J Intell Inf Syst (2014) 42:207–232
DOI 10.1007/s10844-013-0291-2

Hierarchical object-driven action rules

Ayman Hajja · Zbigniew W. Raś ·
Alicja A. Wieczorkowska

Received: 9 May 2013 / Revised: 6 October 2013 / Accepted: 19 November 2013 /
Published online: 12 December 2013

Abstract In this work, we present the hierarchical object-driven action rules; a hybrid action rule extraction approach that combines key elements from both the classical action rule mining approach, first proposed by Raś and Wieczorkowska (2000), and the more recent object-driven action rule extraction approach proposed by Hajja et al. (2012, 2013), to extract action rules from object-driven information systems. Action rules, as defined in Raś and Wieczorkowska (2000), are actionable tasks that describe possible transitions of instances from one state to another with respect to a distinguished attribute, called the decision attribute. Recently, a new specialized case of action rules, namely object-driven action rules, has been introduced by Hajja et al. (2012, 2013). Object-driven action rules are action rules that are extracted from information systems with temporal and object-based nature. By object-driven information systems, we mean systems that contain multiple observations for each object, in which objects are determined by an attribute that assumingly defines some unique distribution; and by temporally-based information systems, we refer to systems in which each instance is attached to a timestamp that, by definition, must have an intrinsic meaning for each corresponding instance. Though the notion of object-driven and temporal-based action rules had its own successes, some argue that the essence of object-driven assumptions, which is in big part the reason for its effectiveness, are imposing few

A. Hajja (✉) · Z. W. Raś
Department of Computer Science, University of North Carolina,
9201 University City Blvd., Charlotte, NC 28223, USA
e-mail: ahajja@uncc.edu

Z. W. Raś
Institute of Computer Science, Warsaw University of Technology,
Nowowiejska 15/19, 00-665 Warsaw, Poland
e-mail: ras@uncc.edu

A. A. Wieczorkowska
Polish-Japanese Institute of Information Technology,
Koszykowa 86, 02-008 Warsaw, Poland
e-mail: alicja@poljap.edu.pl

limitations as well. Object-driven approaches treat entire systems as multi-subsystems for which action rules are extracted from; as a result, more accurate and specific action rules are extracted. However, by doing so, our diverseness of the extracted action rules are much less apparent, compared to the outcome when applying the classical action rule extraction approach, which treats information systems as a whole. For that reason, we propose a hybrid approach which builds a hierarchy of clusters of subsystems; a novel way of clustering through treatments responses similarities is introduced.

Keywords Association rules · Action rules · Grouping · Clustering · Hierarchy · Generalization

1 Introduction

In recent work by Hajja et al. (2012, 2013), the authors proposed a specialized approach for action rules extraction. Ideally, the motivation to propose a specialized action rules extraction system was to extract actionable tasks from information systems of a specialized nature. In Hajja et al. (2012), the authors addressed information systems that have temporal and object-based nature. Temporal-based and object-driven information systems are commonly found in healthcare and economics domains; an example of both an object-driven and temporal-based system are electronic medical records containing multiple observations about various visits of patients, where for each unique patient (representing an object), multiple visits are recorded, and where for each visit (representing an instance) a timestamp is associated with the instance; also, an example of a temporal-based information system is the stock market or foreign exchange prices, where the goal is to detect patterns of price changes. In Hajja et al. (2013), the authors further investigate other approaches to extract action rules from temporal systems. The commonality of both object-driven and temporal-based information systems made us believe that our new proposed approach is highly called for, and it is an important addition to the literature of action rules.

As will be seen in proceeding sections, when dealing with object-driven information systems, the appropriate methodology to extract action rules is to divide the system into multiple independent subsystems that assumingly come from distinguished distributions. Despite the fact that by doing so, the extracted action rules will tend to be more accurate, one must acknowledge that in some cases, mainly when our information system does not contain many observation for unique objects, the extracted action rules may be lacking high degrees of diverseness, which is, admittedly, one of the potentially negative side-effects of treating our system as a collection of multiple independent subsystems.

