

가속도센서를 활용한 회귀모델의 유량 예측

A Flow Rate Prediction of Regression Models Using an Acceleration Sensor

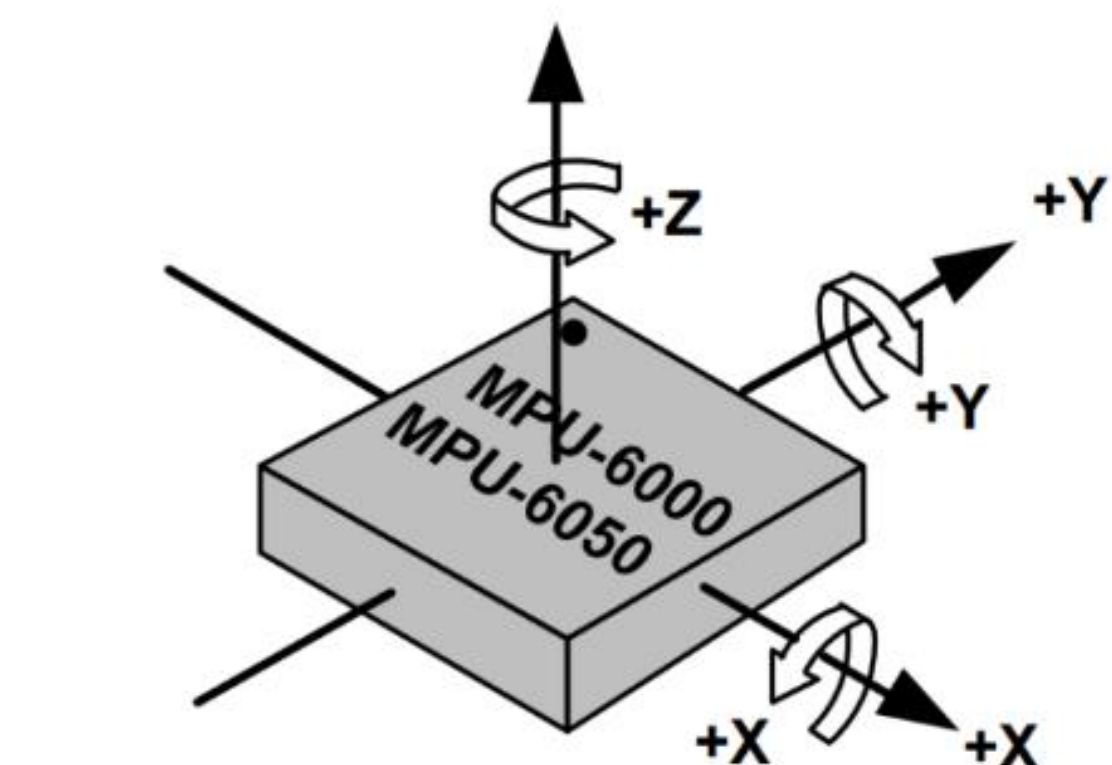
본 논문은 물병으로 섭취한 수분의 양을 측정하는 방법에 대한 것으로, 물병에 부착된 가속도 센서가 측정한 데이터를 이용하여 물병으로부터 흘러나온 유량을 측정하는 회귀분석을 시도하고 제안한 방법의 실효성과 한계에 대해 검토한다.

서론

본 논문에서는 가속도센서를 활용해서 물병으로부터 빠져나간 물의 양을 구하고자 한다. 연구에 사용된 모델들은 가속도 센서로부터 얻은 각도를 feature data로, 물병으로부터 빠져나간 물의 양을 label data로 설정하고 회귀분석을 시도한다. 본 논문은 machine learning과 deep learning을 활용한 회귀분석이 이러한 문제에 얼마나 잘 대답할 수 있는지에 관심이 있다.

