

A Rough Set Approach to Events Prediction in Multiple Time Series

Fatma Ezzahra Gmati¹, Salem Chakhar², Wided Lajoued Chaari¹, and Huijing Chen²

¹ COSMOS, National School of Computer Science, University of Manouba, Manouba, Tunisia.

fatma.ezzahra.gmati@gmail.com, wided.chaari@ensi-uma.tn

² Portsmouth Business School and Centre for Operational Research & Logistics, University of Portsmouth, Portsmouth, UK.

salem.chakhar@port.ac.uk, huijing.chen1@port.ac.uk

Abstract. This paper introduces and illustrates a rough-set based approach to event prediction in multiple time series. The proposed approach uses two different versions of rough set theory to predict events occurrences and intensities. First, classical Indiscernibility relation-based Rough Set Approach (IRSA) is used to predict event classes and occurrences. Then, the Dominance-based Rough Set Approach (DRSA) is employed to predict the intensity of events. This paper presents the fundamental of the proposed approach and the conceptual architecture of a framework implementing this approach.

Keywords: Event Prediction, Multiple Time Series, Rough Sets, Dominance-based Rough Set Approach.

1 Introduction

Event prediction problem is encountered in different research and practical domains. There are several event prediction approaches that have been proposed in the literature. Among these approaches, we may identify the following ones, which are the oldest and also the most used: (i) dynamic systems modeling; (ii) event frequency analysis; (iii) classification based approaches; (iv) events' sequences identification based approaches; and (v) temporal pattern identification based approaches. These approaches are characterized by at least one of the following shortcomings: (1) consider univariate time series only; (2) ignore the temporal dimension; (3) leading to information loss through 'artificial' aggregation of time-varying data; (4) require the use of numerical data only and fail to consider qualitative ones; and (5) ignore the preference that may be associated with the considered variables.

In this paper, a special attention is given to temporal pattern identification based approaches. In this type of approaches, events are predicted based on some characterizing patterns that precede the occurrences of events. This contrasts with the events' sequences identification based approaches where only

the sequences of preceding events are considered. The temporal pattern identification approaches have been applied in several event prediction problems such as flooding [2][4], earthquakes [11], financial events [16] and clinical events [1]. However, existing temporal pattern identification based approaches cannot address all the shortcomings mentioned earlier. For instance, the work of [15] is restricted to univariate time series, while those of [1][12] consider a special type of multiple time series data.

This paper introduces a rough-set based approach to event prediction in multiple time series. This approach adopts a temporal pattern identification strategy and uses two different versions of rough set theory to predict events occurrences and intensities: the event classes and occurrences are predicted based on classical Indiscernibility relation-based Rough Set Approach (IRSA) while the Dominance-based Rough Set Approach (DRSA) is employed to predict the intensity of events. This paper focuses on the theoretical and conceptual aspects of the proposed approach. The application and validation of this approach using real-world data is under progress.

The rest of the paper is organized as follows. Section 2 presents the background on rough set theory. Section 3 introduces the proposed approach. Section 4 provides a framework implementing this approach. Section 5 concludes the paper.

2 Rough Set Theory

2.1 Indiscernibility relation-based Rough Set Approach

The Indiscernibility relation-based Rough Set Approach (IRSA) [13][14] uses equivalence relations to group elements with similar attributes values into indiscernibility classes, and any rough set is characterized by a pair of crisp sets called the lower and the upper approximations. The lower approximation of a rough set contains all the elements that surely belong to the set of interest, where the upper approximation contains all the elements that probably belong to the set of interest.

