Global Conditional Wavenet을 응용한 전력 수요량 예측 모델

• 분석 대상 기간 : ‘23.3.13(월) ~ 3.19(일)(1주일)

※ 모델 간 성능 비교를 위해 반드시 대상 기간에 대해 분석할 것

• 향후 6시간까지의 5분 단위 예측 모델을 제시할 것

• 모델의 성능 지표는 MAPE(Mean Absolute Percentage Error)로 제시

- 제시한 예측모델을 실행하여 과거 수요 데이터로 검증한 예측실적치와 MAPE를 분석 대상 기간에 대해 제시

분석결과 정확도

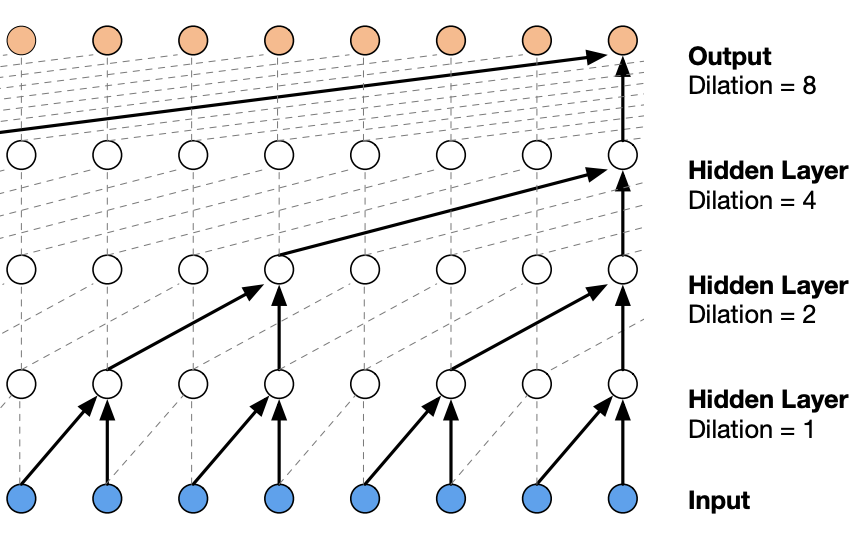
분석방법의 적정성 및 참신성

분석결과의 적용가능성

분석 내용 및 분석 결과

**2. 1. 1. Wavenet 개요**

Wavenet은 2016년 딥마인드(DeepMind)에서 공개한 음성 생성 모델이다. 이 모델은 첫 번째 시점부터 t-1시점까지의 음성 신호를 입력으로 받아 다음 스텝(t)의 음성 신호를 예측하는데, 구체적으로는 Dilated Causal Convolution을 여러 번 반복하여 입력 시계열의 특징 벡터를 효과적으로 추출한 뒤 수 차례의 Convolution을 통해 다음 스텝(t)의 신호를 예측하는 방식이다. Dilated Causal Convolution은 Dilation과 Causal Convolution의 합성어이다. Dilation은 그림 1과 같이 convolution 층을 반복함에 따라 수용장(Receptive Field)을 늘리는 연산으로 가까운 시점의 정보뿐만 아니라 먼 과거의 정보까지 효과적으로 추출할 수 있는 계산 기법이다. Causal Convolution이란 padding 과정에서 단방향성(unidirectionality), 즉 n 시점의 정보는 더 이전 시점의 정보에 영향을 끼치지 않는다는 성질을 반영하기 위해 특정 시점의 특징 벡터를 추출하기 위해 이전 시점의 입력값만 Convolution 연산에 활용하는 기법이다. Wavenet은 Dilation rate, 즉 수용장의 크기를 (2, 4, … , 512)로 점차 늘리면서 Dilated Causal Convolution을 반복해 나감으로써 입력 데이터를 효과적으로 모델링한다.