가속도센서

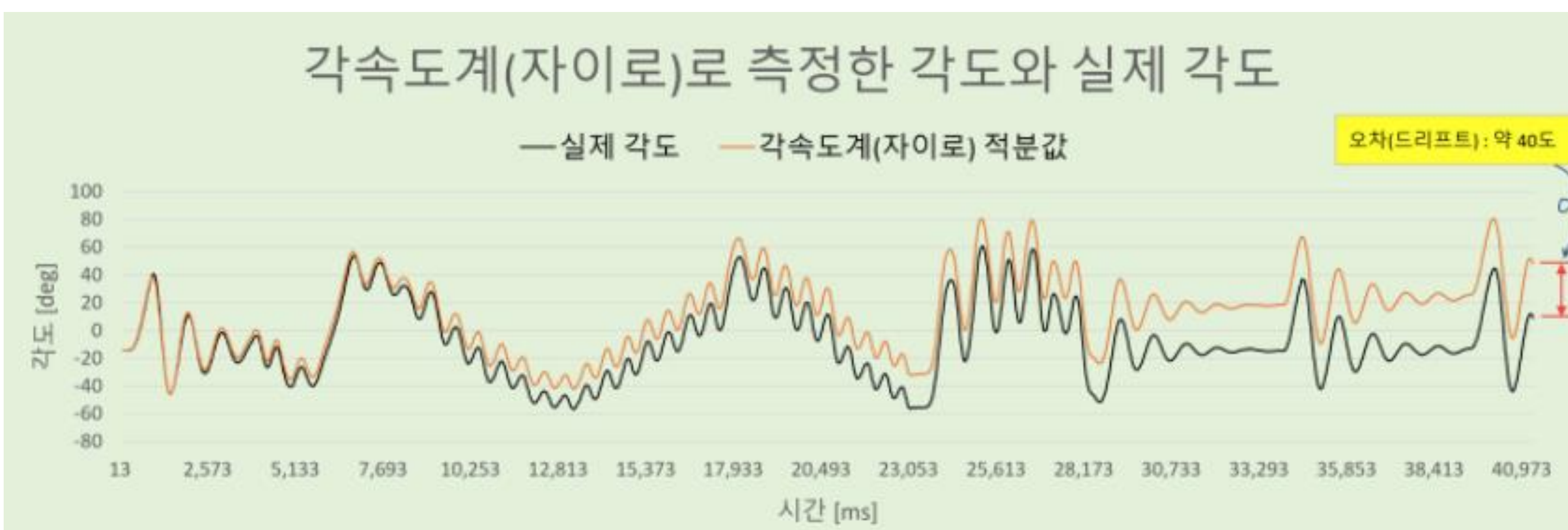
실험에 사용한 가속도센서는 MPU6050이다. 가속도센서란 총 6축 센서로 힘의 가속도 3축(x축, y축, z축)방향과 3차원 축 회전 방향의 각속도 그리고 온도를 측정한다. 그리고 측정값들로부터 센서가 3차원 축 회전 방향으로 얼마나 기울었는지를 파악할 수 있다.



Orientation of Axes of Sensitivity and Polarity of Rotation

$$\theta_x = \tan^{-1} \left(\frac{A_y}{\sqrt{(A_x)^2 + (A_z)^2}} \right)$$
$$\theta_y = \tan^{-1} \left(\frac{-A_x}{\sqrt{(A_y)^2 + (A_z)^2}} \right)$$

MPU6050센서는 몇가지 이유 때문에 측정한 값들을 바로 활용할 수 없다. 첫째, 각속도를 적분해서 얻은 값 thetaX, thetaY, thetaZ를 신뢰하기 어렵다.



본 논문에서는 정확한 각도를 계산하기 위해 상보필터를 사용했다. 상보필터는 저대역은 가속도계 값을, 고대역은 각속도계 값을 가져와 서로 합치는 것이다. 가속도계 값은 오차누적은 없지만 고주파 잡음이 있고 자이로계는 단기간엔 정확하지만 저주파 잡음이 있다. 상보필터는 두 센서의 값들을 합쳐 단점을 보완한다. 본 논문에서는 0.1초 단위의 빠른 계산이 필요했기 때문에 비교적 정확하고 느린 칼만필터보다 정확성은 조금 떨어지지만 계산이 빠른 상보필터를 사용한다.