Let U be a non-empty set of objects (the universe) and D be a non-empty, finite set of attributes such that $q : U \rightarrow V_q$, where V_q is the domain of attribute $q \in D$. With any subset $K \subseteq D$ there is an associated equivalence relation, called K -indiscernibility relation $IND(K)$ such that:

$$IND(K) = \{(x, y) \in U^2 | q(x) = q(y), \forall q \in K\}$$

The relation $IND(K)$ is partitioning U into a set of equivalence classes which is denoted by $U/IND(K)$ or simply U/K . The equivalence classes induced by relation $IND(K)$ are denoted $[x]_K$. Shortly, $[x]_K$ is the equivalence class containing x . In IRSA, any subset $M \subseteq U$ is defined in terms of the elementary sets (equivalence classes) of the partition U/K by lower and upper approximations as follows:

$$- K_*(M) = \{x \in U | [x]_K \subseteq M\}.$$

$$- K^*(M) = \{x \in U | ([x]_K \cap M) \neq \emptyset\}$$

The sets $K_*(M)$ and $K^*(M)$ (or simply M_* and M^*) are called the lower and the upper approximations of M , respectively.

The rough approximations obey the following basic law [14]: $M_* \subseteq M \subseteq M^*$. The difference between the upper and lower approximations is called the boundary of M and is denoted by $Bn_K(M) = M^* - M_*$. If $Bn_K(M) = \emptyset$ then M is crisp (exact) set, otherwise M is rough (inexact) set. Each element $x \in U$ is classified with respect to M as *surely* inside M iff $x \in M_*$ or *probably* inside M , iff $x \in M^*$. Otherwise x is surely outside M .

2.2 Dominance based Rough Sets Approach

The Dominance-based Rough Set Approach (DRSA) [6][7] is an extension of IRSA to multicriteria analysis. The working mechanism of the DRSA is a typical machine learning approach: it uses a subset of data (learning set) to deduce relevant insights that can be used to assess new and unseen datasets (see Figure 1). In this respect, the DRSA is categorized by some authors as a ‘preference learning’ method because it is used to build a preference model based on a sample of past decisions, via preference representation in terms of several if-then rules, for further prescriptive decision purposes.



Fig. 1. Working mechanism of the DRSA

The information regarding the *decision objects* is often structured in a 4-tuple *information table* $\mathbf{S} = \langle U, Q, V, f \rangle$, where U is a non-empty finite set of objects and Q is a non-empty finite set of attributes such that $q : U \rightarrow V_q$ for every $q \in Q$. The V_q is the domain of attribute q , $V = \bigcap_{q \in Q} V_q$, and $f : U \times Q \rightarrow V$ is the *information function* defined such that $f(x, q) \in V_q$ for each attribute q and object $x \in U$. The set Q is often divided into a sub-set $C \neq \emptyset$ of *condition attributes* and a sub-set $D \neq \emptyset$ of *decision attributes*, such that $C \cup D = Q$ and $C \cap D = \emptyset$. In this case, \mathbf{S} is called a *decision table*.

In multicriteria decision making, the domains of the condition attributes are supposed to be ordered according to a decreasing or increasing preference. Such attributes are called *criteria*. The proponents of DRSA assume that the preference is increasing with $f(\cdot, q)$ for every $q \in C$. They also assume that the set of decision attributes $D = \{E\}$ is a singleton. The unique decision attribute E makes a partition of U into a finite number of preference-ordered decision classes $\mathbf{CI} = \{Cl_t, t \in L\}$, $L = \{1, \dots, n\}$, such that each $x \in U$ belongs to one and only one class.

In DRSA the represented knowledge is a collection of *upward unions* $Cl_t^>$ and *downward unions* $Cl_t^<$ of classes defined as follows:

$$Cl_t^> = \bigcup_{s \geq t} Cl_s, Cl_t^< = \bigcup_{s \leq t} Cl_s.$$

The assertion “ $x \in Cl_t^>$ ” means that “ x belongs to at least class Cl_t ” while assertion “ $x \in Cl_t^<$ ” means that “ x belongs to at most class Cl_t ”. The basic idea of DRSA is to replace the indiscernibility relation used in the IRSA with a dominance relation. Let $P \subseteq C$ be a subset of condition criteria. The *dominance relation* Δ_P associated with P is defined for each pair of objects x and y as follows:

$$x \Delta_P y \Leftrightarrow f(x, q) \succeq f(y, q), \forall q \in P.$$

In the definition above, the symbol “ \succeq ” should be replaced with “ \preceq ” for criteria which are ordered according to decreasing preferences. To each object $x \in U$, we associate two sets: (i) the *P-dominating set* $\Delta_P^+(x) = \{y \in U : y \Delta_P x\}$ containing the objects that dominate x , and (ii) the *P-dominated set* $\Delta_P^-(x) = \{y \in U : x \Delta_P y\}$ containing the objects dominated by x .

