \mathbf{A}_{k,m}^{v<t}, \mathbf{A}_{p,m}^{y}, \mathbf{B}_{j,n}^{h}, \mathbf{B}_{k,n}^{v<t}, \mathbf{B}_{p,n}^{y}$ 가중치 모수와 편향 모수들로 편미분하여 얻을 수 있다. 그 값은 식 (4.5)와 같이 CD와 델타(delta) 기법에 의해 다음과 같이 표현한다. k는 마코프 체인의 시행 횟수를 나타낸다.

$$\delta W \propto \left\langle \frac{\partial E}{\partial W} \right\rangle_0 - \left\langle \frac{\partial E}{\partial W} \right\rangle_k, \quad \delta A \propto \left\langle \frac{\partial E}{\partial A} \right\rangle_0 - \left\langle \frac{\partial E}{\partial A} \right\rangle_k, \quad \delta B \propto \left\langle \frac{\partial E}{\partial B} \right\rangle_0 - \left\langle \frac{\partial E}{\partial B} \right\rangle_k$$

또한 예측에 쓰인 설명변수들과 예측결과의 상관관계가 낮을 때, 다른 추가적인 정보의 조합을 통해 예측 정확도를 높여야 하므로 외부 정보로부터의 특징 추출(feature extraction)기법을 통해 차원을 축소하기도 한다. 이 때 RBM이 PCA와 비교했을 때 더욱 효과적이었고, 이를 근거로 Figure 4.4(b)와 같이 가시 층에 가우시안 뉴런을 연결해서 RBM을 Gausian RBM(GRBM)으로 변형하여 적용가능하다 (Hinton과 Salakhutdinov, 2006).

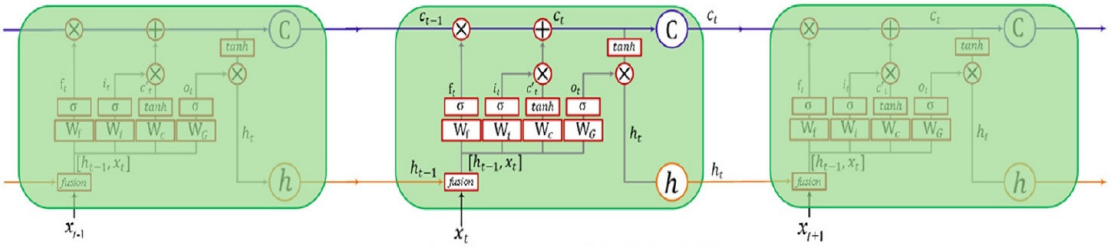


Figure 4.5. Structure of LSTM (Xu *et al.*, 2018).

4.2.RNN 기반 방법

RNN은 순차적인 데이터를 한 번에 하나의 요소씩 처리하면서 시퀀스 단계에서 정보를 선택적으로 전달할 수 있는 연결 모델이다 (Lipton 등, 2015). RNN의 일반적인 구조는 뉴런이 각 층마다 여러 개로 구성되어 있으며 하나의 층에 있는 각 뉴런의 출력 값은 다른 층에 있는 뉴런의 입력 값으로 연결된다는 것이 특징이다. RNN은 하나 이상의 순환 요소를 포함한다. 피쳐(Feature) 벡터를 입력 값으로 갖는 시퀀스 $\mathbf{X} = \{x^1, \cdots, x^T\}$가 주어졌을 때, T는 피쳐 벡터 X의 T개의 차후 관측치 개수이다. 은닉 층은 다음의 함수를 이용해서 표현된다.

$$h^t = \sigma(W[x^t, h^{t-1}] + b)$$

여기서 x는 입력 값의 벡터이고 h는 다음 층의 출력 값의 벡터이다. W와 b는 각각 가중치 값의 행렬, 편향 값의 벡터이며 σ는 활성화 함수이다 (Simoncini 등, 2018).
그러나 기존의 RNN 구조의 단점은 장기 의존성(Long-Term Dependency) 문제이다. 네트워크 구조상 짧은 시퀀스에 대해서만 효과를 보이기 때문에 시점이 길어질수록 앞의 정보가 뒤로 충분히 전달되지 못하는 현상이 발생한다. RNN에서 발생하는 가중치 값의 소실이나 포화 문제로 인해 장기간 데이터셋을 훈련하기에 적합하지 않기 때문에 이 문제를 해결하고자 Hochreiter와 Schmidhuber (1997)은 LSTM을 제안하였다. RNN과 다르게 LSTM 구조는 아래와 같이 표현된다.

$$f^t = \sigma(W_f[x^t, h^{t-1} + b^f])$$

$$i^t = \sigma(W_i[x^t, h^{t-1} + b^i])$$

$$c^{-t} = \tanh(W_c[x^t, h^{t-1} + b^c])$$

$$c^t = f^t c^{t-1} + i^t c^{-t}$$

$$o^t = \sigma(W_o[x^t, h^{t-1} + b^o])$$

$$h^t = o^t \tanh(c^t)$$

LSTM에는 셀 상태인 c^t를 통제하거나 변화하기 위해 세 개의 게이트가 존재한다. 각각 입력 게이트(input gate)인i^t , 망각 게이트(forget gate)인f^t, 출력 게이트(output gate)인o^t이다. Figure 4.5에서 알 수 있듯이 망각 게이트인 f^t는 이전의 셀 상태인 C_{t-1}로부터 어떤 정보를 제거할 것인지 결정한다. 이 때 활성화 함수는 시그모이드 함수이고 현재의 입력 데이터x_t와 마지막 시간 간격인 $t-1$에서의 메모리 셀의 출력 값은 h_{t-1}이다.
RNN과 LSTM을 기반으로 한 변형 기법들로 진행되어진 연구들 중 Marino 등 (2016)은 같은 데이터셋에 분 단위와 시간 단위로 건물의 전기 부하량 예측을 위해 표준 LSTM모델과 S2S LSTM(Sequence to sequence LSTM)모델을 사용했다. RMSE를 기준으로 측정했을 때, 표준 LSTM모델은 분 단위 예측에서 정확도가 좋지 않았지만 변형한 모델인 S2S LSTM 모델은 분 단위와 시간 단위 예측의 정확도가 모두 좋은 것으로 나타났다. 또한 해당 연구를 통해 Mocanu 등 (2016)의 연구에서 제시한 FCRBM과도 버금가는 성능을 갖췄음을 증명하였다. S2S LSTM 모델의 S2S 부분은 서로 다른 길이의 시퀀스를 매핑하기 위해 제안된 구조이다. S2S LSTM 모델은 두 개의 LSTM 네트워크로 가변 길이의 입력 시퀀스를 변환하여 고정 길이 벡터로 인코딩하는 인코더와 다시 입력 상태로 출력하는 디코더로 구성되어있다. 이

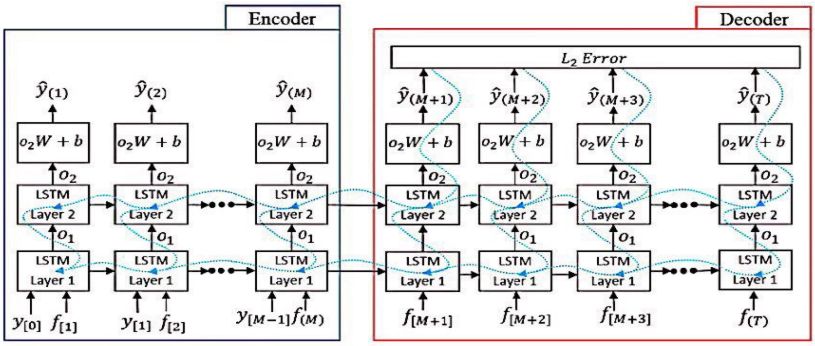


Figure 4.6. Architecture of S2S LSTM (Marino *et al.*, 2016).

