Abstract

전력 소모량, 태양광 발전량, 피아노의 다음 운율 등 다변량 시계열 데이터의 예측은 매우 가치있다. 그러나, time step과 series 간의 복잡하고 비선형적인 상호 의존 관계는 예측을 까다롭게 한다. 정확한 예측을 얻기 위해 시계열 데이터 내 장기 의존성 모델링은 중요한데, 이는 RNN과 Attention mechanism으로 달성될 수 있다. 전형적인 Attention mechanism은 정보를 각 이전의 tiem step을 되새기고, 관련된 정보를 선택함으로써 결과 예측을 돕는다, 그러나, 이는 다양한 tiem step에 걸친 일시적인 패턴을 잡지 못한다. 이 논문에서, 우리는 시계열 데이터를 frequency domain으로 변경하는 것과 유사한, 시간에 강건한 일시적인 패턴을 추출하기 위한 filter 사용을 제안한다. 또한 우리는 관련된 시계열을 선택하는 우리의 고유한 attention mechanism을 제안하고, 다변량 예측을 위해 이의 frequency domain 정보를 사용한다. 우리는 제안한 모델을 현실의 다양한 문제에 적용하고, 거의 대부분의 경우에서 SOTA를 달성했다.

1. Introduction

일상 속에서 시계열 데이터는 어디에든 있다. 우리는 이산적인 time step에 걸쳐 센서로부터 생성되어 발전하는 변수를 관측하고 이를 시계열 데이터로 조직화한다. 예를 들어, 가구 전력 소모량이나, 도로 사용률, 환율, 태양광 발전 등이 있다. 대부분의 경우 수집된 데이터는 종종 다변량 시계열 데이터이다. (aka. MTS). 예를 들어 다양한 고객들의 전력 소모량 등이 있다. MTS에는 서로 다른 series 간의 복잡하고 다양한 상호 의존 관계가 있는데, 이는 중요하지만 인식하고 분석하기 어렵다.

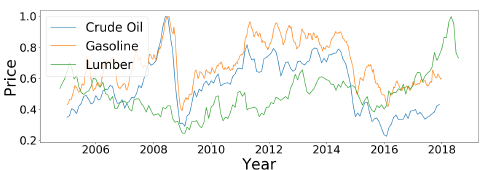


Figure 1.

분석가들은 종종 쌓인 데이터를 활용해 미래를 예측한다. 서로 다른 series 간의 더 나은 상호 의존 관계가 모델링 될수록, 예측률은 더 높아진다. 예를 들어, Figure 1 내의 crude oil 가격은 gasoline 가격에 큰 영향을 받지만, lumber의 가격에는 영향을 덜 받는다. 그러므로 우리는 crude oil의 가격으로부터 gasoline 가격을 예측할 수 있다.

기계학습에서 우리는 데이터 내의 이러한 상호 의존 관계를 모델이 학습하길 원한다. 기계 학습은 시계열 데이터 분석의 분류와 예측 모두에 적용되었다…. 해당 문단 생략

Deep learning을 이용한 MTS 예측에서 RNN이 종종 사용되었다. 그러나, RNN의 시계열 분석은 장기간 의존성에서의 약점이 단점으로 작용한다. Attention mechanism은 이러한 문제를 해결했고, RNN의 효율성을 제고시켰다.

이 논문에서 우리는 MTS 예측을 위한 새로운 Attention mechanism은 temporal pattern attention을 제안한다. 여기서 temporal pattern은 여러 개의 time step에 걸친, 시간에 강건한 패턴과 관련된 모든 것이다. 일반적인 attention mechanism은 예측과 관련된 time step을 명시하고, 이러한 time step에서 정보를 추출하는데 이는 MTS 예측에 있어 명백히 제한이 존재한다. Figure 1에서 gasoline 가격을 예측할 때, 모델은 crude oil에 집중하고, lumber는 무시하고 학습해야 한다. 일반적인 attention mechanism에서와 같이 관련성 있는 time step을 선택하는 대신, TPA는 관련있는 time seires를 선택한다.

또한, 시계열 데이터는 가끔 인식 가능하고 주기적인, 일시적인 패턴을 수반하는데, 이는 예측에 매우 중요하다. 그러나 여러 단계의 time step에 걸친 주기적인 패턴은 전형적인 attention mechanism에 있어서는 파악하기 어려운데, 이는 일반적으로 약간의 time step에 집중하기 때문이다. TPA에서, 우리는 각각 변수로부터 temporal pattern 정보를 추출하기 위한 CNN을 소개한다.

이 논문에서의 주요 내용 요약은 아래와 같다.

