**Long Short Term Memory Networks for Anomaly Detection in Time Series**

**시계열 데이터에서의 이상 탐지를 위한 LSTM의 사용**

**개요**

LSTM은 그 장기간을 기억할 수 있는 특성으로 인하여긴 주기의 패턴을 가지고 길이가 정해지지 않은 데이터에 대하여잘 동작한다는 것이 알려져 있다 재귀적 은닉층을 쌓으면서시간에 따라 변하고 드물게 나타나는 특징에 대하여도 잘 학습하는 모습을 볼 수 있다. 이 종이에서는, 우리는 시계열 데이터에서의 이상/결함 탐지를 위해 LSTM을 사용하려 한다. 모델은 정상 데이터를 사용해 훈련되고 다량의 시계열 데이터 에서의 예측을 시도한다. 예측의 오차는 이상 행동의 유사도를 평가하기 위해 다변수 가우스 분포로 모델링 된다. 이러한 접근의 효율성은 ECG, Space Shuttle, Power Demand, Multi-Sensor engine 데이터셋 4개로 평가된다

**소개**

우리가 사용해오던 모니터링 기술은 분포에 내제된 변화를 탐지하기 위해 일정 시간 간격의 누적합(Cumulative Sum)이나 지수 가중 평균(Exponentially weighted moving average)와 같은 통계적 측정을 사용하였습니다. 이 시간 간격의 길이는 사전에 정의되어야 하고, 결과 또한 이 매개변수에 영향을 받습니다. LSTM신경망은 오차가 유닛의 내부를 통과할 수 있게 하는 통로를 고안함으로써 기존 RNN의 기울기 증발 문제를 극복하였습니다. 입력(I)와 출력(O), 그리고 기억된 정보가 무의미한 입출력에 영향을 받지 않도록 하는 망각 통로(F)로 이루어져 정보의 장기 기억을 가능하게 합니다. 시계열에서의 긴 주기의 패턴을 학습할 수 있는 LSTM의 능력 덕에, 시간 간격을 사전에 정의하지 않아도 되고, 복잡한 다변수 시계열 데이터에 대하여도 어느정도의 정확성을 보장할 수 있게 됩니다. 이 종이에서, 우리는 시계열 데이터에서의 정상 행위를 LSTM을 이용하여 모델링하고, 어떠한 전처리나 사전 정의 없이 정확하게 데이터 흐름에서의 이상을 탐지하는 방법을 설명하려 합니다. 시그모이드 함수를 활성화 함수로써 사용한 적층 재귀 은닉층은 시계열의 구조를 더 잘 포착하고 서로 다른 길이를 가진 시계열 데이터 또한 처리할 수 있다는게 알려져 있습니다. 주목할 만한 이상 탐지를 위한 계층적 시계열 데이터 처리 사례중 하나는 세포, 구역 등의 계층을 가진 두뇌의 신피질을 모방한 계층 시계열 메모리(Hieralchical Temporal Memory)입니다. 또한, 시계열 데이터에서의 이상 탐지는 예측 오차를 사용하여 탐지하는 것이 좋다고 알려서 있습니다. 그러나, 우리의 지식을 극대화시키기 위해서는 LSTM은 이상 탐지를 위한 재귀 계층 처리층과 합쳐지지 않아야 합니다. [5]에서와 같이(제트엔진 장비 이상탐지), 우리는 정상적인 상황을 모델링하고, 그 후에 이상 탐지를 위해 예측 오차를 이용했습니다. (이건 특히 정상 행위가 이상 행위보다 훨씬 많이 발생하는 현실적인 이상 탐지 시나리오에 대하여 실용적입니다). 신경망이 시계열의 구조를 잘 포착하도록 하기 위해서, 우리는 미래의 몇 가지 시간대(time step)는 예측합니다. 그 결과, 일련의 데이터에서 각 시점은 과거의 다른 시점으로부터 추출된 다수의 상응하는 예측을 가지고 있고, 이것이 우리에게 다수의 예측 오차를 가져다줍니다. 오차의 확률 분포는 정상 데이터를 예측하는 동안 만들어지고, 테스트 데이터에서의 정상 행위간의 유사도를 측정하는 데에 사용됩니다. 여전히 통제변인(자동자의 엑셀이나 브레이크같은)이 존재하지만, 신경망은 종속변인뿐만이 아니라 통제변인도 예측 가능하게 만들어졌습니다. 이러한 것들은 신경망으로 하여금 예측 오차의 교차 분포를 통하여 정상적인 사용 패턴을 학습함으로써 통제변인과 종속변인의 오차를 예측을 가능하게 합니다. 그 결과, 통제 입력이 포착되었지만 이상 징후가 탐지되지 않을 경우에 명백한 예측 조차가 발생합니다.



