



A multicriteria approach based on rough set theory for the incremental Periodic prediction

Sarra Bouzayane, Inès Saad

► To cite this version:

Sarra Bouzayane, Inès Saad. A multicriteria approach based on rough set theory for the incremental Periodic prediction. European Journal of Operational Research, 2020, 286, pp.282 - 298. 10.1016/j.ejor.2020.03.024 . hal-03490674

HAL Id: hal-03490674

<https://hal.science/hal-03490674>

Submitted on 22 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

A Multicriteria Approach based on Rough Set Theory for the Incremental Periodic Prediction

Sarra Bouzayane^{a,**}, Inès Saad^{a,b}

^aUniversity of Picardie Jules Verne, MIS Laboratory, Amiens 80039, France

^bAmiens Business School, Amiens 80039, France

Abstract

This paper proposes a Multicriteria Approach for the Incremental Periodic Prediction (*MAI2P*). This approach is periodically applied while considering the sequential evolution of the dynamic information system under the variation of the set of actions in an ever-evolving learning sample. It is based on the Dominance-based Rough Set Approach (DRSA) and consists of three phases. The first aims at constructing a decision table and is based on three steps: (1) constructing a representative learning sample of “Actions of reference”, (2) constructing a coherent criteria family for the actions’ characterization and (3) building a decision table. The second consists in an incremental updating of the DRSA approximations in order to infer a preference model resulting in a set of decision rules. The third consists of classifying the potential actions in one of the predefined decision classes. The first two phases run at the end of the current period and the third phase runs at the beginning of the next period. The approach *MAI2P* has been applied in the context of Massive Open Online Courses (MOOCs). It has been validated on a French MOOC proposed by a Business School in France. Experiments showed that the pessimistic cumulative approach gives the most efficient preference model with an F-measure and an accuracy values reaching 0.66 and 0.89 respectively.

Keywords: Decision support, Dominance-based Rough Set Approach, Incremental learning, Periodic prediction, Massive Open Online Courses

^{*}Fully documented templates are available in the elsarticle package on CTAN.

^{*}Corresponding author

^{**}Corresponding author

Email addresses: sarra.bouzayane@u-picardie.fr (Sarra Bouzayane),
ines.saad@esc-amiens.com (Inès Saad)

1. Introduction

The amount of data collected and stored into databases has significantly increased. Consequently, the traditional data analysis techniques have become unsuitable for the processing of such huge volumes and new techniques have emerged [16]. In this work, we focus on the techniques used for the prediction issue. Prediction is the reasoning from the current events to the future events they will cause [42]. Recently, it was given a special importance in the temporal reasoning literature and was involved into several areas such as those of medicine [16, 50], finance [56, 44] and biology [33, 49].

The prediction process is frequently based on the conventional machine learning techniques such as the Markov chains [41, 54], the neural network [43, 53], the support vector machine [35, 47]. However, these techniques are rather based on the size of the learning sample without considering the quality of the assignment examples it contains that are usually chosen randomly. According to [22], the identification of a learning sample for the prediction model construction represents the central problem in machine learning. In this paper, we use the DRSA (Dominance-based Rough Set Approach) [20] as a technique of supervised learning that is based on the preferences and the expertise of human decision makers in order to define a representative and a quality learning sample. Moreover, we propose a *DRSA-Incremental* algorithm that allows the incremental update of the DRSA decision rules following the insertion of new examples into the learning set. This algorithm permits to minimize the computation time made by the conventional machine learning techniques that resume the calculation from scratch in this case of incremental learning.

Generally, the DRSA approach applies for a purpose of multicriteria classification [24, 28, 40, 46] or prediction [44, 56] in a static information system. The multicriteria classification is an ordinal classification that assigns a set of actions described by a set of criteria to some preference ordered decision classes. Each action is assigned to exactly one class such that the decision classes are defined in an ordinal way [8]. The objective of the multicriteria classification is to guide the decision by a sorting that results from an assignment of each action to a decision class. Indeed, this work presents a

Multicriteria Approach for the Incremental Periodic Prediction (*MAI2P*) based on our *DRSA-incremental* algorithm to deal with the prediction issue in the case of a permanently evolving learning set. This approach is based on the multicriteria classification of the learning actions in order to periodically predict the decision classe to which each potential action will belong.

In this work, we focus on the incremental learning that occurs during the sequential enrichment of the learning set. Incremental learning refers to the situation of continuous adaptation of the model based on a constant arrival of learning examples [45]. This type of learning occurs when the training examples are provided sequentially as a series of observations or measurements. This sequential arrival leads to the sequential enrichment of the learning set which necessarily requires the updating of the previously taken decisions [21]. Thus, we must learn the model after each new observation that may modify the previous prediction. The principle of the *DRSA* approach requires, in this case, the updating of the decision rules or that of the *DRSA* approximations in order to guarantee an updated decisions considering the evolution over time of a real world case (e.g. chess games, weather forecasts). In the remainder of this paper we use the term classification to say multicriteria classification.

The approach *MAI2P* considers the preferences of human decision makers in the decision-making process and consists of three phases. The first aims to build a decision table and is based on three steps: first, the identification of a representative learning sample for the N predefined decision classes; then, the construction of a coherent family of criteria for the characterization of actions; and finally, the classification of each action in one of these decision classes. The second phase concerns the inference of a preference model that results in a set of decision rules. It is based on the incremental update of the upper and lower approximations of the decision classes unions in the *DRSA* approach. The third phase consists of the early classification (prediction) of the “Potential Actions” using the previously inferred preference model. The first two phases run at the end of the current period and the third phase runs at the beginning of the next period. The approach *MAI2P* has been implemented in the context of MOOCs and has been validated on a French MOOC proposed by a Business School in France.

The paper is organized as follows: Section 2 presents the background. Section 3

discusses the related work. Section 4 introduces the approach *MAI2P* proposed for the periodic prediction. Section 5 deals with the algorithm *DRSA-Incremental* proposed for the incremental update of the DRSA approximations at the second phase of the approach *MAI2P*. Section 6 presents a case study and discusses the corresponding results. Section 7 concludes the paper and advances some prospects.

2. Dominance-based Rough Set Approach (DRSA)

The DRSA approach was proposed by Greco et al. [20] and was inspired from the Rough Sets Theory [38]. It allows to compare actions through a dominance relation and takes into account the preferences of a decision maker to extract a preference model resulting in a set of decision rules. According to DRSA, a data table is a 4-tuple $S = \langle A, F, V, f \rangle$, where:

- A is a finite set of reference actions,
- F is a finite set of criteria,
- $V = \bigcup_{g \in F} V_g$ is the set of the possible values of criteria, and
- f denotes an information function $f: A \times F \longrightarrow V$ such that $f(x, g) \in V_g, \forall x \in A, \forall g \in F$.

F is often divided into a subset $C \neq \emptyset$ of condition attributes and a subset $D \neq \emptyset$ of decision attributes such that $C \cup D = F$ and $C \cap D = \emptyset$. In this case, $S \cup D$ is called a decision table. In multicriteria decision making, the scale of condition attributes should be ordered according to a decreasing or an increasing preference of a decision maker. Such attributes are called criteria. We also assume that the decision attribute set $D = \{d\}$ is a singleton that partitions A into a finite number of decision classes $Cl = \{Cl_n, n \in \{1, \dots, N\}\}$, such that each $x \in A$ belongs to one and only one class. Furthermore, we suppose that the classes are preference-ordered, i.e., for all $r, s \in \{1, \dots, N\}$ such that $r > s$, actions from Cl_r are preferred to actions from Cl_s . The approach DRSA introduces a set of parameters that will be defined and explained using the decision table described below (Table 1):

Table 1: Decision table

	g_1	g_2	g_3	g_4	D
A_1	2	2	3	3	3
A_2	1	2	2	3	3
A_3	2	1	1	2	1
A_4	1	2	3	1	2
A_5	2	1	3	3	2
A_6	1	2	1	1	1
A_7	2	2	2	3	3
A_8	1	1	1	2	1
A_9	2	1	2	3	3
A_{10}	1	2	2	2	2
A_{11}	1	2	3	3	2

Example 1: Table 1 gives the decision table used in this paper. The actions set $A = \{A_i : i = 1, 2, \dots, 11\}$ contains eleven actions. Each action is described using four condition criteria $P = \{g_1, g_2, g_3, g_4\}$ and a decision criterion D . Throughout this paper, we consider that the decision table initially contains only the actions $\{A_1, A_2, A_3, A_4, A_5, A_6\}$. Thus, we will not consider for the moment the actions $\{A_7, A_8, A_9, A_{10}, A_{11}\}$.

In DRSA, the represented knowledge is a collection of downward unions Cl_n^{\leq} and upward unions Cl_n^{\geq} of decision classes such that: $Cl_n^{\leq} = \cup_{s \leq n} Cl_s$, $Cl_n^{\geq} = \cup_{s \geq n} Cl_s$; $n \in \{1..N\}$. The assertion “ $x \in Cl_n^{\leq}$ ” means that “ x belongs to at most the decision class Cl_n ”, while “ $x \in Cl_n^{\geq}$ ” means that “ x belongs at least to the decision class Cl_n ”.

Example 2: Consider Table 1, the downward unions are: $Cl_1^{\leq} = Cl_1 = \{A_3, A_6\}$, $Cl_2^{\leq} = Cl_1 \cup Cl_2 = \{A_3, A_4, A_5, A_6\}$ and $Cl_3^{\leq} = Cl_1 \cup Cl_2 \cup Cl_3 = \{A_1, A_2, A_3, A_4, A_5, A_6\}$. The upward unions of decision classes are: $Cl_1^{\geq} = Cl_1 \cup Cl_2 \cup Cl_3 = \{A_1, A_2, A_3, A_4, A_5, A_6\}$, $Cl_2^{\geq} = Cl_2 \cup Cl_3 = \{A_1, A_2, A_4, A_5\}$ and $Cl_3^{\geq} = Cl_3 = \{A_1, A_2\}$.

Dominance relation. Let $P \subseteq C$ be a subset of condition criteria. The dominance relation D_P associated with P is defined for each pair of actions x and y thus:

$$\forall (x, y) \in A^2, x D_P y \Leftrightarrow f(x, g_j) \succcurlyeq f(y, g_j) \forall g_j \in P$$

To each action $x \in A$, two sets are associated:

- P-dominating set $D_P^+(x) = \{y \in A : y D_P x\}$ containing actions that dominate x .
- P-dominated set $D_P^-(x) = \{y \in A : x D_P y\}$ containing actions dominated by x .

Example 3: Consider Table 1 and suppose that all condition criteria are ordered according to increasing preferences. Based on this information, it is easy to establish that A_1 dominates all actions because $f(A_1, g_j) \succcurlyeq f(A_i, g_j) \forall g_j \in P$ and $A_i \in A$. However, there is no dominance relationship between actions A_2 and A_3 , because the evaluation of A_2 is preferred than A_3 on criteria g_2, g_3 and g_4 but the evaluation of A_3 is better than A_2 on criterion g_1 . Consequently, the P-dominating sets are as follows: $D_P^+(A_1) = \{A_1\}$, $D_P^+(A_2) = \{A_1, A_2\}$, $D_P^+(A_3) = \{A_1, A_3, A_5\}$, $D_P^+(A_4) = \{A_1, A_4\}$, $D_P^+(A_5) = \{A_1, A_5\}$ and $D_P^+(A_6) = \{A_1, A_2, A_4, A_6\}$. The P-dominated sets are as follows: $D_P^-(A_1) = \{A_1, A_2, A_3, A_4, A_5, A_6\}$, $D_P^-(A_2) = \{A_2, A_6\}$, $D_P^-(A_3) = \{A_3\}$, $D_P^-(A_4) = \{A_4, A_6\}$, $D_P^-(A_5) = \{A_3, A_5\}$ and $D_P^-(A_6) = \{A_6\}$.

Lower approximation. The P-lower approximations of (Cl_n^{\leq}) and (Cl_n^{\geq}) with respect to $P \subseteq C$, respectively denoted $\underline{P}(Cl_n^{\leq})$ and $\underline{P}(Cl_n^{\geq})$, are defined as follows:

- $\underline{P}(Cl_n^{\leq}) = \{x \in A : D_P^-(x) \subseteq Cl_n^{\leq}, \forall n \in \{1..N\}\}$
- $\underline{P}(Cl_n^{\geq}) = \{x \in A : D_P^+(x) \subseteq Cl_n^{\geq}, \forall n \in \{1..N\}\}$

The P-lower approximation of Cl_n^{\leq} (resp. Cl_n^{\geq}) contains all actions whose P-dominated (resp. P-dominating) set is assigned with certainty to classes that are at most (resp. at least) as good as Cl_n .

Upper approximation. The P-upper approximations of (Cl_n^{\leq}) and (Cl_n^{\geq}) with respect to $P \subseteq C$, respectively denoted $\overline{P}(Cl_n^{\leq})$ and $\overline{P}(Cl_n^{\geq})$, are defined as follows :

- $\overline{P}(Cl_n^{\leq}) = \{x \in A : D_P^+(x) \cap Cl_n^{\leq} \neq \emptyset, \forall n \in \{1..N\}\}$

- $\bar{P}(Cl_n^{\geq}) = \{x \in A : D_P^-(x) \cap Cl_n^{\geq} \neq \emptyset, \forall n \in \{1..N\}\}$

The P-upper approximation of Cl_n^{\leq} (resp. Cl_n^{\geq}) that contains all actions with P-dominating (resp. P-dominated) set is assigned to a class at most (resp. least) as good as Cl_n .

Example 4: Using the results of Example 2 and Example 3; the lower approximation is calculated as follows: $\underline{P}(Cl_1^{\leq}) = \{A_3, A_6\}$, $\underline{P}(Cl_2^{\leq}) = \{A_3, A_4, A_5, A_6\}$, $\underline{P}(Cl_3^{\leq}) = \{A_1, A_2, A_3, A_4, A_5, A_6\}$, $\underline{P}(Cl_1^{\geq}) = \{A_1, A_2, A_3, A_4, A_5, A_6\}$, $\underline{P}(Cl_2^{\geq}) = \{A_1, A_2, A_4, A_5\}$, $\underline{P}(Cl_3^{\geq}) = \{A_1, A_2\}$. For example, $D_P^-(A_3) = \{A_3\} \subseteq Cl_1^{\leq} = \{A_3, A_6\}$ then $A_3 \in \underline{P}(Cl_1^{\leq})$. In this example, all actions were classified with certainty and so the upper approximations still the same that the lowr ones. Hence, we have $\bar{P}(Cl_1^{\leq}) = \{A_3, A_6\}$, $\bar{P}(Cl_2^{\leq}) = \{A_3, A_4, A_5, A_6\}$, $\bar{P}(Cl_3^{\leq}) = \{A_1, A_2, A_3, A_4, A_5, A_6\}$, $\bar{P}(Cl_1^{\geq}) = \{A_1, A_2, A_3, A_4, A_5, A_6\}$, $\bar{P}(Cl_2^{\geq}) = \{A_1, A_2, A_4, A_5\}$, $\bar{P}(Cl_3^{\geq}) = \{A_1, A_2\}$.

P-doubtful region. The P-boundaries (or P-doubtful region) of Cl_n^{\leq} and Cl_n^{\geq} are defined as follows:

- $\text{Bnp}(Cl_n^{\leq}) = \bar{P}(Cl_n^{\leq}) - \underline{P}(Cl_n^{\leq})$
- $\text{Bnp}(Cl_n^{\geq}) = \bar{P}(Cl_n^{\geq}) - \underline{P}(Cl_n^{\geq})$

The P-boundaries group actions that can be ruled neither inside nor outside as members of class Cl_n .