In this work, we introduce a systematic approach to calculate a hierarchical similarity metric between objects in information systems. This will allow us to group objects that have high similarities; consequently, making it feasible to further extract action rules from objects in clusters rather than in isolation, which would in turn result in veritably stronger and more diverse collection of action rules. One of the fundamental motives for object-driven action rules however, is to avoid cross-object learning (Hajja et al. 2012, 2013); by cross-object learning, we mean extracting action rules from instances that come from different distributions. For that reason, our clustering approach introduced in this work must comply with that principle as well; in other words, objects that are similar to each other (according to our similarity metric introduced in Section 5) are fundamentally objects that come from similar distributions. The reasons behind using object behavior as a similarity metric will become rather more justifiable as the notion of action rules is fully explained. Having said that, it is worth stating here that action rules essentially describe some behaviors in term of others; which makes *object behavior* an intrinsically good choice to be used to measure similarities between objects; more details about the clustering phase is discussed in Section 5. Our choice of hierarchical clustering is strictly motivated by the flexibility it provides to system users; as will be shown in later sections, the level of hierarchy chosen by system users determines the degree of specialization (or generalization) prior to action rule extraction; for example, in certain domains such as healthcare, it is always necessary to be entirely confident of the outcome of actions undertaken to patients, which would mean that action rule generalization needs to be done attentively; although by generalizing rules, we tend to cover a wider group of instances (typically patients in this case), the fact that some instances will not behave according to the action rule will make generalization an improper action. Note that this is not the case in other domains; for example, in marketing, we may be interested in sending letters to potential clients to buy a particular product; by generalizing our action rules, we are more likely to target a larger audience, however, the percentage of this larger audience may decrease, though it would still be more profitable to do so. Next, we provide a brief introduction of action rules.

The notion of action rules was first proposed in Raś and Wieczorkowska (2000). The merit for the introduction of action rules was to provide a way for mining appropriate actionable tasks in which if applied, a desired transition would occur. From the knowledge discovery point of view, it is essential to extract and understand patterns in which relations are discovered between attributes of information systems. However, it is perhaps more essential to extract tasks (in the form of solutions) that provide insights about possible ways (or steps to be undertaken) for changing the state of a specific attribute; typically from an undesired state to a more desired one.

Although strongly related, association rules and action rules are two, essentially different data discovery tools; they provide system users with fundamentally different insights about the data, hence demanding different responses. Association rules provide discoveries about relations (or associations) between segments of the data; for example, a useful association rule might imply that most young clients are unhappy with the service, and therefore are more likely to stop using the service. Losing clients is one of the major fears for any company, which makes this information vitally important. However, note that for a company to mend that issue, more analysis is required; it is not clear what caused young clients the lack of satisfaction. Action rules on the other hand provide system users with actionable (or applicable) tasks that can be undertaken to make the disloyal young clients happier; the passive nature of association rules versus the active nature of action rules is a key reason for the popularity of action rules, hence its commonality in diversified fields.

Action rules extraction has been applied in numerous domain areas. Originally in Raś and Wieczorkowska (2000), action rules were intended to be applied in the business field, in areas of the banking industry such as market modeling, risk optimization, and fraud detection. However, and not so long after the introduction of action rules, applications in areas of healthcare domain became one of the major and dominant desired applications of action rules; in Zhang et al. (2010), action rules was applied for tinnitus (or "ringing in the ear") treatment to help discover interesting patterns to understand treatment factors with regard to patients recovery; also, as will be seen throughout this note, action rules were used in Hajja et al. (2012, 2013) to improve the condition of hypernasality speech disorder patients. Other major area of applications is music automatic indexing and retrieval (Raś and Wieczorkowska 2010; Raś and Dardzinska 2011).

In the classical action rule extraction approach proposed by Ras and Wieczorkowska (2000), the goal was to extract actionable patterns from traditional information systems, under the underlying assumption that instances are observed independently and identically from a common distribution. That said, it is not uncommon to come across information systems that follow different assumptions. For example, a temporal information system, in which each instance is associated with a timestamp, for which timestamps have a significant meaning, is an information system that contains instances that are not independent, nor can be treated identically; thus, to extract action rules from such untraditional systems, a new approach need to be proposed. Throughout this paper, a detailed discussion of two untraditional information systems is provided, and an elaborate explanation of how action rule extraction applies to them is presented; the first one of the two is the temporal information system, which can be briefly defined as an information system in which each instance is associated with a timestamp, which implicitly indicates a significant meaning; the second one is the object-driven information system, succinctly defined as an information system that contains an attribute that indicates that some groups of instances belong to particular distributions. Though domain-specific utilization of temporal information have been used in action rules literature in previous work, such as in Zhang et al. (2010); the fact is, until the introduction of the work presented in this note, there has been no systematic approach of which a temporal action rule extraction can be applied structurally to domain-independent applications.