Then, the *P-lower* and *P-upper* approximations of $Cl_t^>$ with respect to P are defined as follows:

- $\underline{P}(Cl_t^>) = \{x \in U : \Delta_P^+(x) \subseteq Cl_t^>\}$,
- $\bar{P}(Cl_t^>) = \{x \in U : \Delta_P^-(x) \cap Cl_t^> \neq \emptyset\}$.

Analogously, the *P-lower* and *P-upper* approximations of $Cl_t^<$ with respect to P are defined as follows:

- $\underline{P}(Cl_t^<) = \{x \in U : \Delta_P^-(x) \subseteq Cl_t^<\}$,
- $\bar{P}(Cl_t^<) = \{x \in U : \Delta_P^+(x) \cap Cl_t^< \neq \emptyset\}$.

The lower approximations group the objects which certainly belong to class unions $Cl_t^>$ (resp. $Cl_t^<$). The upper approximations group the objects which could belong to $Cl_t^>$ (resp. $Cl_t^<$).

The *P-boundaries* of $Cl_t^>$ and $Cl_t^<$ are defined as follows:

- $Bn_P(Cl_t^>) = \bar{P}(Cl_t^>) - \underline{P}(Cl_t^>)$,
- $Bn_P(Cl_t^<) = \bar{P}(Cl_t^<) - \underline{P}(Cl_t^<)$.

The boundaries group objects that can neither be ruled in nor out as members of class Cl_t .

The *quality of approximation* of a partition \mathbf{CI} by means of a set of criteria P is defined as the ratio of all *P*-correctly classified objects to all objects in the system. The accuracy of the rough-set representation of unions of classes is computed as the ratio between the number of objects in the lower approximation and the number of objects in the upper approximation.

The decision attribute induces a partition of U in a way that is independent of the criteria. Hence, a decision table may be seen as a set of ‘if–then’ decision rules. The condition part specifies the values assumed by one or more criteria, and the decision part specifies an assignment to one or more decision classes. Three types of decision rules may be considered: (i) certain rules generated from the lower approximations of unions of classes, (ii) possible rules generated from the upper approximations of unions of classes, and (iii) approximate rules generated from the boundary regions. Only certain decision rules are considered here. The general structures of this kind of decision rules are as follows:

IF *condition*(s), THEN *At Most* Cl_t

IF *condition*(s), THEN *At Least* Cl_t

The decision part of a certain rule takes the form of an assignment to at most class unions or at least class unions.

3 Event Prediction Approach

The proposed approach is composed of three phases: (i) preprocessing, (ii) analysis and inference of prediction rules, and (iii) prediction of events and their intensities. The first phase of the approach concerns the preprocessing of individual time series data in order to transform them into a format adapted to the extraction of prediction rules in the second phase. The main output of the first is an information table. The second phase concerns the inference of decision rules for predicting events and their intensities. The third phase is devoted to events prediction.

3.1 Phase 1: Preprocessing

In multiple time series, complex events result from a combination of different and related individual patterns. Each of these patterns is governed by one (or more) variables. A comprehensive analysis of complex events requires the definition of a common temporal framework. Three steps are required: (i) labeling and segmentation of individual time series; (ii) construction of a common temporal axis; and (iii) construction of the information table.