Decision rule. A decision table may be considered as a set of “if...then...” decision rules, where the condition part specifies values assumed by one or more condition criteria and the decision part specifies an assignment to one or more decision classes. An action $x \in A$ supports a decision rule if its description matches both the condition and the decision parts of the rule. A decision rule covers an action x if the description of x matches at least the condition part of the rule. Decision rules used in this paper are represented as follows:

- **if** $f(x, g_1) \geq r_1 \wedge \dots \wedge f(x, g_m) \geq r_m$ **then** $x \in Cl_n^{\geq}$; such that $(r_1 \dots r_m) \in (v_{g_1} \dots v_{g_m})$.
- **if** $f(x, g_1) \leq r_1 \wedge \dots \wedge f(x, g_m) \leq r_m$ **then** $x \in Cl_n^{\leq}$; such that $(r_1 \dots r_m) \in (v_{g_1} \dots v_{g_m})$.

3. Related work

In this section, we present some works using DRSA for a prediction purpose within a static information system. Then, we discuss two methods proposed for the incremental update of the decision rules when changing the actions set.

3.1. Static prediction models based DRSA

The DRSA approach was used in several areas such as the medicine field [39], the marketing [27] and the finance field [56, 44] for a prediction purpose. The objective was to provide the decision makers with a visibility on a future situation so that they could make a proper decision and avoid undesirable consequences.

Author in [56] used the approach DRSA for a hybrid prediction of a probable company bankruptcy. The classification is based on the discriminator formula $Z = 1.2 * X1 + 1.4 * X2 + 3.3 * X3 + 0.6 * X4 + 0.99 * X5$ proposed by Altman [1] such that $X1$ is $\frac{\text{the Capital Fund}}{\text{Total assets}}$; $X2$ is $\frac{\text{the Retained earnings}}{\text{Total assets}}$; $X3$ is $\frac{\text{the earnings before interest and taxes}}{\text{Total assets}}$; $X4$ is $\frac{\text{the equity market value}}{\text{Total liabilities}}$ and $X5$ is $\frac{\text{the Sales}}{\text{Total assets}}$. Three decision classes are thus identified: The decision class P was assigned to prosperous companies if $Z > 2.67$, the decision class B was dedicated to bankrupt companies if $Z < 0$ and the decision class M was reserved to companies about to bankrupt if $0 < Z < 2.67$. The evaluation criteria were defined in [18] and cover the mechanisms allowing a business to succeed. For each criterion, a scale of seven levels ordered in an ascending order of preference from 1 to 7 was fixed (1: Fort, 6: very low, 7: no idea). The method was tested on data from seven Quebecois companies. The Z value was calculated over a period of 5 years. The preference model obtained was built by five decision rules and is characterized by a quality of approximation of 0.57.

Shen and Tzeng [44] proposed a method based on DRSA for the prediction of the financial performance of commercial banks. The method is based on two phases. The first aims at reducing the complexity of the involved attributes and then at inferring a set of decision rules. It consists of two steps: a discretization step of both the condition and the decision attributes and an inference step of a set of decision rules of the type if-then. The second phase targets to refine the decision rules via the neuro fuzzy inference

technique which would identify the fuzzy intervals of each attribute. It also permits to know whether the criterion is considered as a cause or an effect of performance. The report selected to measure performance is the return on assets (ROA) which indicates the development or the deterioration of the financial performance. The authors identified a set of twenty five criteria such that to each criterion is assigned a three-level scale that varies from low to medium to high. This method is validated through data relative to five banks in Taiwan. Data collected during the period 2008 to 2011 are used for the model training and those which are specific to the year 2012 are applied to the model validation. Based on a descending ranking of the banks' ROA, three decision classes are defined: "Good" for the interval of the first third of banks, "Medium" for the second third interval and "Bad" for the last third interval. The application of this method generated four decision rules refined by the fuzzy technique.

In both methods, the information system is kept stable over several years without updating neither the set of criteria nor that of actions. In this case, the decision taken may be obsolete since it does not consider the factors that may emerge between the learning and the validation periods.

3.2. Models for the incremental update of the approach DRSA under the variation of the set of actions

The DRSA information system comprises the criteria [57], the actions [21, 26] and the criteria values [12]. Sometimes, one or all of these elements can change over time in order to keep a dynamic view of the knowledge to deal with. In such dynamic situation, the traditional rough set methods recalculate the approximations by scratch which makes the computation time too high, a thing that is discouraged especially in the case of real time decisions. In literature, some approaches proposed for the incremental update of the information system under the variation of the set of actions [21, 26], the set of criteria [57] or the criteria values [12].

Authors in [21] propose an algorithm, called Glance, to update the rules base in the case of an incremental learning. This algorithm considers the actions as negative ones. In fact, it establishes decision rules that do not cover the action x in all the unions of the decision classes to which x does not belong. Hence, each rule in a given union does not

need to absolutely satisfy x , if x does not belong to the union, but it cannot also satisfy the action x and remains without any support. Thus, the unsupported decision rules induced by the Glance algorithm are said “non-robust” and so is said the algorithm. The Glance algorithm stores in the computer memory only the decision rules and not the training examples. This makes it economical compared to the use of the memory space. Its complexity is linear considering the number of actions and exponential considering the number of criteria.

The authors in [26] propose an incremental update algorithm of the DRSA approximations when adding or deleting an action in or from the information system. This method is based on the *P-generalized decision* [25] for calculating the P-upper and P-lower approximations of each decision class union. This incremental update requires: first, the update of the upward and the downward of the unions of the decision classes; second, the update of the P-dominated and the P-dominating of each action in the information system and finally, the update of the upper and the lower approximations of the unions of the decision classes. This method minimizes the computational time when an action enters in or exits from the information system. It is also sensitive to the number of actions and to that of attributes. However, the algorithm does not affect the quality of the decision rules since it just updates the approximation and does not make any requirement about the induction algorithm.

In this work, we generalize the algorithm presented in [26] in order to obtain robust decision rules that consider the simultaneous entry of many actions instead of only one.

4. Multicriteria Approach for the Incremental Periodic Prediction

In this section, we propose a multicriteria approach for the incremental and periodic prediction of the decision class Cl_n to which an action x is likely to belong, such that $n \in \{1..N\}$ and N is the number of decision classes. This method consists of three phases: The first concerns the construction of the decision table. The second aims to infer a preference model based on the incremental update of the DRSA approximations. The third is the classification of each “Potential action” into one of the N predefined decision classes using the previously inferred preference model (cf. Figure 1).

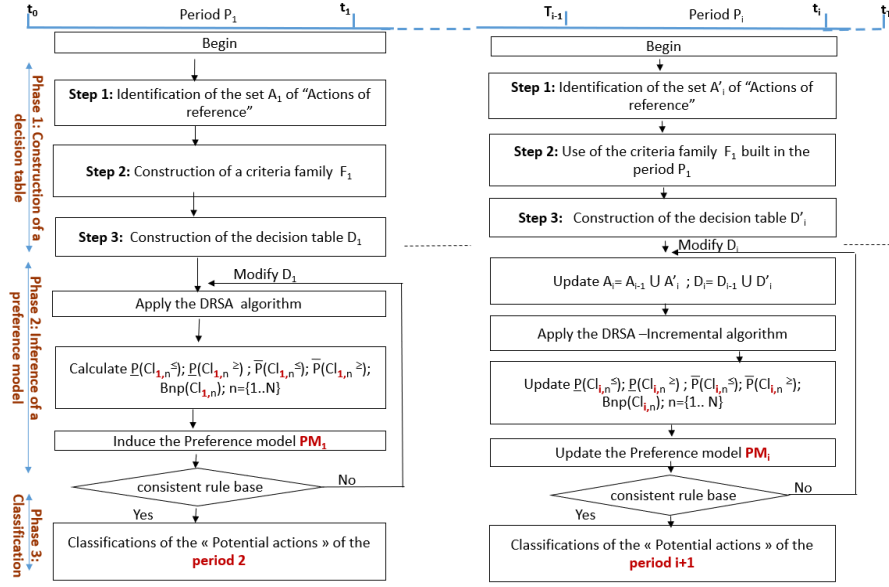


Figure 1: Incremental method for the periodic multicriteria classification

4.1. Phase 1: Construction of the decision table

This phase comprises three steps: (1) the construction of a learning set of actions, called “actions of reference”, (2) the construction of a coherent family of criteria to actions characterization and (3) the construction of the decision table.

Step 1.1: *The identification of a set of “Actions of reference”.* Given the growing mass of data, it is difficult to analyse all of them. Therefore, it is necessary to define a learning sample containing a sufficient number of representative examples for each of the predefined decision classes Cl_n such that $n \in \{1..N\}$. In order to respect the terminology used in the DRSA approach, we call the learning examples “Actions of reference”. This set is identified by decision makers who must be expert in the application field.

However, since the human dimension is strongly involved in the construction phase of the decision table (selection and classification of the “Actions of reference”), the sample size to be proposed must be taken into account. In effect, from a psychological viewpoint [30], a human decision maker is characterized by a cognitive capacity representing the upper limit to which he can associate his responses to the stimuli that

are granted to him. Thus, to construct a learning set, it is sufficient to respect the quality of the selected actions, regardless of their size. Otherwise, a large learning sample may degrade the quality of the assignment decisions made by the decision maker, and consequently that of the decision rules. Thus, this step aims to build a set of “Actions of reference” of both high quality and reasonable size, in harmony with the cognitive capacity of the human decision maker.

Since the approach *MAI2P* is intended to be used in a context of dynamic information systems, it must deal with the non-stable “Actions of reference” that vary periodically throughout the prediction process. Thus, each period P_i , the decision maker identifies a new set A'_i of “Actions of reference” which is added to all of the “Actions of reference”, A_{i-1} , of all the previous periods. The set of “Actions of reference” of the period P_i is therefore $A_i = A'_i + A_{i-1}$; $\forall i \in \{2..T\}$ such that T is the number of periods constituting the prediction process.

The number of “Actions of reference” is to be fixed by the experts decision makers according to their cognitive capacities, preferences and expertises. It can change from one period to another (*i.e.* $A'_i \neq A'_{i-1}$), depending on the context characteristics and the application objectives.

In the case of real-time decision where the saving-time becomes important, it is possible to use one of the clustering techniques to constitute the sets of “reference action”. These techniques allow to automate this step in order to prevent the human intervention which is time consuming.

Step 1.2: *The construction of a coherent family of criteria.* A criteria family is built by the decision makers and must provide the judgment tools for them. It must check the coherence between their expectations and the needs of the concerned actors [4]. Compared to an attribute, a criterion must be used to measure the preferences of a decision maker from a personal viewpoint [34].

Mousseau and his co-authors [34] defined two approaches to construct a criteria family: the top-down approach and the bottom-up approach. The top-down approach consists in declining one or more strategic objectives according to viewpoints and dimensions [15]. Whereas, the bottom-up approach consists in building a family of crite-

ria from a list of indicators that can influence the opinion of decision makers regarding the characterization of actions. In multicriteria decision making these indicators are called “Consequences” [34]. This approach is also based on information gathered during interviews with experts in the field. In this work, we adopt the mixed approach that combines both bottom-up and top-down approaches. Once the list of consequences is established, direct meetings must be held with the decision makers to obtain their preferential information on each criterion. To be coherent, a criteria family must satisfy completeness, cohesion and non-redundancy [13]. In this context, we consider that the criteria family is stable during the prediction process.

Step 1.3: The construction of the decision table. This step is made of two sub-steps: (i) the construction of the information table, and (ii) the construction of the decision table. The information table S_i built at the end of a period P_i is a matrix whose rows form the set of the “ m ” “Actions of reference” identified in step 1.1 and contained in A'_i and whose columns represent the “ p ” evaluation criteria constructed in step 1.2 and contained in F_1 . This matrix contains the evaluation function $f_i(A_{j,i}, g_k)$ of each action $A_{j,i} \in A'_i$ on each criterion $g_k \in F_1$ such that $i \in \{1..T\}$, $j \in \{1..m\}$ and $k \in \{1..p\}$. Variables T , m and p are respectively the number of considered periods, the size $|A'_i|$ of the “Actions of reference” set defined in the i^{th} period and the size $|F_1|$ of the criteria family.

Analogously, variables $A_{j,i}$ and g_k constitute respectively the j^{th} “Action of reference” in the set A'_i and the k^{th} criterion. We remind that A'_i and F_1 represent respectively the set of “Actions of reference” constructed in the i^{th} period and the family of criteria identified at the beginning of the decision process. Once the information table S_i is complete at the end of the i^{th} period, we have to build the decision table with the expert decision makers during some meetings. Thus, we only have to add a column to the end of the information table. The last column concerns the affectation of each action of reference in one of the predefined decision classes (cf. Table 2).

The decision table is thus composed of “ m ” rows and “ $p+1$ ” columns. The assignment decisions made by the decision maker are based on both his expertise and the assessment vectors. We call $D_j = \{d_{1,j}, d_{2,j}, \dots, d_{m,j}\}$ the decision vector of the assign-

Table 2: Example of a decision table built at the $Period_j$

	g_1	...	g_k	...	g_p	\mathbf{D}_j
$A_{1,j}$	$f(A_{1,j}, g_1)$...	$f(A_{1,j}, g_k)$...	$f(A_{1,j}, g_p)$	$d_{1,j}$
$A_{2,j}$	$f(A_{2,j}, g_1)$...	$f(A_{2,j}, g_k)$...	$f(A_{2,j}, g_p)$	$d_{2,j}$
...
$A_{m,j}$	$f(A_{m,j}, g_1)$...	$f(A_{m,j}, g_k)$...	$f(A_{m,j}, g_p)$	$d_{m,j}$

ment of each “Action of reference” in one of the decision classes.

4.2. Phase 2: Inference of a preference model based on the incremental update of the DRSA approximations

The preference model is a set of decision rules permitting to classify each action in one of the defined decision classes. It is inferred by applying an induction algorithm that takes as input the lower and the upper approximations of the upward and the downward unions of the decision classes. In this work, we used the rules induction algorithm *DOMLEM* proposed by the approach DRSA [20]. This algorithm generates a minimal set of dominance-based rules covering all the examples in the decision table. This phase runs at the end of each period P_i during the prediction process and takes as input the decision table build in “Phase 1”. This phase is made of two steps: The first consists in updating the upper and the lower approximations of the DRSA by applying our algorithm *DRSA-Incremental* on the decision table (cf. Section 5). The second step infers a preference model by applying the algorithm *DOMLEM* on the updated approximations.

As quoted above, this approach deals with the case of dynamic information system under the variation of the set of “Reference actions”. Thus, the decision maker, receiving information in portions, should learn how to deal with this situation before all information about it is available. If we consider the traditional application of the DRSA approach, during each period, the new set of the decision rules is inferred by recalculating the approximations from scratch based on the new decision table. This

method is correct but consumes much computational time. According to [29], this situation of sequential flow of information is a primordial reason to incrementally update the DRSA approximations. Hence, instead of recalculating from scratch, we have to update the approximations using our *DRSA-Incremental* algorithm.

The *DRSA-Incremental* algorithm updates the DRSA approximations that will be provided as an input to the algorithm *DOMLEM* in order to generate a preference model resulting in a set of decision rules. The preference model of the period P_i will be given as input to phase 3 in order to classify the “Potential actions” at the beginning of the next period P_{i+1} .

4.3. Phase 3: Classification of the “Potential actions”

The third phase uses the previously inferred decision rules to assign each of the “Potential actions” to one of the N predefined decision classes.

A “potential action” is defined as a generic term used to describe an action or a referent of a decision. An action is considered potential, if it can be implemented or simply if it is considered “fertile” in a decision making context. The notion of potential action clarifies the nature of what constitutes the decision problem and formalizes the decision purpose.