$$\theta = (1 - \alpha) \times \theta_{accel} + \alpha \times \theta_{gyro}$$

GBM

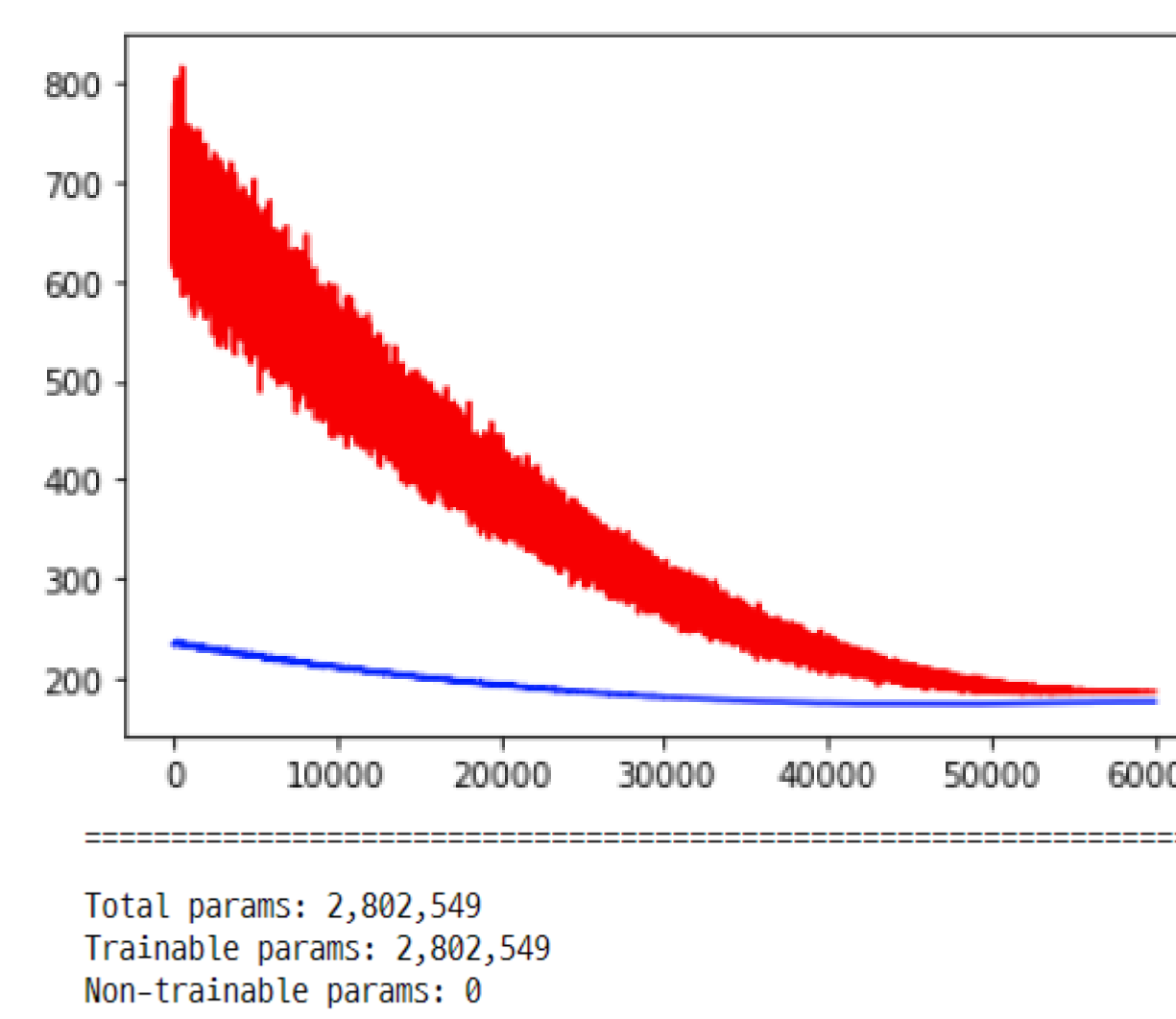
machine learning에 사용한 모델은 GBM이다. GBM이 machine learning algorithm 중 예측성능이 가장 높다고 알려진 algorithm이기 때문이다. 학습 결과 train과 test의 RMSE(Root Mean Square Error)는 9.23ml와 12.81ml였다. train data와 test data의 오차가 10ml 안팎이기 때문에 근사치로 예측한 것 처럼 보이지만 score메소드를 활용하여 R2(R-Squared Score)를 계산하면 각각 0.53, 0.11로 underfitting되어 있음을 알 수 있다. 데이터셋의 부족으로 underfitting이 야기되었을 가능성이 있어 GridSearchCV를 사용하여 hyperparameter tuning과 cross validation을 진행하였지만 Best Test Score는 0.15로 0.04 상승하는데 그쳤다.

	mean	std	25%	50%	75%	max
mean	1.000000	0.920970	0.651904	0.922481	0.975552	0.958724
std	0.920970	1.000000	0.617252	0.751227	0.972187	0.978075
25%	0.651904	0.617252	1.000000	0.524717	0.636510	0.634367
50%	0.922481	0.751227	0.524717	1.000000	0.838458	0.790843
75%	0.975552	0.972187	0.636510	0.838458	1.000000	0.980160
max	0.958724	0.978075	0.634367	0.790843	0.980160	1.000000

