Determination of Event Patterns for Complex Event Processing Using Fuzzy Unordered Rule Induction Algorithm with Multi-Objective Evolutionary Feature Subset Selection

Nijat Mehdiyev, Julian Krumeich, Dirk Werth, Peter Loos
Institute for Information Systems (IWi)
German Research Center for Artificial Intelligence (DFKI)
{nijat.mehdiyev, julian.krumeich, dirk.werth, peter.loos}@iwi.dfki.de

Abstract

Complex Event Processing (CEP) is an emerging technology to process streaming data and to generate response actions in real time. CEP systems treat all sensor data as primitive events and attempt to detect semantically high level events and related actions by matching them using event patterns. These event patterns are the rules which combine primitive events according to temporal, logical, or spatial correlations among them. Although event patterns (decision rules) can be provided by experts in simplistic scenarios, the huge amount of sensor data makes this unfeasible. The main purpose of the underlying paper is replacing manual identification of event patterns. Considering the uncertainty related to the sensor data, Fuzzy Unordered Rule Induction Algorithm (FURIA) was implemented to identify event patterns after selecting the relevant feature subset using Elitist Pareto-based Multi-Objective Evolutionary Algorithm for Diversity Reinforcement (ENORA). The results were compared to the alternative machine learning approaches.

1. Introduction

Recent advancements in both software and hardware technology have resulted in the acquisition and storage of enormous sensor data. Sensor data have found its application in various domains such as healthcare services, industrial production, military, environmental applications, social sensing, biology, financial services, transportation and etc. [1]. The increasing volume of sensor data originating from multiple sources such as wireless sensors networks, pedometers, GPS, accelerometers, etc. offers manifold opportunities for enterprises which comes at a cost. Computational complexities, the lack in ability of traditional mining systems to process sensor data with high volume, variety and velocity, the uncertainty related to data and other factors require advancements in data processing systems [2]. Processing sensor data introduces diverse challenges including various steps ranging from data collection and cleaning, data fusion, data management and sensor mining [1]. Noisy and uncertain nature, redundancies, missing data, impracticability of storing all data, domain specificity are some of reasons for these issues.

The way sensor systems are designed also requires a real time processing of the non-spontaneous data with spatial and temporal relationships in order to make time critical business decision. Event processing, Stream Processing and Complex Event Processing have emerged as novel, effective and agile technologies to tackle the challenges related to real time analysis of streaming sensor data. The basic idea of all these technologies is to derive meaningful and semantically more complex events that trigger critical decision making actions. Although both Event Processing and Stream Processing technologies are capable to handle continuous data, their main difference is the purpose of processing. Stream Processing attempt to optimize the management of large volume of data by minimizing the queries whereas the purpose of Event Processing is response generation considering the effect of sharing across multiple queries [3]. According to [4], Event Processing is analogous to signal processing whereas Complex Event Processing is characterized by high level situational inferencing.

Complex Event Processing is defined as detection and evaluation of meaningful patterns such as potential threats and opportunities out of multitudes of event streams within real time [3]. For current research, and for practical implementation, domain experts manually provide event rule patterns for the detection of complex events. They are expected to choose the appropriate low-level events that are provided as inputs for rule derivation, define the relationship among the selected attributes, and provide the set of rule patterns for deriving more complex events. In the presence of a vast amount of sensor and RFID data, it is unreasonable, and even not possible, to determine the matching rules manually by experts. Additionally, in many scenarios
the rules can change and evolve dynamically depending on the specifications of the application domain due to concept drift [5]. Such uncontrollable changes make it not feasible for experts to constantly update their matching rules. Therefore, there is a need for automation of both processes: the initial identification of rules for detecting complex events, and updating them when the change in the behavior of the active systems is observed.

The main research question of the underlying paper is the investigation of replacing rule derivation process based on human judgments with machine learning approaches. Rule-based machine learning techniques are sustainable approaches in deriving such event patterns. Due to the uncertainties related to conflicting and missed readings of sensors and problems in transmitting to the central node, application of fuzzy rule induction algorithm is supposed to provide better results than conventional crisp classifiers. In the underlying paper, the Fuzzy Unordered Rule Induction Algorithm (FURIA) approach was adopted to derive rule patterns. Alongside to its ability to deal with uncertainty from various sources, shorter rule lengths provided by this algorithm make it attractive in the context of Complex Event Processing.