This approach runs periodically: the first and the second phases run at the end of each period P_i such that $i \in \{1..T - 1\}$ while the third phase runs at the beginning of each period P_i such that $i \in \{2..T\}$. The three phases are chained in a way that each phase inputs the output of the previous one (cf. Figure 1).

4.4. Reduction of the assignment intervals

The DRSA decision rules assign each action to approximate sets of decision classes called downward and upward unions (cf. Section 2). However, x cannot belong to more than one decision class. Thus, the objective of this step is to reduce the assignment interval to a single decision class either by the method proposed in [9] based on the “min”, “max” and “mediane”, or the method proposed in [5] based on a score classification.

Let $I(x) = \{m(x), M(x)\}$ the assignment interval of an action $x \in A$ where $m(x) = \min\{N1(x) \cup N2(x)\}$ and $M(x) = \max\{N1(x) \cup N2(x)\}$ such that $N1(x) = \{Cl_i; x \in Cl_i^{\geq}\}$ and $N2(x) = \{Cl_i; x \in Cl_i^{\leq}\}$. This notation distinguishes two cases:

- Case 1: if $m(x) = M(x)$ then $x \in m(x)$ or $x \in M(x)$.
- Case 2: if $m(x) < M(x)$ then x can be assigned to more than one class. Thus, to reduce the interval $I(x)$ to a single class, we apply one of these three rules:
 - Case 2.1: Use the pessimistic approach by applying the “*min*” operator that assigns the action to the minimal decision class so as $f(x, d) = m(x)$.
 - Case 2.2: Use the optimistic approach by applying the “*max*” operator that assigns the action to the maximal decision class so as $f(x, d) = M(x)$.
 - Case 2.3: Use the median approach. In the case of odd number of decision classes, we find a single class that represents the median . In the case of even number of decision classes, the median is represented in two decision classes, denoted $\{m', M'\}$. In this case, we apply either the “floor” of the median which represents the “min” (*i.e.* $f(x, d) = m'$), or the “ceil” which represents the max (*i.e.* $f(x, d) = M'$).
 - Case 2.4: Use the score classification method. For each decision class Cl_n in which an action is likely to belong, a classification score is calculated. The score represents the quotient between (1) the cardinalities of the set of actions verifying the condition part of the rule that classified the concerned action and the set of actions belonging to class Cl_n , and (2) the cardinality of the set of actions simultaneously verifying this condition part and belonging to class Cl_n . The action is finally assigned to the decision class with the highest score. This method is detailed in [5].

The choice of the reduction method strongly depends on the context characteristics and the classification purpose. The decision makers must study the advantages of each method compared to their expectations for deciding which one is most appropriate.

Example 5: we consider the downward unions of decision classes $Cl_3^{\leq} = \{Cl_1, Cl_2, Cl_3\}$ and x an action such that $x \in Cl_3^{\leq}$. We plan to reduce the assignment interval $I(x) = \{m(x), M(x)\} = \{Cl_1, Cl_3\}$ to affect the action x to a single decision class. If we apply the pessimistic approach we obtain $f(x, d) = m(x) = Cl_1$. If we apply the

optimistic approach we obtain $f(x, d) = M(x) = Cl_3$. If we apply the median approach we obtain $f(x, d) = middle(x) = Cl_2$ since the number of decision classes is even.

Compared to the conventional machine learning techniques that are already used for a prediction purpose, the approach *MAI2P* based-DRSA strongly depends on the experts' preferences to strictly identify both sets of "reference actions" and criteria in order to rigorously characterize the potential actions. Moreover, this approach permits to avoid the unbalanced data problem that occurs because of the presence of a majority class in a prediction context [6].

5. *DRSA-Incremental: algorithm for the incremental update of the DRSA approximations following the addition of a set of actions*

In this section, we present the "P-generalized Decision" defined in [25] and on which the incremental update algorithm is based. Then, we present the approach proposed for the incremental update of the DRSA approximations.

5.1. The "P-generalized Decision"

The "P-generalized Decision" defines an interval of decision classes to which an action may belong. It is presented as follows:

\forall action $x \in A$, $\delta_p(x) = \langle lp(x), up(x) \rangle$ is called the P-generalized decision of the action x , where:

- $lp(x) = \min\{n \in N : D_p^+(x) \cap Cl_n \neq \emptyset\}$
- $up(x) = \max\{n \in N : D_p^-(x) \cap Cl_n \neq \emptyset\}$

Based on this definition, the DRSA approximations can be presented as follows:

- $\underline{P}(Cl_n^{\geq}) = \{x \in A : lp(x) \geq n\}$
- $\overline{P}(Cl_n^{\geq}) = \{x \in A : up(x) \geq n\}$
- $\underline{P}(Cl_n^{\leq}) = \{x \in A : up(x) \leq n\}$
- $\overline{P}(Cl_n^{\leq}) = \{x \in A : lp(x) \leq n\}$

These definitions will be used in parallel with the definitions proposed by Greco et al. [20], in order to incrementally update the DRSA approximations.

5.2. Incremental update of the DRSA approximations following the addition of a set of actions

The entry of an action x^+ in an information system causes the variation of the “P-dominating” as well as the “P-dominated” sets and so it is the case for the lower and the upper approximations of the DRSA approach. Thus, Li et al. [26] proposed an algorithm for the incremental update of these approximations following the entry of only a new action. However, in our context, it is more interesting to consider the simultaneous entry of a set of actions in the information system. To this end, we generalize the algorithm proposed in [26].

We start by defining some notations we will use throughout the approach process. We note that an action x^+ belongs to one and only one decision class $Cl_{n'}$. Let:

- Δ^+ : be the set of k actions to enter in the information system;

$$\Delta^+ = \{x^+\} \text{ and } |\Delta^+| = k$$

- $\Delta_{n<}^+$: be the sub-set of actions in Δ^+ belonging to a decision class $Cl_{n'}$ that is inferior than Cl_n :

$$\Delta_{n<}^+ = x^+ \in (\Delta^+ \cap Cl_{n'}) \text{ such that } n' < n$$

- $\Delta_{n>}^+$: be the sub-set of actions in Δ^+ belonging to a decision class $Cl_{n'}$ that is superior than Cl_n :

$$\Delta_{n>}^+ = x^+ \in (\Delta^+ \cap Cl_{n'}) \text{ such that } n' > n$$

- $\Delta_{n=}^+$: be the sub-set of actions in Δ^+ belonging to a decision class $Cl_{n'}$ that is equal to Cl_n :

$$\Delta_{n=}^+ = x^+ \in (\Delta^+ \cap Cl_{n'}) \text{ such that } n' = n$$

We note that

$$\begin{cases} A_{i+1} = A_i + \Delta^+ \\ \Delta_{n<}^+ \cup \Delta_{n>}^+ \cup \Delta_{n=}^+ = \Delta^+ \\ \Delta_{n<}^+ \cap \Delta_{n>}^+ \cap \Delta_{n=}^+ = \emptyset \end{cases}$$

To consider the evolution of the information system over many periods, we adopt the notation presented in Table 5.2.

Table 3: Symbolic representations of notations in DRSA in the $Period_i$ and the $Period_{i+1}$

	$Period_i$	$Period_{i+1}$
Information table	$\langle A_i, F_i, V_i, f_i \rangle$	$\langle A_{i+1}, F_{i+1}, V_{i+1}, f_{i+1} \rangle$
P-dominating set of the action x	$D_i^+ p(x)$	$D_{i+1}^+ p(x)$
P-dominated set of the action x	$D_i^- p(x)$	$D_{i+1}^- p(x)$
The decision class number n	$Cl_{i,n}$	$Cl_{i+1,n}$
Upward union of Cl_n	$Cl_{i,n}^{\geq}$	$Cl_{i+1,n}^{\geq}$
Downward union of Cl_n	$Cl_{i,n}^{\leq}$	$Cl_{i+1,n}^{\leq}$
Lower approximation of Cl_n^{\geq}	$\underline{P}(Cl_{i,n}^{\geq})$	$\underline{P}(Cl_{i+1,n}^{\geq})$
Upper approximation of Cl_n^{\geq}	$\overline{P}(Cl_{i,n}^{\geq})$	$\overline{P}(Cl_{i+1,n}^{\geq})$
Lower approximation of Cl_n^{\leq}	$\underline{P}(Cl_{i,n}^{\leq})$	$\underline{P}(Cl_{i+1,n}^{\leq})$
Upper approximation of Cl_n^{\leq}	$\overline{P}(Cl_{i,n}^{\leq})$	$\overline{P}(Cl_{i+1,n}^{\leq})$

Example 6: We consider that in the week $i+1$, the actions set $\Delta^+ = \{A_7, A_8, A_9, A_{10}, A_{11}\}$ is inserted to the decision table D_i as described in Table 1. Thus, the sub-sets that we can construct from Δ^+ are as follows: $\Delta_{1<}^+ = \{\emptyset\}$, $\Delta_{2<}^+ = \{A_8\}$, $\Delta_{3<}^+ = \{A_8, A_{10}, A_{11}\}$, $\Delta_{1>}^+ = \{A_7, A_8, A_9, A_{10}, A_{11}\}$, $\Delta_{2>}^+ = \{A_7, A_9\}$, $\Delta_{3>}^+ = \{\emptyset\}$, $\Delta_{1=}^+ = \{A_8\}$, $\Delta_{2=}^+ = \{A_{10}, A_{11}\}$, $\Delta_{3=}^+ = \{A_7, A_9\}$.

Example 7: First, we have to calculate the P-dominated and the P-dominating sets of each inserted action as described in section 2. We obtain then: $D_{i+1}^+ p(A_7) = \{A_1, A_7\}$, $D_{i+1}^+ p(A_8) = \{A_1, A_2, A_3, A_5, A_7, A_8, A_9, A_{10}, A_{11}\}$, $D_{i+1}^+ p(A_9) = \{A_1, A_5, A_7, A_9\}$, $D_{i+1}^+ p(A_{10}) = \{A_1, A_2, A_7, A_{10}, A_{11}\}$ and $D_{i+1}^+ p(A_{11}) = \{A_1, A_{11}\}$. Similarly, we calculate $D_{i+1}^- p(A_7) = \{A_3, A_6, A_7, A_8, A_9, A_{10}\}$, $D_{i+1}^- p(A_8) = \{A_8\}$, $D_{i+1}^- p(A_9) = \{A_3, A_8, A_9\}$, $D_{i+1}^- p(A_{10}) = \{A_6, A_8, A_{10}\}$ and $D_{i+1}^- p(A_{11}) =$

$\{A_2, A_4, A_6, A_8, A_{10}, A_{11}\}$. As well, we need to calculate the P-generalized decision of each inserted action. The obtained results are as follows: $l_{i+1}p(A_7) = \min\{3, 3\} = 3$, $l_{i+1}p(A_8) = \min\{3, 3, 1, 2, 3, 1, 3, 2, 2\} = 1$, $l_{i+1}p(A_9) = \min\{3, 2, 3, 3\} = 2$, $l_{i+1}p(A_{10}) = \min\{3, 3, 3, 2, 2\} = 2$, $l_{i+1}p(A_{11}) = \min\{3, 2\} = 2$, $u_{i+1}p(A_7) = \max\{1, 1, 3, 1, 3, 2\} = 3$, $u_{i+1}p(A_8) = \max\{1\} = 1$, $u_{i+1}p(A_9) = \max\{1, 1, 3\} = 3$, $u_{i+1}p(A_{10}) = \max\{1, 1, 2\} = 2$, $u_{i+1}p(A_{11}) = \max\{3, 2, 1, 1, 2, 2\} = 3$. The information about the inserted actions will be used to update those of the initial actions.

The update of the DRSA approximations is based on three steps (cf. Algorithm 1): first, the update of the upward and the downward sets of decision classes. Second, the update of the P-dominated and the P-dominating sets. Third, the update of the approximation sets. The proofs of different propositions defined below are detailed in Appendix C.

5.2.1. Incremental update of the upward and the downward decision classes

Each action $x_i^+ \in \Delta^+$ belongs to one and only one decision class. We suppose that $x_i^+ \in \Delta_{n<}^+$ or $x_i^+ \in \Delta_{n>}^+$ or $x_i^+ \in \Delta_{n=}^+$. The upward and the downward of the unions of the decision classes are updated according to the lemma 1:

Lemma 1. Let $\Delta_{n<}^+ \cup \Delta_{n>}^+ \cup \Delta_{n=}^+ = \Delta^+$ the set of entering actions.

$$Cl_{i+1,n}^{\geq} = \begin{cases} Cl_{i,n}^{\geq}, & \text{if } \Delta_{n>}^+ = \emptyset \text{ and } \Delta_{n=}^+ = \emptyset \\ Cl_{i,n}^{\geq} \cup \Delta_{n>}^+ \cup \Delta_{n=}^+ & \text{otherwise} \end{cases} \quad (1)$$

$$Cl_{i+1,n}^{\leq} = \begin{cases} Cl_{i,n}^{\leq}, & \text{if } \Delta_{n<}^+ = \emptyset \text{ and } \Delta_{n=}^+ = \emptyset \\ Cl_{i,n}^{\leq} \cup \Delta_{n<}^+ \cup \Delta_{n=}^+ & \text{otherwise} \end{cases} \quad (2)$$

Example 8: According to lemma 1, the updated downward and upward of the unions of decision classes are as follows: $Cl_{i+1,1}^{\leq} = \{A_3, A_6, A_8\}$, $Cl_{i+1,2}^{\leq} = \{A_3, A_4, A_5, A_6, A_8, A_{10}, A_{11}\}$, $Cl_{i+1,3}^{\leq} = \{A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8, A_9, A_{10}, A_{11}\} = Cl_{i+1,1}^{\geq}$, $Cl_{i+1,2}^{\geq} = \{A_1, A_2, A_4, A_5, A_7, A_9, A_{10}, K_{11}\}$, $Cl_{i+1,3}^{\geq} = \{A_1, A_2, A_7, A_9\}$.

5.2.2. Incremental update of the P-dominating and the P-Dominated sets

When a set of actions enters into the information system, the dominance relation D_p according to the set of criteria may be also updated. The relation $D_{i+1}p$ can be presented as follows: $D_{i+1}p = D_i p \cup \mathcal{A} \cup \mathcal{B}$, such that $\mathcal{A} = \{(x_i^+, y) \in \Delta^+ \times A_{i+1} : f_{i+1}(x_i^+, g) \geq f_{i+1}(y, g), \forall g \in P\}$ and $\mathcal{B} = \{(y, x_i^+) \in A_{i+1} \times \Delta^+ : f_{i+1}(y, g) \geq f_{i+1}(x_i^+, g), \forall g \in P\}$.

It is evident that $D_i p \subseteq D_{i+1} p$. So, $D_{i+1}^+ p(x_i^+)$ and $D_{i+1}^- p(x_i^+)$ can be calculated using the sets \mathcal{A} and \mathcal{B} respectively. For each $x \in A_i$, it may exists $(x_i^+, x) \in \mathcal{A}$ or $(x, x_i^+) \in \mathcal{B}$. Thus $D_i^+ p(x) \subseteq D_{i+1}^+ p(x)$ and $D_i^- p(x) \subseteq D_{i+1}^- p(x)$. Thus the sets P-dominating and P-dominated may be updated incrementally by the lemma 2. First $\forall x_i^+ \in \Delta^+$ we calculate the sets $D_{i+1}^+ p(x_i^+)$ and $D_{i+1}^- p(x_i^+)$.