때, 디코더에서 출력된 길이 n의 출력 시퀀스에서의 에너지 부하를 예측하는 것이 목적이다. 이 구조의 장점은 임의의 길이의 입력이 가능하다는 것이다.
M개의 과거의 데이터를 바탕으로 $T - M$만큼의 시점동안 다양한 기간에 대한 에너지 부하량의 예측을 위해 인코더에 입력으로 사용되는 시간 별 부하량 측정값은 다음과 같다.

$$y = \{y_{[0]}, y_{[1]}, \cdots, y_{[M-1]}\} \tag{4.16}$$

여기에서 $y_{[t]}$는 시간 단계 t에 대해 실제 측정된 전기 부하량 값이다. 예측해야 할 에너지 부하량 값들은 다음과 같다.

$$\hat{y} = \{\hat{y}_{[M]}, \hat{y}_{[M+1]}, \cdots, \hat{y}_{[T]}\} \tag{4.17}$$

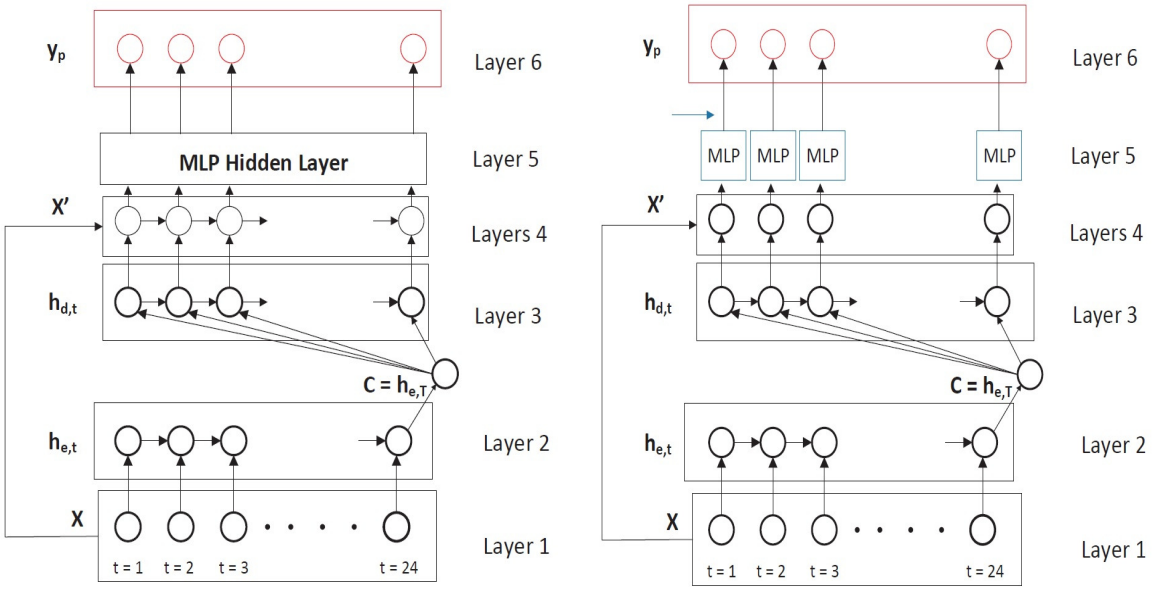
여기에서 $\hat{y}_{[T]}$는 시간 단계 t에 대해 예측한 전기 부하량 값이다. 식 (4.16)과 같은 입력 값들의 훈련을 위해 인코더 네트워크는 다음과 같은 오차 L_E를 최소화하도록 사전 훈련된다.

$$L_E = \sum_{t=1}^{M} (y_[t] - \hat{y}_{[t]})^2 \tag{4.18}$$

Figure 4.6에 의해 훈련 중 역전파가 일어났을 때 신호는 디코더에서 인코더 방향인 반대방향으로 흐르는 것을 알 수 있다. 사전 훈련된 인코더 네트워크는 디코더 네트워크에 연결되면서 두 네트워크를 훈련하여 아래의 목적함수 L_D값 를 최소화하면서 인코더와 디코더에 대한 가중치가 모두 업데이트 된다.

$$L_D = \sum_{t=M+1}^{T} (y_[t] - \hat{y}_{[t]})^2 \tag{4.19}$$

Kong 등 (2017a, 2017b)은 LSTM 모델을 기반으로 가계의 전력 부하량 단기 예측 연구를 진행하였다. 가계의 전력 부하량 예측은 거주자 활동의 다양성으로 인해 건물의 전력 부하량 예측보다 정확한 예측이 어려운 것이 특징이다. 온도, 습도, 주중, 특별한 날과 같은 많은 상황 변수뿐만 아니라 거주자의 행동까지 고려해야하기 때문이다. 이때 거주자의 일정한 생활 패턴을 파악 하는 것은 예측력 향상의 핵심이다. 해당 연구에서는 이를 반영하여 예측이 가능한 건조기, 세탁기, 식기 세척기, 텔레비전, 오븐 등의 모든 가전 제품 에너지 시퀀스가 LSTM의 입력으로 들어가고 해당 모델을 훈련했으며 state-of-art 모델인 FFNN과 K-nearest neighbor(KNN)과 비교하여 뛰어난 성능을 증명하였다.
Orozco 등 (2018)은 시계열 예측에 대해 Memory-endowed Ordianal Regression Deep Neural Network(MOrdRed) 모델을 제안한다. 이 모델은 2개의 RNN으로 구성되는데 관찰된 시퀀스를 매핑하는 인코더 $f^{(enc)}$를 통해 입력 시퀀스인 $\mathbf{X}$는 $\mathbf{h}_0^{(dec)}$와 $\mathbf{C}_0^{(dec)}$로 반환되며 소프트맥스 함수를 거치면서 $\hat{\mathbf{y}}_t = Softmax(f^{(enc)}(\mathbf{h}_{t-1}^{(dec)}, \mathbf{C}_{t-1}^{(dec)}, x_t))$ 같은 식을 얻게 된다.