2 Classical action rules

The goal of action rules extraction is to describe possible transitions of objects from one state to another with respect to a specific attribute, called the decision attribute. By doing so, system users will be provided with actionable tasks that, when applied, will result in a shift in the decision attribute; ideally, from a less desired state to a more desired one.

Let $S = (X, A, V)$ denotes an information system (Pawlak 1981), where:

1. X is a nonempty, finite set of instances,
2. A is a nonempty, finite set of attributes; $a : X \rightarrow V_a$ is a function for any $a \in A$, where V_a is called the domain of a,
3. $V = \bigcup\{V_a : a \in A\}$.

Information systems are commonly represented in the form of *attribute-value table*, where each row in the table represents one complete record of instance $x \in U$, and where each column represents states for one attribute $a \in A$ for all $x \in U$.

Table 1 shows an example of an information system S_1. Despite the fact that elements of X are sometimes referred to as objects, in this paper we will not use the two terms interchangeably, objects will possibly consist of, as will be discussed in future sections, multiple instances. The distinction between instances and objects is necessary for the understanding of the research presented in this work.

For example, Table 1 shows an information system S_1 with a set of instances $X = \{x_0, x_1, x_2, x_3, x_4, x_5, x_6\}$, set of attributes $A = \{e, f, g, d\}$, and a set of their values $V = \{e_1, e_2, f_1, f_2, g_1, g_2, d_1, d_2\}$. Each row in Table 1 shows one complete observation about its corresponding instance; the first row for example, shows values for instance x_0; its state for attribute e is e_1, which would also be denoted by the expression (e, e_1); attribute f has state f_1, or (f, f_1); attribute g has state g_1, or (g, g_1); and the state of attribute d is d_1, or (d, d_1).

Table 1 Information System S_1

	e	f	g	d
x_0	e_1	f_1	g_1	d_1
x_1	e_1	f_1	g_1	d_1
x_2	e_2	f_2	g_2	d_2
x_3	e_1	f_2	g_1	d_1
x_4	e_2	f_1	g_1	d_2
x_5	e_1	f_2	g_1	d_2
x_6	e_2	f_2	g_1	d_1

By a decision system (table), we mean an information system that makes a clear explicit distinction between attributes in A, and will therefore label each attribute as either a *decision attribute*, or a non-decision attribute, called *condition attribute*. The decision attribute(s), which normally but not necessarily is a single attribute, is the attribute that we are interested in the most. For system users, the eventual goal would be to change the decision attribute from a less desirable to a more desirable state. For example, a company would be interested in moving clients' states of loyalty from lower to higher.

All non-decision, or condition, attributes are further partitioned into two mutually exclusive sets; the first one is the *stable* attributes set, and the second one is the *flexible* attributes set. By stable attributes set we mean the set that contains attributes that we have no control over; their values cannot be changed by the users of our system. An example of a *stable* attribute is the place where the person was born. On the other hand, values of *flexible* attributes can be influenced and changed; an example of a *flexible* attribute is the patient's prescribed medication. In this paper, A_{St}, A_{Fl}, and $\{d\}$ will represent the set of stable attributes, the set of flexible attributes, and the decision attribute, respectively. Hence, the set of attributes A can be redefined as $A = A_{St} \cup A_{Fl} \cup \{d\}$.

An *atomic action set* is an expression that defines a change of state for a single distinct attribute. For example, $(e, e_1 \rightarrow e_2)$ is an atomic action set which defines a change of state for the attribute e from e_1 to e_2, where $e_1, e_2 \in V_e$. Clearly in this case, the attribute e is a flexible attribute, since it changes its state from e_1 to e_2. In the case when there is no change, we omit the right arrow sign, so for example, (g, g_1) means that the value of attribute g is g_1 and remains g_1, where $g_1 \in V_g$.

