	リカルド　モラエス　ムニズ　ダ　シルバ
氏　　　　名	RICARDO MORAES MUNIZ DA SILVA
学位の種類	博士（工学）
学位記番号	博第1213号
学位授与の日付	2021年3月31日
学位授与の条件	学位規則第４条第１項該当　課程博士
学位論文題目	Accuracy Improvement of Temporal Series Prediction by Analysis of Memory Dependency Through Linear Regression Functions （線形回帰関数を用いたメモリ独立性の解析による時系列信号の予測精度の改善）
論文審査委員	主査　教授　本谷　秀堅 准教授　南角　吉彦 准教授　田口　亮 教授　梅崎　太造（中部大学）

論文内容の要旨

Time series forecasting models are some of the most important type of quantitative models, in which past observations are collected and analyzes to describe their relationship. This modeling approach is particularly useful when little knowledge is available on the generation process or when there is no accurate model that relates the prediction variable to other explanatory variables. When modeling time series, two well-known methodologies are used for prediction in linear time series, the autoregressive integrated moving average (ARIMA) and the autoregressive fractionally integrated moving average (ARFIMA). The prediction of future events depends on the analysis of past values. In case of ARIMA it is limited to a short memory dependency, due to model characteristics on the modelling procedures with stationarity time series. This means that the statistical proprieties such as mean, variance and autocorrelation are all constant over time. However, depending on the dataset, the amount of past values necessary for an accurate prediction may vary, as the correlation between data and their parameter deviate along the time, leading to a long memory dependency. In these cases, the ARFIMA model can be used, since it provides a solution for the

tendency over-differentiation on stationary series that exhibit a long-run dependence. This uncertainty on the amount of necessary past values necessary make an accurate prediction defines the classical case of long and short memory dependency. When modeling a time series, the statistical proprieties can reveal characteristics on the dataset that demonstrates memory dependency that decays exponentially (ARIMA) or hyperbolically (ARFIMA). Thus, is not always clear if a process is stationary or what are the influence of the past samples on the future values and, consequently, which of the two models is the best choice for a given time series. As different datasets contain different characteristics, a common approach to improve the accuracy is the use of preprocessing techniques or the modification on a specific component of the aforementioned models in order to tackle the memory dependency problem. Nevertheless, such solutions are specific to certain datasets, in which different scenarios can generate different results in the performance and accuracy of the model. This research proposes a new methodology which combines the two models and their parameters in order to keep a low accumulative error. The proposed model is applied on a variety of time series datasets without resorting to model modifications or the use of preprocessing techniques. Twelve different datasets of different contexts were selected from the UCI Machine Learning repository, Brazilian stock market and the Time Series Data Library. The proposed method was compared with the original ARIMA and ARFIMA models, as well as wavelet techniques that also tackle the memory dependency problem. Thus, performed better accuracy in the majority of the datasets, except in multivariate biomedical time-series. In addition, the proposed model also presented better accuracy in regions with abrupt changes when compared with other models, being a promise alternative in applications requiring fast decision making based on data forecast. This indicates that the number of features in a dataset can have a direct influence in the correlation between values.

Key words: *ARIMA, ARFIMA, combined methodology, time series forecasting, accumulative error, memory dependency.*

論文審査結果の要旨

Time series forecasting models are some of the most important type of quantitative models, in which past observations are collected and analyzes to describe their relationship. This modeling approach is particularly useful when little knowledge is available on the generation process or when there is no accurate model that relates the prediction variable to other explanatory variables. When modeling time series, two well-known methodologies are used for prediction in linear time series, the autoregressive integrated moving average (ARIMA) and the autoregressive fractionally integrated moving average (ARFIMA). The prediction of future events depends on the analysis of past values. In case of ARIMA it is limited to a short memory dependency, due to model characteristics on the modelling procedures with stationarity time series. This means that the statistical proprieties such as mean, variance and autocorrelation are all constant over time. However, depending on the dataset, the amount of past values necessary for an accurate prediction may vary, as the correlation between data and their parameter deviate along the time, leading to a long memory dependency. In these cases, the ARFIMA model can be used, since it provides a solution for the tendency of over-differentiation on stationary series that exhibit a long-run dependence. This uncertainty on the amount of necessary past values necessary make an accurate prediction defines the classical case of long and short memory dependency. When modeling a time series, the statistical proprieties can reveal characteristics on the dataset that demonstrates memory dependency that decays exponentially (ARIMA) or hyperbolically (ARFIMA). Thus, is not always clear if a process is stationary or what are the influence of the past samples on the future values and, consequently, which of the two models is the best choice for a given time series. As different datasets contain different characteristics, a common approach to improve the accuracy is the use of preprocessing techniques or the modification on a specific component of the aforementioned models in order to tackle the memory dependency problem. Nevertheless, such solutions are specific to certain datasets, in which different scenarios can generate different results in the performance and accuracy of the model. This research proposes a new methodology which combines the two models and their parameters in order to keep a low accumulative error. The proposed model is applied on a variety of time series datasets without resorting to model modifications or the use of preprocessing techniques. Twelve different datasets of different contexts were selected from the UCI Machine Learning repository, Brazilian stock market and the Time Series Data Library. The proposed method was compared with the original ARIMA and ARFIMA models, as well as wavelet techniques that also tackle the memory dependency problem. Thus, performed better accuracy in the majority of the datasets, except in multivariate biomedical time-series. In addition, the proposed model also presented better accuracy in regions with abrupt changes when compared with other models, being a promise alternative in applications requiring fast decision making based on data forecast. This indicates that the number of features in a dataset can have a direct influence in the correlation between values.