(a) Structure of Model A (b) Structure of Mdodel B

Figure 4.7. Deep RNN models proposed by Rahman *et al.* (2018).

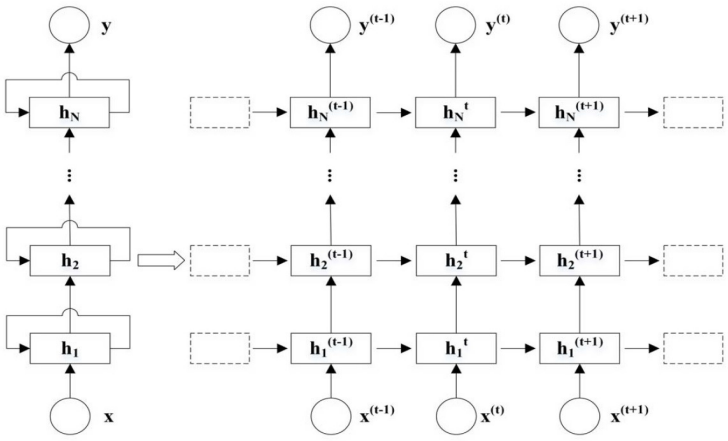


Figure 4.8. Architecture of N layer deep-RNN (Shi *et al.*, 2018).

Rahman 등 (2018)은 1시간 단위로 상업용 및 주거용 건물의 중장기 전력 소비량 예측 연구에 최적인 모델로서 LSTM 모델이 기반인 Deep RNN 모델로 Model A와 Model B의 2개 모델을 제안하였다. 제안하고 있는 모델들은 S2S 모델링에 적합한 인코더-디코더 아키텍처를 사용하는 LSTM 기반 모델 위에 적층된 multi-layer perceptron 신경망을 사용한다. 또한 실제로 예측에 쓰이는 전력 데이터에는 결측값이 많은데 해당 모델을 사용해서 결측값 대체를 수행하였다. Figure 4.7은 두 모델의 구조 그림이다. Model A의 구조와 Model B의 구조를 비교해보면 3번째 층까지는 같지만 4번째 층부터는 다른 구조를 지니고 있다. ○은 각 시간 단계를 뜻하는데, Model A의 4번째 층의 출력은 각 시간 단계 스칼라 값들의 연결

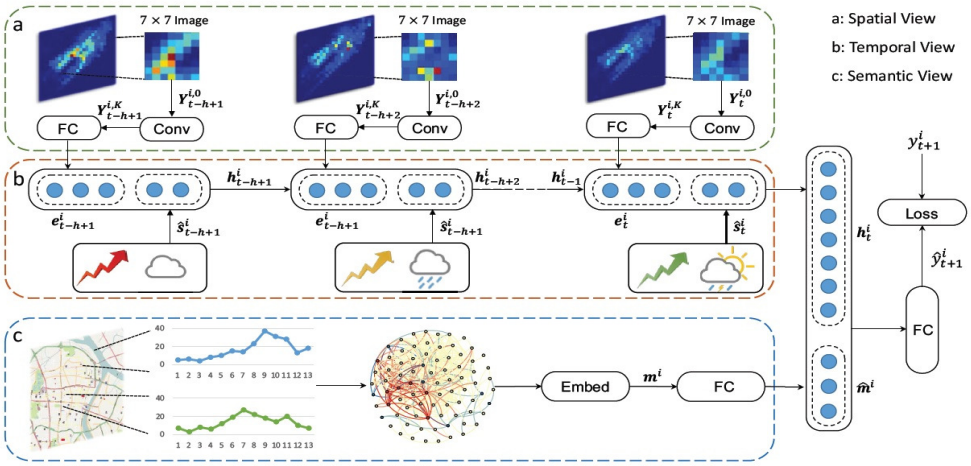


Figure 4.9. Flowchart of PDRNN (Shi *et al.*, 2018).

(concatenation)로 이루어진다. Model B의 4번째 층은 각 시간 단계에서 벡터를 내보내는 연결 층이다. Model A의 5번째 층은 3계층 MLP 은닉 층이다. Model B의 5번째 층은 각 시간 단계에 적용되어 공유되는 MLP 은닉 층이다. 마지막으로 두 모델의 6번째 층은 출력층에 해당한다. Model B의 5번째 층은 3개의 MLP 은닉 층과 비교했을 때 1년 동안의 전력 소비량 예측에는 적합하지 않지만 결측값 대체에 의한 정확도는 Model A보다 높은 결과를 얻었다.

Shi 등 (2018)은 가계의 전력 부하량 예측 연구에 네트워크 층의 갯수를 증가하면서 발생하게 되는 과적합을 방지하고자 RNN을 기반으로 하여 풀링을 적용한 Pooling Deep RNN(PDRNN) 모델을 제안하였다. Figure 4.8은 이 모델에 대한 구조 그림이고 Figure 4.9를 통해 PDRNN의 흐름을 알 수 있다. 새로 도입한 방법은 이웃 가계들의 부하량 데이터를 추가함으로써 새로운 데이터 차원을 만들고 과적합 문제를 해결하기 위해 인접한 가구들 간의 상관관계를 통해 상호 연결된 공간정보를 불충분한 시간 정보로 사용하는 것이다. 결론적으로 RMSE를 기준으로 비교했을 때 ARIMA 모델과 SVR 모델보다 더 나은 성능을 나타냈음을 보였다.

또 Yao 등 (2018)은 이전의 연구들과 달리 공간과 시간의 관계를 동시에 고려하여 모델링 하기 위해 Deep Multi-View Spatial-Temporal Network(DMVST-Net)를 제안하였다. 제안하는 네트워크 방법은 공간(spatial), 시간(temporal), 의미(semantic)의 총 3가지 뷰(view)를 가진다. 여기에서 의미 뷰는 단순히 영역 간 거리가 아닌 기능적 유사성을 나타내는 위치의 그래프를 구성하기 때문에 비슷한 지역이 항상 공간적으로도 가깝지 않다는 것을 고려한다. 예를 들어 주거 지역의 주중 아침은 사람들이 출근하기 때문에 높은 수요량을 가지지만, 상업 지역은 주말에 높은 수요량을 가질 수 있는 것이다.