The *action set* is defined as follows:

1. If t is an atomic action set, then t is an action set.
2. If t_1, t_2 are action sets and $\wedge$ is a 2-argument functor called composition, then $t_1 \wedge t_2$ is a candidate action set.
3. If t is a candidate action set and for any two atomic action sets $(e, e_1 \rightarrow e_2)$, $(f, f_1 \rightarrow f_2)$ contained in t we have $e \neq f$, then t is an action set.
4. No other sets are called action sets.

The *domain* $Dom(t)$ of an action set t is the set of attributes of all atomic action sets contained in t. For example, $t = (e, e_1 \rightarrow e_2) \wedge (g, g_1)$ is an action set that consists of two atomic action sets, namely $(e, e_1 \rightarrow e_2)$ and (g, g_1). Therefore, the domain of t is $\{e, g\}$.

Action rules are expressions that take the following form: $r = [t_1 \Rightarrow t_2]$, where t_1, t_2 are action sets. The interpretation of the action rule r is that by applying the action set t_1, we would get, as a result, the changes of states in action set t_2. We also assume that $Dom(t_1) \cup Dom(t_2) \subseteq A$, and $Dom(t_1) \cap Dom(t_2) = \phi$.

For example, $r = [[(e, e_1 \rightarrow e_2) \wedge (g, g_2)] \Rightarrow (d, d_1 \rightarrow d_2)]$ means that by changing the state of the attribute e from e_1 to e_2, and by keeping the state of the attribute g as g_2, we would observe a change in the attribute d from the state d_1 to d_2, where d is commonly referred to as the *decision attribute*.

Standard interpretation N_s of action sets in S is defined as follows:

1. If $(a, a_1 \rightarrow a_2)$ is an atomic action set, then $N_s((a, a_1 \rightarrow a_2)) = [\{x \in X : a(x) = a_1\}, \{x \in X : a(x) = a_2\}]$.
2. If $t_1 = (a, a_1 \rightarrow a_2) \wedge t$ and $N_s(t) = [Y_1, Y_2]$, then $N_s(t_1) = [Y_1 \cap \{x \in X : a(x) = a_1\}, Y_2 \cap \{x \in X : a(x) = a_2\}]$.

Let us define $[Y_1, Y_2] \cap [Z_1, Z_2]$ as $[Y_1 \cap Z_1, Y_2 \cap Z_2]$ and assume that $N_s(t_1) = [Y_1, Y_2]$ and $N_s(t_2) = [Z_1, Z_2]$. Then, $N_s(t_1 \wedge t_2) = N_s(t_1) \cap N_s(t_2)$.

If t is an action set and $N_s(t) = [Y_1, Y_2]$, then the support of t in S is defined as $supp(t) = min\{card(Y_1), card(Y_2)\}$.

Let $r = [t_1 \Rightarrow t_2]$ be an action rule, $supp(t_1) > 0$, $N_s(t_1) = [Y_1, Y_2]$, and $N_s(t_2) = [Z_1, Z_2]$. Support $supp(r)$ and confidence $conf(r)$ of r are defined as:

$$supp(r) = min\{card(Y_1 \cap Z_1), card(Y_2 \cap Z_2)\},$$

$$conf(r) = \left[\frac{card(Y_1 \cap Z_1)}{card(Y_1)}\right] * \left[\frac{card(Y_2 \cap Z_2)}{card(Y_2)}\right].$$

3 Association action rules

There has been considerable research on the varied methodologies for extracting action rules from information systems. In general however, we can categorize all methodologies into two groups; the first one being when classification rules are required for the construction of action rules (Raś et al. 2005, 2008; Raś and Wieczorkowska 2000; Tsay and Raś 2006), and the second, more recent approach, being when action rules are directly extracted from an information system (Raś et al. 2008). To extract action rules, we used the algorithm described in Raś et al. (2008). The idea of the algorithm is to start by constructing all possible action sets that have occurred more than a pre-defined number, called the minimum support. Then, in accordance to our desired change in the decision attribute, action rules are formed.