Labeling and segmentation of individual time series First, each time series will be segmented into different elementary ‘trendings’ and labeled accordingly. Each elementary trending models a single type of variability such as linear increase, linear decrease, stability, etc. Figure 2 presents some elementary trendings. The identification of elementary trendings will rely on existing time series labeling and segmentation algorithms. Figure 3 illustrates graphically the segmentation of individual time series into a collection of elementary ‘trendings’. The segmentation operation should be applied with two important constraints: all the temporal axis is covered and the absence of ‘holes’ between elementary ‘trendings’.

Time series labeling techniques assign labels to the different segments of a time series. It can be handled using different classical machine learning algorithms such as c-Means and K-Nearest Neighborhood (KNN) (see e.g. [9]). Time series segmentation is a method of time series analysis in which an input time series is divided into a sequence of discrete segments in order to reveal the underlying properties of its source. Examples of time series segmentation techniques are change-point detection methods including sliding windows, bottom-up, and top-down methods [10], and probabilistic methods based on hidden Markov models [5].

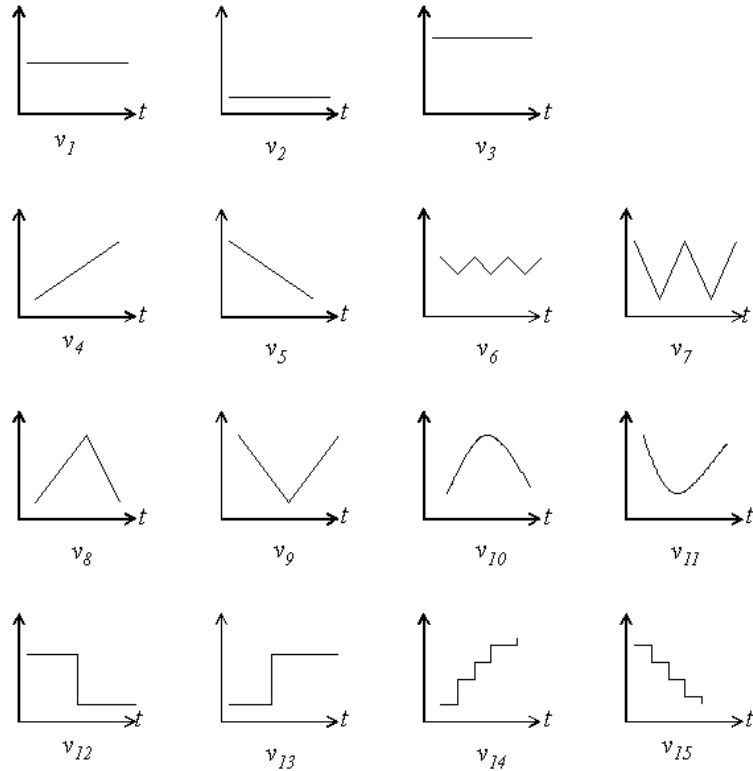


Fig. 2. Examples of elementary trendings

Construction of a common temporal axis The next step is the construction of a common temporal axis. The common temporal axis is constructed by intersecting the temporal axes associated with the segmented individual times series. The obtained axis is then divided into a series of ordered time segments such that in each time segment, each variable will have only one trending. Figure 4 illustrates graphically the construction of a common temporal axis T .