Lemma 2. Let $x_i^+ \in \Delta^+$ an action in Δ^+ such that $i \in \{1..k\}$.

$$D_{i+1}^+ p(x) = \begin{cases} D_i^+ p \text{ if } x \notin \bigcup_{i=1}^k D_{i+1}^- p(x_i^+) \\ D_i^+ p(x) \cup \bigcup_{i=1}^{k'} \{x_i^+\}, \forall x_i^+ \in \Delta^+ \text{ such that } x \in D_{i+1}^- p(x_i^+) \end{cases} \quad (3)$$

k' is the number of x_i^+ in Δ^+ such that $x \in D_{i+1}^- p(x_i^+)$

$$D_{i+1}^- p(x) = \begin{cases} D_i^- p \text{ if } x \notin \bigcup_{i=1}^k D_{i+1}^+ p(x_i^+) \\ D_i^- p(x) \cup \bigcup_{i=1}^{k'} \{x_i^+\}, \forall x_i^+ \in \Delta^+ \text{ such that } x \in D_{i+1}^+ p(x_i^+) \end{cases} \quad (4)$$

k' is the number of x_i^+ in Δ^+ such that $x \in D_{i+1}^+ p(x_i^+)$

Example 9: According to lemma 2, the P-dominated and the P-dominating sets are calculated as follows: $D_{i+1}^- p(A_1) = \{A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8, A_9, A_{10}, A_{11}\}$, $D_{i+1}^- p(A_2) = \{A_2, A_6, A_8, A_{10}\}$, $D_{i+1}^- p(A_3) = \{A_3, A_8\}$, $D_{i+1}^- p(A_4) = \{A_4, A_6\}$, $D_{i+1}^- p(A_5) = \{A_3, A_5, A_8, A_9\}$, $D_{i+1}^- p(A_6) = \{A_6\}$, $D_{i+1}^+ p(A_1) = \{A_1\}$, $D_{i+1}^+ p(A_2) = \{A_1, A_2, A_{11}\}$, $D_{i+1}^+ p(A_3) = \{A_1, A_3, A_5, A_7, A_9\}$, $D_{i+1}^+ p(A_4) = \{A_1, A_4, A_{11}\}$, $D_{i+1}^+ p(A_5) = \{A_1, A_5\}$, $D_{i+1}^+ p(A_6) = \{A_1, A_2, A_4, A_6, A_7, A_{10}, A_{11}\}$,

5.2.3. Incremental update of the upper and the lower approximations of DRSA

In this section we present four propositions for the incremental update of the approximations DRSA.

Proposition 1. Let $x_i^+ \in Cl_{i+1,n'}$; $n, n' \in N$. We have:

$$\underline{P}(Cl_{i+1,n}^{\geq}) = \begin{cases} \underline{P}(Cl_{i,n}^{\geq}), \text{ if } \forall x_i^+ \in \Delta^+, l_{i+1}p(x_i^+) < n \leq n' \\ \underline{P}(Cl_{i,n}^{\geq}) \cup \cup_{i=1}^{k'} \{x_i^+\}, \forall x_i^+ \in (\Delta_{n>}^+ \cup \Delta_{n=}^+), l_{i+1}p(x_i^+) \geq n \\ \underline{P}(Cl_{i,n}^{\geq}) - \Delta\mathbb{X}_n, \forall x_i^+ \in \Delta_{n<}^+ \end{cases} \quad (5)$$

Where

1. k' is the number of $x_i^+ \in (\Delta_{n>}^+ \cup \Delta_{n=}^+)$ such that $l_{i+1}p(x_i^+) \geq n$
2. $\Delta\mathbb{X}_n = \{x \in \underline{P}(Cl_{i,n}^{\geq}) \cap \cup_{i=1}^{k'} D_{i+1}^- p(x_i^+); l_{i+1}p(x) < n, \forall i \in \{1..k\}\}$

Example 10: According to the proposition 1 and using the results of Example 6, the lower approximation of the $Cl_{i+1,n}^{\geq}$ is as follows: $\underline{P}(Cl_{i+1,1}^{\geq}) = \{A_1, A_2, A_3, A_4, A_5, A_6\} \cup \{A_8\}$, $\underline{P}(Cl_{i+1,2}^{\geq}) = \{A_1, A_2, A_4, A_5\} \cup \{A_{10}, A_{11}\}$, $\underline{P}(Cl_{i+1,3}^{\geq}) = \{A_1, A_2\} \cup \{A_7\}$.

Proposition 2. Let $x_i^+ \in Cl_{i+1,n'}$; $n, n' \in N$. We have:

$$\underline{P}(Cl_{i+1,n}^{\leq}) = \begin{cases} \underline{P}(Cl_{i,n}^{\leq}), \text{ if } \forall x_i^+ \in \Delta^+, u_{i+1}p(x_i^+) > n \geq n' \\ \underline{P}(Cl_{i,n}^{\leq}) \cup \cup_{i=1}^{k'} \{x_i^+\}, \forall x_i^+ \in (\Delta_{n<}^+ \cup \Delta_{n=}^+), u_{i+1}p(x_i^+) \leq n \\ \underline{P}(Cl_{i,n}^{\leq}) - \Delta\mathbb{M}_n, \forall x_i^+ \in \Delta_{n>}^+ \end{cases} \quad (6)$$

Where

1. k' is the number of $x_i^+ \in (\Delta_{n<}^+ \cup \Delta_{n=}^+)$ such that $u_{i+1}p(x_i^+) \geq n$
2. $\Delta\mathbb{M}_n = \{x \in \underline{P}(Cl_{i,n}^{\leq}) \cap \cup_{i=1}^{k'} D_{i+1}^+ p(x_i^+) : u_{i+1}p(x) > n, \forall i \in \{1..k\}\}$

Example 11: According to the proposition 2 and using the results of Example 6, the lower approximation of the $Cl_{i+1,n}^{\leq}$ is as follows: $\underline{P}(Cl_{i+1,1}^{\leq}) = \{A_3, A_6\} \cup \{A_8\}$, $\underline{P}(Cl_{i+1,2}^{\leq}) = \{A_3, A_4, A_5, A_6\} \cup \{A_8, A_{10}\} - \{A_5\}$, $\underline{P}(Cl_{i+1,3}^{\leq}) = \{A_1, A_2, A_3, A_4, A_5, A_6\} \cup \{A_7, A_8, A_9, A_{10}, A_{11}\}$. Here, we note that the action A_5 is no longer classified with certainty in $Cl_{i+1,1}^{\leq}$ because it dominates the action A_9 while $A_9 \in Cl_3$ and $A_5 \in Cl_2$. However, A_5 still belonging to $\bar{P}(Cl_{i+1,2}^{\leq})$.

Algorithm 1: Incremental algorithm for the update of the approximations of DRSA when adding a set of actions in the information system

Data:
 $\forall x \in K, D_i^+ p(x)$ and $D_i^- p(x)$ of the week_{*i*}
 $\forall n \in T, \bar{P}(Cl_{i,n}^{\geq}), \bar{P}(Cl_{i,n}^{\leq}), \underline{P}(Cl_{i,n}^{\geq})$ and $\underline{P}(Cl_{i,n}^{\leq})$ of the week_{*i*}
A set Δ^+ containing new actions
Result:
 $\bar{P}(Cl_{i+1,n}^{\geq}), \bar{P}(Cl_{i+1,n}^{\leq}), \underline{P}(Cl_{i+1,n}^{\geq})$ and $\underline{P}(Cl_{i+1,n}^{\leq})$ of the week_{*i+1*}
Begin
for *i* **from** 1 **to** $|\Delta^+|$ **do**
 Calculate $D_{i+1}^+ p(x_i^+)$ and $D_{i+1}^- p(x_i^+)$;
end
for *n* **from** 1 **to** *N* **do**
 Calculate $\Delta_{n<}^+, \Delta_{n>}^+$ and $\Delta_{n=}^+$;
end
for *i* **from** 1 **to** *K* **do**
 for *j* **from** 1 **to** $|\Delta^+|$ **do**
 Update $D_{i+1}^+ p(x_i)$ and $D_{i+1}^- p(x_i)$
 end
end
for *i* **from** 1 **to** $|\Delta^+|$ **do**
 Calculate $l_{i+1} p(x_i^+)$ and $u_{i+1} p(x_i^+)$
end
for *n* **from** 1 **to** *N* **do**
 for *i* **from** 1 **to** $|\Delta_{n>}^+|$ **do**
 Calculate $\Delta_{n>}^+$;
 Update $\underline{P}(Cl_{i+1,n}^{\geq}), \underline{P}(Cl_{i+1,n}^{\leq}), \bar{P}(Cl_{i+1,n}^{\geq})$ and $\bar{P}(Cl_{i+1,n}^{\leq})$;
 end
 for *i* **from** 1 **to** $|\Delta_{n=}^+|$ **do**
 Update $\underline{P}(Cl_{i+1,n}^{\geq}), \underline{P}(Cl_{i+1,n}^{\leq}), \bar{P}(Cl_{i+1,n}^{\geq})$ and $\bar{P}(Cl_{i+1,n}^{\leq})$;
 end
 for *i* **from** 1 **to** $|\Delta_{n<}^+|$ **do**
 Calculate $\Delta_{n<}^+$;
 Update $\underline{P}(Cl_{i+1,n}^{\geq}), \underline{P}(Cl_{i+1,n}^{\leq}), \bar{P}(Cl_{i+1,n}^{\geq})$ and $\bar{P}(Cl_{i+1,n}^{\leq})$;
 end
end
Return $\bar{P}(Cl_{i+1,n}^{\geq}), \bar{P}(Cl_{i+1,n}^{\leq}), \underline{P}(Cl_{i+1,n}^{\geq})$ and $\underline{P}(Cl_{i+1,n}^{\leq})$

Proposition 3. Let $x_i^+ \in Cl_{i+1,n'}$; *n, n' ∈ N*. We have:

$$\bar{P}(Cl_{i+1,n}^{\leq}) = \begin{cases} \bar{P}(Cl_{i,n}^{\leq}), \text{ if } \forall x_i^+ \in \Delta^+, l_{i+1} p(x_i^+) > n \\ \bar{P}(Cl_{i,n}^{\leq}) \cup \cup_{i=1}^{k'} \{x_i^+\}, \forall x_i^+ \in \Delta_{n>}^+ : l_{i+1} p(x_i^+) \leq n < n' \\ \bar{P}(Cl_{i,n}^{\leq}) \cup \cup_{i=1}^{k''} D_{i+1}^- p(x_i^+), \forall x_i^+ \in (\Delta_{n<}^+ \cup \Delta_{n=}^+), \end{cases} \quad (7)$$

Where

1. k' is the number of $x_i^+ \in \Delta_{n>}^+$ such that $l_{i+1} p(x_i^+) \leq n$
2. k'' is the number of x_i^+ such that $n' \leq n$; i.e. $k'' = |\Delta_{n<}^+ \cup \Delta_{n=}^+|$

Example 12: According to the proposition 3 and using the results of Example 6, the upper approximation of the $Cl_{i+1,n}^{\leq}$ is as follows: $\bar{P}(Cl_{i+1,1}^{\leq}) = \{A_3, A_6\} \cup \{A_8\}$, $\bar{P}(Cl_{i+1,2}^{\leq}) = \{A_3, A_4, A_5, A_6\} \cup \{A_8\} \cup \{A_{10}\} \cup \{A_2, A_{11}\}$, $\bar{P}(Cl_{i+1,3}^{\leq}) = \{A_1, A_2, A_3, A_4, A_5, A_6\} \cup \{A_7, A_8, A_9, A_{10}, A_{11}\}$.

Proposition 4. Let $x_i^+ \in Cl_{i+1,n'}$; $n, n' \in N$. We have:

$$\bar{P}(Cl_{i+1,n}^{\geq}) = \begin{cases} \bar{P}(Cl_{i,n}^{\geq}), \text{ if } \forall x_i^+ \in \Delta^+, u_{i+1}p(x_i^+) < n \\ \bar{P}(Cl_{i,n}^{\geq}) \cup \bigcup_{i=1}^{k'} \{x_i^+\}, \forall x_i^+ \in \Delta_{n<}^+ : u_{i+1}p(x_i^+) \geq n > n' \\ \bar{P}(Cl_{i,n}^{\geq}) \cup \bigcup_{i=1}^{k''} D_{i+1}^+ p(x_i^+), \forall x_i^+ \in (\Delta_{n>}^+ \cup \Delta_{n=}^+), \end{cases} \quad (8)$$

Where

1. k' is the number of $x_i^+ \in (\Delta_{n<}^+)$ such that $u_{i+1}p(x_i^+) \geq n > n'$
2. k'' is the number of x_i^+ such that $n' \geq n$; i.e. $k'' = |\Delta_{n>}^+ \cup \Delta_{n=}^+|$

Example 13: According to the proposition 4 and using the results of Example 6, the upper approximation of the $Cl_{i+1,n}^{\geq}$ is as follows: $\bar{P}(Cl_{i+1,1}^{\geq}) = \{A_1, A_2, A_3, A_4, A_5, A_6\} \cup \{A_7, A_8, A_9, A_{10}, A_{11}\}$, $\bar{P}(Cl_{i+1,2}^{\geq}) = \{A_1, A_2, A_4, A_5\} \cup \{A_7, A_9, A_{10}, A_{11}\}$, $\bar{P}(Cl_{i+1,3}^{\geq}) = \{A_1, A_2\} \cup \{A_7, A_9, A_{11}\}$.

The algorithm of the incremental update of the DRSA approximations following the entry of a set of actions in the information system is detailed in Appendix A.

6. Case study

This section presents the application and the validation of the proposed approach using real-world data.

6.1. Decision problem

The problem considered here concerns the dropout rate in the context of MOOC. A MOOC is a model of educational delivery that is massive, with theoretically no limit to enrolment; open, allowing anyone to participate; online, with learning activities taking place over the web; and a course structured around a defined set of learning goals [17]. The MOOC is characterized by a dynamic information system since it is open, so the learners can enter and leave it at any time during its broadcasting. Since 2008, the number of MOOCs has rapidly grown around the world [37]. However, despite their proliferation, the MOOCs still suffer from a high dropout rate that usually

reaches 90% [55]. According to [36], the excessive dropout rate is one of the major recurring issues in the MOOCs.

To this end, several prediction models were proposed to early identify the learners who intend to leave the course using the conventional machine learning techniques such as the Hidden Markov Model [2, 11], the neural network [10], the decision tree [52], the Support Vector Machine [31] and the General Bayesian Network [52].

However, these models suffer from three major limitations: First, the intuitive selection of the indicators used to characterize the MOOC participants. Second, the random identification of the learning set. Third, the strong presence of the unbalanced data problem because of the majority class of learners who intend to leave the course.

The objective of this case study is to apply the proposed incremental approach based DRSA in the context of MOOC and to demonstrate how it could avoid the above-mentioned limitations. This approach targets the weekly and early classification of the MOOC learners in these three decision classes:

- Cl_1 . The decision class of the “At-risk learners” corresponding to learners who are likely to dropout the course in the next week of the MOOC.
- Cl_2 . The decision class of the “Struggling learners” corresponding to learners who have some difficulties but are still active on the MOOC environment and don’t have the intention to leave it at least in the next week of the MOOC.
- Cl_3 . The decision class of the “Leader learners” corresponding to learners who are able to lead a team of learners by providing them with an accurate and an immediate response.

6.2. Data set

The data set used in this application concerns a French-speaking MOOC about the “Design Thinking” proposed by a Business School in France and lasted “ $T=5$ ” weeks. It started with 2565 learners where only 1535 are considered in these experiments. The neglected learners are those who didn’t complete the registration form. Data was saved in CSV (Comma-Separated Values) files. Each week, the recovered file was carefully processed and cleaned to save only the data we need. However, for

the time experiments we defined in addition an artificial data set containing data about 10000 actions that are generated randomly by a computer program. This data set permits to demonstrate how our approach permits to reduce the computational complexity especially in a bigdata context.