Let t_a be an action set, where $N_s(t_a) = [Y_1, Y_2]$ and $a \in A$. We say that t_a is a *frequent action set* (Raś et al. 2008) if $card(Y_1) \geq \lambda_1$ and $card(Y_2) \geq \lambda_1$, where λ_1 is the minimum support. Another way of interpreting the frequent action sets would be that all frequent action sets have support greater than or equal to the minimum support λ_1. By specifying λ_1, we make sure that the extracted action rules have support greater than or equal to the minimum support λ_1. The algorithm presented below is similar to Agrawal et al. (1993). To extract action rules, we start by generating atomic action sets that have support greater than or equal to the minimum support value λ_1 pre-defined by the user; we will refer to this set as *1-element frequent action set*. The term *frequent* will be used to indicate that an action set has support greater than or equal to the minimum support, and the term *k-element* will be used to indicate the number of elements (or atomic action terms) in an action set. Both *frequent atomic action set* and *1-element frequent action set* refer to exactly the same set, since from the definition of atomic action sets, they consist of only one element. After generating all frequent atomic action sets, we undertake the following two-step process initially for $k = 1$:

1. **M**erge step: Merge pairs (t_1, t_2) of k-element action sets into all $(k + 1)$-element candidate action sets.
2. **D**elete step: Delete all $(k + 1)$-element candidate action sets that are either not action sets, or contain a non-frequent k-element action set, or that have support less than the minimum support λ_1.

We keep iterating the above two steps until we cannot generate new frequent action sets anymore. At this point, we have generated all $(k + 1)$-element frequent action sets, which will allow us to generate action rules that are guaranteed to have support greater than or equal to the minimum support λ_1. Last step is to further filter the desired action rules based on their confidence, where we only consider action rules with confidence greater than or equal to a pre-defined minimum confidence λ_2. For example, from the frequent action set $t_1 = (a, a_1 \to a_2) \wedge (d, d_1 \to d_2)$, we can generate the following two action rules:

1. $r_1 = [(a, a_1 \to a_2) \Rightarrow (d, d_1 \to d_2)]$.
2. $r_2 = [(d, d_1 \to d_2) \Rightarrow (a, a_1 \to a_2)]$.

where both r_1 and r_2 have support greater than or equal to the minimum support λ_1. However, we will only be interested in specific changes of the decision attribute, e.g., in changing the decision attribute d from state d_1 to d_2. Therefore, we will only consider r_1.

4 Object-based action rules

The drive behind the introduction of object-driven action rules in Hajja et al. (2012) was to bring forth an adapted approach to extract action rules from systems of temporal and object-driven nature. In Hajja et al. (2012), the authors proposed an object-independency assumption approach that suggests extracting patterns from subsystems defined by unique objects, and then aggregating similar patterns amongst all objects. The motivation behind this approach is based on the fact that same-object observations share similar features that are not shared with other objects, and these features are possibly not explicitly included in our dataset. Therefore, by individualizing objects prior to calculating action rules, variance is reduced, and over-fitting is potentially avoided. In addition to the object independency assumption, temporal information is exploited by taking into account only the state transitions that occurred in the valid direction. In the next subsection, we limit our discussion to the object independency assumption, and in subsequent subsections we discuss the temporal assumption.

4.1 Object-driven assumption

Information systems with object-driven nature are distinct from the classical information system, for that they have an attribute which we call the *object attribute*. Prior to illustrating object-driven information systems with an example, we provide three main conditions that must be satisfied in any object attribute: 1) for each distinct state of an object attribute, there must be multiple observations, 2) the possible values for an object attribute should be arbitrarily assigned; meaning that it is both unique, and does not have any meaning associated with the object; a good example of an object attribute would be the *employee number*, and 3) all observations for a particular object attribute state should share the same distribution;

this essentially means that instances observed from the same object must have similar characteristics that are not necessarily provided in the information system; so although *patient (or employee) number* do not have any significant meaning associated to the patient (or employee), they still imply that all instances with identical *number* were observed from the same person; hence, share common characteristics.

Let **O** be the set of object ID's which instances belong to X. We define object-driven action rules to be rules that are extracted from a subsystem $S_p = (X_p, A, V)$ of S, where X_p contains all instances in X of the object p, $p \in \mathbf{O}$.

Here, we provide an example to demonstrate the advantages of object-driven action rules extraction, using the information system S_2 shown in Table 2. As a company, we are interested in extracting action rules that change the state of attribute *Loyalty* from *Low* to *High*. We have two objects with 4 instances each.