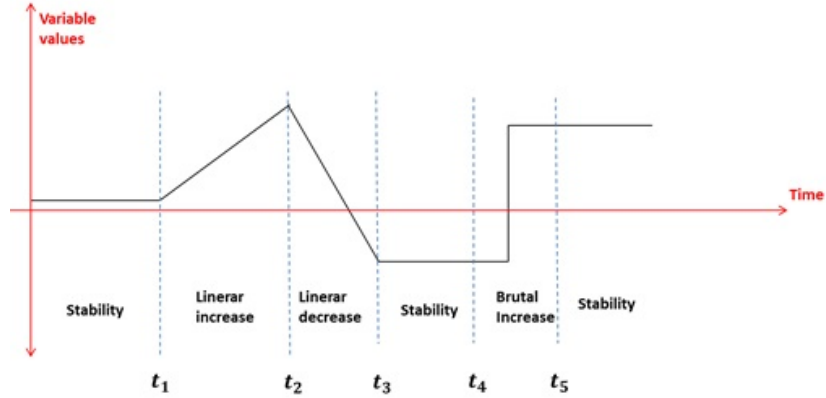


Fig. 3. Identification of elementary trendings

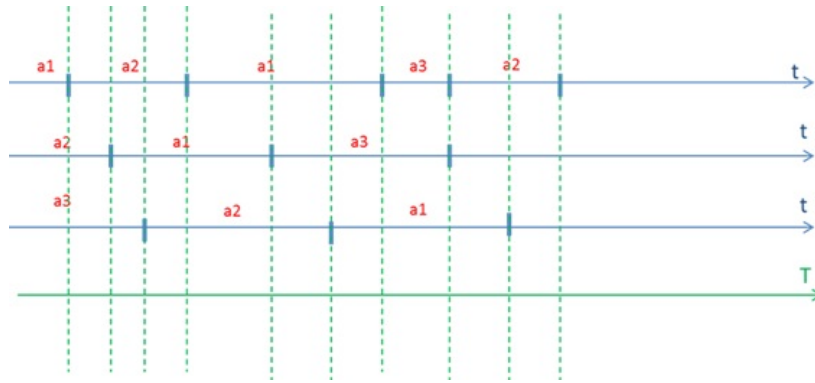


Fig. 4. Construction of a common temporal axis

Construction of information table The third step concerns the transformation of the raw data into an information table. The latter is a matrix data structure that summarizes the raw input data and permits its analysis using IRSA and DRSA. A generic representation of an information table is given in Table 1. The first column specifies the observations O_1, \dots, O_p . The other columns correspond to a set of pairs of the form (T_j, A_k) where T_j ($j = 1, \dots, N$; $k = 1, \dots, M$) is the j th time period in the common time axis T and A_k is the k th variable (condition attribute); N and M are the number of time periods in T and the number of variables, respectively. The entries of the information table are elementary trendings.

3.2 Phase 2: Inference of Prediction Rules

The objective of the second phase is to infer a collection of decision rules that will be used later (in phase 3) to design the prediction algorithms. There are

Table 1. Generic representation of an information table

Observation	(T_1, V_1)	\dots	(T_1, V_M)	\dots	(T_N, V_1)	\dots	(T_N, V_M)
O_1	\dots		\dots		\dots		\dots
\dots	\dots		\dots		\dots		\dots
\dots	\dots		\dots		\dots		\dots
O_p	\dots		\dots		\dots		\dots

two types of prediction rules that we need to infer. The first type permits to predict the occurrence of events while the second type is devoted to predict the intensity of each event. The prediction of occurrence of events relies on IRSA combined with similarity measures, while the prediction of events intensity relies on DRSA.

Decision Rules for Predicting Events Occurrence The inference of events' occurrence rules will rely on the IRSA combined with different similarity measures. In IRSA data analysis starts from a data table called a decision table, which columns are labelled by attributes, rows by objects of interest and entries of the table are attribute values. Attributes of the decision table are divided into two disjoint groups called condition and decision attributes, respectively. Each row of a decision table induces a decision rule, which specifies decision if some conditions are satisfied. If a decision rule uniquely determines decision in terms of conditions the decision rule is certain. Otherwise the decision rule is uncertain. Decision rules are closely connected with approximations. Roughly speaking, certain decision rules describe lower approximation of decisions in terms of conditions, whereas uncertain decision rules refer to the boundary region of decisions. The definition of the lower and upper approximations relies on an equivalence relation, called indiscernibility relation, which ensures that objects having the same descriptions are assigned to the same equivalence class. The indiscernibility relation is very restrictive and thus several extension of this relation have been proposed, most of them rely on a similarity relation.