6.3. Application

In what follows, we explain how the approach *MAI2P* was applied to weekly predict the decision class to which belongs each learner participating in the MOOC “Design Thinking”. In this work, the actions are the learners and the decision makers are the members of the pedagogical team.

Let $W = \{W_1, \dots, W_i, \dots, W_T\}$ be the set of weeks making up a MOOC such that $T = 5$ is the number of weeks a MOOC holds and W_i is the i^{th} week of the MOOC.

6.3.1. Phase 1: Construction of the decision table

As presented, this phase comprises three steps.

- *Step 1.1: The identification of sets of “Learners of reference”.* At the end of each week W_i , the pedagogical team defined a new set A'_i of 30 “Learners of reference” that was directly appended to the sets of “Learners of reference”, A_{i-1} , of the previous weeks. Thus, the set of “Learners of reference” of the week W_i becomes $A_i = A'_i + A_{i-1}, \forall i \in \{1..4\}$. The pedagogical team saw that 30 is a sufficient number to constitute the learning set in this context especially, by applying the cumulative approach that gives 30 actions for the first week, 60 for the second week, 90 for the third and 120 for the fourth week.
- *Step 1.2: The construction of a coherent family of criteria.* First, we identified a list of indicators that would permit to characterize a MOOC learner. These indicators would give sign about the learner’s skills, profile and motivation. Among the indicators mentioned in literature, we quote the study level of a learner [32], the MOOC language mastery and the motivation to participate in the MOOC [3], the cultural background [48], the level of technical skills and the lack of time [19]. In the e-learning field, Wolff et al. [51] distinguished three types

of activities that enable to predict the learner dropout which are the access to a course material; the publishing of a message on the forum and the access to the evaluation space. Second, to validate a final family of criteria we conducted some meetings with the pedagogical team that retained 11 criteria and applied a preference order on each of them. For example, for the criterion “Study level”, four increasing ordered scales are defined: 1: Scholar student; 2: High school student; 3: PhD Student; 4: Doctor. This step is detailed in [7]. [The criteria family is described in Table 4 and remains stable over weeks.](#)

- *Step 1.3: The construction of the decision table.* At the end of each week W_i of the MOOC, a matrix, whose rows form the set of the 30 “Learners of reference” and whose columns represent the 11 evaluation criteria, is built. This matrix contains the evaluation function $f_i(L_{j,i}, g_k)$ of each learner $L_{j,i} \in A'_i$ on each criterion $g_k \in F_1$ such that $i \in \{1..4\}$, $j \in \{1..30\}$ and $k \in \{1..11\}$. Then, based on its expertise and the information table, the pedagogical team assign each of the “Learners of reference” in one of the three decision classes. [Four extracts from the decision tables built each week of the MOOC are presented in Table 5.](#)

6.3.2. Phase 2: The inference of a preference model based on the incremental update of the DRSA approximations

At the end of each week W_i during the MOOC broadcast such that $i \in \{1..4\}$, we applied the algorithm *DRSA-Incremental* on the decision table built at the same week W_i in order to update the approximations of the upward and the downward of the unions of all the decision classes. Then, the updated approximations were provided as an input to the algorithm *DOMLEM* [20] in order to infer the decision rules. The output of this phase is a preference model representing the input to phase 3. [Four extracts from the preference models inferred each week of the MOOC are presented in Table 6.](#)

6.3.3. Phase 3: The classification of the “Potential learners”

At the beginning of each week W_i of the MOOC; such that $i \in \{2..5\}$ we used the inferred decision rules in order to assign the “Potential learners” in one of the three

Table 4: List of the constructed criteria family

Code	Label	Description	Scale	Type	Pref
g_1	The study level	Indicates the actual study level of the learner or the last diploma he obtained	1: Scholar student; 2: High school student; 3: PhD Student; 4: Doctor	Ordinal	Gain
g_2	Level of technical skills	Indicates the extent to which the learner masters the use of computer tools	1: Basic; 2: Average; 3: Expert	Ordinal	Gain
g_3	Level of proficiency in MOOC language	Indicates the extent to which the learner masters the MOOC language	1: Basic; 2: Average; 3: Good	Ordinal	Gain
g_4	Motivation for MOOC registration	Indicates the motivation behind the participation of the learner in the MOOC	1: Just to discover the MOOCs; 2: To exchange ideas with the other learners or to have a certificate ; 3: To exchange ideas with the other learners and to have a certificate	Ordinal	Gain
g_5	Previous experience with MOOCs	Indicates whether the learner has a previous experience on learning via MOOCs or not	0: No experience at all; 1: At least one experience	Ordinal	Gain
g_6	Mastery level of the subject of a MOOC	Indicates to which extent the learner masters both the topic and the theme of the MOOC	0 : No knowledge at all; 1: Average knowledge; 2: Deepened knowledge	Ordinal	Gain
g_7	Probability to finish the MOOC	Indicates the probability for a learner to follow the MOOC until the end	1: Very weak; 2: Weak; 3: Average; 4: Strong; 5: Very strong	Ordinal	Gain
g_8	Weekly availability	Indicates the estimative weekly availability of the learner to follow the MOOC	1: Less than one hour ; 2: From one to two hours; 3: From two to three hours; 4: Four hours or more	Ordinal	Gain
g_9	Weekly number of forum posts	Indicates the number of posts published on the forum per week	1 : $n = 0$; 2 : $n \in \{1, 2\}$; 3 : $n \in \{3, 4\}$; 4 : $n \geq 5$	Ordinal	Gain
g_{10}	Weekly number of viewed or downloaded resources	Indicates the weekly number of the viewed and / or downloaded resources and material courses	1 : $n < 10$; 2 : $10 \leq n < 20$; 3 : $20 \leq n < 30$; 4 : $30 \leq n < 40$; 5 : $n \geq 40$	Ordinal	Gain
g_{11}	Weekly score	Indicates the weekly score the learner got on the set of activities he made	1 : $0 \leq Score < 6$; 2 : $6 \leq Score < 7$; 3 : $7 \leq Score < 8$; 4 : $8 \leq Score \leq 10$	Ordinal	Gain

Table 5: Extracts from the periodic decision tables

Extract of the decision table of the $Week_{1,2}$												
Reference actions	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}	$Decision_1$
3486	2	2	3	3	1	2	4	3	1	2	1	1
18182	2	1	3	1	1	0	5	3	1	1	2	2
37526	2	1	3	3	0	2	5	3	4	5	4	3
Extract of the decision table of the $Week_{2,3}$												
Reference actions	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}	$Decision_2$
37193	2	3	3	2	0	2	4	2	1	2	4	1
3486	2	2	3	3	1	2	4	3	3	4	4	2
37105	2	3	2	3	1	0	5	3	2	5	3	3
Extract of the decision table of the $Week_{3,4}$												
Reference actions	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}	$Decision_3$
1631	2	2	1	2	1	0	3	2	1	5	1	1
28912	4	3	1	1	1	0	4	3	2	4	4	2
36056	2	2	3	3	1	0	5	3	2	3	2	3
Extract of the decision table of the $Week_{4,5}$												
Reference actions	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}	$Decision_4$
1346	2	2	2	3	1	0	4	2	1	3	4	1
15748	2	3	2	1	1	2	2	2	1	2	4	2
39175	2	3	3	2	1	2	3	3	2	3	4	3

Table 6: Extracts from the periodic preference models

Preference model $PM_{1,2}$		
Decision class	Decision rule	Strength
At most Cl_1 (Cl_1^{\leq})	If $f(A_{i,2}, g_7) \leq 2$ Then $A_{i,2} \in Cl_1^{\leq}$	42%
At most Cl_2 (Cl_2^{\leq})	If $f(A_{i,2}, g_9) \leq 1$ Then $A_{i,2} \in Cl_2^{\leq}$	85%
At least Cl_2 (Cl_2^{\geq})	If $f(A_{i,2}, g_3) \geq 3 \wedge f(A_{i,2}, g_{10}) \geq 3 \wedge f(A_{i,2}, g_{11}) \geq 2$ Then $A_{i,2} \in Cl_2^{\geq}$	52%
At least Cl_3 (Cl_3^{\geq})	If $f(A_{i,2}, g_9) \geq 3$ Then $A_{i,2} \in Cl_3^{\geq}$	40%
Preference model $PM_{2,3}$		
Decision class	Decision rule	Strength
At most Cl_1 (Cl_1^{\leq})	If $f(A_{i,3}, g_{10}) \leq 1 \wedge f(A_{i,3}, g_7) \leq 2$ Then $A_{i,3} \in Cl_1^{\leq}$	20%
At most Cl_2 (Cl_2^{\leq})	If $f(A_{i,3}, g_9) \leq 1$ Then $A_{i,3} \in Cl_2^{\leq}$	80%
At least Cl_2 (Cl_2^{\geq})	If $f(A_{i,3}, g_3) \geq 2 \wedge f(A_{i,3}, g_7) \geq 3 \wedge f(A_{i,3}, g_8) \geq 2 \wedge f(A_{i,3}, g_{10}) \geq 3 \wedge f(A_{i,3}, g_{11}) \geq 2 \wedge f(A_{i,3}, g_1) \geq 2$ Then $A_{i,3} \in Cl_2^{\geq}$	70%
At least Cl_3 (Cl_3^{\geq})	If $f(A_{i,3}, g_2) \geq 3 \wedge f(A_{i,3}, g_9) \geq 2 \wedge f(A_{i,3}, g_{10}) \geq 4$ Then $A_{i,3} \in Cl_3^{\geq}$	30%
Preference model $PM_{3,4}$		
Decision class	Decision rule	Strength
At most Cl_1 (Cl_1^{\leq})	If $f(A_{i,4}, g_4) \leq 2 \wedge f(A_{i,4}, g_7) \leq 2 \wedge f(A_{i,4}, g_8) \leq 2 \wedge f(A_{i,4}, g_{11}) \leq 2$ Then $A_{i,4} \in Cl_1^{\leq}$	23%
At most Cl_2 (Cl_2^{\leq})	If $f(A_{i,4}, g_9) \leq 1$ Then $A_{i,4} \in Cl_2^{\leq}$	71%
At least Cl_2 (Cl_2^{\geq})	If $f(A_{i,4}, g_{11}) \geq 3 \wedge f(A_{i,4}, g_{10}) \geq 3 \wedge f(A_{i,4}, g_4) \geq 2 \wedge f(A_{i,4}, g_6) \geq 1 \wedge f(A_{i,4}, g_7) \geq 3 \wedge f(A_{i,4}, g_8) \geq 2$ Then $A_{i,4} \in Cl_2^{\geq}$	28%
At least Cl_3 (Cl_3^{\geq})	If $f(A_{i,4}, g_9) \geq 4$ Then $A_{i,4} \in Cl_3^{\geq}$	16%
Preference model $PM_{4,5}$		
Decision class	Decision rule	Strength
At most Cl_1 (Cl_1^{\leq})	If $f(A_{i,5}, g_2) \leq 2 \wedge f(A_{i,5}, g_7) \leq 3 \wedge f(A_{i,5}, g_8) \leq 2 \wedge f(A_{i,5}, g_9) \leq 1 \wedge f(A_{i,5}, g_{10}) \leq 2 \wedge f(A_{i,5}, g_{11}) \leq 2$ Then $A_{i,5} \in Cl_1^{\leq}$	10%
At most Cl_2 (Cl_2^{\leq})	If $f(A_{i,5}, g_9) \leq 1$ Then $A_{i,5} \in Cl_2^{\leq}$	75%
At least Cl_2 (Cl_2^{\geq})	If $f(A_{i,5}, g_3) \geq 3 \wedge f(A_{i,5}, g_6) \geq 1 \wedge f(A_{i,5}, g_7) \geq 3 \wedge f(A_{i,5}, g_8) \geq 2 \wedge f(A_{i,5}, g_{10}) \geq 3 \wedge f(A_{i,5}, g_{11}) \geq 3$ Then $A_{i,5} \in Cl_2^{\geq}$	27%
At least Cl_3 (Cl_3^{\geq})	If $f(A_{i,5}, g_9) \geq 4$ Then $A_{i,5} \in Cl_3^{\geq}$	20%

decision classes: Cl_1 of the “At-risk learners”, Cl_2 of the “Struggling learners” or Cl_3 of the “Leader learners”. The “Potential learners” are those who filled their registration form and, therefore, they were likely to be classified in one of the three decision classes. A summary of the classification results is presented in table 7.

Table 7: [Summary table of the periodic classifications](#)

$Period_i$	Number of “Actions of reference”	Number of criteria	Number of decision rules	Number of actions assigned to Cl_1	Number of actions assigned to Cl_2	Number of actions assigned to Cl_3
$Week_{1,2}$	30	11	12	430	855	249
$Week_{2,3}$	60	11	9	899	527	108
$Week_{3,4}$	90	11	9	1256	262	16
$Week_{4,5}$	120	11	7	1361	130	43

6.4. Results and discussion

In this section we compare, at first, the computational time made by the non incremental algorithm and the *DRSA-Incremental* one when updating the DRSA approximations. Then, we discuss the quality of the weekly inferred preference model.

All the algorithms in this paper are coded by Java and run on a personal computer with Windows 7, Intel (R) *CoreTM* i3-3110M CPU @ 2.4 GHz and 4.0 GB memory.

6.4.1. Incremental vs non-incremental algorithm: Comparative study

This section shows a comparative study of computational time between the incremental and the non-incremental algorithms. Two types of experiments are considered: real data experiments coming from the MOOC “Design Thinking” and artificial data experiments generated by a computer program to create the criteria values. The real dataset contains 1535 actions representing the MOOC learners and the artificial dataset

contains 10000 actions. Both datasets are assessed on eleven criteria. The incremental and the non-incremental algorithms are running on each of the datasets.

The experiments shown below aim at demonstrating the time efficiency of our DRSA-Incremental algorithm for the update of DRSA approximations when (1) increasing the dataset size, (2) changing the inserted actions ratios and (3) increasing both criteria and decision classes numbers. In all figures, the y-coordinate concerns the computational time in seconds and the x-coordinate concerns the changing variable.

Comparison of running time on increasing data size. Figure 2 shows a comparison of running time when increasing the data size. In the case of real data, we have four sets that represent the “actions of reference” of each week that are identified by the pedagogical team of the MOOC “Design Thinking”. The pedagogical team fixed the number of actions of reference to 30 so that they can decide later their assignment in the decision classes. Since we used a cumulative approach, the “actions of reference” of each week were merged with those of previous weeks which gives the sets of 30, 60, 90 and 120 “actions of reference”. In the case of artificial data, we divided the original set to five subsets containing 2000, 4000, 6000, 8000 and 10000 actions respectively. Each time, we insert in the considered subset 2% of actions, compared to its size, that are generated by randomly assigning criteria values in the current criteria domains.

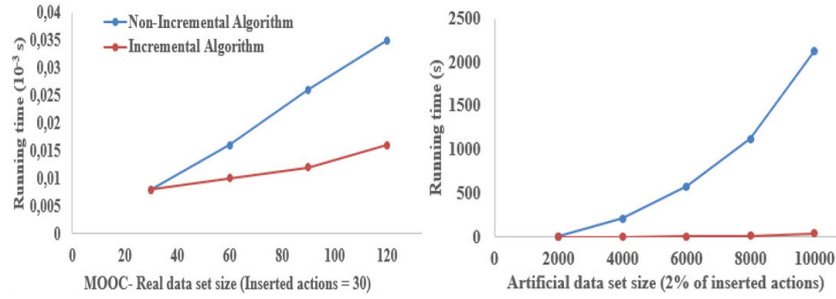


Figure 2: Comparison of computation time between the incremental and the non-incremental algorithms on increasing data sets

In two cases, we note that running time of both non-incremental and incremental algorithms grows up with the continuous increasing in data size. However, the incremental algorithm is always faster than the non-incremental one. Moreover, the differences

between two algorithms, called enhancing ratios, become more and more distinct with the increasing scale of these data sets. Thus, updating the DRSA-approximations permits to reduce the computational complexity compared to calculating them by scratch especially in big data contexts.

