

A Versatile Approach to Classification of Multivariate Time Series Data

Adam Zagorecki

Centre for Simulation and Analytics
Cranfield University

Defence Academy of the United Kingdom
Shrivenham, SN6 8LA, United Kingdom

Email: a.zagorecki@cranfield.ac.uk

Abstract—During the recent decade we have experienced a rise of popularity of sensors capable of collecting large amounts of data. One of most popular types of data collected by sensors is time series composed of sequences of measurements taken over time. With low cost of individual sensors, multivariate time series data sets are becoming common. Examples can include vehicle or machinery monitoring, sensors from smartphones or sensor suites installed on a human body. This paper describes a generic method that can be applied to arbitrary set of multivariate time series data in order to perform classification or regression tasks. This method was applied to the 2015 AAIA Data Mining Competition concerned with classifying firefighter activities and consecutively led to achieving the second-high score of nearly 80 participant teams.

I. INTRODUCTION

IN THIS paper I present a generic approach to classification of multivariate time series data. This approach was developed and evaluated in the context of the 2015 AAIA Data Mining Competition, where it led to the second highest score of nearly 80 solutions.

During the recent decade we have experienced a rise of popularity of sensors capable of collecting large amounts of data. One of most popular types of data collected by sensors is time series composed of sequences of measurements taken over time. With low cost of individual sensors, multivariate time series data sets are becoming common. Examples can include vehicle or machinery monitoring, sensors from smartphones or sensor suites installed on human body. The collected measurement data is typically not directly useful to the users, as it consists of typically a large number of data points and is very noisy. It should be processed and transformed into knowledge that can be useful to the user. Because of the sheer volume of the data and typically non-trivial patterns present in data this task is suitable for data-mining approaches. In fact in recent years we observe a significant increase of applications that rely on data mining to interpret sensor data and provide useful and actionable knowledge to the users. One of such areas is human body monitoring that can be valuable for healthcare applications, such as post-surgery patient monitoring, monitoring patients with chronic diseases and general well-being promotion, among others.

In this paper the time series data was generated by a sensor suite worn by firefighters during training sessions. The

main focus will be on time series generated by a set of accelerometers and gyroscopes installed on different parts of human body. The data generated by sensors will be used to fully automatically identify activities performed by a subject such as running, climbing a ladder, etc.

The rest of the paper is composed as follows: in the next section the competition task will be introduced with details of the sensors, available data and the evaluation. In the following section I will discuss the proposed approach to classification of multivariate time series data. Consequently each step in of the proposed approach will be discussed in more detail: feature engineering, feature selection, and actual classification. I will finish the paper with a short discussion.

II. THE COMPETITION TASK

This paper describes a solution to the AAIA'15 data mining competition [1] was organized using the Knowledge Pit competition platform [2]. The objective of the competition was to develop efficient methods for automatic labeling of short series of the sensory data in the context of firefighter training activities.

The basic task of the competition was to create a data mining model to predict training activities performed by a firefighter based on data collected from sensor readings installed on the firefighter body. For this purpose a commercial off-the-shelf body sensor suite was used to generate the data.

A. Data

The data for the competition was generated using *smart jacket* – a wearable set of body sensors for monitoring kinematics and psycho-physical condition of firefighters. For each record the data was divided into two subsets.

The first subset consisted of 42 columns that represented aggregations of data from sensors monitoring firefighter's vital functions. Examples of measurements taken are ECG, heart rate, respiration rate, skin temperature, etc. The data for those measurements was pre-processed by the organizers and made available in the form of statistics (mean, standard deviation, skewness, etc.) rather than time series.

The second subset of data consisted of a set of 42 time series, each consisting of 400 data points. The time series was generated by a set of accelerometers and gyroscopes.



Fig. 1. Distribution of sensors on a firefighter body.

There were 7 pairs of accelerometer-gyroscope installed on firefighter's body. The locations of the sensors are shown in Figure 1. Each pair of sensors generated 6 data streams (x , y , and z axes for an accelerometer and x , y , and z axes for a gyroscope). The 400 points corresponded to approximately 1.8 second period of continuous measurement. Since the measurements in the time series were not taken in equal intervals (but all of them were taken at the same time for a given set of 42 time series), the organizers provided a set of 400 time stamps that corresponded to time of measurements.

In total a data record consisted of 17,242 columns, all of them were real numbers.

There were two class attributes associated with data record. They related to activities during the firefighter's training. The first attribute was the body posture that had 5 states: *standing*, *stooping*, *moving*, *crawling* and *crouching*. The second attribute related to the main activity with 16 different activities, such as *no action*, *walking*, *running*, *searching*, *stairs up*, *manipulating*, *throwing hose*, etc.