With respect to multiple time series data analysis, decision objects correspond to observations, condition attributes are the considered decision variables and the decision attributes are the event types. One strong difference between the input data used in the IRSA and the one we needed in multiple time series data analysis, concerns the presence of time-varying attributes. The generic structure of the decision table required to be used in multiple time series data analysis is given in Table 2. It consists of the information table obtained at the end of phase 1 with an additional column (last column in Table 2) corresponding to the types of events. This decision table is different to the conventional one used in IRSA. Indeed, in multiple time series data analysis we need to explicitly take into account the temporal dimension with respect to the values of attributes. Accordingly, an extended version of IRSA is needed.

The extended version of IRSA requires first the definition of a new similarity measure permitting to evaluate the similarity between two multidimensional time series. The similarity relation is a function from $W \times W$ to the range $[0, 1]$ where

Table 2. Generic representation of decision table

Observation	(T_1, V_1)	\cdots	(T_1, V_M)	\cdots	(T_N, V_1)	\cdots	(T_N, V_M)	(Event,Time)
O_1	\cdots		\cdots		\cdots		\cdots	(e_x, t_x)
\cdots	\cdots		\cdots		\cdots		\cdots	
\cdots	\cdots		\cdots		\cdots		\cdots	
O_p	\cdots		\cdots		\cdots		\cdots	(e_y, t_y)

W is the set of rows in the decision table. The new similarity relation takes as input two vectors W_1 and W_2 of the form

$$((T_1, V_1), \cdots, (T_1, V_M), \cdots, (T_N, V_1), \cdots, (T_N, V_M))$$

and provides a value in $[0, 1]$ indicating the similarity relationship between the vectors W_1 and W_2 . This new temporal similarity relation will be used to define the rough approximations of the events in the decision table.

The extended version of IRSA requires also the design of new rule inference algorithms. The inference algorithm should in fact be able to explicitly take into account the temporal dimension that relates the condition attributes.

Decision Rules for Predicting Events Intensity The inference of events' intensities rules will rely on DRSA. The DRSA is an extension of IRSA to multicriteria classification and looks to assign a set of objects described by a set of criteria (attributes with preference-ordered domains) to some pre-defined decision classes or categories, such that each object is assigned to exactly one class. The decision table used as input to DRSA is similar to the once used in IRSA but the condition attributes and the decision attribute are assumed to be preference-ordered. In this research project, the decision classes correspond to intensities of events as shown in Table 3.

Table 3. Generic representation of decision table for events intensity prediction

Observation	(T_1, V_1)	\cdots	(T_1, V_M)	\cdots	(T_N, V_1)	\cdots	(T_N, V_M)	(Event,Intensity,Time)
O_1	\cdots		\cdots		\cdots		\cdots	(e_x, i_x, t_x)
\cdots	\cdots		\cdots		\cdots		\cdots	
\cdots	\cdots		\cdots		\cdots		\cdots	
O_p	\cdots		\cdots		\cdots		\cdots	(e_y, x, t_y)

The dominance relation used in the DRSA is not appropriate for the analysis of multiple times series data. Thus, we need first to extend the classical dominance relation into a new relation that we will call *temporal dominance* relation. This relation permits to specify the preference relations between vectors of the form

$$((T_1, V_1), \cdots, (T_1, V_M), \cdots, (T_N, V_1), \cdots, (T_N, V_M))$$

This vector is similar to the one given above but the definition of the temporal relation requires that each attribute should be associated with a preference direction, which may be gain or cost.

The new temporal domain relation will be used to design a new method as a temporal extension of the DRSA. The design of the TDRSA requires the definition of new concepts of temporal lower approximations and temporal upper approximations. Furthermore, there will be a need to design new decision rules inference algorithms. The inferred decision rules will relate temporal patterns to events and their intensities.

