

## ガウシアンプロセスを用いた株価予測の検討

肱岡 浩大

現在、機械学習による数値計算を用いた株価予測の研究が数多く行われている。本研究で使用するガウシアンプロセスは、統計学や地質統計学などの分野では以前から知られてきたが、機械学習の分野で知られるようになったのは比較的最近であり、株価予測への応用もあまり進んでいない。そこで、本研究では、ガウシアンプロセスの共分散関数の選択やハイパーパラメータ値の設定に着目しながら、株価予測への適用性を検討した。

本研究では、休場日を考慮しない場合と考慮した場合とで、異なる業種から選択した 5 銘柄でそれぞれ株価予測を行った。「休場日を考慮しない」とは、株式市場の休場日を考慮せず、休場日を挟んだ 2 日を連続した日とする考え方である。株式市場が休場の間、株式の売買は行われないが、企業や社会全体は流動しているため、どちらの考え方も理に適っていると考えた。

ガウシアンプロセスでは、モデル選択として共分散関数を選んで事前分布を決め、訓練データによって条件づけることで予測分布を得る。本研究では、共分散関数として次の(1)から(5)を組み合わせて使用した：(1) Squared Exponential Covariance Function (2) Matern Class of Covariance Function( $\nu = 3/2, \nu = 5/2$ ) (3) Exponential Covariance Function (4) Rational Quadratic Covariance Function (5) Squared Exponential Covariance Function と Periodic Covariance Function の積。それぞれの関数に正の係数を乗じて足合わせ、訓練データにはノイズがあることも考慮した。期待値を時間に依存しない定数として含め、計 16 個のハイパーパラメータを設定した。それらの値は負の周辺対数尤度を最小化することで設定した。予測の精度は平均誤差、最大誤差、騰落の方向一致率で評価した。

結果であるが、訓練データに合わせたモデル選択においては、複数の共分散関数を組み合わせると負の周辺対数尤度の取得値は複雑となった。つまり大きく異なるハイパーパラメータ値で複数の局所最適値が存在し、その局所最適値の多くは非常に近い値をとっていた。共分散関数は上述した関数(1),(2),(3)にかかる重みが大きく、株価の細かい変動を表現できる関数だけでなく滑らかな変動を表現する関数も必要であることが分かった。休場日を考慮しない場合と考慮した場合における負の周辺対数尤度でどちらが小さくなるかは各銘柄で異なっていたが、考慮しない場合の方が値が小さくなるものが多かった。しかし、考慮した場合の方が値が小さくなるものもあった為、データに適合したモデルを作成するという観点では、休場日を考慮した方が良い銘柄もあることが分かった。

テストデータでの平均誤差、最大誤差、方向一致率に関しては、良くないものが多かった。しかし、休場日を考慮しない場合の野村ホールディングスでの予測株価は良い結果となっていた為、本研究での手法に適した銘柄が存在する可能性があると言える。

(指導教員 鎮目浩輔)