Using the classical action rules extraction method, two action rules will be extracted from the system:

1. $r_1 = [[(Income, Low \rightarrow Medium) \wedge (Children, \leq 3)] \Rightarrow (Loyalty, Low \rightarrow High)]$; $conf(r_1) = 100\,\%$.
2. $r_2 = [[(Income, Medium \rightarrow High) \wedge (Children, > 3)] \Rightarrow (Loyalty, Low \rightarrow High)]$; $conf(r_2) = 100\,\%$.

Now let us assume that the attribute *Children* is missing in S_2. The following action rules will be extracted instead:

1. $\wp_1 = [(Income, Low \rightarrow Medium) \Rightarrow (Loyalty, Low \rightarrow High)]$;
   $conf(\wp_1) = 50\,\%$.
2. $\wp_2 = [(Income, Low \rightarrow High) \Rightarrow (Loyalty, Low \rightarrow High)]$;
   $conf(\wp_2) = 100\,\%$.
3. $\wp_3 = (Income, Medium) \Rightarrow (Loyalty, Low \rightarrow High)$;
   $conf(\wp_3) = 25\,\%$.
4. $\wp_4 = [(Income, Medium \rightarrow High) \Rightarrow (Loyalty, Low \rightarrow High)]$;
   $conf(\wp_4) = 50\,\%$.

We can observe that the rules $\wp_1, \wp_4$ are weaker than r_1, r_2 and also the condition and decision part in both rules $\wp_2$ and $\wp_3$ are referring to different objects, so we should not consider them valid. By using object-driven action rules extraction, we would not get $\wp_2$ and $\wp_3$.

Table 2 Information system S_2

	ObjectID	Observation	Loyalty	Income	Children
x_0	1	1	High	High	More than 3
x_1	1	2	High	High	More than 3
x_2	1	3	Low	Medium	More than 3
x_3	1	4	Low	Medium	More than 3
x_4	2	1	High	Medium	Less than or equal to 3
x_5	2	2	High	Medium	Less than or equal to 3
x_6	2	3	Low	Low	Less than or equal to 3
x_7	2	4	Low	Low	Less than or equal to 3

Each one of the two employees was observed four different times

Coming back to the action sets, we define the p^{th} standard interpretation $N_{s(p)}$, where p is the object's unique ID, of action sets in $S = (X, A, V)$ as follows:

1. If $(a, a_1 \rightarrow a_2)$ is an atomic action set, then

$$N_{s(p)}((a, a_1 \rightarrow a_2)) = [\{x \in X_p : a(x) = a_1\}, \{x \in X_p : a(x) = a_2\}] ,$$

   where X_p is the set of all instances of the p^{th} object.
2. If $t_1 = (a, a_1 \rightarrow a_2) \wedge t$ and $N_{s(p)}(t) = [Y_1, Y_2]$, then

$$N_{s(p)}(t_1) = [Y_1 \cap \{x \in X_p : a(x) = a_1\}, Y_2 \cap \{x \in X_p : a(x) = a_2\}] ,$$

   where X_p is the set of all instances of the p^{th} object.

4.2 Classical temporal-driven approach

In this section, we show a systematic approach for extracting object-driven action rules using the three previously mentioned assumptions. We will base our algorithm on association action rules (Raś et al. 2008) described in Section 3.

Here, we make the assumption that the only valid change of attribute value, is the change that happens between two instances of the same object. Accordingly, the standard interpretation N_s^{TC} that complies with the temporal constraint of an action set in $S = (X, A, V)$ is redefined as follows:

1. If $(a, a_1 \rightarrow a_2)$ is an atomic action set, then

$$N_s^{TC}((a, a_1 \rightarrow a_2)) = [\{x_1 \in X : a(x_1) = a_1\}, \{x_2 \in X : a(x_2) = a_2\}] = [X_1, X_2],$$

   where for each $x_1 \in X_1$ there exist $x_2 \in X_2$ such that instance x_2 occurred after x_1, and there are no other instances in X_2 that do not have a matching instance in X_1.
2. If $t_1 = (a, a_1 \rightarrow a_2) \wedge t$ and $N_s^{TC}(t) = [Y_1, Y_2]$, then

$$N_s^{TC}(t_1) = [Y_1 \cap \{x_1 \in X : a(x_1) = a_1\}, Y_2 \cap \{x_2 \in X : a(x_2) = a_2\}] = [X_1, X_2],$$

   where for each $x_1 \in X_1$ there exist $x_2 \in X_2$ such that instance x_2 occurred after x_1, and there are no other instances in X_2 that do not have a matching instance in X_1.