One of the most important aspects of Complex Event Processing is using temporal correlations to match the primitive events alongside with casual and logical pattern templates. Thus, it is important to define the time windows of correlating events before providing them as feature to classification algorithm. This in term requires the inclusion of lagged values of sensor input data which results in growing number of inputs. Considering the presence of a vast amount of input sources in the sensor and RFID mining, using all available data as input may overload the classification algorithm which would lead to computational complexities. Reducing input dimension accelerates the classification approach as the size of the search space diminishes significantly. Therefore, feature subset selection is a very crucial step in complex event processing. Selected smaller feature set with strong explanation capability leads to (i) improved classification accuracy, (ii) reliability of the results, (iii) gaining ability to interpret the processes and (iv) smaller computational complexity. Thus, before inducing the rules, the input data have to be pre-processed.

Summarizing, we propose a hybrid model consisting of multi-evolutionary feature subset selection algorithm and a fuzzy rule induction system. This model is assumed to provide accurate rules (event patterns) with significantly reduced computational complexity.

The remainder of the paper is structured as follows. Section 2 provides a brief overview to the principal elements of Complex Event Processing systems. Section 3 provides related work to both feature subset selection and rule induction approaches. Section 4 discusses the structure of proposed model and mathematical background of selected approaches. Section 5 presents data structure, pre-processing outcomes and final empirical results of proposed model. These results were compared to alternative approaches. Section 6 concludes the paper.

2. Complex Event Processing

Complex Event Processing has three core phases: (i) Filtering step: attempts to define the lists of the relevant simple events which can be attributed as predictors, (ii) Matching step: tries to identify the subset of simple events provided from the previous step by analyzing if they can fulfill the specific conditions of defined rule patterns, and (iii) Derivation step: detects more complex events using the information provided from matching subsets [6].

The research problem of the present paper covers the first two stages. In the offline mode the streaming sensor data are pre-processed and the relevant patterns are identified using rule based classifiers. The obtained patterns then are inserted to the CEP engines using expressive languages. Such novel languages support the specification of sequencing, negation operation and sliding windows [3]. An example for such SQL like language is as follows:

```sql
from s in all Temperature {sensorID="T001"}
within 60.0
select Out (last(s), mean(s), stdev(s)) : o {
    //define the upper/lower bands as mean+/-2*stdev
    if (o.value>(o.mean+o.stdv*2.0)) or
    (o.value<(o.mean-o.stdv*2.0)) then {
        Log "Unusual Temperature for T001"
    }
}
```

After providing the event pattern obtained from experts or machine learning algorithms, the CEP engine driven by special algorithms, analyze the streaming data in order to record the complex events. RETE, TREAT, LEAPS and GATOR are such algorithms which can be considered as the execution kernel of CEP engines. These algorithm push notifications, alerts or reminders when the complex events are detected [7].

3. Related Work

As mentioned in the introductory section, the main purpose of the present paper at hand is the determination of the event patterns after processing the feature set. Therefore, in the following subsections we present the related work to feature subset selection techniques and rule-based classification approaches.
3.1. Feature Subset Selection

Including unnecessary features also results in the degradation of prediction system precision. Selecting the influential variables is one of the crucial and challenging issues in determination of event patterns from streaming data and has gained popularity and importance in forecasting modelling in the last two decades [8, 9]. The main purpose of feature subset selection is defining the optimal candidate features from the high dimensional data set. The list of the optimal features subset is identified by removing the redundant (also mutually), irrelevant, erroneous variables with low explanatory abilities [10]. The advantages introduced by the implementation of feature subset selection come at a price, as the search for optimal subset from the candidate features has to be embedded as an additional layer to the modelling task [11]. Identification of optimal feature subsets has to be modelled separately as an alternative to optimizing the full subset and it is not certain if the determined set of variables can improve the accuracy of the model. If the feature subset selection is inefficiently conducted, the noisy, inappropriate and suboptimal variable set could be obtained. Consequently, the speed of classification algorithm can be negatively affected and the preciseness of the results will be threatened.