Comparison of running time when changing inserted actions ratios. Figure 3 shows a comparison of the computational time when changing the ratio of inserted actions in the original data set. The first set concerns the 1535 real actions and the second is a subset of 4000 actions selected randomly from the original artificial dataset. In both two sets we insert from 10% to 100% of data compared to their size in order to compare the incremental and the non-incremental algorithms with these different changing ratios. We recall that the inserted actions are generated randomly and must satisfy the domain criteria values.

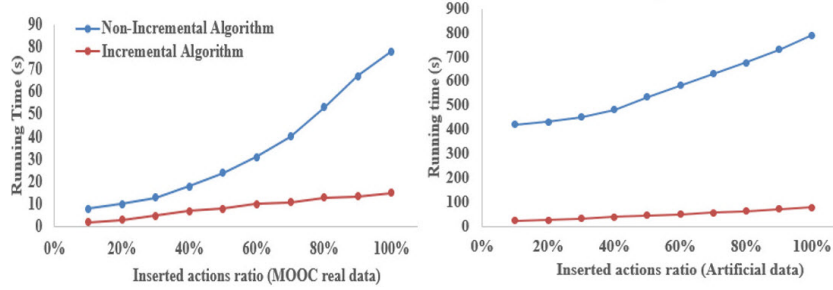


Figure 3: Comparison of computation time between the incremental and the non-incremental algorithms on increasing inserted actions ratios

We note that the incremental algorithm significantly reduces the computational time compared to the non-incremental algorithm. Moreover, the enhancing ratio is increasingly important when the changing ratio increases. Finally, we note that the rate of change of the incremental algorithm is much more stable than that of the non-incremental one.

Comparison of running time of the incremental algorithm when increasing the decision classes and the criteria numbers. Figure 4 shows a comparison of running time of the incremental algorithm when increasing the number of decision classes and that

of criteria. The curves (a) and (c) concern the real dataset containing the 120 actions of reference. The curves (b) and (d) concern a subset of 4000 actions selected randomly from the original artificial dataset. Both two sets are divided into 4 subsets containing 30, 60, 90, and 120 for real data, and 1000, 2000, 3000, 4000 for artificial data. The incremental algorithm is applied on each of these eight datasets to compare its computational time by increasing either the number of decision classes (curves (a) and (b)) or that of criteria (curves (c) and (d)). In curves (a) and (b) the number of criteria is fixed to eleven and for the curves (c) and (d) the number of decision classes is fixed to three.

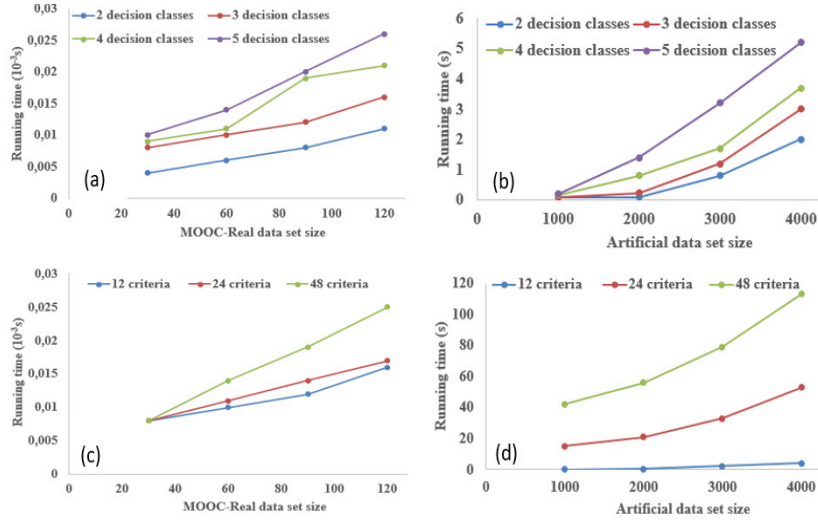


Figure 4: Comparison of computation time of the incremental algorithm on increasing decision classes number (Curves (a) and (b)) and the criteria number (Curves (c) and (d))

Results show that the computational time increases proportionally with both the number of decision classes and the number of criteria. The higher the number of decision classes or criteria, the greater the time required to update the approximations. Moreover, the enhancing ratios become more and more large with the increasing scale of changing variables. Thus, the incremental algorithm is sensitive, not only to the number of actions, but also to the number of decision classes and that of criteria.

The lessons drawn from these experiments show that the incremental algorithm is significantly faster than the non-incremental one considering the computational com-

plexity. This contribution is very interesting in the case of large databases or real-time decision making where the saving-time would be very important such as the crisis management and the healthcare domains.

In this context of MOOCs, the gain in time is obviously worthwhile but not necessary since the decision makers intervention for the assessment and the assignment of actions is time consuming. However, the human experts opinions still also significant to obtain stronger and more meaningful decision rules. Despite this, some works in literature as [14] proposed methods to minimize the human intervention and so to automatize the decision process in order to permit a real-time decision in such contexts based on DRSA which makes the incremental approach increasingly important.

6.4.2. Evaluation metrics

In this subsection we define some evaluation metrics (F-measure, accuracy, MAE, RMSE) that will be used in order to assess the effectiveness of the proposed model.

The F-measure is the summary indicator commonly used to evaluate classification algorithms. It is calculated as follows:

$$F - measure = \frac{2 * recall * precision}{(recall + precision)}$$

The F-measure is the harmonic value between recall and precision. The recall is the number of relevant answers predicted in relation to the total number of relevant answers. The precision is the number of the relevant answers predicted among all predicted answers. These two measures are calculated as follows:

$$Recall = \frac{TP}{(TP+FN)}; precision = \frac{TP}{(TP+FP)} \text{ such that as}$$

- True positive (TP): element of the positive class correctly predicted.
- True negative (TN): element of the negative class correctly predicted.
- False positive (FP): element of the positive class poorly predicted.
- False negative (FN): element of the negative class poorly predicted.

However, both recall and precision measures can not, together, give an indication about the classification model performance since, generally, when one is high the other

is low. Therefore, the F-measure is proposed to harmonize these two measures. The F-measure has a quality which is to drop sharply when one of its parameters is low and to be increased when the two parameters are close to each other.

The accuracy is the proportion of correct results obtained by a classifier. It is calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy usually gives values that are better than the F-measure. Indeed, compared to an accuracy measure, the F-measure allows the distribution of errors in the prediction sets. However, accuracy only allows to know if the prediction or the classification is acceptable in general or not. It remains much more superficial than the F-measure. Accuracy is more efficient in the case of symmetric datasets where values of false positive and false negatives are almost same.

The root mean square error (RMSE) and the mean absolute error (MAE) are another useful measures widely used in prediction model evaluations.

RMSE is a quadratic scoring rule that measures the average magnitude of the error. It's the square root of the average of squared differences between prediction and real observation. It is calculated as follows:

$$RMSE = \sqrt{\frac{1}{N} * \sum_{j=1}^N (y_j - y'_j)^2}$$

MAE is the average of the difference between the real and the predicted values without specifying whether we are under or over predicting the real data. MAE is represented as follows:

$$MAE = \frac{1}{N} * \sum_{j=1}^N |y_j - y'_j|$$

6.4.3. *Quality of predictions made by the inferred preference model*

This section presents the results given when applying the preference model on real MOOC dataset. To asses the predictive quality, we have to weekly compare between the predicted and the real classifications. Experiments are based on two approaches:

- *The pessimistic, optimistic and [score classification](#) approaches*: as already stated in section 4, a decision rule can classify an action in different unions such that

Cl_1^{\leq} that means “at-most Cl_1 ”, Cl_2^{\leq} that means “at-most Cl_2 ”, Cl_2^{\geq} that means “at-least Cl_2 ” and Cl_3^{\geq} that means “at-least Cl_3 ”. An action classified in the union “at-most Cl_2 ” can belong either to the class Cl_1 or to the class Cl_2 . Similarly, an action classified in the union “at-least Cl_2 ” may belong either to the class Cl_2 or to the class Cl_3 . However, an action has to be classified in one and only one decision class. In this case, the pessimistic approach assigns the actions of the union “at-most Cl_2 ” to the class Cl_1 and those of the union “at-least Cl_2 ” to the class Cl_2 . Likewise, the optimistic approach assigns them respectively to the decision classes Cl_2 and Cl_3 . The score classification proposes an harmonic assignment between the “min” and the “max” according to an affectation score.

- *The cumulative/ non-cumulative approach*: the model learning phase can be non-cumulative, so based only on the “actions of reference”) defined in the current week of the MOOC in order to make predictions concerning the following week during this same MOOC. Otherwise, this sample can be cumulative if we add to the learning sample of the current week, samples of all the previous weeks.

Based on these approaches, five cases were considered: the non-cumulative pessimistic approach, the cumulative pessimistic approach, the non-cumulative optimistic approach, the cumulative optimistic approach and the cumulative approach based on the score classification. The prediction model effectiveness is assessed using the F-measure evaluation throughout the five weeks of the MOOC broadcast. The results are summarized in Figure 5.

Consequently, we note that the pessimistic cumulative approach gives the highest prediction effectiveness. Indeed, to meet the cognitive capacity of the decision maker when using DRSA, we generally select learning samples of a limited size. Yet, since the sample size is so important for the diversification of the examples in the learning set, the cumulative approach has addressed this limitation and has eventually led to better results compared to the non-cumulative one. Moreover, in our case the pessimistic approach gives the more efficient results and the classification score method is always more effective than the optimistic approach and less effective than the pessimistic approach. Therefore, if we consider certain external factors leading to the

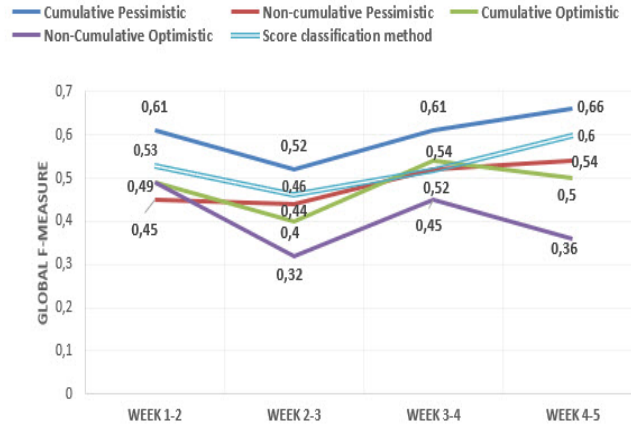


Figure 5: The global F-measure calculated for the four approaches during the MOOC broadcast

MOOCs dropout such as the lack of time and the absence of a concrete commitment, the “Struggling learners” are more likely to abandon the MOOC than to be “Leader learners”. Hence, it seems obvious that the pessimistic approach is more efficient compared to the optimistic and the score classification ones in this context.

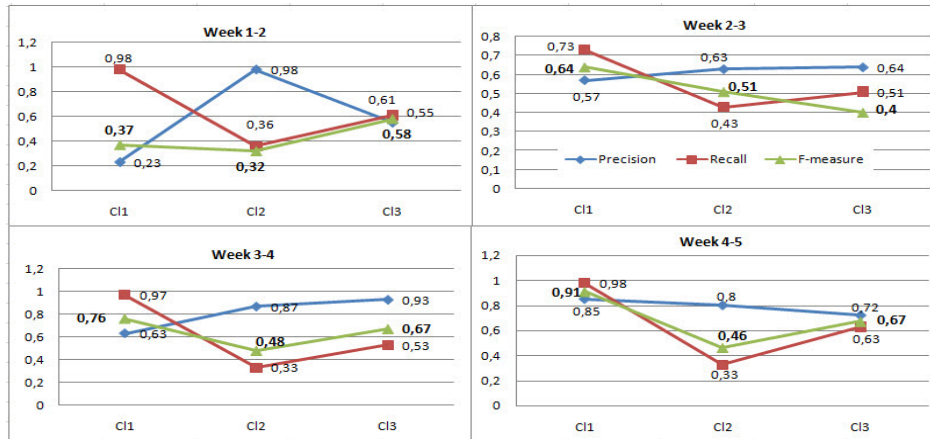


Figure 6: The effectiveness measures of the prediction model for each decision class over weeks

In Figure 6 we focus on the pessimistic cumulative approach and we present the precision, the recall and the F-measure measures relative to each decision class Cl_1 , Cl_2 and Cl_3 for each week during the MOOC broadcast. The predictive performance

results from the application of the preference model on all of the 1535 learners.

- The F-measure corresponding to the decision class Cl_1 of the “At-risk learners” increases over time. Thus, the effectiveness of the Cl_1 class prediction increases from a week to another. In effect, a MOOC is known by the presence of what we call “lurkers”. These are the participants who register just to discover the MOOC concept and who leave it at the first evaluation. And in spite of their activity during the first week of the MOOC, they keep having the prior intention to abandon it. This type of learners degrades the quality of the prediction model which is based on the profile and the behaviour of the learner and not on his intention. Consequently, the fewer the number of lurkers gets, the higher the prediction quality becomes.
- The F-measure relative to the Cl_3 decision class of the “Leader learners” increases over time too. In fact, from one week to another, these learners enhance their participation in the forum, a thing which gives us more information concerning their profile. In addition, the assessment activities provided by the MOOC are increasingly complex over the weeks. Obviously, if compared to a simple Quiz, a complex assessment such as the peer-to-peer activity permits more to better assess a learner. This is justified by the deterioration of the F-measure of the Cl_3 class in the week 2-3. In fact, our MOOC proposed a quiz at the end of week 2 and a peer to peer activity at the end of week 3. However, students who pass a quiz may hang at a peer-to-peer activity or even abandon the MOOC because of its complexity. This may also affect the prediction quality.

Finally, Figure 7 shows a comparison between the overall F-measure, accuracy, MAE and RMSE of the weekly prediction model. We recall that the F-measure and the accuracy are better when they are close to 1 whereas the MAE and the RMSE are better when they are close to 0. We notice that all of the measures gave the best result in week 4 and the worst in week 1. For the F-measure and the accuracy we obtained 0.66 and 0.89 what are satisfactory results. Concerning the MAE and the RMSE we obtained 0.16 and 0.42 respectively. These results are not very close to 0 because MAE and

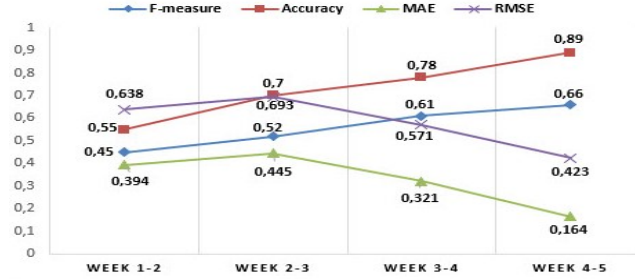


Figure 7: Comparison between the average F-measure and the overall accuracy of the decision classes during the four weeks of the MOOC

RMSE are rather used in the case of quantities prediction and therefore of continuous values. Whereas for our case the prediction values are not continuous since they belong to the finite set $CI = \{1, 2, 3\}$. Thus, in our case, we consider that the F-measure is the most adequate since it analyzes the difference between the real data and the predicted data based on of the false-positive and the false-negative values.