The data sets consisted of 20,000 training records and 20,000 test cases that were collected during firefighter training. Multiple firefighters participated in data collection.

B. Evaluation

The evaluation of the model performance was determined using the score s which was defined in the following manner:

$$s(p, y) = \frac{1}{3}(BAC_p(p, y) + 2 * BAC_a(p, y)),$$

where $BAC_p(p, y)$ and $BAC_a(p, y)$ are *balanced accuracies* for posture and main activity respectively, determined for a set of predictions p given true labels y . The balanced accuracy BAC is defined as the average of accuracies for individual labels. Let l be the number of all possible labels, then the balanced accuracy is defined as follows:

$$BAC(p, y) = \frac{1}{l} \sum_{i=1}^l ACC_i(p, y),$$

where ACC is accuracy for a given label i and it is defined as:

$$ACC_i(p, y) = \frac{|j : p_j = y_j = i|}{|j : y_j = i|}.$$

The goal was to propose a model that would generate a set of predictions p for the cases for which are known true labels y in order to maximize the score s . The true labels were known only to the organizers, but not to the competitors. The competition platform was used to present provisional evaluation results based on the subset of the actual evaluation set. The final evaluation was made on the remaining test data set.

III. SOLUTION OVERVIEW

In this section I present overview of the solution to the competition task that I developed. The basic steps are presented in Figure 2.

The first, and probably the most critical step was the feature engineering step. At this step the original data set was converted to a secondary data set that consisted of the features generated from the time series data. This step is discussed in detail in the Section IV. It is important to note, that I decided to reject the features related to firefighter's vital measurements and I completely relied on data generated by kinetic sensors. That meant that the data used consisted entirely of a set of 42 time series, each of the same length and all of coupled. The next decision was to ignore time stamp data and to assume that measurements were taken in equal intervals. This decision was dictated entirely by desire to simplify the data processing.

There was another important decision related to data pre-processing I made: I decided to collapse two class attributes into one. Initially, I approached the two class classification problem as two independent classification problems – building two two models one for body posture and the other for activity, with no information shared between the two models. However, I noticed that even though theoretically there were $5 \cdot 16 = 80$ possible states of combined class attributes, in practice only 24 were present in the training data set, which was only slightly higher than the number of states for the second class attribute. Using one class attribute led to dramatic increase of classification performance.

As the competition progressed and the number of features increased, it has become clear that feature selection step would provide benefit. Toward the end of competition a typical features data set would consist of 4,000 to 8,000 attributes. By experimentation it has become clear that reducing the number of attributes to the number between 200 and 600 would clearly improve classification performance. I used a feature selection algorithm to reduce the number of features. It turned out that selecting different number of features from the same feature data set can have quite profound effect on the classification performance. Feature selection led to generation of the reduced feature data set that was used for actual classification task.

As the basic classifier I used combination of the Random Forest classifier with the Multi-Class classifier that converted multiple class problem into set of forests each corresponding

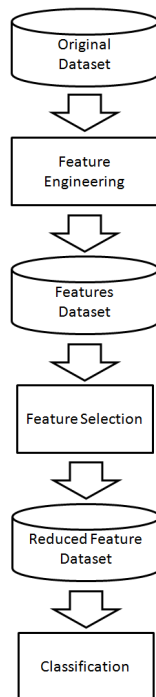


Fig. 2. The outline of the basic tasks used during the competition.

to a set of binary problem Random Forests. I experimented with other classification algorithms available in Weka such as Neural Networks, Logistic Regression, Naive Bayes, Decision Trees, Support Vector Machines, but all other algorithms seemed to perform significantly worse or were taking too long to finish. I did not attempt to compensate for imbalanced distribution of attribute classes.

For feature selection and classification I used Weka software [3]. The feature engineering step was performed using my own code written especially for the purpose of the competition.

For the sake of competition I decided to ignore checking my results for over-fitting. This decision was made strictly for pragmatic reasons – initial attempts to cross-validation did not seem to be representative to the results obtained on the leaderboard. The models that were achieving 100% accuracy on the test set seemed to perform better on the leaderboard than those with lower accuracies on the test set. Obviously, getting the right prediction of the accuracy error would likely lead to improved results, but because of the limited time I wanted to spend on the competition and the fact that it was possible to test results using the submission system I did not focus on getting proper handling of over-fitting.

IV. FEATURE ENGINEERING

The first step in data pre-processing was transformation time series data into a set of numerical values that would summarize different aspects of the time series data. This step is commonly referred as *feature engineering*. The features can be derived from individual time series (e.g. mean, standard deviation) or from some form of a function that can take more than one data

series (such example can be a correlation coefficient between two time series).