The feature subset selection approaches may be categorized into three groups: filter, wrapper and hybrid approaches [8]. Filter approaches, mainly assign the scores for features and retain the ones with higher relevancy. The main advantage of filter approaches is their independency from the classification algorithm which makes them fast and scalable for large datasets. A serious weakness with the univariate filter approaches is that, they ignore the interdependency among the variables. This problem can be avoided by applying the multivariate versions of these techniques but it would lead to slowing down the algorithm. Correlation Approach [12], Information Gain [13], t-test [14], Principal Component Analysis [15] are some examples for filter approaches. In contrast to filter approaches, wrapper approaches first apply an optimization algorithm for feature subset selection and embed one of the various classification algorithms to evaluate the obtained results. The accuracy of the predictions provided by the selected classification algorithm identifies the subset of the features. Wrapper techniques can be divided into two groups: deterministic and stochastic wrappers. Sequential Forward Selection [16], Sequential Backward Selection [17], Beam Search [18] are main deterministic or greedy wrapper techniques. Stochastic wrapper algorithms include Simulated Annealing [19], Genetic Algorithms [8], Particle Swarm Optimization [20], Ant Colony Optimization [21] and other evolutionary algorithms. Finally, hybrid approaches attempt to combine the filter and wrapper approaches in order to overcome the shortcoming of both models. Various combinations have already been successfully implemented in the literature [22]. A detailed review analysis of filter, wrapper and hybrid approaches can be found in [9].

Direct interaction with the selected classifier and the ability to consider the dependencies among explanatory variables are the advantages of wrapper approaches. The filter approaches, which evaluate the suitability of the features individually often eliminate the variables which have strong explanatory ability within a group when interacting other features. The literature analysis reveals that in spite of computational costs due to the iterative interaction with the classifier, the wrapper approaches provide more accurate results than filter approaches as they can consider the interdependency [9, 10, 15]. [23] suggest that stochastic wrapper approaches, particularly Genetic Algorithm technique, outperforms both deterministic (greedy) Sequential Forward Selection and Sequential Backward Selection approaches. The authors revealed both computational and accuracy superiority of Genetic Algorithm to alternative approaches using both synthetic (24 dimensional data set) and real world data (30 dimensional data set).

Considering the superiority of wrapper approaches over filter approaches and preciseness of stochastic wrapper approaches compared to greedy wrapper approaches, in the underlying paper we adopt the evolutionary approach for feature subset selection.

3.2. Rule-based Classifiers

Various machine learning classifiers can be implemented for determination of event patterns from streaming sensor or RFID data. As mentioned above, the main goal of this paper is induction of rules that can be easily processed by CEP engines. Therefore, the rule-based classifiers were chosen in the underlying paper for identification of rule patterns to match events. Rule-based classifiers have already been successfully implemented in various information research domains such as intrusion detection [24], content-based image retrieval [25], fingerprint identification [26], and machinery fault diagnosis [27], with each application domain delivering promising results.

Alongside direct rule-based classifiers, the rules can be extracted from various machine learning approaches such as Neural Networks [28], Support Vector Machines [29], Self-Organizing Networks [30] and etc. In the last two decades, fuzzy rule based classifiers have gained popularity. Their ability to deal uncertainty in high dimensional data, applicability of gradient optimization methods, gradual allocation of instances to
the class labels are some advantages over crisp rule-based classifiers. Wang-Mendel approach [31], Fuzzy Decision Trees [32], Fuzzy Unordered Rule Induction Algorithm (FURIA) [33], Adaptive Neuro Fuzzy Inference Systems (ANFIS) [34] are examples to Fuzzy Rule Based Classification Systems.

Considering the relative shortness of the rule lengths and numbers, its ability to deal with diverse type uncertainties and its stretching function to handle the uncovered instances, the FURIA approach was selected as the rule induction algorithm in the underlying paper. The main disadvantage of this classifier is the complexity in interpretation of derived rules.

4. Proposed Model

Proposed model consists of two stages: In the first stage the feature subset with most influential variables is selected using Elitist Pareto-based Multi-Objective Evolutionary Algorithm for Diversity Reinforcement (ENORA) approach (See Fig. 1). This wrapper approach is integrated to FURIA classifier in order to evaluate the relevancy of the selected features iteratively. Finally, selected feature subset is provided as input to FURIA prediction system to derive the rules (event patterns). The following subsections provide a detailed overview to the mathematical background of the both selected approaches.