To sum up, our multicriteria approach for the incremental periodic prediction proved many advantages in the context of MOOC. First, it permitted to rigorously characterize the learners' profiles based on a bibliographic study and direct meetings with the experts to elicit their preferences. Second, the incremental approach allowed saving more time especially in a bigdata domain or real-time decision making. The incrementation helped also the pedagogical team capitalize on their previous preferences in order to revise their previous decisions. Indeed, this approach considered the evolution of the learner profiles and the characteristics of the MOOC in order to update the classification rules from one week to the next. For example, the learner 3486 in the second week has no longer belong to the decision class of "At-risk learners" since it becomes just in difficulty (cf. Table 5). Third, the human intervention for the learning sets identification permitted to avoid the unbalanced data problem that strongly occurs in the MOOC context because of the presence of the majority class of "At-risk learners" [6]. Finally, the DRSA approach has achieved very satisfactory results (overall F-measure reached 0.66 for the week 4-5 and 0.61 for the week 1-2; and the accuracy reached 0.89). Moreover, the cumulative approach allowed to obtain a more diverse

sample and the pessimistic approach permitted to take into account the external factors of a MOOC that can incite active learners to abandon the course. All these advantages concerning the predictive quality and the time saving, will help us provide the “AT-risk learners” by an accurate assistance in the right moment to encourage them not leaving the course.

7. Conclusion

In this paper, a new multicriteria classification approach based DRSA, called *MAI2P*, and a *DRSA-Incremental* algorithm are proposed for the incremental periodic prediction. The *MAI2P* approach consists of three phases: The first is the construction of a decision table. The second infers a preference model by applying our *DRSA-Incremental* algorithm for the update of the upper and the lower approximations of the DRSA. The third phase uses the inferred preference model to predict the decision class to which each action will belong. The first phase is based on three steps: (i) the construction of a representative learning sample for each of the N decision classes called “Actions of reference”, (ii) the construction of a coherent family of criteria for the characterization of the actions, and (iii) the classification of all the “Reference Actions” in the N predefined decision classes. The decision-making process is based on the preferences of the human decision makers.

The approach *MAI2P* addresses the problem of dynamic learning samples that evolve over time. It is based on the incremental learning algorithm *DRSA-Incremental* that we defined for the incremental update of the upper and lower approximations following the sequential enrichment of the learning sample. It is periodically applied at the end of each period to predict classifications for the next period.

The approach *MAI2P* is generic and can be carried out in several areas for incremental and periodic prediction in a time-varying information system. However, the intervention of human decision makers to select and to classify the “Actions of reference” may lead to temporal constraints, which prevents the application of the *MAI2P* approach in a context of prediction where the decision must be made in real time.

Validated in a MOOC context, experiments showed that the pessimistic cumulative

approach gives the most efficient preference model with an F-measure that reaches 0.66 and an accuracy that reaches 0.89. Our perspectives aim to conduct a strict comparative study between our predictive method based-DRSA and the other conventional machine learning techniques, especially the neural network. Indeed, the neural network is the most frequently used technique in the context of MOOC and it proved many limits compared to our proposed approach. A such comparative study will allow to more promote the effectiveness of our approach and, especially, the importance of the experts intervention in the decision process.

References

- [1] Altman, E., 1968. Financial ratios, discriminate analysis and the prediction of corporate bankruptcy. *Journal of Finance* 23, 589–609.
- [2] Balakrishnan, G., May 2013. Predicting student retention in massive open online courses using hidden markov models. Master’s thesis, EECS Department, University of California, Berkeley, retrieved from <http://www2.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-109.html>.
- [3] Barak, M., 2015. The same mooc delivered in two languages: Examining knowledge construction and motivation to learn. *Proceedings of the EMOOCS*, 217–223.
- [4] Beccali, M., Cellura, M., Mistretta, M., 2003. Decision-making in energy planning. application of the electre method at regional level for the diffusion of renewable energy technology. *Renewable energy* 28 (13), 2063–2087.
- [5] Błaszczyszński, J., Greco, S., Słowiński, R., 2007. Multi-criteria classification—a new scheme for application of dominance-based decision rules. *European Journal of Operational Research* 181 (3), 1030–1044.
- [6] Bouzayane, S., Saad, I., 2017. Dominance-based rough set approach to learn in the presence of class imbalance. In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI) Workshop on Learning in the Presence of Class Imbalance and Concept Drift (LPCICD’17)*. pp. 24–31.

- [7] Bouzayane, S., Saad, I., 2017. A preference ordered classification to leader learners identification in a mooc. *Journal of Decision Systems* 26 (2), 189–202.
- [8] Chakhar, S., Ishizaka, A., Labib, A., Saad, I., 2016. Dominance-based rough set approach for group decisions. *European Journal of Operational Research* 251 (1), 206–224.
- [9] Chakhar, S., Saad, I., 2012. Dominance-based rough set approach for groups in multicriteria classification problems. *Decision Support Systems* 54 (1), 372–380.
- [10] Chaplot, D., Rhim, E., Kim, J., 2015. Predicting student attrition in moocs using sentiment analysis and neural networks. *Proceedings of AIED 2015 fourth workshop on intelligent support for learning in groups*, 7–12.
- [11] Chaturvedi, S., Goldwasser, D., Daume II, H., 2014. Predicting instructor’s intervention in mooc forums. *Proceedings of the 52nd annual meeting of the association for computational linguistics (ACL)*, 1501–1511.
- [12] Chen, H., Li, T., Ruan, D., 2012. Maintenance of approximations in incomplete ordered decision systems while attribute values coarsening or refining. *Knowledge Based Systems* 31, 140–161.
- [13] Chevalier, J., Rousseaux, P., 1999. Classification in lca: building of a coherent family of criteria. *The International Journal of Life Cycle Assessment* 4 (6), 352–356.
- [14] Dau, H. N., Chakhar, S., Ouelhadj, D., Abubahia, A. M., 2019. Construction and refinement of preference ordered decision classes. In: *UK Workshop on Computational Intelligence*. Springer, pp. 248–261.
- [15] David, A., Hatchuel, A., Laufer, R., 2001. *Les nouvelles fondations des sciences de gestion*. Vuibert.
- [16] Delen, D., Walker, G., Kadam, A., 2005. Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial intelligence in medicine* 34 (2), 113–127.

- [17] Educause, 2013. 7 things you should know about moocs II. Web log post, Educause Learning Initiative, retrieved from <https://library.educause.edu/media/files/library/2013/6/eli7097-pdf.pdf>.
- [18] Fillion, L., 1991. Visions et relations: clefs du succès de l'entrepreneur. Editions de l'entrepreneur, 27–70.
- [19] Fini, A., 2009. The technological dimension of a massive open online course: The case of the cck08 course tools. *International Review of Research in Open and Distance Learning* 10 (5), retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/643/1402>.
- [20] Greco, S., Matarazzo, B., Slowinski, R., 2001. Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research* 129 (1), 1–45.
- [21] Greco, S., Slowinski, R., Stefanowski, J., Zurawski, M., 2004. Incremental versus nonincremental rule induction for multicriteria classification. *Transaction on Rough Sets II*, 33–53.
- [22] Hall, M. A., 1999. Correlation-based feature selection for machine learning. Ph.D. thesis, The University of Waikato.
- [23] Hone, K., El Said, G., 2016. Exploring the factors affecting mooc retention: A survey study. *Computers & Education* 98, 157–168.
- [24] Hu, Q., Chakhar, S., Siraj, S., Labib, A., 2017. Spare parts classification in industrial manufacturing using the dominance-based rough set approach. *European Journal of Operational Research* 262 (3), 1136–1163.
- [25] Kotłowski, W., Dembczyński, K., Greco, S., Słowiński, R., 2008. Stochastic dominance-based rough set model for ordinal classification. *Information Sciences* 178 (21), 4019–4037.
- [26] Li, S., Li, T., Liu, D., 2013b. Dynamic maintenance of approximations in dominance-based rough set approach under the variation of the object set. *Int. J. Intell. Syst.* 28, 729–751.

- [27] Liou, J. H., Tzeng, G. H., 2010. A dominance-based rough set approach to customer behavior in the airline market. *Information Sciences* 180 (11), 2230–2238.
- [28] Liu, J., Liao, X., Yang, J.-b., 2015. A group decision-making approach based on evidential reasoning for multiple criteria sorting problem with uncertainty. *European Journal of Operational Research* 246 (3), 858–873.
- [29] Michalski, R. S., Reinke, R. E., 1988. Incremental learning of concept descriptions: A method and experimental results. *Machine Intelligence XI*. Clarendon Press, 263–288.
- [30] Miller, G., 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63, 81–97.
- [31] Moon, S., Potdar, S., Martin, L., 2014. Identifying student leaders from mooc discussion forums thorough language influence. *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)*, 15–20.
- [32] Morris, N., Hotchkiss, S., Swinnerton, B., 2015. Can demographic information predict mooc learner outcomes? In: *Proceedings of the EMOOC Stakeholder Summit*, 199–207.
- [33] Moul, J., Pedersen, J. T., Judson, R., Fidelis, K., 1995. A large-scale experiment to assess protein structure prediction methods. *Proteins: Structure, Function, and Bioinformatics* 23 (3), 2–4.
- [34] Mousseau, B. R., VINCENT, Roy, B., 1996. A theoretical framework for analysing the notion of relative importance of criteria. *J. Multi-Criteria Decis. Anal* 5, 145–159.
- [35] Najafi, G., Ghobadian, B., Moosavian, A., Yusaf, T., Mamat, R., Kettner, M., Azmi, W., 2016. Svm and anfis for prediction of performance and exhaust emissions of a si engine with gasoline–ethanol blended fuels. *Applied Thermal Engineering* 95, 186–203.

- [36] Onah, D. F. O., Sinclair, J., Boyatt, R., 2014. Dropout rates of massive open online courses: behavioral patterns. 6th international conference on education and new learning technologies, published in: Edulearn14 proceedings, 5825–5834.
- [37] Patru, M., Balaji, V., 2016. Making sense of moocs: a guide for policy makers in developing countries. Tech. rep., UNESCO, retrieved from <http://unesdoc.unesco.org/images/0024/002451/245122E.pdf>.
- [38] Pawlak, Z., 1992. Rough sets. *International Journal & Computer Sciences*, 341–356.
- [39] Ryu, Y., Chandrasekaran, R., Jacob, V., 2007. Data classification using the isotonic separation technique: Application to breast cancer prediction. *European Journal of Operational Research*, 842–854.
- [40] Saad, I., Rosenthal-Sabroux, C., Grundstein, M., 2005. Improving the decision making process in the design project by capitalizing on company's crucial knowledge. *Group Decision and Negotiation* 14 (2), 131–145.
- [41] Sarukkai, R. R., 2000. Link prediction and path analysis using markov chains. *Computer Networks* 33 (1), 377–386.
- [42] Shanahan, M., 1989. Prediction is deduction but explanation is abduction. *International Joint Conf on Artificial Intelligence*, 1055–1060.
- [43] Sharma, A., Sahoo, P. K., Tripathi, R., Meher, L. C., 2016. Artificial neural network-based prediction of performance and emission characteristics of ci engine using polanga as a biodiesel. *International Journal of Ambient Energy* 37 (6), 559–570.
- [44] Shen, K., Tzeng, G., 2014. DRSA-based neuro-fuzzy inference system for the financial performance prediction of commercial bank. *Int. J. Fuzzy Syst* 16 (2), 173–183.
- [45] Silver, D. L., 2011. Machine lifelong learning: challenges and benefits for artificial general intelligence. In: *International Conference on Artificial General Intelligence*. Springer, pp. 370–375.

- [46] Stańczyk, U., 2011. On performance of drsa-ann classifier. *Hybrid Artificial Intelligent Systems*, 172–179.
- [47] Sun, Y., Leng, B., Guan, W., 2015. A novel wavelet-svm short-time passenger flow prediction in beijing subway system. *Neurocomputing* 166, 109–121.
- [48] Suriel, R. L., Atwater, M. M., 2012. From the contributions to the action approach: White teacher. *Journal of Research in Science Teaching* 49 (10), 1271–1295.
- [49] Thusberg, J., Olatubosun, A., Vihinen, M., 2011. Performance of mutation pathogenicity prediction methods on missense variants. *Human mutation* 32 (4), 358–368.
- [50] Ture, M., Kurt, I., Kurum, A. T., Ozdamar, K., 2005. Comparing classification techniques for predicting essential hypertension. *Expert Systems with Applications* 29 (3), 583–588.
- [51] Wolff, A., Zdrahal, Z., Herrmannov, D., Kuzilek, J., Hlosta, M., 2014. Developing predictive models for early detection of at-risk students on distance learning modules. In: *Workshop: Machine Learning and Learning Analytics at LAK*.
- [52] Xing, W., Chen, X., Stein, J., Marcinkowski, M., 2016. Temporal predication of dropouts in moocs: Reaching the low hanging fruit through stacking generalization. *Computers in Human Behavior* 58, 119–129.
- [53] Yadav, A. K., Chandel, S., 2014. Solar radiation prediction using artificial neural network techniques: A review. *Renewable and Sustainable Energy Reviews* 33, 772–781.
- [54] Yan, H., DeChant, C. M., Moradkhani, H., 2015. Improving soil moisture profile prediction with the particle filter-markov chain monte carlo method. *IEEE Transactions on Geoscience and Remote Sensing* 53 (11), 6134–6147.
- [55] Yang, D., Sinha, T., Adamson, D., Rose, C., 2013. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In: *Nips workshop on data driven education*.

- [56] Zaras, K., 2011. The dominance-based rough set approach (DRSA) applied to bankruptcy prediction modeling for small and medium businesses. *Multiple Criteria Decision Making/The University of Economics in Katowice*, 287–295.
- [57] Zhang, J., Li, T., Ruan, D., 2012. Rough sets based matrix approaches with dynamic attribute variation in set-valued information systems. *International Journal of Approximate Reasoning* 53, 620–635.

