1　データとモデルに関する考察

1.1　前提条件

データとして用いたものは，四本値，四本値と出来高，四本値と出来高とニュース記事のスコア，四本値と出来高とニュース記事のスコアと為替の 4 通りである．これらは，それぞれ input_size の 4,5,6,7 と対応している．モデルとして用いたものは，NN，NN(2)，RNN，LSTM の 4 種である．

1.2　考察

input_size	best_accuracy	best_f1
4	51.11	0.49
5	51.47	0.51
6	51.77	0.5
7	52.29	0.52

表 1　単層 NN の平均正解率，F 値

input_size	best_accuracy	best_f1
4	51.18	0.51
5	51.52	0.5
6	51.54	0.5
7	52.11	0.52

表 2　2 層 NN の平均正解率，F 値

input_size	best_accuracy	best_f1
4	51.14	0.45
5	51.14	0.48
6	52.14	0.51
7	51.83	0.51

表 3　RNN の平均正解率，F 値

input_size	best_accuracy	best_f1
4	49.48	0.36
5	51.5	0.51
6	52.64	0.52
7	52.18	0.51

表 4　LSTM の平均正解率，F 値

1.2.1　データに関する考察

■出来高の有無の比較　input_size が 4 の場合と，5 の場合を比較する．今回用いた 4 通りのモデルに対して，4 つのモデルのうち 3 つのモデルで平均 0.91% の best_accuracy の向上が見られた．また，4 つのモデルのうち 3 つのモデルで平均 0.67 の best_f1 の向上が見られた．このことから出来高をデータに加えることで，予測精度の向上が見られたと言える．

これは，取引量の多い日にはサンプル数が増え，データの予測がしやすくなったことが理由として考えられる．

■ニュースデータの有無の比較　input_size が 5 の場合と 6 の場合を比較する．今回用いた 4 通りのモデルに対して，4 つのモデルすべてにおいて，平均 0.62% の best_accuracy の向上が見ら

れた．また，4 つのモデルのうち 2 つのモデルで，平均 0.02 の best_f1 の向上が見られた．このことから，ニュースデータを予測データに加えることで，予測精度の向上が見られたといえる．

これは，株式を売買する人が株を売買するにあたり，ニュース記事のデータを参考にしているためニュース記事の影響が出ていることが理由として考えられる．

■為替データの有無の比較　input_size が 6 の場合と 7 の場合を比較する．今回用いた 4 通りのモデルに対して，NN を用いたモデルに対してのみ，平均 0.55% の best_accuracy の向上が見られ，RNN や LSTM などの時系列データを扱うモデルに対しては，0.75% の精度の減少がみられた．このことから，為替データは時系列データを得意としないデータに対して用いられたとき予測精度が向上しているといえる．

これは，為替データは 1 日や 2 日でそこまで大きく変わるものではなく，また，為替データが影響を与えるとしても，輸出入量や，業界によって為替データの影響の受け方が異なるので，今回，企業を細かく分類できていない状態で為替データを時系列的に扱ったため，精度が下がったのではないかと考えられる．また一方で，為替データを時系列的に扱わなかったことによって精度が向上した原因としては，為替が日本国内の景気を反映しているため，長期的な視点で見た時に，景気がいい時には株価が上がりやすいなどの特徴を学習することができたことが理由として考えられる．

1.2.2　モデルに関する考察

■NN,NN(2) と RNN,LSTM の比較　NN,NN(2) と RNN，LSTM を比較すると，input_size が，4,5,7 の時に，NN,NN(2) のほうが精度がよく，input_size が 6 の時に，RNN,LSTM のほうが精度がよくなっている．だが一方で最も精度が良かったのは RNN,LSTM を用いて，四本値，出来高，News のデータを用いた時でその時の予測精度は 52.39% であった．このことから，株価の予測を行うのに適したデータは，LSTM や RNN などの時系列を扱うことを得意とするモデルであるが，その日ごとの特徴を表すデータを適切に与えなければ精度は向上しないことがわかる．しかしながら，NN を用いて学習した場合でも，四本値，出来高，News，為替のデータを用いた時の予測精度が，RNN,LSTM においての最高の予測精度より 0.19% しか下がっていないため，今回のデータセットを用いた株価予測において，株を時系列データとして扱っても，そうでなくても，それほど大きな違いはないものと考えられる．

input_size	best_accuracy	best_f1
4	51.15	0.5
5	51.49	0.51
6	51.66	0.5
7	52.2	0.52

表 5　NN,NN(2) の平均正解率，F 値

input_size	best_accuracy	best_f1
4	50.31	0.4
5	51.32	0.49
6	52.39	0.52
7	52.0	0.51

表 6　RNN,LSTM の平均正解率，F 値

■RNN と LSTM の比較　LSTM(表 4) は，RNN(表 3) の改善版であり，より長期の依存関係を保持することができるように改良されたものである．表 3 と表 4 を比較すると，input_size が 4 の場合は RNN の方が予測精度がよく，input_size が 5,6,7 の場合は LSTM の方が精度がよくなっている．今回は，データとして 5 日分のデータを使っており，そこまで長期依存性は必要ないほどの長さではあるが，LSTM の方が精度がよくなっている．

これは，LSTM の方が RNN よりもより必要な情報を取捨選択することができたということが理由として考えられる．つまり，LSTM の忘却ゲートによって，5 日分のデータという短い長さであっても，長期依存性の対策になったのではないかと考えられる．