4.1. ENORA

Proposed first by [35], the initial versions of genetic algorithm approach consider single criterion, mainly estimated error rate, when choosing the subset of variables. In order to obtain more promising results, feature subset selection can be defined as a multi criteria optimization problem which encompasses the optimization of subset cardinality and one or more of the other manually defined criterion such as classification accuracy, error rate (RMSE, MAE etc.), f-measure, or AUC (area under ROC curve). It is very important to note that multiple objective optimizations provide not a unique optimal solution but a set of solution based on Pareto optimality [36]. Niched Pre-Selection Multi-Objective Algorithm, Elitist Pareto-based Multi-Objective Evolutionary Algorithm for Diversity Reinforcement (ENORA), Multi-objective Genetic Algorithm (MOGA), Weight-based Genetic Algorithm (WBGA), Random Weighted Genetic Algorithm (RWGA), Non-dominated Sorting Genetic Algorithm (NSGA), Strength Pareto Evolutionary Algorithm (SPEA), improved SPEA (SPEA2), Pareto-Archived Evolution Strategy (PAES), Pareto Envelope-based Selection Algorithm (PESA), Region-based Selection in Evolutionary Multi-objective Optimization (PESA-II), Fast Non-dominated Sorting Genetic Algorithm (NSGA-II), Multi-objective Evolutionary Algorithm (MEA), Micro-GA, Rank-Density Based Genetic Algorithm (RDGA) and Dynamic Multi-objective Evolutionary Algorithm (DMOEA) are the some of the multi-objective evolutionary approaches for feature subset selection.

![](image)

**Figure 1. Structure of proposed model**

Niched Pre-Selection Multi-Objective Algorithm, Elitist Pareto-based Multi-Objective Evolutionary Algorithm for Diversity Reinforcement (ENORA) and Non-dominated Sorting Genetic Algorithm (NSGA) are three pareto based approaches can handle constraints through repair algorithms and have adaptive variation operators which make them more advanced than other approaches. A comparative analysis conducted by [37] using patient's data set from an intensive care burn unit and Iris dataset suggest that ENORA approach outperforms other two multi-objective optimization algorithm in terms of accuracy and achieves a significantly better computational cost level. Considering its superiority over both conventional and evolutionary wrapper approaches, we adopt Elitist Pareto-based Multi-objective Evolutionary Algorithm for diversity reinforcement (ENORA) to diminish the size of features candidates in our sensor data set.

In the ENORA approach the central problem is the following generic multi-objective optimization problem:

\[
\text{Minimize } f_i(x), \quad i = 1, \ldots, n \\
\text{s.t. } g_j(x) \leq 0, \quad j = 1, \ldots, m
\]
where \( x = (x_1, \ldots, x_p) \) is a vector of real parameters and \( f_j(x) \) and \( g_j(x) \) are arbitrary linear or non-linear functions.

First, the following terms have to be defined in order to describe the steps of ENORA approach [37]:

- \( f_j \): Objective function’s value vector for individual \( j \), \( f_j = f_j^1, \ldots, f_j^n \) \( j = 1, \ldots, N \) where \( N \) is the population size;
- \( f_{\text{max}} \): Maximum value vector for feasible features \( f_{\text{max}} = f_{\text{max}}^1, \ldots, f_{\text{max}}^n \) where \( f_{\text{max}}^i = \max_j \{ f_j^i \} \);
- \( f_{\text{min}} \): Minim value vector for feasible features \( f_{\text{min}} = f_{\text{min}}^1, \ldots, f_{\text{min}}^n \) where \( f_{\text{min}}^i = \min_j \{ f_j^i \} \);
- \( h_j \): Objective normalized values vector for function for feature \( j \), \( h_j = h_j^1, \ldots, h_j^n \) where \( h_j^i = \frac{f_j^i - f_{\text{min}}^i}{f_{\text{max}}^i - f_{\text{min}}^i} \);
- \( c_{\text{in}} \): The set of feasible features which fulfill the following requirement: \( \forall i = 1, \ldots, n, \ f_{\text{min}}^i \leq h_j^i \leq f_{\text{max}}^i \);
- \( c_{\text{out}} \): The set unfeasible features which don’t line in the range described above.

After introducing the parameters, the steps of the ENORA approach for generating a new population can be presented as follows:

Step 1: For each generation two parents are selected using Binary Tournament Selection. The selection criterion is the Rank-Crowding-Better function. According to this function (i) a feasible feature outperforms an unfeasible one, (ii) a feasible feature outperforms the other one if it dominates the second one and (iii) an unfeasible feature outperforms another one if its \( \max \{ g_j^i \} \) is better.