A. Original signals

For the feature engineering I used the original 42 time series (generated by accelerometers and gyroscopes). I ignored the time stamps provided and assumed that the measurements are taken in equal intervals.

B. Derived Signals

For the feature generation I decided to use additional time series data that were derived from the original time series. In particular, I combined x , y , and z coordinates using Euclidean norm for each of the accelerometers and gyroscopes, which led to additional 14 derived time series.

C. Extracted Features

For each of the time series (either original or derived) the following features were extracted:

- the mean value
- the maximal value
- the minimal value
- the range (difference between the maximal and minimal values)
- the sum of squared values (mean power)
- the logarithm of the sum of squared values (log mean power)
- the standard deviation
- skewness
- kurtosis
- the 5th central moment
- the maximal difference between two consecutive measurements
- autocorrelation taken at $t=1,2,5,20$, and 50
- power for the bin with the maximal value (power) Fast Fourier Transform (excluding the zeroth frequency)
- maximal value of frequency (in the form of an index) for the bin with maximal value for Fast Transform (excluding the zeroth frequency)
- slope and intercept for the linear regression
- mean square error for the linear regression
- parameters for polynomial fitting with $n = 2$ (a_0, a_1, a_2)

Each of the above features generated a single number that was used as an individual feature for further analysis. This produced 1400 features.

D. Correlations

Finally, I decided to add correlation coefficients between time series. I did it for the original and derived signals separately, that led to 861 features and 91 features, respectively.

V. FEATURE SELECTION

The feature selection has quickly become a necessity as the number of features in the feature set increased. I tested various feature selection algorithms available in Weka. The best results were achieved with the `CfsSubsetEval` algorithm.

The algorithm determines the worth of attribute' subsets by considering the individual predictive ability of each attribute along with the degree of redundancy between attributes in the subset. Subsets of features that are highly correlated with the class while having low inter-correlation are preferred. The Weka's default best first search method was used with default parameters. I used 10-cross validation.

One of the challenges was to decide on the actual number of features to be used. For the winning solution, the feature data set had 2352 features. The 10 runs of cross-validation for feature selection resulted with 541 features that were selected by the feature selection in at least 1 fold. However by trying only those features that were selected at least in 8 out of 10 folds, turned out to result with better prediction score. This resulted in the reduced set having 394 attributes for the best score I could achieve. I did not have chance to explore the effect of the number of features further, but clearly it may have been an important factor.

VI. CLASSIFICATION

I used Random Forest [5] as the basic classifier. An interesting twist was applying a multi-class meta classifier which resulted with a classifier that had multiple Random Forests, one for each class. This approach was effectively comparing the class records vs. remaining records for each class. This step, although not strictly required, resulted in improved classification score.

One of the challenges with applying Random Forest effectively is selection of optimal number of features used for each tree. In the case of competitions it is typically done by trial and error approach. That was the case in this case – I experimented with different numbers of features per tree and for the particular feature set the numbers between 40 and 80 features seemed to work well. For the best score I could

achieve, each Random Forest had 1000 trees. The number of features for each tree was limited to 40.

VII. CONCLUSIONS

In this paper I presented an approach to classification of multivariate time series. The approach was developed for the data mining competition and this approach led to scoring the second high result from nearly 80 solution.

I believe that the approach presented in this paper can be easily generalized to similar problems for which multiple measurements in form of time series are available. It should be expected that different features may turn out to be more predictive or even different classifier may prove to be more suitable. Actually, the author used this approach to another competition where the method allowed to achieve the highest score of nearly 50 submitted solutions.

ACKNOWLEDGMENT

The author would like to thank to Krasin Georgiev for inspiring discussions.

REFERENCES

- [1] Meina, M., Janusz, A., Rykaczewski, K., Ślęzak, D., Celmer, B., and Krasuski, A., "Tagging Firefighter Activities at the Emergency Scene: Summary of AIAA15 Data Mining Competition at Knowledge Pit", Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, 2015.
- [2] Janusz, A., Krasuski, A., Stawicki, S., Rosiak, M., Slezak, D., and Hung Son Nguyen, "Key risk factors for Polish State Fire Service: A Data Mining Competition at Knowledge Pit," Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on pp.345–354, 7-10 Sept. 2014, doi: 10.15439/2014F507.
- [3] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., "The WEKA Data Mining Software: An Update", SIGKDD Explorations, Volume 11, Issue 1, 2009.
- [4] Hall, M. A., "Correlation-based Feature Subset Selection for Machine Learning". Hamilton, New Zealand. 1998.
- [5] Breiman, L., "Random Forests", Machine Learning, Volume 45, Issue 1, pp. 5-32. October 2001.