**Appendix A. Incremental algorithm for the update of the approximations of DRSA
when adding a set of actions in the information system**

Data:
 $\forall x \in K, D_i^+ p(x)$ and $D_i^- p(x)$ of the *week_i*
 $\forall n \in T, \bar{P}(CI_{i,n}^{\geq}), \bar{P}(CI_{i,n}^{\leq}), \underline{P}(CI_{i,n}^{\geq})$ and $\underline{P}(CI_{i,n}^{\leq})$ of the *week_i*
A set Δ^+ containing new actions
Result:
 $\bar{P}(CI_{i+1,n}^{\geq}), \bar{P}(CI_{i+1,n}^{\leq}), \underline{P}(CI_{i+1,n}^{\geq})$ and $\underline{P}(CI_{i+1,n}^{\leq})$ of the *week_{i+1}*
Begin
for i *from* 1 *to* $|\Delta^+|$ **do**
 Calculate $D_{i+1}^+ p(x_i^+)$ and $D_{i+1}^- p(x_i^+)$;
end
for n *from* 1 *to* N **do**
 Calculate Δ_n^+, Δ_n^+ and Δ_n^+ ;
end
for i *from* 1 *to* K **do**
 for j *from* 1 *to* $|\Delta^+|$ **do**
 if $x_i \in D_{i+1}^+ p(x_j^+)$ **then**
 $D_{i+1}^- p(x_i) \leftarrow D_{i+1}^- p(x_i) \cup \{x_j^+\}$;
 end
 if $x_i \in D_{i+1}^- p(x_j^+)$ **then**
 $D_{i+1}^+ p(x_i) \leftarrow D_{i+1}^+ p(x_i) \cup \{x_j^+\}$;
 end
 end
end
for i *from* 1 *to* $|\Delta^+|$ **do**
 Calculate $l_{i+1} p(x_i^+)$ and $u_{i+1} p(x_i^+)$
end
for n *from* 1 *to* N **do**
 for i *from* 1 *to* $|\Delta_n^+|$ **do**
 Calculate $\Delta \mathbb{M}_n$;
 $\underline{P}(CI_{i+1,n}^{\leq}) \leftarrow \underline{P}(CI_{i,n}^{\leq}) - \mathbb{M}_n$;
 $\bar{P}(CI_{i+1,n}^{\geq}) \leftarrow \bar{P}(CI_{i,n}^{\geq}) \cup D_{i+1}^+ p(x_i^+)$;
 if $n < l_{i+1} p(x_i^+)$ **then**
 $\underline{P}(CI_{i+1,n}^{\geq}) \leftarrow \underline{P}(CI_{i,n}^{\geq}) \cup \{x_i^+\}$;
 end
 if $n = l_{i+1} p(x_i^+)$ **then**
 $\underline{P}(CI_{i+1,n}^{\geq}) \leftarrow \underline{P}(CI_{i,n}^{\geq}) \cup \{x_i^+\}$;
 $\bar{P}(CI_{i+1,n}^{\leq}) \leftarrow \bar{P}(CI_{i,n}^{\leq}) \cup \{x_i^+\}$;
 end
 if $n > l_{i+1} p(x_i^+)$ **then**
 $\bar{P}(CI_{i+1,n}^{\leq}) \leftarrow \bar{P}(CI_{i,n}^{\leq}) \cup \{x_i^+\}$;
 end
 end
 for i *from* 1 *to* $|\Delta_n^+|$ **do**
 if $n < l_{i+1} p(x_i^+)$ **then**
 $\underline{P}(CI_{i+1,n}^{\geq}) \leftarrow \underline{P}(CI_{i,n}^{\geq}) \cup \{x_i^+\}$;
 end
 if $n = l_{i+1} p(x_i^+)$ **then**
 $\underline{P}(CI_{i+1,n}^{\geq}) \leftarrow \underline{P}(CI_{i,n}^{\geq}) \cup \{x_i^+\}$;
 end
 if $n > l_{i+1} p(x_i^+)$ **then**
 $\underline{P}(CI_{i+1,n}^{\leq}) \leftarrow \underline{P}(CI_{i,n}^{\leq}) \cup x_i^+$;
 end
 if $n = u_{i+1} p(x_i^+)$ **then**
 $\underline{P}(CI_{i+1,n}^{\leq}) \leftarrow \underline{P}(CI_{i,n}^{\leq}) \cup x_i^+$;
 end
 $\bar{P}(CI_{i+1,n}^{\geq}) \leftarrow \bar{P}(CI_{i,n}^{\geq}) \cup D_{i+1}^+ (x_i^+)$
 $\bar{P}(CI_{i+1,n}^{\leq}) \leftarrow \bar{P}(CI_{i,n}^{\leq}) \cup D_{i+1}^- (x_i^+)$
 end
 for i *from* 1 *to* $|\Delta_n^+|$ **do**
 Calculate $\Delta \mathbb{X}_n$;
 $\underline{P}(CI_{i+1,n}^{\geq}) \leftarrow \underline{P}(CI_{i,n}^{\geq}) - \mathbb{X}_n$;
 $\bar{P}(CI_{i+1,n}^{\leq}) \leftarrow \bar{P}(CI_{i,n}^{\leq}) \cup D_{i+1}^- p(x_i^+)$;
 if $n < u_{i+1} p(x_i^+)$ **then**
 $\bar{P}(CI_{i+1,n}^{\geq}) \leftarrow \bar{P}(CI_{i,n}^{\geq}) \cup \{x_i^+\}$;
 end
 if $n = u_{i+1} p(x_i^+)$ **then**
 $\bar{P}(CI_{i+1,n}^{\geq}) \leftarrow \bar{P}(CI_{i,n}^{\geq}) \cup x_i^+$;
 $\underline{P}(CI_{i+1,n}^{\leq}) \leftarrow \underline{P}(CI_{i,n}^{\leq}) \cup x_i^+$;
 end
 if $n > u_{i+1} p(x_i^+)$ **then**
 $\underline{P}(CI_{i+1,n}^{\leq}) \leftarrow \underline{P}(CI_{i,n}^{\leq}) \cup x_i^+$;
 end
 end
end
Return $\bar{P}(CI_{i+1,n}^{\geq}), \bar{P}(CI_{i+1,n}^{\leq}), \underline{P}(CI_{i+1,n}^{\geq})$ and $\underline{P}(CI_{i+1,n}^{\leq})$

Appendix B. Proofs of the incremental update of the DRSA approximations

Lemma 1: Proof

We suppose that $(\Delta_{n>}^+ \cup \Delta_{n=}^+) = \emptyset$, thus we have $\forall x_i^+ \in \Delta^+, n' < n$, then $Cl_{i+1,n'} \not\subseteq Cl_{i,n}^{\geq}$. Thus, $Cl_{i+1,n}^{\geq} = Cl_{i,n}^{\geq}$. Otherwise, we have $\forall x^+ \in (\Delta_{n>}^+ \cup \Delta_{n=}^+)$ such that $x_i^+ \in \Delta_{n>}^+$ or $x_i^+ \in \Delta_{n=}^+$, $Cl_{i+1,n'} \subseteq Cl_{i,n}^{\geq}$. Thus $Cl_{i+1,n}^{\geq} = Cl_{i,n}^{\geq} \cup \Delta_{n>}^+ \cup \Delta_{n=}^+$. Similarly, equation (6) holds.

Lemma 2: Proof

At the $week_{i+1}$, $\forall x \in A_i$, we have $D_{i+1}^+ p(x) = \{y \in A_i \cup \Delta^+ : y D_{i+1} p(x)\} = \{y \in A_i : y D_{i+1} p(x)\} \cup \{y \in \Delta^+ : y D_{i+1} p(x)\} = D_i^+ p(x) \cup \{y \in \Delta^+ : y D_{i+1} p(x)\} = D_i^+ p(x) \cup \{x_i^+ : x_i^+ D_{i+1} p(x)\}$. Thus, $\forall x_i^+ \in \Delta^+$ such that $i \in \{1..k\}$, we have $\{x_i^+ : x_i^+ D_{i+1} p(x)\} = \{x_i^+ : x \in D_{i+1}^-(x_i^+)\} = \bigcup_{i=1}^{k'} \{x_i^+\}$ such that $x \in D_{i+1}^-(x_i^+)$. Otherwise, if $x \notin D_{i+1}^-(x_i^+)$, then $\{x_i^+ : x \in D_{i+1}^-(x_i^+)\} = \emptyset$. Thus, $\forall x \in A_i, \forall x_i^+ \in \Delta^+$, if $x \in D_{i+1}^-(x_i^+)$ then $D_{i+1}^+ p(x) = D_i^+ p(x) \cup \bigcup_{i=1}^{k'} \{x_i^+\}$. Otherwise, $D_{i+1}^+ p(x) = D_i^+ p(x)$. Similarly, equation (8) holds.

Proposition 1: Proof

We have $\underline{P}(Cl_{i+1,n}^{\geq}) = \{x \in A_{i+1} : D_{i+1}^+ p(x) \subseteq Cl_{i+1,n}^{\geq}\} = \{x \in A_i \cup \Delta^+ : D_{i+1}^+ p(x) \subseteq Cl_{i+1,n}^{\geq}\} = \{x \in A_i : D_{i+1}^+ p(x) \subseteq Cl_{i+1,n}^{\geq}\} \cup \{x \in \Delta^+ : D_{i+1}^+ p(x) \subseteq Cl_{i+1,n}^{\geq}\} = \{x \in A_i : D_{i+1}^+ p(x) \subseteq Cl_{i+1,n}^{\geq}\} \cup \{x_i^+ \in \Delta^+ : D_{i+1}^+ p(x_i^+) \subseteq Cl_{i+1,n}^{\geq}\} = \{x \in A_i : D_{i+1}^+ p(x) \subseteq Cl_{i+1,n}^{\geq}\} \cup \{x_i^+ \in \Delta^+ : l_{i+1} p(x_i^+) \geq n\} = \underline{P}(Cl_{i,n}^{\geq}) \cup \{x_i^+ \in \Delta^+ : l_{i+1} p(x_i^+) \geq n\}$. According to the equation (5), we have $Cl_{i+1,n}^{\geq} = Cl_{i,n}^{\geq}$, if $(\Delta_{n>}^+ \cup \Delta_{n=}^+) = \emptyset$ and $Cl_{i+1,n}^{\geq} = Cl_{i,n}^{\geq} \cup \Delta_{n>}^+ \cup \Delta_{n=}^+$ if $\Delta_{n>}^+ \cup \Delta_{n=}^+ \neq \emptyset$.

1. If $\Delta_{n>}^+ \neq \emptyset$ or $\Delta_{n=}^+ \neq \emptyset$, thus $\exists x_i^+ \in (\Delta_{n>}^+ \cup \Delta_{n=}^+)$ such that $n \leq n'$. Thus, if $\forall x_i^+ \in (\Delta_{n>}^+ \cup \Delta_{n=}^+)$, $l_{i+1} p(x_i^+) < n \leq n'$, thus $\nexists x_i^+ \in (\Delta_{n>}^+ \cup \Delta_{n=}^+)$ such that $l_{i+1} p(x_i^+) \geq n$ thus $\underline{P}(Cl_{i+1,n}^{\geq}) = \underline{P}(Cl_{i,n}^{\geq})$. As well, if $l_{i+1} p(x_i^+) \geq n$ then $\{x \in \Delta_{n>}^+ \cup \Delta_{n=}^+ : l_{i+1} p(x_i^+) \geq n\} = \{\bigcup_{i=1}^{k'} \{x_i^+\} \subseteq (\Delta_{n>}^+ \cup \Delta_{n=}^+) : l_{i+1} p(x_i^+) \geq n\}$. Thus $\underline{P}(Cl_{i+1,n}^{\geq}) = \underline{P}(Cl_{i,n}^{\geq}) \cup \{\bigcup_{i=1}^{k'} \{x_i^+\} : l_{i+1} p(x_i^+) \geq n\}$.
2. If $\exists \{x_i^+\} \in \Delta_{n<}^+$, i.e. $\{x_i^+\} \in \Delta^+$ such that $n' < n$, then $\{x \in \Delta^+ : l_{i+1} p(x_i^+) \geq n\} = \emptyset$ and $\{x \in A_i : D_{i+1}^+ p(x) \subseteq Cl_{i+1,n}^{\geq}\} = \{x \in A_i : D_{i+1}^+ p(x) \subseteq Cl_{i,n}^{\geq}\} = \{x \in A_i : l_{i+1} p(x) \geq n\} = \{x \in A_i : l_i p(x) \geq n\} - \{x \in \underline{P}(Cl_{i,n}^{\geq}) : l_{i+1} p(x) <$

$n\} = \underline{P}(Cl_{i,n}^{\geq}) - \{x \in \underline{P}(Cl_{i,n}^{\geq}) : l_i p(x) < n\}$. As noted, $D_i^+ p(x) \subset D_{i+1}^+ p(x)$ and $l_{i+1} p(x) \leq l_i p(x)$. Thus $\{x \in \underline{P}(Cl_{i,n}^{\geq}) : l_{i+1} p(x) < n\} = \{x \in \underline{P}(Cl_{i,n}^{\geq}) \cap D_i^- p(x_i^+) : l_{i+1} p(x) < n\}$.

Let $\Delta \mathbb{X}_n = \{x \in \underline{P}(Cl_i^{\leq}) \cap \bigcup_{i=1}^{k'} D_i^- p(x_i^+) : l_{i+1} p(x) < n, \forall i \in \{1..k\}\}$. Thus, if $n > n'$, then $\underline{P}(Cl_{i+1,n}^{\geq}) = \underline{P}(Cl_{i,n}^{\geq}) - \Delta \mathbb{X}_n$.

The same reasoning holds for the proposition 2.

Proposition 3: Proof

We have $\overline{P}(Cl_{i+1,n}^{\leq}) = \{x \in A_{i+1} : D_{i+1}^+ p(x) \cap Cl_{i+1,n}^{\leq} \neq \emptyset\} = \{x \in A_i : D_{i+1}^+ p(x) \cap Cl_{i+1,n}^{\leq} = Cl_{i,n}^{\leq} \neq \emptyset\} \cup \{x_i^+ \in \Delta^+ : D_{i+1}^+ p(x) \cap Cl_{i+1,n}^{\leq} \neq \emptyset\} = \{x \in A_i : D_{i+1}^+ p(x) \cap Cl_{i+1,n}^{\leq} \neq \emptyset\} \cup \{x_i^+ \in \Delta^+ : l_{i+1} p(x_i^+) \leq n\}$. According to the equation (6), we have $Cl_{i+1,n}^{\leq} = Cl_{i,n}^{\leq}$ if $\Delta_{n<}^+ = \emptyset$ and $\Delta_{n=}^+ = \emptyset$ (i.e. $n' > n$, $\forall x_i^+ \in \Delta^+$); and $Cl_{i+1,n}^{\leq} = Cl_{i,n}^{\leq} \cup \Delta_{n<}^+ \cup \Delta_{n=}^+$ else. Thus $\{x \in K_i : D_{i+1}^+ p(x) \cap Cl_{i+1,n}^{\leq} \neq \emptyset\} \cup \{x_i^+ \in \Delta^+ : l_{i+1} p(x_i^+) \leq n\} = (Cl_{i,n}^{\leq}) \cup \{x_i^+ \in \Delta^+ : l_{i+1} p(x_i^+) \leq n\}$

1. If $n < n'$: If $\exists x_i^+ \in \Delta_{n>}^+$ such that $l_{i+1} p(x_i^+) \leq n < n'$, then $\{x_i^+ \in \Delta_{n>}^+ : l_{i+1} p(x_i^+) \leq n\} = \{\bigcup_{i=1}^{k'} x_i^+ : x_i^+ \in \Delta_{n>}^+ \text{ and } l_{i+1} p(x_i^+) \leq n\}$. Else, if $\forall x_i^+ \in \Delta_{n>}^+$ we have $l_{i+1} p(x_i^+) > n$ then $\{x \in \Delta_{n>}^+ : l_{i+1} p(x_i^+) \leq n\} = \emptyset$, thus $\overline{P}(Cl_{i+1,n}^{\leq}) = \overline{P}(Cl_{i,n}^{\leq})$.
2. If $n \geq n'$, then $\{x_i^+ \in (\Delta_{n<}^+ \cup \Delta_{n=}^+) : l_{i+1} p(x_i^+) \leq n\} = \{\bigcup_{i=1}^{k'} x_i^+ : x_i^+ \in (\Delta_{n<}^+ \cup \Delta_{n=}^+) \text{ and } l_{i+1} p(x_i^+) \leq n\}$ and $\{x \in A_i : D_{i+1}^+ p(x) \cap Cl_{i+1,n}^{\leq} \neq \emptyset\} = \overline{P}(Cl_{i,n}^{\leq}) \cup \{x \in A_i \cap D_{i+1}^- p(x_i^+) : D_{i+1}^+ p(x) \cap Cl_{i+1,n}^{\leq} \neq \emptyset\}$. Thus, we have $\{x \in A_i \cap D_{i+1}^- p(x_i^+) : D_{i+1}^+ p(x) \cap Cl_{i+1,n}^{\leq} \neq \emptyset\} \cup \{\bigcup_{i=1}^{k''} x_i^+ : x_i^+ \in \Delta_{n<}^+ \cup \Delta_{n=}^+\} = \{\bigcup_{i=1}^{k''} D_{i+1}^- p(x_i^+)\}$, thus $\overline{P}(Cl_{i+1,n}^{\leq}) = \overline{P}(Cl_{i,n}^{\leq}) \cup \bigcup_{i=1}^{k'} D_{i+1}^- p(x_i^+), \forall x_i^+ \in (\Delta_{n<}^+ \cup \Delta_{n=}^+)$.

The same reasoning holds for the proposition 4.