

時系列モデルを用いた遅延相関分析の評価

Evaluation of time-delayed correlation analysis using time series model

蘆田 誠人* 大平 雅雄†

あらまし 本研究の目的は、従来の相関分析をベースとし時間の遅延を考慮した遅延相関分析と時系列モデルとの比較評価することである。本稿では、比較及び評価の方法について述べる。

1 はじめに

ソフトウェア開発の現場では、プロジェクトを定量的に管理することが必要不可欠となっており、定量的なプロジェクト管理を目的としたソフトウェア開発データの計測及び分析が行われている。ソフトウェア開発データの分析手法の代表例として、相関分析が挙げられる。相関分析では同時点での事象 A-B 間の相関係数を求めているが、実際のソフトウェア開発では事象 A が発生してからある程度遅延して事象 B が発生することもあり得る。例えば、開発要員を増やしたとしても生産性に寄与するには一定程度の時間を要するなどといった関係は、従来の相関分析では事象 A-B 間の関係を正確に捉えることが難しい。相関分析の欠点を補完することを目的として、山谷らは遅延相関分析 [1] を用いて従来の相関分析との比較評価 [2] を行っている。遅延相関分析とは、時間的順序関係を考慮した相関分析の手法である。ただし、[2] では従来の相関分析以外の時系列データ分析手法との比較がなされておらず有用性が十分明らかになっていないと言いがたい。そこで本研究では、遅延相関分析の有用性を検討するために時系列モデルとの比較を行う。

2 時系列データ分析

本章では、本研究で用いる時系列データ分析の手法（遅延相関分析及び時系列モデル）について説明する。

2.1 遅延相関分析

遅延相関分析とは、2つの時系列データを用意し、時間を遅延させながら観察される関係を探る分析手法である。遅延相関分析のイメージを図1に示す。遅延相関分析では、一方の時系列データを説明変数、もう一方の時系列データを目的変数とし、説明変数の一定期間の平均値と目的変数の変化量との相関を求める。

まず、説明変数及び目的変数の時系列データの処理を行う。説明変数の処理には加算係数を用いる。加算係数とは説明変数の値を平均する期間のことを指す。説明変数の一定期間の平均は加算係数の値に従って求める。 e_i を時刻 i における説明変数の値、 e_j を時刻 j における説明変数の値、 e_{ij} を説明変数の値の平均とすると、 e_{ij} は式 (1) で表される。

$$e_{ij} = \frac{e_i + e_{i-1} + \dots + e_j}{i - j + 1} \quad (1)$$

目的変数の変化量は目的変数の値の差分を求めることで導く。時刻 n における目的変数の値を r_n 、時刻 n における、目的変数の変化量を r_{nn} とすると、 r_{nn} は式 (2) で表される。

$$r_{nn} = r_n - r_{n-1} \quad (2)$$

次にスピアマンの順位相関を用いて相関係数を算出する。

*Makoto Ashida, 和歌山大学

†Masao Ohira, 和歌山大学

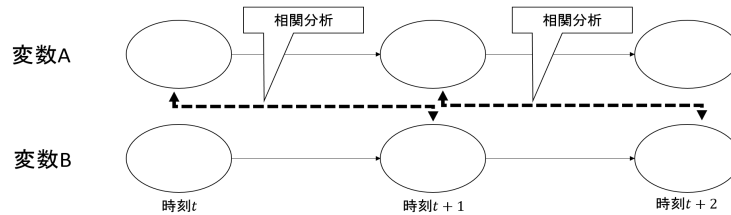


図1 遅延相関分析のイメージ

最後に、各パラメータの値の最適な組み合わせを求めるために、最大相関係数の更新を行う。出力された最大相関係数、並びにパラメータ値により、説明変数がどれくらいの期間変化すると、目的変数にどれくらい遅延して影響を与えているかについての情報が得られる。

2.2 時系列モデル

時系列モデルとは、過去のデータから現在のデータを予測するモデルのことを指す。時系列モデルを用いて分析するには、データは定常でなければならない。定常とは確率的な変動の性質が時点に依存せず一定である確率過程のことである。

時系列モデルの代表的な例として、AR (Auto Regressive : 自己回帰) モデルがある。基本式は式 (3) に示す。

$$Y_t = \sum_{k=1}^p a_k Y_{t-k} + \epsilon_k \quad (3)$$

Y_t は時点 t におけるデータの値、 a_k はデータに対してのパラメータ値、 ϵ_k はそのデータの誤差値を表している。パラメータ値は、パラメータ推定で用いられるユールウォーカー法を適用して算出する。

3 評価方法

本研究では、AR モデルをはじめとして様々な時系列モデルとの比較を行う時系列モデルを適用した際、自己相関係数が出力される。自己相関係数は、ある時系列データにおいて、過去のデータと時間をずらしたデータでの間での関連性があるかどうかを表す指標である。自己相関係数を用いて、遅延相関分析で出力された結果を比較し、評価を行う。

4 おわりに

時間の遅延を考慮した遅延相関分析の有用性を明らかにするために、遅延相関分析と時系列モデルの比較実験を行う。また、現在は2変数間の関係を分析するだけに留まっているが、将来的には複数の説明変数を用いて遅延を伴う関係についても分析できるようにしたいと考えている。

謝辞 本研究の一部は、文部科学省科学研究補助金（基盤（A）：JP17H00731，基盤（C）：JP15K00101）による助成を受けた。

参考文献

- [1] 竹内裕之，児玉直樹，”生活習慣と健康状態の時系列データ解析手法の開発”，Proc. 3rd Forum on Data Engineering and Information Management (DEIM'08), E1-5, 2008
- [2] 山谷陽亮，大平雅雄，パソコンパンナチッタ，伊原彰紀，”OSS システムとコミュニティを目的としたデータマイニング手法”，情報処理学会論文誌 vol.56, No.1, pp59-71, Jan 2015