For example, if we consider the following action rule: $t = (Income, Medium \rightarrow Low) \Rightarrow (Loyalty, High \rightarrow Low)$ for only $(ObjectID, 2)$, we can observe that the *standard interpretation* $N_s^{TC}(t)$ is $[\{x_4, x_5\}, \{x_6, x_7\}]$; note here that there is no condition that states that the cardinality of X_1 and X_2 must be equal; if for example, the instance x_4 did not exist, we would still have $X_2 = \{x_6, x_7\}$ since both x_6 and x_7 occurred after x_5; also, if x_6 did not exist, we would still have $X_1 = \{x_4, x_5\}$ since both x_4 and x_5 occurred before x_7. The definition of support of an action set and the definitions of support and confidence of an action rule are all the same as in the previous subsection.

Following both the definition of the **T**emporal **C**onstraint standard interpretation N_s^{TC} and the p^{th} standard interpretation $N_{s(p)}$, it becomes apparent that the definition of the p^{th} standard interpretation that complies with the Temporal Constraint, $N_{s(p)}^{TC}$, where p is the object's unique ID (in the following sections, p will be called an object), of action sets in $S = (X, A, V)$ can be defined as follows:

1. If $(a, a_1 \to a_2)$ is an atomic action set, then
$$N^{TC}_{s(p)}((a, a_1 \to a_2)) = [\{x_1 \in I_{s(p)} : a(x_1) = a_1\}, \{x_2 \in I_{s(p)} : a(x_2) = a_2\}] = [X^1_p, X^2_p],$$
where $\forall x_1 \exists x_2$ such that x_2 is after x_1, and $I_{s(p)}$ is the set of all instances for the p^{th} object. Additionally, we assume that there are no other instances in X^2_p.
2. If $t_1 = (a, a_1 \to a_2) \wedge t$ and $N^{TC}_{s(p)}(t) = [Y_1, Y_2]$, then
$$N^{TC}_{s(p)}(t_1) = [Y_1 \cap \{x_1 \in I_{s(p)} : a(x_1) = a_a\}, Y_2 \cap \{x_2 \in I_{s(p)} : a(x_2) = a_2\}] = [X^1_p, X^2_p],$$
where $\forall x_1 \exists x_2$ such that x_2 is after x_1, and $I_{s(p)}$ is the set of all instances for the p^{th} object. Additionally, we assume that there are no other instances in X^2_p.

If t is an action set and $N^{TC}_{s(p)}(t) = \{Y_1, Y_2\}$, then the support of t in S is defined as: $supp^{TC}_p(t) = min\{card(Y_1), card(Y_2)\}$.

Let $r = [t_1 \Rightarrow t_2]$ be an action rule, where $N^{TC}_{s(p)}(t_1) = [Y_1, Y_2]$, $N^{TC}_{s(p)}(t_2) = [Z_1, Z_2]$. The p^{th} support $supp^{TC}_p(r)$ and the p^{th} confidence $conf^{TC}_p(r)$ of r are defined as follows:

$$supp^{TC}_p(r) = min\{card(Y_1 \cap Z_1), card(Y_2 \cap Z_2)\},$$

$$conf^{TC}_p(r) = \left[\frac{card(Y_1 \cap Z_1)}{card(Y_1)}\right] \cdot \left[\frac{card(Y_2 \cap Z_2)}{card(Y_2)}\right].$$

Now, assume that by O we mean the set of all objects' unique IDs. After all objects-driven action rules are extracted and their p^{th} support and p^{th} confidence are computed for each $p \in \mathbf{O}$, we then calculate their total support $supp^{TC}_{\mathbf{O}}(r)$ (called support) and total confidence $conf^{TC}_{\mathbf{O}}(r)$ (called confidence) following the definitions below:

$$supp^{TC}_{\mathbf{O}}(r) = \sum_{p \in \mathbf{0}} supp^{TC}_p(r),$$

$$conf^{TC}_{\mathbf{O}}(r) = \sum_{p \in \mathbf{0}} \left(\frac{supp^{TC}_p(r) \cdot conf^{TC}_p(r)}{supp^{TC}_{\mathbf{O}}(r)}\right).$$