**LSTM-AD: LSTM 기반 이상 탐지**

Consider a time series *X* = *{***x**(1)*,* **x**(2)*, ...,* **x**(*n*)*}*, where each point **x**(*t*) *∈ Rm* in the time series is an *m*-dimensional vector *{x*(*t*)1 *, x*(*t*)2 *, ..., x*(*t*)*m }*, whose elements correspond to the input variables. 예측 모델은 입력으로부터 1 <= d <= m인 d개의 데이터를 통해 다음 l개의 값을 예측하는 방법을 학습합니다. 정상 데이터의 학습은 4개의 그룹으로 나뉘어집니다. 정상 훈련, 정상 검증1, 정상 검증2, 정상 테스트. 이상 데이터의 학습은 이상 검증, 이상 테스트 2가지의 분류로 나뉘어 있습니다. 우리는 우선 적층된 LSTM을 사용하여 예측 모델을 학습시키고, 이상을 탐지하는데 사용하는 오차 분포를 예측을 계산합니다.

**Stacked LSTM 기반 예측 모델**

우리는 아래의 LSTM 신경망 구조를 고안했습니다. 각 m차원에서 한 개의 입력 벡터와, d x l 개의 출력층에서 하나를 가져옵니다. 각 d차원의 벡터 내부의 각l차원의 각 벡터는 하나의 유닛으로 봅니다. LSTM의 아래 은닉층은 fully connected하며 그 위의 은닉층은 feedforward계층을 통과합니다. 예측 모델은 Sn에 존재하는 sequence(s)를 이용하여 학습하고, v(n1)은 가중치를 학습하는 동안 조기 종료에 사용됩니다.

**오차 분포를 통한 이상 탐지**

With a prediction length of *l*, each of the selected *d* dimensions of **x**(*t*) *∈ X* for *l < t ≤ n − l* is predicted *l* times. We compute an *error vector* **e**(*t*) for point **x**(*t*) as **e**(*t*) = [*e*(*t*)11 *, ..., e*(*t*)1*l , ..., e*(*t*)*d*1 *, ..., e*(*t*)*dl* ], where *e*(*t*)*ij* is the difference between *x*(*t*) *i* and its value as predicted at time *t − j*. its value as predicted at time t − j. *sn*을 사용해 학습된 예측 모델은 검증과 테스트 시계열의 각 시점에서의 오차 벡터를 계산하는 데에 사용됩니다. 오차 벡터는 다변수 가우스 분포 *N* = *N*(*μ,* **Σ**)에 맞추어 모델링됩니다. 가능성 탐색된 오차 벡터 **e**(*t*)에 대한 가능성 *p*(*t*)는 **e**(*t*)에서의 *N*값에 의해 주어집니다. (칼만 필터를 이용한 이상 탐지 동적 예측 모델에서 사용되는 정규오차제곱(Normalized Innovations Squared)과 비슷합니다.) *vN*1, 어느 한 지점에서의 오차 벡터는 최대 가능도를 이용하여 매개변수 *μ와* **Σ**을 정하는데에 사용됩니다. 관측 **x**(*t*)는 *p*(*t*) *< τ*, 일 때 이상 징후로 탐지되고, 그렇지 않으면 정상으로 평가됩니다. 집합 *vN*2와 *vA*는 *Fβ*-score를 최대화하는 방식으로 *매개변수 τ*를 정하는 데에 사용합니다. (이상 행동을 보이는 시간은 양성 클래스, 정상적인 시간은 음성 클래스).