3.3 Phase 3: Prediction of Events and their Intensities

The prediction of future events and their intensities using future trendings and estimated data requires the design of new prediction algorithms. These algorithms take the form of rough classifiers; each is defined in terms of a collection of decision rules. The basic hypothesis on which the prediction algorithm will rely is stated as follows: “similar behaviour (i.e. temporal pattern) will most probably lead to the same event”. The idea is to scan future times series in order to identify future potential temporal patterns that are similar to the temporal patterns that have preceded (and/or followed) by past events.

A critical question that should be considered at this level concerns the time window that should be considered in order to identify the patterns leading to different events. This window should be defined in such a way that the events can be fully differentiated between each other. The idea that will be investigated in this research project consists in starting by using the smallest time window and then progressively extend this window until all the events have been fully predicted.

4 Framework

A multiple times series analysis framework supporting the different conceptual, theoretical and algorithmic solutions that been indicted in the previous section is being designed and implemented. The conceptual architecture of this framework is given in Figure 5. As shown in this figure, the framework is composed of two main layers. The first layer is devoted to predict event occurrences while the second layer is aimed to predict the intensity of events. An additional layer of overall validation is added to enhance the framework.

5 Conclusion

We presented the first results of a rough-set based approach to event prediction in multiple time series. This approach adopts a temporal pattern identification strategy and uses two different and extended versions IRSA and DRSA, which are rough set based approaches. The major advantage of using IRSA and DRSA

relies on the fact that they are able to deal with almost any kind of data (symbolic, binary, ordinal and numerical) and their ability to handle missing data.

The application and validation of the proposed approach using real-world data is under progress. The computational behavior and performance of the different algorithms with very large datasets will also be investigated. Additionally, the proposed approach can easily be enhanced by using recent extensions of IRSA and DRSA such as the Variable Consistency Dominance-based Rough Set Approach (VC-DRSA) [8] that enables relaxation of the conditions for assigning decision objects to the lower approximations by accepting a limited proportion of negative examples, which is particularly useful for large decision tables and the Stochastic DRSA [3] relaxes the rough approximations in the DRSA by allowing inconsistencies to some degree.

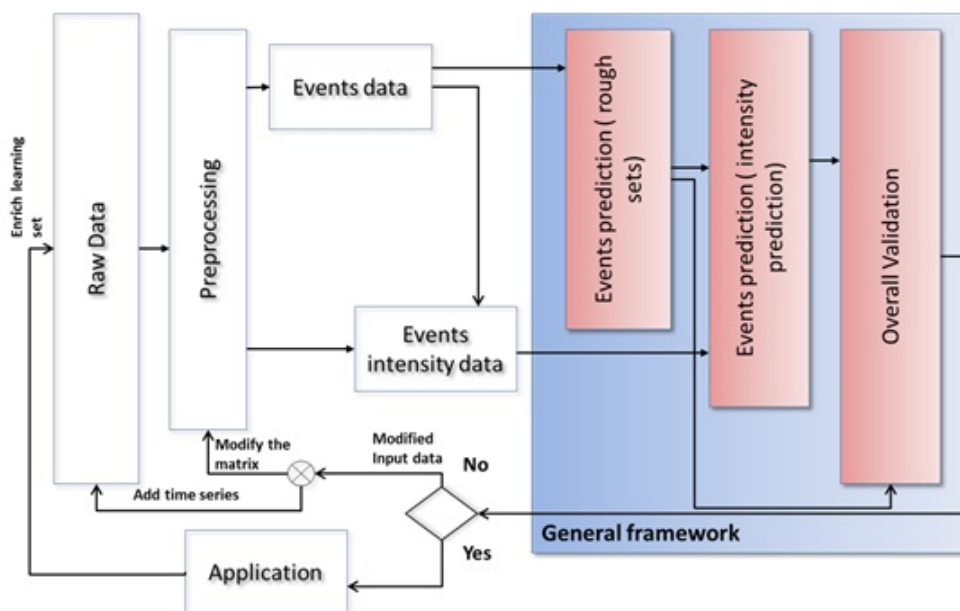


Fig. 5. Architecture of the proposed framework

References

1. I. Batal, G.F. Cooper, D. Fradkin, J.H.Jr.F., Moerchen, and M. Hauskrecht. An efficient pattern mining approach for event detection in multivariate temporal data. *Knowledge Information Systems*, 2015.
2. C. Damle and A. Yalcin. Flood prediction using time series data mining. *Journal of Hydrology*, 333:305–316, 2006.