Step 2: The ranking of the individuals in this list are defined using the following formula which determines the slot (t) of inclusion:

\[
    t = \sum_{j=1}^{n-1} d^{j-1} \left\lfloor d \frac{\alpha_i}{\pi} \right\rfloor
\]

where \( d = \left\lfloor n^{-\frac{\sqrt{N}}{2}} \right\rfloor \) and

\[
    \alpha_i = \begin{cases} 
        \frac{\pi}{2} & \text{if } h_j^i = 0 \\
        \arctan \left( \frac{h_j^{i+1}}{h_j^i} \right) & \text{if } h_j^i \neq 0
    \end{cases}
\]

Step 3: Cross, mutation and reparation over the selected parents are carried out and the children are added to a population \( Q \) until this population list covers all individuals.

Step 4: The populations \( Q \) and \( P \) are united into the population list \( R \) and the Rank-Crowding-Better function is executed again on the new list \( R \) in order to identify the elements which survive to next generation.

### 4.2. Evaluation of Feature Subset Selection

One of the most important stages of wrapper approaches is the evaluation of the feature subset performance. The computational costs are strongly related to the selected induction learning algorithm which iteratively evaluates the performance of feature subsets provided from genetic algorithm. [36] suggest that such an evaluation can be carried out using two different approaches. First, one can use an induction algorithm which is simple and have low training requirement. This method leads to better computational costs but the selected subset cannot fit to the main model which is used to make final forecasts. Second, the master model algorithm can be used directly for whole feature selection process. Although the computational complexity may increase compared to the simplistic induction algorithms, the feature selection is optimized according to the specifications of master model once from the beginning of the process which lead to the selection of more precise feature subsets.

Elitist Pareto-based Multi-objective Evolutionary Algorithm for Diversity Reinforcement (ENORA) can be flexibly embedded to any rule induction algorithm. In order to obtain perfectly tailored induce rules which can be provided as event patterns to Complex Event Processing Engine, the ENORA approach is embedded to FURIA algorithm in the present paper.

### 4.3. FURIA

Incorporation of diverse multi-objective evolutionary algorithms to various fuzzy rule induction algorithms has gained a popularity in recent years. There are diverse purposes of incorporation of multi-objective evolutionary algorithms to fuzzy rule induction system: (i) reducing the input variable size, (ii) increasing the classification accuracy, (iii) decreasing the rule length and etc.

[38] proposed a model which attempt to reduce the number of the rules generated by fuzzy inference
systems by using multi-objective generic search. The objective of the optimization were defined as maximization of classification accuracy, minimization of the number and length of the rules. [39] embedded the proposed fuzzy rule modelling approach to three-objectives evolutionary algorithm in order to define the most influential variables. The extension of this paper attempt to integrate HILK (Highly Interpretable Linguistic Knowledge) fuzzy modeling approach to Non-dominated Sorting Genetic Algorithm (NSGA-II) in order to maximize the readability and comprehensibility of the induced rules and the accuracy of the classifier [40], [41] evaluated the performance of six different multi-objective evolutionary algorithms which attempt to optimize the accuracy and interpretability of fuzzy rule induction approaches. A detailed review of the multi-objective genetic fuzzy systems, their computational performance and future research direction are introduced by [42, 43].

The main reason for choosing FURIA approach is its ability to provide compact rules in terms of the rule length even if the interpretation of the rules are more complicated than other fuzzy rule induction approach alternatives [44]. CEP engines match the streaming sensor events in real time according to the provided event patterns which can be referred as decision rules from rule induction algorithm. Therefore, the number and compactness of the provided rules (event patterns) are utmost important.

FURIA is the extension and modification of a crisp rule induction algorithm, Repeated Incremental Pruning to Produce Error Reduction (RIPPER). The traditional rules and rule lists provided by RIPPER algorithm are replaced by fuzzy rules and unordered rule sets in FURIA. Rule stretching methodology offered by FURIA also enables to handle uncovered examples [45]. Before introducing the elements of FURIA, we present the working principle of RIPPER approach. RIPPER was proposed by [46] to produce easily readable, fast, and accurate rules from noisy and large data sets. The main idea of RIPPER approach is for seeking an initial set of rules and iteratively improving it by applying an optimization algorithm. Such modelling with determination of initial rule sets makes this approach effective and fast. The training set used in the rule induction process of this approach is split into two parts: growing set and pruning set.