<그림 1. Dilated Causal Convolution 개요>

텍스트, 스크린샷, 도표, 폰트이(가) 표시된 사진

자동 생성된 설명

<그림 2. Wavenet 내 Residual Block 개요>

Wavenet은 <그림 2>와 같이 2의 배수로 증가하는 Dilation rates을 가지는 각기 다른 ResidualBlock들을 순차적으로 결합하여 특정 시점(t)의 특징 벡터를 추출한다. 이 모델이 긴 음성 신호를 모델링하는 방식은 자기회귀(Autoregressive, 이하 AR)적이다. AR이란 출력 변수가 자신의 이전 값에 의존한다는 것인데, Wavenet은 한 스텝(t)의 신호를 예측한 뒤 해당 예측 신호를 입력 시계열에 추가하여 그 다음 스텝(t+1)의 신호를 예측하는 과정을 반복하여 긴 음성 신호를 만들어내기 때문이다. 학습은 단일 시점 예측 데이터와 실측치 간의 L2-loss로 이루어진다. 외생변수를 조건으로 추가하지 않고 오로지 내생변수만을 활용한 Wavenet을 Unconditional Wavenet이라고 한다.

**2.1.2. Global Conditional Wavenet 개요**

기존의 Wavenet에 각종 외생변수(Exogeneous Variable)를 조건(Condition)으로 입력한 Conditional Wavenet의 Residual Block의 구조는 다음과 같다. 매 Residual Block의 연산에서 해당 시점의 외생변수에 Convolution을 시행해 추출한 특징 벡터가 기존의 Dilated Causal Convolution을 거친 내생변수와 더해진다. 시점에 관계없이 동일한 정보를 조건으로 입력하는 경우를 시간 독립적인(Time-independent) Global Conditional Wavenet이라고 하고 매 시점마다 변화하는 조건을 담는 경우를 시간 의존적인(Time-dependent) Local Conditional Wavenet이라고 한다.

텍스트, 도표, 스크린샷, 라인이(가) 표시된 사진

자동 생성된 설명

<그림 3. Conditional Wavenet의 개요>

**2.1.3. Global Conditional Wavenet을 활용한 전력 수요 예측 모델 골자**

본 분석과제 4는 향후 6시간까지의 5분 단위 전력 수요 예측 모델을 개발하는 것으로, 6시간, 즉총 72(5분X72) 시점의 전력수요를 예측해야 한다. Wavenet 기반 모델들은 단일 스텝 예측을 수행하므로 이를 본 대회에 응용하기 위해서는 2.1.1절에서 소개한 AR을 바탕으로 72개의 Wavenet 기반 모델들을 직렬 연결하여야 한다.

12개의 외생변수들을 모델링하는 과정에서는 Global Conditional Wavenet을 활용한다. 기온, 습도 등의 기상 정보 역시 시간에 따라 변화하는 지역 조건(Local Condition)이다. 그러나 72개 시점의 전력수요를 예측해야 하는 본 과제의 특성상 Local Conditional Wavenet을 이용하게 되면 내생변수에 대한 예측 뿐만 아니라 외생변수들(Exogeneous Variable)에 대한 예측 역시 매 시점 수행해야 하고 이로 인해 불필요하게 오류가 누적되거나 파라미터의 개수가 기하급수적으로 증가하게 되는 문제가 발생한다. 기존의 Multi-step Prediction에 해당하는 분석과제가 Multi-step, Multi-Dimension Prediction으로 변하게 되고 예측 정확도가 불가피하게 하락하는 결과를 초래할 수 있다.

Global Conditional Wavenet을 활용해 본 과제를 수행하기 위해서는 외생변수들의 시점을 고정해 둔 채로 특징 벡터를 추출하여 입력해야 한다. 이를 위해서는 t시점부터 t+71시점까지 총 72개 시점의 전력수요를 순차적으로 예측하는 72개의 Wavenet들 중 t번째 스텝을 예측하는 첫 번째 Wavenet의 Residual Block에만 기상 정보 등 외생변수의 특징 벡터를 입력한다. 즉 t시점 이전의 내, 외생변수를 모두 활용한 1개의 Global Conditional Wavenet으로 t번째 스텝의 전력수요를 예측해 내고 이후 시점의 예측부터는 외생변수를 입력으로 넣지 않고 오로지 내생변수만을 활용하는 Unconditional Wavenet으로 t+1번째 스텝의 전력수요를 예측해 내는 과정을 71번 반복하는 것이다. 자세한 내용은 3.1절 모델 세부 정보에서 설명한다.