3. K. Dembczynski, S. Greco, W. Kotłowski, and R. Slowiński. Statistical model for rough set approach to multicriteria classification. In J.N. Kok, J. Koronacki, R. Lopez de Mantaras, S. Matwin, D. Mladenic, and A. Skowron, editors, *Knowledge Discovery in Databases: PKDD 2007*, volume 4702 of *Lecture Notes in Computer Science*, pages 164–175. Springer Berlin Heidelberg, 2007.
4. M.G. Erechtkhoukova, P.A. Khaitey, and S. Saffarpour. Short-term predictions of hydrological events on an urbanized watershed using supervised classification. *Water Resource Management*, 30:4329–4343, 2016.
5. E.B. Fox, E.B. Sudderth, M.I. Jordan, and A.S. Willsky. An HDP-HMM for systems with state persistence. In *Proceedings of the 25th International Conference on Machine Learning*, pages 312–319, New York, NY, USA, 2008. ACM.
6. S. Greco, B. Matarazzo, and R. Slowiński. The use of rough sets and fuzzy sets in MCDM. In T. Gal, T. Hanne, and T. Stewart, editors, *Advances in Multiple Criteria Decision Making*, pages 14.1–14.59. Kluwer Academic Publishers, Dordrecht, Boston, 1999.
7. S. Greco, B. Matarazzo, and R. Slowiński. Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, 129(1):1–47, 2001.
8. S. Greco, B. Matarazzo, R. Slowiński, and J. Stefanowski. Variable consistency model of dominance-based rough sets approach. In W. Ziarko and Y. Yao, editors, *Rough Sets and Current Trends in Computing*, volume 2005 of *Lecture Notes in Computer Science*, pages 170–181. Springer Berlin / Heidelberg, 2001.
9. J. Harguess and J.K. Aggarwal. Semantic labeling of track events using time series segmentation and shape analysis. In *The 16th IEEE International Conference on Image Processing (ICIP 2009), Cairo, Egypt*, pages 4317–4320. IEEE, 2009.
10. E. Keogh, S. Chu, D. Hart, and M. Pazzani. *Segmenting time series: A survey and novel approach*, volume 57 of *Series in Machine Perception and Artificial Intelligence*, pages 1–21. World Scientific, 2011.
11. A. Morales-Esteban, F. Martinez-Alvarez, A. Troncoso, J.L. Justo, and C. Rubio-Escudero. Pattern recognition to forecast seismic time series. *Expert Systems with Applications*, 37:8333–8342, 2010.
12. F. Morchen and A. Ultsch. Discovering temporal knowledge in multivariate time series. In C. Weihs and W. Gaul, editors, *Classification — the Ubiquitous Challenge: Proceedings of the 28th Annual Conference of the Gesellschaft für Klassifikation e.V. University of Dortmund, March 9-11, 2004*, pages 272–279. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
13. Z. Pawlak. Rough sets. *International Journal of Information & Computer Sciences*, 11:341–356, 1982.
14. Z. Pawlak. *Rough set. Theoretical aspects of reasoning about data*. Kluwer Academic Publishers, Dordrecht, 1991.
15. J.R. Povinelli and X. Feng. A new temporal pattern identification method for characterization and prediction of complex time series events. *IEEE Transactions On Knowledge And Data Engineering*, 15(2):339–352, 2003.
16. R.J. Povinelli. *Time Series Data Mining: Identifying Temporal Patterns for Characterization and Prediction of Time Series Events*. PhD thesis, Marquette University, Milwaukee, WI, 1999.