* 우리는 관련된 time step을 선택하는 것이 아닌, 관련된 변수를 선택하는 새로운 개념의 attention을 소개한다. 이 방법은 간단하며 RNN에 일반적으로 적용 가능하다.
* 우리는 test example을 이용해 우리의 attention mechanism이 모델로 하여금 temporal pattern을 추출할 수 있게 하고, 서로 다른 time series에 대한 서로 다른 time step에 집중할 수 있게 한다는 것을 입증했다.
* 주기적인 데이터, 부분 선형 데이터, 비주기적이고 비선형적인 현실 데이터에 대해 인증된 결과로 우리는 제안된 attention mechanism이 SOTA를 달성했음을 보여준다.
* TPA 내의 학습된 CNN filter는 흥미롭고 해석 가능한 행동을 보여준다.

나머지 문단 생략

1. 관련 연구

가장 잘 알려진 단변량 선형 시계열 예측 모델은 ARIMA이다. 이는 다른 자기 회귀 모델인 AR, MA, ARMA를 포함한다. 추가적으로, Linear Support Vector Regression은 시간에 변화하는 parameter들로 예측 문제를 전형적인 회귀 문제로 풀었다. 그러나, 이러한 모델들은 대개 단변량 시계열에 제한되어있어 MTS에 적합하지 않다. MTS 에측을 위해 AR 기반의 일반화 모델인 Vector Autoregression이 제안되었다. VAR은 MTS 예측에서 아마 가장 잘 알려진 모델일 것이다. 그럼에도 불구하고 AR 기반 모델과 VAR 기반 모델 모두 비선형성에 대한 한계가 있었다. 이러한 이유로, 시계열 예측을 위한 비선형 모델이 연구되었다. 이에 kernel method 기반과 ensemble 기반, gaussian process 기반 또는 regime switching 등이 있다. 여전히, 이러한 접근은 비선형성을 가정함으로써, 다른 MTS에 대한 비선형성을 잡지 못했다.

최근 Deep neural network는 비선형적인 상호 의존 관계를 잡아내는 능력으로 많은 관심을 받았다. RNN의 변종인 LSTM은 수많은 NLP 문제에서 전망이 좋은 결과를 보여주었고, MTS 예측에도 적용되었다. 이 분야는 기본 RNN의 사용을 시작으로 ARIMA와 MLP의 결합으로 발전하였고, 최근에는 Dynamic Boltzmann Machine과 RNN의 결합으로까지 발전했다. 이러한 모델들이 MTS에 적용될 수 있음에도 불구하고, 이들은 일반적으로 단변량 또는 이변량 시계열 데이터에 주로 적용되었다.

알려진 바로는 Long- and Short-term time-series network (LSTNet)이 다양한 time series에 대해 MTS 예측을 목표로 고안된 첫 번째 모델이다. LSTNet은 CNN을 이용해 short-term pattern을 인식하고, LSTM 또는 GRU가 상대적으로 긴 패턴을 기억한다. 그러나 현실적으로 LSTM과 GRU는 gradient vanishing 문제와 학습의 불안정성으로 인해 매우 긴 기간의 상호 의존 관계를 기억하지 못한다. 이를 다루기 위해 LSTNet은 recurrent-skip layer 또는 전형적인 attention mechanism을 사용했다. 모델의 전반적인 부분은 전통적인 autoregression이었는데, 이는 neural network의 scale의 무감각함을 완화시키는데 사용됐다. 그렇더라도 LSTNet은 우리가 제안한 Attention mechanism에 비해 세 가지 주요 단점이 있다. (1) recurrent-skip layer의 skip length를 데이터의 기간에 맞추기 위해 수동으로 조절해야 했지만 우리의 제안된 방법은 스스로 주기적인 패턴을 찾아내 학습한다. (2) LSTNet-Skip 모델은 주기적인 패턴이 있는 MTS 데이터에 맞춰 고안되었지만, 우리의 제안된 모델은 간단하고 다양한 데이터에 적용 가능하며, 심지어 비주기적인 데이터에도 적용 가능하다. (3) LSTNNet-Attn 모델 내 attention layer는 일반적인 attention mechanism의 경우와 같이, 관련된 hidden state를 선택하지만 우리의 제안된 attention mechanism은 관련있는 time series를 선택한다.

1. 사전 지식

논문 내 모델의 두 가지 필수적인 module : RNN, 일반적인 Attention mechanism

1. RNN

주어진 정보의 sequence에 대해 RNN은 일반적으로 재귀 함수 F로 정의되며, 각 time step t에 대해 ht를 계산한다.