**2. 2. 데이터 수집**

**2. 2. 1. 내생변수(Endogeneous Variable, 약칭 endog) 수집**

본 분석과제의 예측대상 변수인 전력수요를 내생변수로 정의한다. 한국전력거래소에서 제공하는 5분단위 전력수급현황을 데이터로 활용하여 2014년 1월 1일부터 2023년 4월 30일까지의 데이터를 수집하였다.

**2. 2. 2. 외생변수(Exogeneous Variable, 약칭 exog) 수집**

기상정보, 전력기본정보 등 전력수요를 예측하기 위해 활용되는 여러 변수를 외생변수로 정의한다. 먼저 기상정보 데이터를 수집하기 위해 기상자료개방포털에서 제공하는 기상청 예보 및 실적을 크롤링해 서울, 수원, 대전, 대구 등 17개 구역의 시간별 기온, 강수량, 풍속 정보를 수집했다. 내생변수는 5분 단위로 정렬되어 있기 때문에 기존의 시간 단위 기상정보를 선형보간(Linear Interpolation)을 활용해 5분 단위로 업샘플링(Upsampling)하여 그 차원을 같게 하였다. 공급능력, 최대예측수요, 공급예비력, 공급예비율, 운영예비력, 운영예비율 정보는 한국전력거래소의 5분단위 전력수급현황에서 수집하였다.

**2. 3. 데이터 전처리**

**2. 3. 1. 내생변수 전처리: 추세 정보 반영**

Kim et al. (2004)에 따르면 인공 신경망(Artificial Neural Network)을 활용한 모델은 추세(Trend)나계절성(Seasonality)등 존재 시 시계열 데이터의 정상성(Stationarity)을 해치는 성질에 상당히 robust한 특징이 있기에 별도의 계절성 제거 등이 필요하지 않다. 다만 데이터의 샘플링 주기가 5분으로 매우 짧고 전체 데이터셋의 기간이 약 10년 정도이기 때문에 72 시점의 전력수요를 예측하는 비교적 미시적인 과제를 수행할 때 추세와 같은 거시적인 정보를 놓칠 수 있다. 이에 추세 정보를 외생변수로 활용하였다.

MSTL(Multiple Seasonal-Trend Decomposition using LOESS)은 다중 계절성이 있는 데이터의 계절성과 추세를 효과적으로 모델링하는 알고리즘이다. 예측 대상 변수인 전력수요는 하루 24시간의 날씨 변화에 반응하기에 일별 계절성, 4계절의 변화에 반응하기에 계절별 계절성을 모두 가지는 다중 계절성 시계열 데이터에 해당한다. 따라서 MSTL 알고리즘을 활용해 추세 정보를 추출하는 것이 적합하다. MSTL 알고리즘으로 전체 데이터셋의 추세 정보를 얻을 수 있었고 이를 ‘trend’라는 이름의 외생변수로 저장했다. <그림 4>는 MSTL 알고리즘을 바탕으로 추출한 추세 정보이다. 추세 그래프가 주기성을 가지면서 점차 상승하는 것을 확인할 수 있고 이는 시간이 지남에 따라 전력 수요가 점차 늘어나는 사회적인 현상과 맞닿아 있다.

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명

<그림 4. 전력 수요 추세>

**2. 3. 2. 외생변수 전처리: 차원 축소**

기상자료개방포털에서 얻은 기상 데이터는 서울, 세종, 제주 등 17개의 주요 거점 지역에 대한 기온, 강수량, 풍속 데이터로 총 51종류이다. 이에 희소성 문제(Sparsity problem)으로 인한 학습 효율성 저하를 방지하고자 외생변수의 개수를 줄이고자 한다.