The rule growing is carried out using FOIL algorithm proposed by [47]. The instances for the selected class label are denoted as positive instances if the rule can cover them, otherwise they are labeled as negative instances. The FOIL algorithm starts with an empty set and add selectors until no negative instances are covered by rule. The second rule attempts to maximize the Information Gain criterion of FOIL algorithm which can be formulated as follows:

$$IG_r \equiv p_r \cdot \left( \log_2 \left( \frac{p_r}{p + n} \right) - \log_2 \left( \frac{n}{p + n} \right) \right)$$

where $p_r$ and $n_r$ are the positive and negative instances covered by the rule and $p$ and $n$ are the positive and negative rules covered by the default rule.

Once the rule is grown using the data from the first set, the instances from the pruning data set are applied to advance the performance of the obtained set by pruning it. The criterion for the pruning position is defined as:

$$V(r) \equiv \frac{p_r - n_r}{p_r + n_r}$$

After the rules are grown and pruned, the optimization algorithm proposed by [46] used to generate the final list.

As mentioned above the FURIA approach is built upon RIPPER algorithm by introducing some modifications and extensions. The first modification is related to using the unordered version of RIPPER algorithm. Another modification is related to splitting the dataset used for training algorithm. [33] has found out that using the pruning step discussed above influences the performance of FURIA approach negatively. Therefore, the dataset to be used for pruning is used for growing the rules in FURIA.

The selectors in FURIA approach can be presented as $(A_i \in I)$ where $I \subseteq R$ is an interval: $I = (-\infty; v]$ if the rule has the selector $(A_i \leq v)$; $I = [u; \infty)$ if the rule has the selector $(A_i \geq u)$; $I = [u, v]$ if the rule contains the both selectors.

In FURIA approach, the rules are expressed fuzzy intervals, fuzzy sets with trapezoidal membership function [33]:

$$I^F(v) \equiv \begin{cases} 1 & \phi^{c.L} \leq v \leq \phi^{c.U} \\ \frac{v - \phi^{s.L}}{\phi^{c.L} - \phi^{s.L}} & \phi^{s.L} < v < \phi^{c.U} \\ \frac{\phi^{s.U} - v}{\phi^{s.U} - \phi^{c.U}} & \phi^{c.U} < v < \phi^{s.U} \\ 0 & \text{else} \end{cases}$$

Where $\phi^{c.L}$ and $\phi^{c.U}$ are the lower and upper bound of the instance with membership 1 and $\phi^{s.L}$ and $\phi^{s.U}$ are the lower and upper bounds of elements with membership greater than 0.

For the elements $x$ the membership function can be calculated with the following formula [33]:

$$\mu_r^F(x) = \prod_{i=1}^{k} I^F_i(x_i)$$
The next step is the fuzzification of the obtained rules. The main idea of this stage is seeking for best fuzzy extension of each individual rules. The training dataset \( D_t \) is used for fuzzification:

\[
D_t = \{ x = (x_1, \ldots, x_k) \in D_T \mid f_j(x_j) > 0 \text{ for all } j \neq i \} \in D_T
\]

To purity formula is used to evaluate the quality of the obtained rules:

\[
pur = \frac{p_i}{p_i + n_i}
\]

where

\[
p_i \equiv \sum_{x \in D_{t+}} \mu_{A_i}(x)
\]

and

\[
n_i \equiv \sum_{x \in D_{t-}} \mu_{A_i}(x)
\]

\( D_{t+} \) and \( D_{t-} \) are the subset of positive and negative instances respectively.

Once the fuzzy rules for the selected class have been learned, the support of this class for the new instance \( x \) is defined by:

\[
s_j(x) \equiv \sum_{i=1}^{k} \mu_{r_j(i)}(x) \ast CF(r_j(i))
\]

where

\[
CF(r_j(i)) = \frac{2 \cdot |D_{t+}^{(i)}| + \sum_{x \in D_{t+}^{(i)}} \mu_{r_j(i)}(x)}{2 + \sum_{x \in D_{t+}} \mu_{r_j(i)}(x)}
\]

The instance is assigned to the class with maximal support.

