SHORT TERM SCIENTIFIC MISSION (STSM) – SCIENTIFIC REPORT

The STSM applicant submits this report for approval to the STSM coordinator

Action number: CA15140: ImAppNIO

STSM title: Adapting Singular Spectrum Analysis to Noisy evolutionary search and Graceful Scaling. Application to prediction of glucose-insulin interaction.

STSM start and end date: 15/01/2018 to 31/03/2018
Grantee name: Jose Manuel Velasco Cabo

PURPOSE OF THE STSM/

Type 1 Diabetes Mellitus (T1DM) is a chronic autoimmune disorder in which the immune system attacks the insulin-secreting cells of the pancreas. Without insulin, cells do not assimilate sugar and, as a consequence, there is a rise in blood glucose levels which is known as hyperglycemia. If this situation extends for a long period of time, the patient can develop serious long-term complications including heart diseases, blindness, kidney failure or foot ulcers. Currently, more than 40 million people suffer from T1DM in the world. T1DM can be treated with synthetic insulin injected into the blood stream. However an excessive dose of insulin can produce hypoglycemia. If hypoglycemia is very severe, it can lead to unconsciousness and a diabetic coma. Due to this fact, an accurate modelling and prediction of blood glucose levels is needed.

In our previous research, we employed Grammatical Evolution (GE) and Genetic Programming (GP) to predict glucose levels in diabetic patients with a forecasting horizon within a range of 1-2 hours. Currently, we are trying to extend this horizon to a range of 2-4 hours which represents the working period of a good part of commercial synthetic insulin.

Our initial goal was to improve the classical Seasonal Trend decomposition technique for time series based on Loess (STL). In the case of a very noisy signal, as glucose time series is, this algorithm is not capable of finding a significative seasonal component (the signal to noise ratio is very small). With a better decomposition, we expect to enhance the GE and GP forecasting models.

DESCRIPTION OF WORK CARRIED OUT DURING THE STSMS

During this STSM, we have developed a framework for the exploratory analysis, modelling and prediction of glucose time series using Singular Spectrum Analysis (SSA). We have employed real data from patients which were provided by Hospital Principe de Asturias at Alcala de Henares.

The SSA technique provides us with a number of EigenTriples (eigenvalues and eigenvectors) which can be used to reconstruct the original signal. Our framework is capable of grouping these eigentriples using different distance methods and clustering techniques in order to experiment and compare them:

– Distances: Euclidean, Frechet, Autocorrelation, Partial Autocorrelation, Periodogram, Integrated Periodogram, Pearson's Correlation, Chouakria et al's correlation, Wasserstein and wavelet transform.

– Methods: K-means, Dynamic Time Warping, Symbolic Aggregation Approximation, Permutation Distribution and Maharaj et al's ARMA.

Once we have the eigentriples grouped in a specific number of clusters, the framework can experiment with two options:

– First, we can reconstruct the original signal using a subset of the clusters and, then, proceed to its prediction using GE and GP, or

– Second, we can make a prediction of each cluster individually and get the final prediction as a combination of them.

Besides the prediction provided by the Grammatical Evolution algorithm, the framework is prepared to experiment with the following classical techniques:

– Naive, Simple Exponential Smoothing, Holt smoothing, Holt and Winters smoothing, ARMA, ARIMA, ARIMA using Fourier coefficients, Seasonal and Trend decomposition using Loess and Artificial Neural Network with one hidden layer.

The framework has been implemented using R language and the following packages:

– *shiny, hashmap, cluster, factoextra, BBmisc, spectral, TSclust, Rssa and pdc.*

DESCRIPTION OF THE MAIN RESULTS OBTAINED

At this moment, we are gathering the experimental results of the different forecasting techniques jointly with the different clustering methods. So far, the results are promising and show that we can extend the forecasting horizon beyond two hours using the SSA decomposition into eigentriples clusters and using an evolutionary technique for selecting the best prediction method for each cluster.

FUTURE COLLABORATIONS (if applicable)

We expect that this research will produce a journal paper in which we present a wide comparison that integrates all the clustering, spectral and forecasting methods mentioned.