먼저 기온의 경우 <그림 5>에서 확인할 수 있듯 모든 지역의 온도의 상관관계가 최소 0.9를 넘는다. 이는 지역별 온도 간 강한 비례성이 존재한다는 것을 의미하기에 전체 지역의 기온을 산술평균하여 ‘mean\_temp’이라는 한 개의 외생변수로 그 차원을 축소했다.

강수의 경우 <그림 6>에서 볼 수 있듯 인천과 서울, 울산과 창원 등의 몇몇 지역을 제외하고는 유의미한 상관관계를 찾을 수 없었다. 주성분 분석(Principal Component Analysis, 이하 PCA)을 활용하여 17개 지역의 강수량을 차원 축소한다. <그림 7>의 설명 분산비(Explained variance ratio)를 기준으로 하여 축소 후 차원의 개수를 결정하였고 실험 결과 4개의 성분으로 축소하는 것이 가장 합리적이라 판단하여 ‘rain\_1’, ‘rain\_2’, ‘rain\_3’, ‘rain\_4’로 이루어진 네 개의 외생변수로 그 차원을 축소했다.

풍속의 경우는 <그림 8>에서 볼 수 있듯 기온보다는 상관관계가 낮으나 강수보다는 상관관계가 높았다. 대부분의 지역들의 1대1 상관계수가 약한 양의 상관 값을 가졌다. 이에 PCA를 진행하였고 <그림 9>의 설명 분산비를 기준으로 실험한 결과 4개의 성분 ‘wind\_1’, ‘wind\_2’, ‘wind\_3’, ‘wind\_4’로 축소하기로 결정했다.

텍스트, 스크린샷, 패턴, 사각형이(가) 표시된 사진

자동 생성된 설명

<그림 5. 지역간 온도 상관관계>

텍스트, 스크린샷, 패턴, 사각형이(가) 표시된 사진

자동 생성된 설명

<그림 6. 지역 간 강수량 상관관계>

텍스트, 라인, 그래프, 도표이(가) 표시된 사진

자동 생성된 설명

<그림 7. 지역 간 강수량 설명 분산비>

텍스트, 스크린샷, 패턴, 다채로움이(가) 표시된 사진

자동 생성된 설명

<그림 8. 지역 간 풍속 상관관계>

텍스트, 스크린샷, 라인, 도표이(가) 표시된 사진

자동 생성된 설명

<그림 9. 풍속 PCA 설명 분산비>

**2. 3. 3. 외생변수 전처리: 변수 제거**

전력수요 예측에 필요하지 않은 변수들을 제거하기 위해 전력수요와의 상관관계 분석을 실시하였고 강수량 변수들은 상관계수가 -0.1에서 0.1사이로 전력수요와 큰 상관관계를 가지지 않는 것으로 판단하여 강수량 관련 외생변수들을 모두 삭제했다.

**2. 3. 4. 데이터 전처리: 정규화**

마련한 데이터는 ‘현재수요(MW)’라고 하는 내생변수와 12개의 외생변수들로 이루어져 있다. 그러나 내생변수는 단일 입력으로, 외생변수는 해당 입력에 더해지는 조건으로 분리 처리되기 때문에 내생변수에 대해서는 별도의 표준화(Standardization)나 정규화(Normalization)를 수행하지 않아도 학습에 지장을 끼치지 않는다. 외생변수는 MW, 섭씨, 퍼센트 등 다양한 범위(scale)을 가지기 때문에 그 범위를 0부터 1로 정규화한다.

2.3.4. 데이터 전처리: 학습 데이터 범위 설정

본 공모전의 분석과제 4에서 규정하는 분석 대상 기간은 2023년 3월 13일 00시부터 2023년 3월 19일 24시까지의 일주일이다. 따라서 분석의 타당성을 위해 해당 기간 내 정보는 학습 과정에서 이용할 수 없으며, 오로지 모델 성능 평가 시에만 접근할 수 있다. 전체 데이터셋에서 분석 대상 기간 이전인 2023년 3월 12일 이전의 데이터만 분리해 학습을 위한 데이터셋을 마련했다. 이렇게 전처리된 데이터는 다음과 같이 요약할 수 있다.