5. Experimental Settings and Results

5.1. Data and Pre-processing

In order to evaluate the ability of the selected machine learning approaches for inducing CEP rule patterns from large datasets, in this paper we use sensor data generated in phone-based accelerometers, which identifies the physical activities of users. The real-world dataset presented by the Department of Computer and Information Science of Fordham University, which was conducted by gathering accelerometer sensor data from 29 users and indicates their daily activities, was used for our empirical analysis [48].

Accelerometer sensors provide three important measures, namely (i) z-axis values, which record the forward movement of the leg, (ii) y-axis, which captures the upward and downward motion, and (iii) x-axis, which captures horizontal movement of the user’s leg. Daily routine activities such as walking, jogging, ascending stairs, descending stairs, sitting, and standing were selected as outputs of the classification problem. The accelerometer sensor data were collected every 50ms, which makes 20 measurements per second. In order to capture the temporal semantics in derivation of rules for recognition of complex processes, we extended the single time points of sensor values by including their lagged values. The values from the previous five periods for each three-sensor inputs were included to our dataset. By adding these attributes, the number of input variables reached 18.

Before inducing rule patterns using the classifiers, the sensor data set have to be pre-processed. One of the most widely recognized and challenging issues deals with the missing values. The failure to measure the values due to abrasion effects and the problems related to the transfer of the data from sensor device to databases are two popular reasons for missing values. [49] propose various approaches to deal with missing values in sensor data. Deleting the values of the missing instances, applying an algorithm based on iteratively reweighed least squares, and replacing missing values with the mean values of the variables are some typical approaches. In the underlying research, we adopt the later approach due to its simplicity and effectiveness.

Another important issue affecting the performance of the classifiers is data outliers. Outliers are sensor data that deviate from the relevant and logical values of the variables. These outliers can be obvious, which refer to the incorrect measurements of the sensor devices. Using the catalogue of meaningful ranges for variables, these outliers can be easily detected and removed. A challenging issue is identification of hidden outliers, which lay in the predetermined range but cannot be representative for the affected variable. In our research we implemented the “outliers” package of the R software and determined the list of instances with the most difference of the affected variable and removed them from the list.

5.2. Results and Comparison

In this section we present the rules and classification results obtained from FURIA approach after selecting feature subsets with ENORA methodology. The evaluator of ENORA algorithm was set as FURIA as well. The obtained results are then compared to alternative machine learning techniques with and without feature subset selection. After preprocessing data, the WEKA tool was used for classification of instances [50]. Various error measures, such as Root Mean Squared Error, Mean Absolute Error, Relative
Absolute Error, and Root Squared Relative Absolute Error have been used to analyze and compare the performance of the selected rule-based classification approaches.

ENORA approach was run 100 times with the following parameters (See Table 1):

<table>
<thead>
<tr>
<th>Table 1. ENORA parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
</tr>
<tr>
<td><strong>N_eval</strong></td>
</tr>
<tr>
<td><strong>M_min</strong></td>
</tr>
<tr>
<td><strong>M_max</strong></td>
</tr>
<tr>
<td><strong>γ_1</strong></td>
</tr>
<tr>
<td><strong>γ_2</strong></td>
</tr>
<tr>
<td><strong>p_v</strong></td>
</tr>
<tr>
<td><strong>g_a</strong></td>
</tr>
<tr>
<td><strong>NC</strong></td>
</tr>
<tr>
<td><strong>NS_min</strong></td>
</tr>
<tr>
<td><strong>NS_max</strong></td>
</tr>
</tbody>
</table>

These values are standard parameters reported in the literature and accepted by the scientific community [37]. According to the results of feature subset selection phase the number of features are reduced from 18 to 11. After determination of feature subset, the FURIA combined with ENORA was executed to induce decision rules. In this paper, we adopted the 10-fold cross validation methodology to validate the performance results. The performance of the proposed hybrid model was compared to RIPPER, Multilayer Perceptron and Random Forest approaches with and without feature subset selection. According to test results, FURIA classifier combined with ENORA methodology outperforms all other machine learning approaches in terms of both classification accuracy and error measures by classifying 98.39% of instances correctly. Random Forest and Multilayer Perceptron approaches combined with ENORA obtain close results with accuracy rate of 98.19% and 97.99% respectively (See Table 2). The experimental results also suggest that the feature subset selection with multi-objective evolutionary algorithm improves the performance of all classifiers. It is very important to figure out whether the classification results are reliable and were not simply obtained by chance or guesswork.