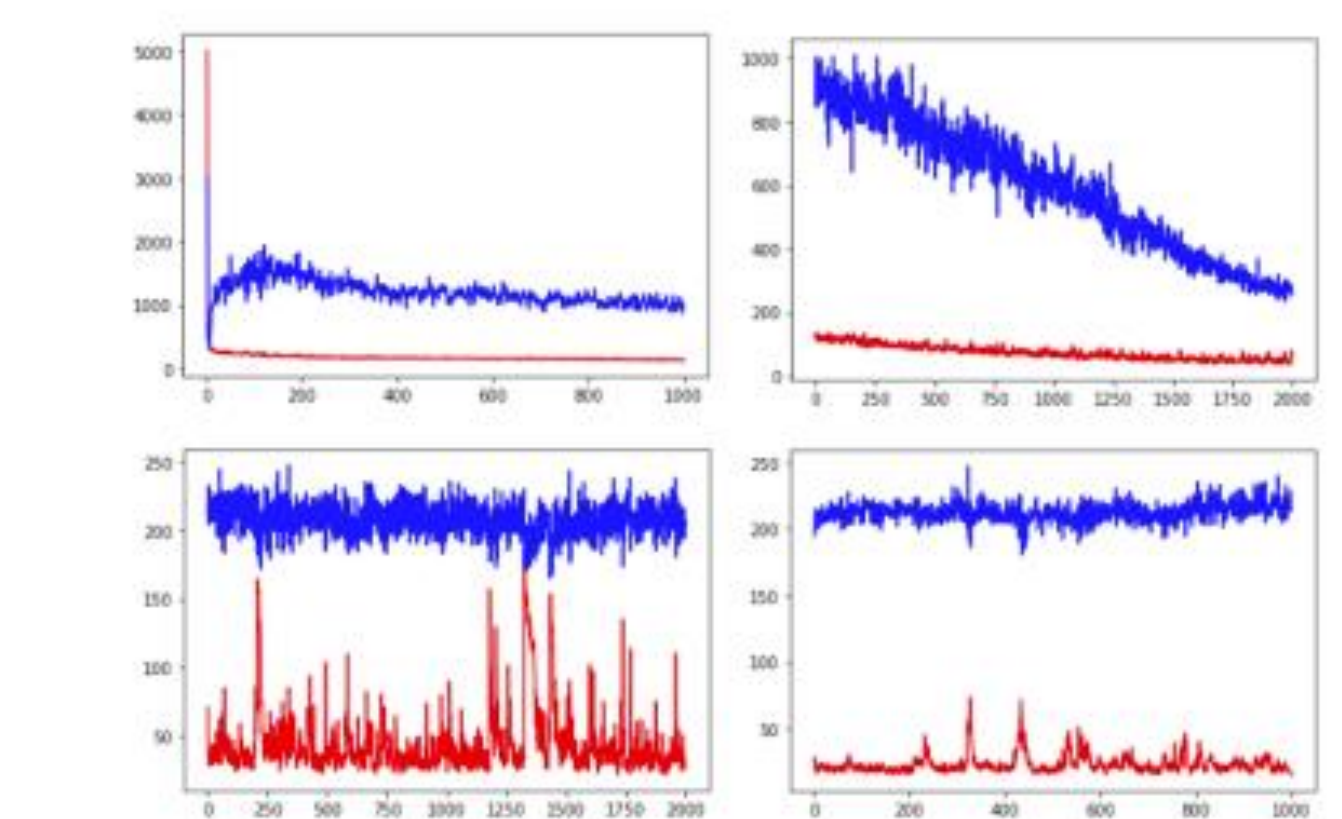


MLP

deep learning에 사용한 모델은 MLP이다. 앞선 machine learning 학습에서 underfitting하는 모습을 보였기 때문에 model-1은 은닉층을 두텁게 쌓아서 학습을 진행하였다. model-1의 RMSE는 train data의 경우 1.08ml로 매우 낮은 오차를 보인 반면 test data는 15.66ml로 다소 높았다. model-1이 overfitting을 보였으므로 model-2는 은닉층을 조금 더 얇게 구성하고 dropout을 추가하여 overfitting을 방지하고 학습을 진행하였다. model-2의 RMSE는 아래 그래프와 같이 학습이 진행되다가 train data와 test data 각각 6.30ml, 15.66ml 정도에서 더 이상 학습이 진행되지 않았다.



Total params: 2,802,549
Trainable params: 2,802,549
Non-trainable params: 0



Total params: 351,137
Trainable params: 351,137
Non-trainable params: 0

결론

본 논문에서 실시한 연구 결과에 따르면 model-1에서는 overfitting이 GBM과 model-2에서는 underfitting 현상이 나타났다. 데이터 셋의 부족을 감안하더라도 오차가 높게 나타났다. 따라서 GBM과 MLP 모두 시간에 따른 가속도센서의 값만으로는 유량을 예측하기 어려운 것으로 보인다. 물병에 담겨있는 물의 수위와 같은 변수들이 통제되지 않았고 feature data에도 반영되지 못한 것이 그 원인이라 생각된다.