

INTISARI

SELEKSI FITUR BERBASIS ALGORITMA GENETIKA PADA MODEL KLASIFIKASI DATA *MULTIVARIATE TIME SERIES*

Oleh

Rr Ani Dijah Rahajoe
13/351362/SPA/00470

Multivariate Time Series (MTS) banyak digunakan pada multimedia, keuangan, medis, pola dan pengenalan suara. Data MTS terdiri dari m variabel dan n waktu pengamatan untuk setiap subyek, apabila data pengamatan dan data besar maka klasifikasi data MTS atau *clustering* mempunyai dimensi yang tinggi. Tujuan dari seleksi fitur ini untuk mereduksi dimensional data tanpa menghapus informasi dari data original.

Penelitian ini menghasilkan metode seleksi fitur *hybrid filter wrapper*. Seleksi fitur metode *filter* berdasarkan faktor *loadings* dari analisis faktor, sedangkan seleksi fitur metode *wrapper* menggunakan *Genetic Algorithm-Support Vector Machine* (GASVM). Seleksi fitur metode *filter* ini menggunakan hasil faktor analisis dari pembentukan analisa komponen utama. Idanya adalah untuk menyeleksi fitur berdasarkan waktu pengamatan yang paling kuat pengaruhnya di antara seluruh waktu pengamatan yang ada dengan menggunakan PCA analisis faktor. Waktu pengamatan yang paling berpengaruh saja dari setiap komponen utama yang digunakan untuk matriks fiturnya. Seleksi fitur metode *filter* yang pertama ini dinamakan FSBLF (*Feature Selection based on Loadings Factor*) dan yang kedua dinamakan VecTOR (*Vectorization for Time of Observation based on high correlation*). Fitur yang lebih sedikit akan meminimalkan waktu komputasi daripada menggunakan seluruh fitur yang ada. Hasil metode FSBLF-GASVM dan VecTOR-GASVM diujicobakan pada data CMU dan data *wafer*.

Hasil seleksi fitur FSBLF-GASVM menghasilkan 2 fitur untuk data set CMU dengan akurasi maksimal, sedangkan untuk data set *Wafer* menghasilkan fitur atau sensor yang kesatu dengan akurasi sebesar 98,3 persen. VecTOR-GASVM menghasilkan sebelas fitur atau sensor terseleksi dengan akurasi yang maksimal untuk data set CMU. Pada data set *Wafer* menghasilkan dua fitur atau sensor dengan akurasi sebesar 97,98 persen. Kedua metode ini mempunyai hasil akurasi yang lebih baik dan *time processing* lebih cepat dibandingkan dengan tanpa seleksi fitur dan algoritma pembanding lainnya. Algoritma pembanding untuk FSBLF adalah langsung klasifikasi dengan SVM, GABayes, FSBLF-SVM dan FSBLF-GABayes. Sedangkan algoritma pembanding untuk VecTOR-GASVM adalah VecTOR-SVM, VecTOR-GABayes, VecTOR-ForwardBayes, VecTOR-BackwardBayes.

Kata kunci: *multivariate time series*, seleksi fitur, *filter*, *wrapper*, GASVM.

ABSTRACT

FEATURE SELECTION BASED ON GENETIC ALGORITHM FOR MULTIVARIATE TIME SERIES CLASSIFICATION MODEL

by

Rr Ani Dijah Rahajoe
13/351362/SPA/00470

Multivariate Time Series (MTS) is mostly used in multimedia, finance, and voice pattern and recognition. MTS data consist of m variable and n observation time for each subject. If observation data and obtained data are big enough, the classification of MTS data has a high dimension. The main purpose of feature selection is to reduce data dimensional data without eliminating information from the original data.

This research results in hybrid filter wrapper feature selection method. This filter feature selection method is based on loadings factor from factor analysis. Meanwhile, Wrapper feature selection method uses Genetic Algorithm-Support Vector Machine (GASVM). Filter feature selection method uses the result of analysis factor gained from the analysis of principal component. The primary idea is to select feature based on the most influential observation time by using PCA factor analysis. It only involves the most influential observation time from each principal component in order to establish the feature matrix. The first filter feature selection method is FSBLF (Feature Selection based on Loadings Factor) and the second method is VecTOR (Vectorization for Time of Observation based on high correlation). Less feature will minimize time processing instead of using the overall features. The result of FSBLF-GASVM and VecTOR-GASVM is tested on CMU and Wafer data.

FSBLF-GASVM result in 2 features for CMU set data with the maximum accuracy while Wafer set data result in the first feature with the accuracy of 98,3 percent. VecTOR-GASVM has eleven selected features with the maximum accuracy for CMU data set. This method results in two features for the accuracy of 97,98 percent for Wafer data set. These methods have better accuracy and faster time processing than without feature selection and other comparing algorithm. Comparing algorithm for FSBLF is direct classification with SVM, GABayes, FSBLF-SVM and FSBLF-GABayes. Comparing algorithm for VecTOR-GASVM is VecTOR-SVM, VecTOR-GABayes, VecTOR-ForwardBayes, VecTOR-BackwardBayes.

Key Words: multivariate time series, feature selection, filter, wrapper, GASVM.