공간 뷰는 목표로 하는 지역을 예측하기 위해 상대적으로 상관관계가 적은 지역을 배제하고 Local CNN을 통해 인접한 지역만을 고려하게 된다. Figure 4.10과 같이 t라는 시점 간격마다 하나의 장소 i는 채널이 1인 $S \times S \times 1$의 크기인 이미지가 되고 그 장소의 경계는 제로 패딩을 사용한다. 이미지를 하나의 텐서로서 갖게 되고 각 장소 t와 여기에서 $*$는 합성곱을 뜻하며 함수 f는 활성화 함수를 뜻하는데 이 연구에서는 ReLU 함수를 사용하였다. W_t^k, W_t^k는 각각 k번째 컨볼루션 레이어의 가중치와 편차를 뜻하는 파라미터들이다. 출력값 $Y_t^{i,K} \in R^{S \times S \times \lambda}$은 장소$i$와 시간 간격 t에 대해 피쳐 벡터로서 $s_t^i \in R^{s^2 \lambda}$가 된다. 최종적으로 fully connected 층에서 아래와 같은 값을 얻는다.

$$\hat{s}_t^i = f(W_t^{fc} s_t^i + b_t^{fc}), \hat{s}_t^i \in R^d$$

다음으로 시간 뷰는 LSTM네트워크를 통해 모델링 되는데 Figure 4.10에서 알 수 있듯이 공간 뷰로부터 가져와서 컨텍스트 피쳐와 연결하고 다음과 같이 나타낸다.

$$g_t^i = \hat{s}_t^i \oplus e_t^i, g_t^i \in R^{r+d}$$

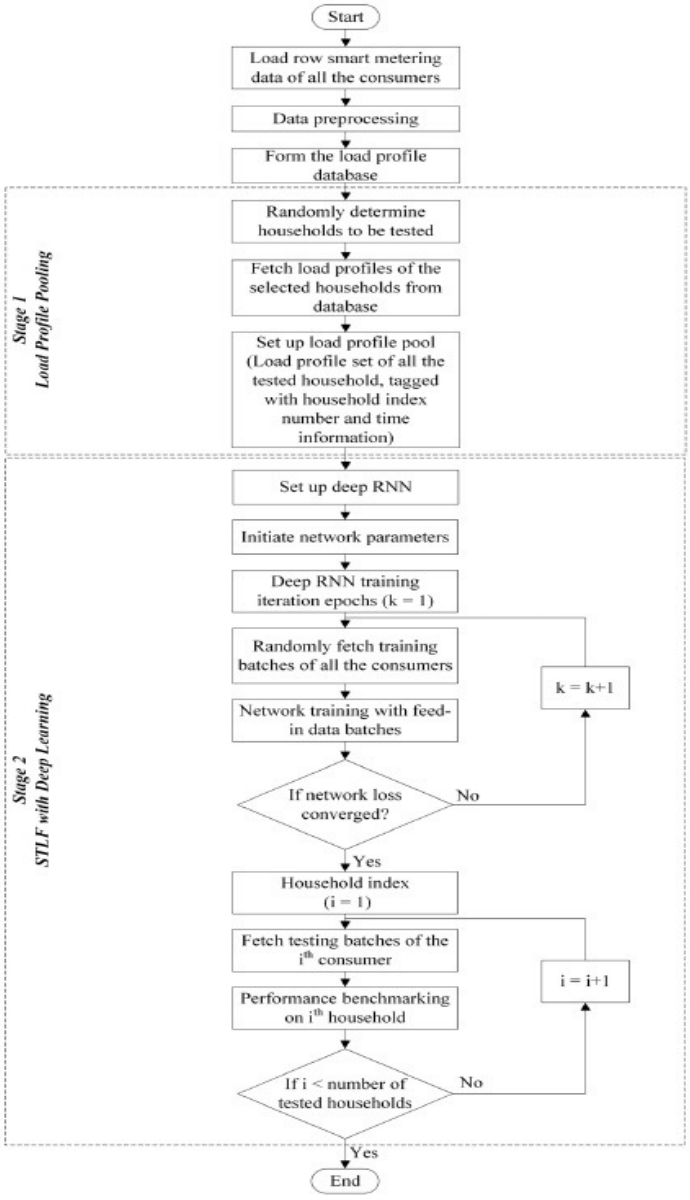


Figure 4.10. Structure of DMVST-Net (Yao *et al.*, 2018).

시간 간격 t에 대해 $Y_t^i \in R^{S\times S\times 1}$이 된다. local CNN은 Y_t^i를 입력 $Y_t^{i,0}$으로 취하고 K개의 컨볼루션 레이어(convolution layer)로 보내며 다음과 같은 식으로 표현된다.

$$Y_t^{i,k} = f(Y_t^{i,k-1} * W_t^k + b_t^k) \tag{4.20}$$

마지막으로 의미 뷰에는 두 개의 시계열 데이터가 존재한다고 할 때 서로간의 유사도를 알아내는 알고리즘 중 하나인 Dynamic Time Wraping(DTW)을 적용했다. 지역i 의 수요패턴과 지역 j의 수요패턴 간의

유사도 ω_{ij}는 $\exp(-\alpha DTW(i,j))$를 통해 구할 수 있으며 해당 연구에서는 거리의 감소율을 제어하기 위한 모수 α의 값을 1로 지정하였다. 주간 수요 시계열의 평균을 수요 패턴으로 사용하고 출력된 각 그래프에 Large-scale Information Network(LINE)임베딩 방법을 적용해서 지역 에 대해 특징벡터 m^i를 출력한다. 그리고 이 벡터는 전체 네트워크 아키텍처와 co-training 하기 위해서 fully connected 층으로 보내지며 다음과 같이 표현된다.

$$\hat{m}^i = f(W_{fc}\mathbf{m}^i + b_{fc}) \tag{4.21}$$

DMVST-Net의 최종 목표인 $t+1$시점에서의 수요예측을 위해 세 개의 뷰로 얻은 $\hat{m}^i$과 LSTM의 출력 값 h_t^i를 연결하여 q_t^i를 얻는다. 마지막으로 예측 값 $\hat{y}_{t+1}^i$은 활성화 함수로 시그모이드 함수를 사용해서 다음과 같이 얻는다.

$$\hat{y}_{t+1}^i = \sigma(W_{ff}\mathbf{q}_t^i + b_{ff}) \tag{4.22}$$

이 때 이 값은 0과 1 사이에서 존재하지만 실제 수요량을 예측하기 위해 다시 비정규화한다. 또 손실함수는 다음과 같이 정의한다.

$$L(\theta) = \sum_{i=1}^{N}\left[(y_{t+1}^i - \hat{y}_{t+1}^i)^2 + \gamma_2(\frac{y_{t+1}^i - \hat{y}_{t+1}^i}{y_{t+1}^i})^2\right]$$

여기에서 θ는 학습 가능한 모든 모수들이고 γ_2는 하이퍼파라미터이다. 실험 결과 해당 네트워크 모델은 다른 모델들, 한 개나 두 개의 뷰로만 이루어진 모델들과 MAPE, RMSE를 기준으로 비교했을 때 더 좋은 성능을 보였다. 이와 같은 원리의 모델인 Conv-LSTM를 적용해서 Ai 등 (2019)은 자전거 대여 시스템의 수요예측을 연구한 사례가 있다.