Kappa’s coefficient (Cohen’s Kappa) is a statistical measure used for this purpose that is assumed to give the rating of the magnitude of agreement between observers. The value of Kappa statistics is measured as the difference between the observed agreement and expected agreement, which refers to occurrence by chance. The formula of the coefficient is as follows:

\[ k = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \]

where \(Pr(a)\) is the observed agreement and \(Pr(e)\) is expected agreement among raters. According to Landis and Koch, if Kappa’s coefficient is equal to 0, there is chance, while if it is equal to 1, there is a perfect agreement among observers [51]. The results reveal that all approaches exceed 0.9 threshold suggesting that all approaches are reliable (See Fig. 2).

Table 2. Classification results

<table>
<thead>
<tr>
<th>Approach</th>
<th>RMSE</th>
<th>MAE</th>
<th>REA</th>
<th>RRSE</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>FURIA with ENORA</td>
<td>0.071</td>
<td>0.008</td>
<td>2.78%</td>
<td>17.89%</td>
<td>98.39%</td>
</tr>
<tr>
<td>FURIA</td>
<td>0.159</td>
<td>0.025</td>
<td>7.94%</td>
<td>39.79%</td>
<td>93.63%</td>
</tr>
<tr>
<td>RIPPER with ENORA</td>
<td>0.105</td>
<td>0.018</td>
<td>5.87%</td>
<td>26.28%</td>
<td>96.82%</td>
</tr>
<tr>
<td>RIPPER</td>
<td>0.178</td>
<td>0.034</td>
<td>10.61%</td>
<td>44.46%</td>
<td>92.02%</td>
</tr>
<tr>
<td>Multilayer Perceptron with ENORA</td>
<td>0.079</td>
<td>0.011</td>
<td>3.55%</td>
<td>19.83%</td>
<td>97.99%</td>
</tr>
<tr>
<td>Multilayer Perceptron</td>
<td>0.169</td>
<td>0.038</td>
<td>11.94%</td>
<td>42.08%</td>
<td>92.17%</td>
</tr>
<tr>
<td>Random Forest with ENORA</td>
<td>0.077</td>
<td>0.010</td>
<td>3.15%</td>
<td>19.25%</td>
<td>98.19%</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.155</td>
<td>0.0382</td>
<td>11.92%</td>
<td>38.80%</td>
<td>93.58%</td>
</tr>
</tbody>
</table>

Figure 2. Kappa values

Another criterion to compare the classifiers is the runtime. All experiments are performed on Intel I7 processor. According to the results, the runtime for FURIA was 33 seconds, whereas the RIPPER approach
needed just 11 seconds to build model on 10000 data points. Increase of runtime by factor can be explained with more complicated steps involved in FURIA algorithm. Generation of unordered rule lists, fuzzification process and rule stretching are reasons for this difference. Random Forest and Multilayer Perceptron required almost the same time as FURIA to build the model.

6. Conclusion

Determination of event patterns is considered as the core task of CEP systems. Inferencing rule patterns from high volume of semantically low level sensor data is a challenging task which requires the involvement of machine learning techniques. In this paper we applied the Fuzzy Unordered Rule Induction Algorithm (FURIA) approach to determine event patterns from accelerometer sensor data. Before implementing the classification algorithm, a multi-objective evolutionary algorithm, ENORA, was used to determine the influential subset of features. The obtained promising results were compared against RIPPER, Multilayer Perceptron and Random Forest approaches with and without feature subset selection.

Although the proposed model achieves high accuracy and shows computational advantages, the complexity of rule interpretations can be considered as its main drawback. [52] proposed a novel graphical approach to ease the interpretability of fuzzy rules provided by FURIA. The future research direction can be set as the analysis of alternative fuzzy rule based classification system which deliver simply interpretable rules. Empirical evaluation and comparison of alternative multi-objective evolutionary algorithms can be considered as another research question. Combining filter feature subset selection approaches with selected wrapper approach may also deliver more precise results. Expressing rules in fuzzy terms gives a new challenge for transferring rules to CEP engines. Incorporation of fuzzy SQL queries to the CEP systems can also be a future technical investigation point.

This research was funded in part by the German Federal Ministry of Education and Research under grant numbers 01IS12050 (project IDENTIFY) and 01IS14004A (project iPREDICT).

7. References