|  |  |  |
| --- | --- | --- |
| 변수 종류 | 내생변수 | 외생변수 |
| 차원 | (965573, 1) | (965573, 12) |
| 명칭 | 현재수요(MW) | 공급능력(MW), 최대예측수요(MW), 공급예비력(MW), 공급예비율(퍼센트), 운영예비력(MW), 운영예비율(퍼센트), mean\_temp, wind\_1, wind\_2, wind\_3, wind\_4, trend |

<표 1. 데이터 전처리 후 변수 차원 및 세부 정보>

**2. 4. 데이터 분리**

**2. 4. 1. 하이퍼파라미터 설명**

데이터를 학습 과정에 활용될 학습데이터(Train data) 및 검증 데이터(Validation data)와, 성능 평가 과정에 활용될 실험 데이터(Test data)로 나눈다. 데이터셋 분리 및 제작 과정 내 주요 하이퍼파라미터는 다음과 같이 정의할 수 있다. 먼저 train\_size는 한 개의 스텝을 예측하는 데 필요한 입력 시퀀스의 길이를 의미한다. 필자는 2의 배수로 늘어나는 dilation rates와의 호환성을 높이기 위해 train\_size를 512와 1024로 각각 정의해 실험을 진행하였다. 일례로 train\_size가 512라는 것은 2023년 2월 6일 18시 40분의 수요 예측을 위해서 활용된 데이터가 2023년 2월 4일 00시 00분~2023년 2월 6일 18시 35분까지(512 스텝)라는 의미이다.

두 번째는 step\_size이다. step\_size는 데이터의 양이 약 97만 개로 아주 방대하기에 가상 환경 용량 확보 차원에서 몇몇 시점의 데이터만 샘플링하기 위해 건너뛰는 시계열의 길이를 의미한다. 예를 들어 step\_size가 3이라고 한다면 데이터 샘플링은 15분(5분X3)에 한번씩 이루어진다는 의미이다. step\_size가 커질수록 샘플링하는 데이터의 개수가 적다. 본 모델에서 step\_size는 GPU의 연산 가능 범위 내에서 최대한 많은 데이터를 활용하기 위해 최대한 작게 설정하였고 train\_size가 512인 경우 5, 1024인 경우 13으로 설정했다. 만일 데이터가 시간 주기성(Hourly Seasonality)을 가진다면 step\_size가 6, 12 등이라고 할 때 데이터의 시간 주기성을 편향되게 학습할 수 있기에 주기성을 없앨 수 있는 무작위한 사이즈로 설정하였다.

마지막은 test\_size이다. test\_size는 한 번에 출력하는 출력 시퀀스의 길이로 본 분석과제에 따라 이는 72로 고정된다. test\_size는 곧 몇 개의 Wavenet 모델을 잇는지를 결정하는. 척도이며 모델의 파라미터 개수 및 연산 복잡도에 직접적인 영향을 끼친다.

**2. 4. 2. 데이터 분리 과정**

이렇게 설정한 하이퍼파라미터를 바탕으로 데이터셋을 분리 및 제작하는 과정은 다음과 같다. 먼저 사전 정의된 입력 시퀀스 길이(train\_size)와 출력 시퀀스 길이(test\_size)를 기준으로 데이터셋을 분리한 뒤 시계열 스킵 길이(step\_size)에 맞추어 (입력 시퀀스, 출력 시퀀스)의 순서쌍을 마련한다. <그림 4>는 {train\_size = 4, test\_size = 2, step\_size = 1}로 파라미터 설정 후 데이터를 분리하는 과정을 나타내는 도식이다.

(입력 시퀀스, 출력 시퀀스) 순서쌍을 마련한 후 학습 데이터(train data)와 검증 데이터(validation data) 간 독립성을 보장하기 위해 시간 인덱스에 독립적으로 랜덤 샘플링을 진행해 80%의 학습 데이터와 20%의 검증 데이터로 분리한다. 가령 학습 데이터를 2022년 이전 순서쌍, 검증 데이터를 2022년 이후 순서쌍이라고 하게 되면 데이터가 과거 데이터의 추세에 과적합할 수 있다. 따라서 시간에 독립적인 랜덤 샘플링에 의거해 데이터를 분리한다.