5.결론

과거보다 다양해진 분야의 시장형태와 소비패턴으로 인해 수요예측 데이터는 더욱 복잡한 관계를 이루고 있다. 그러나 수요예측은 중요한 요소이기 때문에 더 낮은 오차와 더 높은 정확도를 갖춘 수요예측 모델을 적용할 필요가 있다. 본 연구는 머신러닝 또는 딥러닝을 기반으로 수요예측이 활발히 이루어지고 있는 분야들 중 수 전력, 현금수요, Ride-services, 관광, 단기수요, 예비부품 및 신제품 분야로 나누어 기술하였다. 또한, 최근에는 전통적으로 사용되어진 시계열모델이나 회귀모델에서 갖는 한계를 극복하고자 인공지능을 기반으로 한 수요예측 기법을 적용하는 추세이다. 이러한 흐름에 맞게 수요예측이 이루어지는 해당 분야의 데이터 특성을 고려하여 인공지능을 기반으로 한 수요예측 기법을 각각 머신러닝 기법과 딥러닝 기법으로 나누어 설명하였다. 본 논문에서는 이전 리뷰들과는 달리 한 분야에만 그치지 않고 수요예측이 이루어지고 있는 전반적인 분야를 소개하였고, 인공지능 기법에 기반한 다양한 기법들을 소개하였다. 또한, 분야와 기법을 정리하는 것에 그치지 않고 이론적인 설명을 첨가하여 이해를 돕고자 하였다.
눈에 띄는 점으로 국내외 모두 전력수요 분야는 머신러닝 및 딥러닝 기법을 이용한 단기예측에 대한 연구가 많았다는 것이다. 추후에는 머신러닝 및 딥러닝 기법들을 적용하여 전력수요 분야의 중장기예측에 대한 연구가 더욱 활발히 이루어지면 좋을 것 같다. 또한 대부분의 분야들에서 SVR 모델을 기반으로 수요예측 연구를 진행한 사례가 더 많았기 때문에 다양한 딥러닝 기법들을 적용해보면 좋을 것 같다. 그리고 수요예측에는 시계열 데이터를 사용해왔기 때문에 딥러닝 기법으로 이루어진 연구들은 주로 RNN 모델을 기반으로 이루어졌으며 RBM 기반 기법을 적용한 연구들은 전력 수요예측이 대부분이었고 상대적으로 신제품 분야의 수요예측에 관한 연구는 미비했음을 알 수 있었다.
그러나 앞으로는 비정형 데이터인 이미지, 텍스트나 음성 같은 이종의 데이터들을 함께 써서 CNN계열의 최신 모델이나 시퀀스 데이터에 적합한 어텐션 기반의 딥러닝 모델들을 적용할 수도 있을 것으로 생각한다. 뿐만 아니라 이러한 딥러닝 모델들의 결합을 통한 앙상블 기법으로 이루어진 통합예측모델로도 더욱 활발한 수요예측 연구가 이루어질 수 있음을 기대하는 바이다.

References

길소현, 이루다, 양보원, and 최광순. (2016). 딥러닝을 이용한 에너지 수요 예측 방법에 관한 연구,〈 한국통신학회 학술대회논문집〉, 1014–1015.

김인중, 나기현, 양소희, 장재민, 김윤종, 신원영, and 김덕중. (2017). 딥러닝과 통계 모델을 이용한 T-커머스 매출 예측,〈정보과학회논문지〉. 제 44권 8호. 803–812.

김정경. (2016). 국내 · 외 빅데이터 동향 및 성공사례,〈 ie 매거진〉. 제 23권 1호. 47–52.

김태영, and 조성배. (2018). Particle Swarm Optimization 기반 CNN-LSTM 신경망을 이용한 주거용 에너지 소비 예측,〈 한국정보과학회 학술발표논문집〉, 737–739.

김항석, and 신현정. (2013). 준지도 학습 및 신경망 알고리즘을 이용한 전기가격 예측,〈 대한산업공학회지〉. 제 39권 1호. 30–45.

남수현, and 권오병. (2017). 텍스트마이닝과 딥러닝을 활용한 영화 시나리오 분석 및 사전 흥행 예측,〈 한국 IT 서비스학회 학술대회 논문집〉, 515–518.

박수지, 신진옥, 송상헌, and 정철. (2017). 텍스트 마이닝을 통한 관광지 수요예측-온라인 검색 엔진을 중심으로, 〈 관광학연구〉. 제 41권 1호. 13–27.

박준호, 신동하, and 김창복. (2017). 특수일 분리와 예측요소 확장을 이용한 전력수요 예측 딥 러닝 모델, 〈 한국항행학회논문지〉. 제 21권 4호. 365–370.

박지현, 나완수, and 쑤얀. (2017). 딥러닝 알고리즘과 데이터 전처리에 따른 수요 전력 예측 성능 평가, 〈 한국정보과학회 학술발표논문집〉, 1882–1884.

서한석, and 신광섭. (2018). 딥러닝을 이용한 열 수요예측 모델 개발, 〈 한국빅데이터학회지〉. 제 3권 2호. 59–70.

신동하, and 김창복. (2016). 하계 전력수요 예측을 위한 딥 러닝 입력 패턴에 관한 연구,〈 한국정보기술학회논문지〉. 제 14권 11호. 127–134.

안성원, and 조성배. (2010). 뉴스 텍스트 마이닝과 시계열 분석을 이용한 주가예측, 〈 한국정보과학회 학술발표논문집〉. 제 37권 1C호. 364–369.

안준영, 박상민, and 김창복. (2017). 동계 전력수요예측을 위한 신경망 모델에 관한 연구, 〈 한국정보기술학회논문지〉. 제 15권 9호. 1–9.

윤현민, and 김수환. (2017). 데이터마이닝을 활용한 해군함정 수리부속 수요예측, 〈 산업경영시스템학회지〉. 제 40권 4호.253–259.

이상훈, 조장식, 강창완, and 최승배. (2015). 텍스트 마이닝을 활용한 영화흥행 예측 연구, 〈 한국데이터정보과학회지〉.제 26권 6호. 1259–1269.

이오준, 박승보, 정다울, and 유은순. (2014). 소셜 빅데이터를 이용한 영화 흥행 요인 분석, 〈 한국콘텐츠학회논문지〉. 제 14권 10호. 527–538.

이종민, 이종아, and 정준호. (2017). 뉴스 빅데이터를 이용한 전세 가격 예측, 〈 부동산학보〉, 43–57.

이형로, and 신현정. (2011). Support Vector Regression 에 기반한 전력 수요 예측, 〈 IE Interfaces〉. 제 24권 4호.351–361.

임준엽, and 황병연. (2014). 트위터를 이용한 기계학습 기반의 영화흥행 예측, 〈 정보처리학회논문지〉. 제 3권 7호. 263–270.

정호철, 선영규, 이동구, 김수현, 황유민, 심이삭, ... and 김진영. (2019). 에너지인터넷에서 1D-CNN 과 양방향 LSTM 을 이용한 에너지 수요예측, 〈 전기전자학회논문지〉. 제 23권 1호. 134–142.

