LSTMで発電量の予測を行なった

翌日のデータを予測するには前日の20:00までのデータしか使えないので

8/17の00:00 ~ 23:30を予測するとき

8/17の23:30を8/15 20:00 ~ 8/16 20:00

8/17の23:00を8/15 19:30 ~ 8/16 19:30

....

8/17の00:00を8/14 20:30 ~ 8/15 20:30で予測する

というモデルを構築した

使ったデータは富士河口湖のアメダスの日射時間データ(直達日射量が120W/m^2以上の時間)

LSTMに入れた

self.input\_layer\_size = 1 #入力層の数

self.hidden\_layer\_size = 100 # 隠れ層の数、適当

self.output\_layer\_size = 1 #出力層の数

self.batch\_size = 200 #バッチサイズ、適当

self.chunk\_size = 49 # 一回の系列データの長さ

self.learning\_rate = 0.01 # 学習率 適当

self.forget\_bias = 0.9 # 忘却率 適当

self.epochs = 100 #エポック数 適当

LSTMに入れる前にmin-max正規化させた

あんまりうまく学習できてないっぽい

・そもそも使うべきアメダスデータの地点が違う？

・日射時間データは多くが600とかになってしまう

(元のデータは120W/m^2以上の秒数を記録しているもので、データの処理の時気候データに関しては30mあたり平均でまとめている)

・日射時間データだけじゃきつい？

入力の次元を上げることは可能

・outputデータのスケールが0 ~ 3500なのが割と辛い

LSTMに突っ込む際に0~1あるいは-1 ~ 1するべきはずだがスケールを戻す時にLSTMの微妙な出力の違いが誇張されてしまう。

Lossを計算する際にスケールを戻して計算するのもありかもしれない

・outputの次元を増やすことも検討中

[0かどうかのフラグ, 発電量の予測]

output 0は割と多い、かつMAEを用いてスコアリングされるので0をバシッと0で予測できればスコアは伸びそう