텍스트, 스크린샷, 도표, 라인이(가) 표시된 사진

자동 생성된 설명

<그림 10. 데이터 분리 과정>

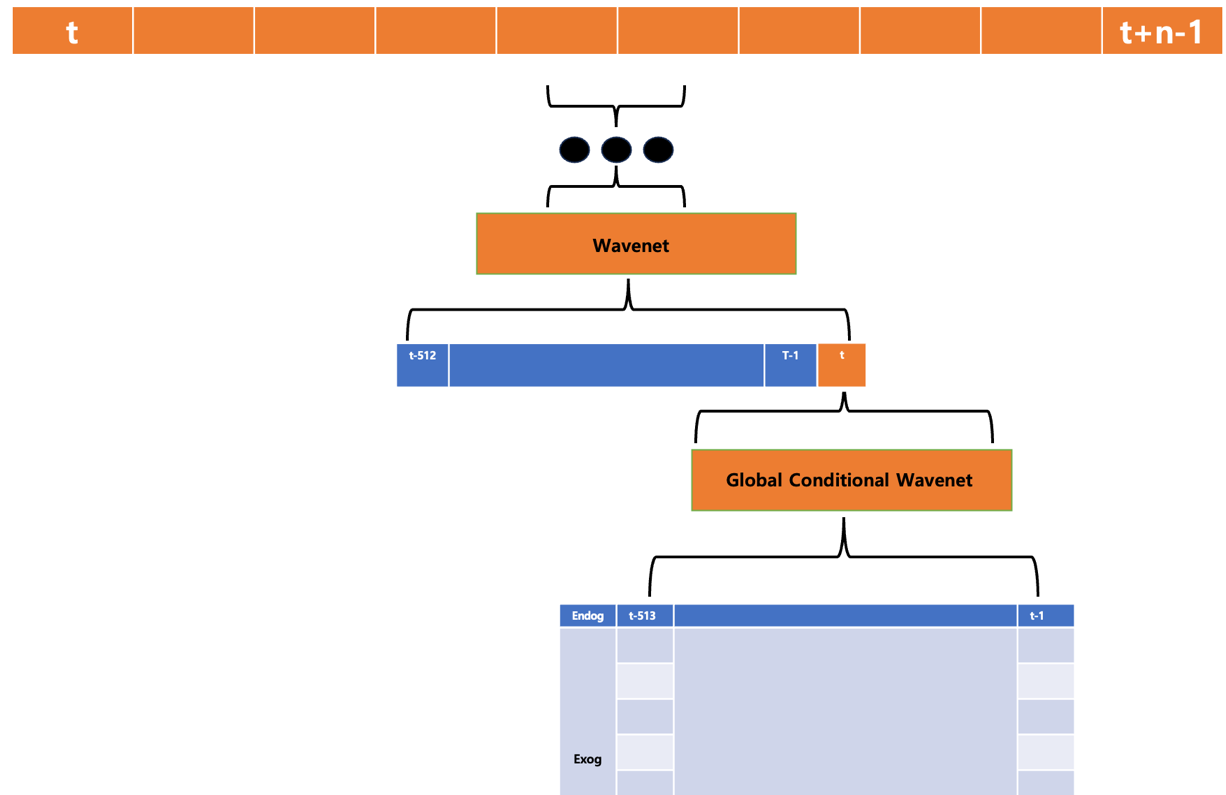
이렇게 분리된 데이터셋은 내생변수에 해당하는 ‘현재수요(MW)’의 입력 시퀀스 모음, ‘현재수요(MW)’의 출력 시퀀스 모음, 외생변수들의 입력 시퀀스 모음, 외생변수들의 출력 시퀀스 모음의 네 가지로 분리되게 된다. 이들 중 외생변수들의 출력 시퀀스 모음은 불필요하기에 아래의 <표 1>과 같이 데이터를 정리할 수 있게 된다. <표 1>은 {train\_size = 512, test\_size = 72, step\_size = 5, validation\_size = 0.2}로 데이터를 분리한 상황을 나타내는 표이다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 데이터명 | 데이터 종류 | 데이터 설명 | 데이터 차원  (#sample, #window size, #feature size) | 변수 종류 |
| X\_endog\_train | 학습(train) | 내생변수의 입력 시퀀스 | (154398, 512, 1) | 설명변수 |
| X\_exog\_train | 외생변수의 입력 시퀀스 | (154398, 512, 12) |
| y\_endog\_train | 내생변수의 출력 시퀀스 | (154398, 72, 1) | 반응변수 |
| X\_endog\_test | 검증(validation) | 내생변수의 입력 시퀀스 | (38600, 512, 1) | 설명변수 |
| X\_exog\_test | 외생변수의 입력 시퀀스 | (38600, 512, 12) |
| y\_endog\_test | 내생변수의 출력 시퀀스 | ( 38600, 72, 1) | 반응변수 |