최석재, 이재웅, and 권오병. (2016). 온라인 뉴스 제목 분석을 통한 특정 장소 이벤트 성과 예측을 위한 형태소 분석 방법, 〈 한국전자거래학회지〉. 제 21권 1호. 15–32.

Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., ... and Wilby, R. L. (2012). Two decades of anarchy? emerging themes and outstanding challenges for neural network river forecasting, *Progress in Physical Geography*, **36(4)**, 480–512.

Ai, Y., Li, Z., Gan, M., Zhang, Y., Yu, D., Chen, W., and Ju, Y. (2019). A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system., *eural Computing and Applications* ,**31**, 1665—1677.

Amirkolaii, K. N., Baboli, A., Shahzad, M. K., and Tonadre, R. (2017). emand forecasting for irregular demands in business aircraft spare parts supply chains by using artificial intelligence(AI), *IFAC-PapersOnLine*, **50(1)**, 15221–15226.

Ansuj, A. P., Camargo, M. E., Radharamanan, R., and Petry, D. G. (1996). ales forecasting using time series and neural networks, *Computers Industrial Engineering,*, **31(1-2)**, 421–424.

Apostolopoulos, P. A., Tsiropoulou, E. E., and Papavassiliou, S. (2018). Demand response management in smart grid networks: a two-stage game-theoretic learning-based approach, *Mobile Networks and Applications*, 1–14.

Bajari, P., Nekipelov, D., Ryan, S. P., and Yang, M. (2015). Machine learning methods for demand estimation, *American Economic Review*, **105(5)**, 481–35.

Bergman, J. J., Noble, J. S., McGarvey, R. G., and Bradley, R. L. (2017). A Bayesian approach to demand forecasting for new equipment programs, *obotics and Computer-Integrated Manufacturing*, **47**, 17–21.

Bier, I. J. (1984). Boeing commercial airplane group spares department: simulation of spare parts operations, *ORSA/TIMS Joint National Meeting, Detroit, MI*

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers., *Proceedings of the fifth annual workshop on Computational learning theory*, **36(4)**, 144–152. ACM.

Brentan, B. M., Luvizotto Jr, E., Herrera, M., Izquierdo, J., and Pérez-García, R. (2017). Hybrid regression model for near real-time urban water demand forecasting, *ournal of Computational and Applied Mathematics,*, **309**, 532–541.

Chae, Y. T., Horesh, R., Hwang, Y., and Lee, Y. M. (2016). rtificial neural network model for forecasting sub-hourly electricity usage in commercial buildings, *Energy and Buildings,*, **111**, 134–194.

Chen, K. Y., and Wang, C. H. (2007). Support vector regression with genetic algorithms in forecasting tourism demand, *Tourism Management*, **28(1)**, 215–226.

Chen, K., Jiang, J., Zheng, F., and Chen, K. (2018). A novel data-driven approach for residential electricity consumption prediction based on ensemble learning, *Energy*, **150,**, 49–60.

Chen, M., Mao, S., Zhang, Y., and Leung, V. C. (2014). Big data: related technologies, challenges and future prospects, *Progress in Physical Geography*

Chintagunta, P. K., Gopinath, S., and Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets., *arketing Science*, **29(5)**, 944–957.

Chu, B. S., and Cao, D. B. (2011). Dynamic cubic neural network with demand momentum for new product sales forecasting., *nformation-An International Interdisciplinary Journal*, **14(4),**, 1171–1182.

Chu, C. W., and Zhang, G. P. (2003). A comparative study of linear and nonlinear models for aggregate retail sales forecasting, *International Journal of production economics*, **86(3)**, 217–231.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands, *Journal of the Operational Research Society*, **23(3)**, 289–303.

Duan, K., Keerthi, S. S., and Poo, A. N. (2003). valuation of simple performance measures for tuning SVM hyperparameters., *Neurocomputing*, **51**, 41–59.

Fan, S., and Chen, L. (2006). Short-term load forecasting based on an adaptive hybrid method, *IEEE Transactions on Power Systems*, **21(1)**, 392–401.

Fischer, A., and Igel, C. (2012). An introduction to restricted boltzmann machines, *iberoamerican congress on pattern recognition*, 14–36, Springer, Berlin, Heidelberg.

Frank, C., Sztandera, L., Vemulapali, B., Garg, A., and Raheja, A. (2004). A fuzzy forecasting model for women's casual sales, *National Textile Center Annual Report*.

Freund, Y. (1995). Boosting a weak learning algorithm by majority, *nformation and computation*, **121(2)**, 256–285.

Friedman, J. H. (2001). greedy function approximation: a gradient boosting machine, *Annals of statistics,*,1189–1232.

Friedman, J. H. (2002). Stochastic gradient boosting, *Computational statistics and data analysis*, **38(4)**, 367–378.

Friedman, J., Hastie, T., and Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors), *The annals of statistics*, **28(2)**, 337–407.

Ghods, L., and Kalantar, M. (2011). Different methods of long-term electric load demand forecasting; a comprehensive review., *Iranian Journal of Electrical Electronic Engineering*, **7(4)**, 249–259.

Hinton, G. E. (2002). Training products of experts by minimizing contrastive divergence, *Neural computation*, **14(8)**, 1771–1800.

Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks, *science*, **313(5786)**, 504–507.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory, *Neural computation*, **9(8)**, 1735–1780.

Holland, J. H. (1992). Adaptation in natural and artificial systems: an introductory analysis with applica-

tions to biology, control, and artificial intelligence,MIT press.

Hong, W. C., Dong, Y., Chen, L. Y., and Wei, S. Y. (2011). SVR with hybrid chaotic genetic algorithms for tourism demand forecasting, *Applied Soft Computing*, **11(2)**, 1881–1890.

Hu, Y. C. (2017). Predicting foreign tourists for the tourism industry using soft computing-based grey–markov models, *Sustainability*, **9(7)**, 1228.

Hua, Z., and Zhang, B. (2006). A hybrid support vector machines and logistic regression approach for forecasting intermittent demand of spare parts, *pplied Mathematics and Computation*, **181(2)**, 1035–1048.

Huang, G. B., Zhu, Q. Y., and Siew, C. K. (2004). Extreme learning machine: a new learning scheme of feedforward neural networks, *Neural networks*, **2**, 985–990.

Jegede, C. A. (2014). Effects of automated teller machine on the performance of nigerian banks, *Journal of Applied Mathematics and Statistics*, **2(1)**, 40–46.

Jeon, S., and Son, Y. S. (2016). Prediction of box office using data mining, *Korean Journal of Applied Statistics*, **29(7)**, 1257–1270.

i, G., Wang, J., Ge, Y., and Liu, H. (2014). Urban water demand forecasting by LS-SVM with tuning based on elitist teaching-learning-based optimization, *The 26th Chinese Control and Decision Conference (2014 CCDC)*, 3997–4002.

