

Model selection of ensemble forecasting using weighted similarity of time series

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20335483&lokasi=lokal>

Abstrak

Beberapa metode telah diajukan untuk menggabungkan beberapa hasil forecasting dalam single forecast yang diberi nama simple averaging, pemberian rata-rata dengan bobot pada tahap validasi kinerja, atau skema kombinasi non-parametrik. Metode ini menggunakan kombinasi tetap pada individual forecast untuk mendapatkan hasil final dari forecast. Dalam paper ini, pendekatan berbeda digunakan untuk memilih metode forecasting, di mana setiap titik dihitung dengan menggunakan metode terbaik yang digunakan oleh dataset pelatihan sejenis. Dengan demikian, metode yang dipilih dapat berbeda di setiap titik perkiraan. Similarity measure yang digunakan untuk membandingkan deret waktu untuk pengujian dan validasi adalah Euclidean dan Dynamic Time Warping (DTW), di mana setiap titik yang dibandingkan diberi bobot sesuai dengan keterbaruannya. Dataset yang digunakan dalam percobaan ini adalah data time series yang didesain untuk NN3 Competition dan data time series yang di-generate dari paten-paten USPTO dan publikasi ilmiah PubMed di bidang kesehatan, yaitu pada Apnea, Aritmia, dan Sleep Stages. Hasil percobaan menunjukkan bahwa pemberian kombinasi bobot dari metode yang dipilih berdasarkan kesamaan antara data pelatihan dan data pengujian, dapat menyajikan hasil yang lebih baik dibanding salah satu kombinasi metode unweighted yang dipilih berdasarkan similarity measure atau kombinasi tetap dari individual forecast terbaik.

<hr>

Abstract

Several methods have been proposed to combine the forecasting results into single forecast namely the simple averaging, weighted average on validation performance, or non-parametric combination schemas. These methods use fixed combination of individual forecast to get the final forecast result. In this paper, quite different approach is employed to select the forecasting methods, in which every point to forecast is calculated by using the best methods used by similar training dataset. Thus, the selected methods may differ at each point to forecast. The similarity measures used to compare the time series for testing and validation are Euclidean and Dynamic Time Warping (DTW), where each point to compare is weighted according to its recentness. The dataset used in the experiment is the time series data designated for NN3 Competition and time series generated from the frequency of USPTO's patents and PubMed's scientific publications on the field of health, namely on Apnea, Arrhythmia, and Sleep Stages. The experimental result shows that the weighted combination of methods selected based on the similarity between training and testing data may perform better compared to either the unweighted combination of methods selected based on the similarity measure or the fixed combination of best individual forecast.