<표 2. 데이터 활용 개요>

**3. 1. 모델 세부 정보**

이 모델은 2.1.3. 절에서 소개했듯 1개의 Global Conditional Wavenet과 71개의 Unconditional Wavenet으로 이루어져 있다. <그림 11>은 예측 모델의 구조를 간단하게 표현한 그림이다.



<그림 11. 모델 개괄>

먼저 Global Conditional Wavenet은 dilation rates가 각각 다른 Residual Block들의 시퀀스와 이후 수 차례의 Convolution을 통해 스텝 t의 값을 예측하는 모델이다. <그림 3>에서 Residual Block의 형태를 확인해 볼 수 있는데, Residual Block에서는 내생변수에 Dilated causal convolution 연산을 시행해 시간 순차적으로 연산된 특징 벡터를 계산한다. 이후 별개의 convolution 연산을 통해 추출된 외생변수의 특징 벡터와 더해진 뒤 기타 연산을 수행한다. Residual connection으로 입력값을 한번 더 더하는 과정을 통해 장기기억 의존도(Long-term dependency)를 고려하고자 했다. 이 층을 2의 배수로 늘어나는 dilation rates를 달리하여 반복하고 최종 층에서는 Convolution과 Fully Connected Layer 연산을 통해 스텝 t의 전력수요를 예측하게 된다. 이렇게 예측된 스텝 t의 전력수요는 이전 시점들의 전력수요들과 합쳐져 다음 Wavenet의 입력이 된다. 두 번째 시점의 Wavenet 부터는 Global Conditional Wavenet과 다르게 외생변수를 입력하지 않는다는 특징이 있다. <그림 11>에서 나와있듯, 이 과정을 반복하게 되면 총 72 스텝의 값을 순차적으로 예측할 수 있게 된다. 이후 y\_endog\_train, 즉 내생변수의 출력 시퀀스와의 MAPE를 계산하여 손실을 순차적으로 역전파하며 adam optimizer에 의거해 학습한다.

**3.2. 실험 결과**

다음 <표 3>은 각각 train\_size 등 데이터셋 하이퍼파라미터와 filters라는 모델 내 하이퍼파라미터를 달리하여 실험한 결과이다. 실험 결과에 의거해 최종 분석과제 수행을 위해 활용할 모델은 하이퍼파라미터가 {train\_size = 512, test\_size = 72, step\_size = 5, #filters = 18}인 모델이다.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Train\_size | Test\_size | Step\_size | #filters | #Parameters | Training time(s/epoch) | MAPE |
| 512 | 72 | 5 | 9 | 67000+- | 2787 s | 9.1313 |
| 512 | 72 | 5 | 18 | 84000+- | 3000 s | 9.0931 |
| 1024 | 72 | 13 | 9 | 82000+- | 2878 s | 9.1734 |

<표 3. 모델 실험 결과>

3.3. 분석 대상 데이터 실험

3.2.절의 실험 결과는