Jung, G., Park, J., Kim, Y., and Kim, Y. B. (2017). A modified bootstrap method for intermittent demand forecasting for rare spare parts, *International Journal of Industrial Engineering,*, **24(2)**.

Kang-Lin, P., Wu, C. H., and Yeong-Jia, J. G. (2004). The development of a new statistical technique for relating financial information to stock market returns, *International Journal of Management*, **21(4)**,492.

Kaytez, F., Taplamacioglu, M. C., Cam, E., and Hardalac, F. (2015). Forecasting electricity consumption: a comparison of regression analysis, neural networks and least squares support vector machines, *International Journal of Electrical Power and Energy Systems*, **67**, 431–438.

Kim, B., and Lim, C. (2018). Prediction of movie audience numbers using hybrid model combining GLS and bass models, *Korean Journal of Applied Statistics*, **31(4)**, 447–461.

Kong, W., Dong, Z. Y., Hill, D. J., Luo, F., and Xu, Y. (2017). Short-term residential load forecasting based on resident behaviour learning, *IEEE Transactions on Power Systems*, **33(1)**, 1087–1088.

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., and Zhang, Y. (2017). hort-term residential load forecasting based on LSTM recurrent neural network, *IEEE Transactions on Smart Grid*

Lee, H., Kim, S. G., Park, H. W., and Kang, P. (2014). Pre-launch new product demand forecasting using the Bass model: a statistical and machine learning-based approach, *echnological Forecasting and Social Change*, **86**, 49–64.

Lipton, Z. C., Berkowitz, J., and Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning, arXiv preprint arXiv:1506.00019.

Liu, Y. (2006). Word of mouth for movies: its dynamics and impact on box office revenue, *ournal of marketing*, **70(3)**, 74–89.

Kong, W., Dong, Z. Y., Hill, D. J., Luo, F., and Xu, Y. (2017). Short-term residential load forecasting based on resident behaviour learning, *IEEE Transactions on Power Systems*, **33(1)**, 1087–1088.

Liu, Y., Szeto, W. Y., and Ho, S. C. (2018). A static free-floating bike repositioning problem with multiple heterogeneous vehicles, multiple depots, and multiple visits., *Transportation Research Part C: Emerging Technologies,*, **92**, 208–242.

Lolli, F., Gamberini, R., Regattieri, A., Balugani, E., Gatos, T., and Gucci, S. (2017). Single-hidden layer neural networks for forecasting intermittent demand, *International Journal of Production Economics*, **183**, 116–128.

Lopez, R., Flores, J. J., and Puig, V. (2015). Multi-model forecasting based on a qualitative and quantitative decomposition with nonlinear noise filter applied to the water demand, *ROPEC 2015 (2015 IEEE International Autumn Meeting on Power, Electronics and Computing)*.

Lu, R., and Hong, S. H. (2019). Incentive-based demand response for smart grid with reinforcement learning and deep neural network, *Applied Energy*, **236**, 937–949.

Marino, D. L., Amarasinghe, K., and Manic, M. (2016). uilding energy load forecasting using deep neural networks, *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*, 7046–7051.

Mayrink, V., and Hippert, H. S. (2016). A hybrid method using exponential smoothing and gradient boosting for electrical short-term load forecasting, *2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 1–6

Mnih, V., Larochelle, H., and Hinton, G. E. (2012). Conditional restricted boltzmann machines for structured output prediction,arXiv preprint arXiv:1202.3748.

Mocanu, E., Nguyen, P. H., Gibescu, M., and Kling, W. L. (2014). Comparison of machine learning methods for estimating energy consumption in buildings, *2014 international conference on probabilistic methods applied to power systems (PMAPS)*, 1–6

Mocanu, E., Nguyen, P. H., Gibescu, M., and Kling, W. L. (2016a). Deep learning for estimating building energy consumption, *ustainable Energy, Grids and Networks*, **6**, 91–99.

Mocanu, E., Nguyen, P. H., Gibescu, M., Larsen, E. M., and Pinson, P. (2016b). Demand forecasting at low aggregation levels using factored conditional restricted boltzmann machine, *2016 Power Systems Computation Conference (PSCC)*, 1–7.

Msiza, I. S., Nelwamondo, F. V., and Marwala, T. (2007). Artificial neural networks and support vector machines for water demand time series forecasting, *2007 IEEE International Conference on Systems, Man and Cybernetics*, 638–643.

Nenni, M. E., Giustiniano, L., and Pirolo, L. (2013). Demand forecasting in the fashion industry a review, *International Journal of Engineering Business Management*, **5**, 37.

Orozco, B. P., Abbati, G., and Roberts, S. (2018). Mordred: Memory-based ordinal regression deep neural networks for time series forecasting,arXiv preprint arXiv:1803.09704.

Pai, P. F., Wei-Chiang, H., Ping-Teng, C., and Chen-Tung, C. (2006). The application of support vector machines to forecast tourist arrivals in Barbados: an empirical study, *International Journal of Management*, **23(2)**,375.

Pai, P., and Hong, W. (2005). An improved neural network model in forecasting arrivals, *Annals of Tourism Research*, **32(4)**, 1138–1141.

Pal, A., and Zhang, Y. (2017). Free-floating bike sharing: solving real-life large-scale static rebalancing problems,*Transportation Research Part C: Emerging Technologies*, **80**, 92–116.

Parry, M. E., Cao, Q., and Song, M. (2011). Forecasting new product adoption with probabilistic neural networks, *Journal of Product Innovation Management*, **28(s1)**, 78–88.

Rahman, A., Srikumar, V., and Smith, A. D. (2018). Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks, *Applied energy,*, **212**, 372–385.

Ramanathan, R., Engle, R., Granger, C. W., Vahid-Araghi, F., and Brace, C. (1997). Short-run forecasts of electricity loads and peaks, *International journal of forecasting*, **13(2)**, 161–174.

Regue, R., and Recker, W. (2014). Proactive vehicle routing with inferred demand to solve the bikesharing rebalancing problem, *Transportation Research Part E: Logistics and Transportation Review,*, **72**, 192–209.

Schapire, R. E. (1990). The strength of weak learnability, *Machine learning*, **5(2)**, 197–227.

Schölkopf, B., Smola, A. J., and Bach, F. (2002). Learning with kernels: support vector machines, regularization, optimization, and beyond, *MIT press*.

Şen, A. (2008). The US fashion industry: a supply chain review, *International Journal of Production Economics*, **114(2)**, 571–593.

Setiawan, A., Koprinska, I., and Agelidis, V. G. (2009). Very short-term electricity load demand forecasting using support vector regression, *IEEE 2009 International Joint Conference on Neural Networks*, 2888–2894.

Shalev-Shwartz, S. (2012). Online learning and online convex optimization, *Foundations and Trends® in Machine Learning*, **4(2)**, 107–194.

Shi, H., Xu, M., and Li, R. (2017). Deep learning for household load forecasting—a novel pooling deep RNN, *IEEE Transactions on Smart Grid*, **9(5)**, 5271–5280.

Simoncini, M., Taccari, L., Sambo, F., Bravi, L., Salti, S., and Lori, A. (2018). Vehicle classification from low-frequency GPS data with recurrent neural networks, *ransportation Research Part C: Emerging Technologies*, **91**, 176–191.

Simutis, R., Dilijonas, D., and Bastina, L. (2008). Cash demand forecasting for ATM using neural networks and support vector regression algorithms, *20th International Conference, EURO Mini Conference,*

"Continuous Optimization and Knowledge-Based Technologies"(EurOPT-2008), Selected Papers, Vilnius , 416–421.

Singh, A. K., Ibraheem, S. K., Muazzam, M., and Chaturvedi, D. K. (2013). n overview of electricity demand forecasting techniques, *Network and Complex Systems*, **3(3)**, 38–48.

Song, H., and Liu, H. (2017). Predicting tourist demand using big data, *Analytics in smart tourism design*,13–29.

Sun, X., Sun, W., Wang, J., Zhang, Y., and Gao, Y. (2016). Using a grey–markov model optimized by cuckoo search algorithm to forecast the annual foreign tourist arrivals to China, *Tourism Management*, **52**, 369–379.

Sun, Z. L., Choi, T. M., Au, K. F., and Yu, Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing, *Decision Support Systems*, **46(1)**, 411–419.

Taieb, S. B., and Hyndman, R. J. (2014). gradient boosting approach to the kaggle load forecasting competition., *International journal of forecasting*, **30(2)**, 382–394.

Tang, Z., De Almeida, C., and Fishwick, P. A. (1991). Time series forecasting using neural networks vs. box-jenkins methodology, *Simulation*, **57(5)**, 303–310.

homassey, S., and Happiette, M. (2007). A neural clustering and classification system for sales forecasting of new apparel items, *Applied Soft Computing*, **7(4)**, 1177–1187.

Tiwari, M. K., and Adamowski, J. (2013). Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models, *Water Resources Research*, **49(10)**, 6486–6507.

Tiwari, M. K., and Adamowski, J. F. (2014). Medium-term urban water demand forecasting with limited data using an ensemble wavelet–bootstrap machine-learning approach, *Journal of Water Resources Planning and Management*, **141(2)**, 04014053.

Tiwari, M. K., and Adamowski, J. F. (2017). An ensemble wavelet bootstrap machine learning approach to water demand forecasting: a case study in the city of Calgary, Canada, *Urban Water Journal*, **14(2)**, 185–201.

Venkatesh, K., Ravi, V., Prinzie, A., and Van den Poel, D. (2014). Cash demand forecasting in ATMs by clustering and neural networks, *European Journal of Operational Research*, **232(2)**, 383–392.

Voitcu, O., and Wong, Y. S. (2006). On the construction of a nonlinear recursive predictor, *ournal of Computational and Applied Mathematics*, **190(1-2)**, 393–407.

Wang, C. H. (2004). redicting tourism demand using fuzzy time series and hybrid grey theory, *Tourism management*, **25(3)**, 367–374.

Wang, D., Cao, W., Li, J., and Ye, J. (2017). DeepSD: supply-demand prediction for online car-hailing services using deep neural networks., *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, 243–254.

Wang, Z., and Srinivasan, R. S. (2017). A review of artificial intelligence based building energy use prediction: contrasting the capabilities of single and ensemble prediction models, *Renewable and Sustainable Energy Reviews*, **75**, 796–808.

Wong, W. K., and Guo, Z. X. (2010). A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm, *International Journal of Production Economics*, **128(2)**, 614–624.

Wu, Q. (2009). The forecasting model based on wavelet v-support vector machine, *Expert Systems with Applications*, **36(4)**, 7604–7610.

Wu, Q. (2010). A hybrid-forecasting model based on Gaussian support vector machine and chaotic particle swarm optimization, *xpert Systems with Applications*, **37(3)**, 2388-2394.

Wu, Q., Law, R., and Xu, X. (2012). A sparse gaussian process regression model for tourism demand forecasting in Hong Kong, *Expert Systems with Applications*, **39(5)**, 4769–4774.

Xu, C., Ji, J., and Liu, P. (2018). The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets, *Transportation research part C: emerging technologies*, **95**, 47–60.

Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., ... and Li, Z. (2018). eep multi-view spatial-temporal network for taxi demand prediction, *Thirty-Second AAAI Conference on Artificial Intelligence*.

Yuan, X., Yuan, Y., and Zhang, Y. (2002). A hybrid chaotic genetic algorithm for short-term hydro system scheduling, *Mathematics and Computers in Simulation*, **59(4)**, 319–327.

인공지능 기반 수요예측 기법의 리뷰

정혜린[a] · 임창원[a,1]

[a]중앙대학교 응용통계학과

요 약

최근 다양한 분야에서 '빅데이터'가 생성되었다. 많은 기업들은 인공지능(AI)을 기반으로 빅데이터 분석이 가능한 시스템을 구축하여 이익 창출을 시도하고 있다. 인공지능 기술을 접목함으로써 방대한 양의 데이터를 효율적으로 분석하고 효과적으로 활용하는 것은 점점 더 중요해지고 있다. 특히 재무, 조달, 생산 및 마케팅과 같은 다양한 분야에서 국가 및 기업 경영 관리에있어 최소의 오차와 최대의 정확도를 갖춘 수요예측은 절대적으로 중요한 요소이다. 이 때 각 분야의 수요패턴을 고려한 적절한 모델을 적용하는 것이 중요하다. 전통적으로 쓰이는 시계열모델이나 회귀모델로도 비대해진 실제 데이터의 복잡한 비선형적인 패턴을 분석할 수 있다.
그러나 다양한 비선형 모델들 중에서 적절한 모델을 선택하는 것은 사전 지식 없이는 어려운 일이다. 최근에는 인공지능 기반의 기법들인 머신러닝이나 딥러닝 기법을 중심으로 이루어진 연구들이 이를 극복할 수 있음을 증명하고 있다. 뿐만 아니라 정형데이터와 이미지나 텍스트의 비정형 데이터 분석을 통한 수요예측도 높은 정확도를 갖춘 결과를 보이고 있다. 따라서 본 연구에서는 수요예측이 비교적 활발하게 일어나는 중요한 분야들을 나누어 설명하였다. 그리고 각 분야별로 갖는 특징적인 성격을 고려한 인공지능 기반의 수요예측 기법에 대해 머신러닝과 딥러닝 기법으로 나누어 소개하였다.

주요용어: 빅데이터, 인공지능, 수요예측, 머신러닝, 딥 러닝

이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단-차세대정보 컴퓨팅기술개발사업의 지원을 받아 수행된 연구임 (NRF-2017M3C4A7083281).

[1]교신저자: (06974) 서울특별시 동작구 흑석로 84, 중앙대학교 응용통계학과. E-mail: clim@cau.ac.kr