4.3 Pair-based approach

As defined previously, temporal object-driven datasets consist of numerous unique objects, where each object is comprised of multiple instances that have assigned corresponding timestamps. Previously in Hajja et al. (2012); as explained in Subsection 4.2, the object p based standard interpretation of an action set $t = (a, a_1 \to a_2)$ was defined as the pair of two sets $[Y_1, Y_2]$ where Y_1 is the set of instances of the object p that satisfy the left side, or condition side, of the action set, and Y_2 is the set of instances of the object p that satisfy the right side, or decision side, of the action set, with the addition that for every instance in Y_1, there exist a matching instance in Y_2 that occurred after it. This definition resembles the definition of standard interpretation for classical action rules while restricting valid transitions to only one direction. In this section however, we present an approach in which the nature of the object-driven temporal dataset allows us to redefine the standard interpretation into a more intuitive pair-based structure which we believe is more appropriate for object-driven temporal systems.

Let us first assume that $I_s(p)$ denotes the set of all instances of the object p in an information system S. Also, the relation $\angle \subseteq I_s(p)$ is defined as:

$(x_1, x_2) \in \angle$ iff x_2 has occurred after x_1.

The pair-based standard interpretation $N^{TC}_{s(p)}$ in $S = (X, A, V)$ for an object p is redefined as:

1. If $(a, a_1 \rightarrow a_2)$ is an atomic action set, then $N^{TC}_{s(p)}((a, a_1 \rightarrow a_2)) = \{(x_1, x_2) \in \angle : a(x_1) = a_1, a(x_2) = a_2\}$
   where $\angle \subset I_s(p)$.
2. If $t_1 = (a, a_1 \rightarrow a_2)\wedge$ and $N^{TC}_{s(p)}(t) = Y_1$, then $N^{TC}_{s(p)}(t_1) = Y_1 \cap \{(x_1, x_2) \in \angle : a(x_1) = a_1, a(x_2) = a_2\}$
   where $\angle \subset I_s(p)$.

In other words, our standard interpretation will consist of all valid transitions from the left side of an action set to the right side, represented as pairs. The motivation behind this new interpretation is due to the fact that the instances within one object are not observed independently, which will allow us to relax the minimum assumption previously used. Our object-independency assumption states that the whole system consists of multiple independent subsystems, each one marked by a unique object. Although it confines the system to extract action rules only from instances of the same object, it provides more flexibility to be applied within unique objects.

If t is an action set and $N^{TC}_{s(p)}(t) = Y_1$, then the support of t in S is defined as: $supp^{TC}_p = card(Y_1)$.

Let $r = [t_1 \Rightarrow t_2]$ be an action rule, where $N^{TC}_{s(p)}(t_1) = Y_1$, $N^{TC}_{s(p)}(t_2) = Y_2$. The p^{th} support $supp^{TC}_p(r)$ and the p^{th} confidence $conf^{TC}_p(r)$ of r are defined as follows:

$$supp^{TC}_p(r) = card(Y_1 \cap Y_2),$$

$$conf^{TC}_p(r) = \left[\frac{card(Y_1 \cap Y_2)}{card(Y_d)}\right].$$

To define Y_d, let us first assume that $\zeta(Y)$ denotes the set of first elements of the set of pairs Y. For instance, if $Y = \{(x_1, x_3), (x_3, x_4), (x_1, x_2)\}$, then $\zeta(Y) = \{x_1, x_3\}$. We define $Y_d = \{(x_1, x_2) \in Y_1 : x_1 \in \zeta(Y_2)\}$. The interpretation of this definition means that to calculate the confidence of the action rule $r = (a, a_1 \rightarrow a_2) \Rightarrow (d, d_1 \rightarrow d_2)$, the pairs that we are considering are the ones that have first elements that satisfy $a = a_1$ and $d = d_1$. Since the transition from a_1 to a_2 could possibly trigger other states of decision attribute d, we are only interested in the states of our action rule.

After all object-driven action rules are extracted and their p^{th} support and p^{th} confidence are computed for all $p \in O$, we then calculate their total support $supp^{TC}_O(r)$ (called support) and total confidence $conf^{TC}_O(r)$ (called confidence) following the definition below:

$$supp^{TC}_O(r) = \sum_{p \in O} supp^{TC}_p(r) \ ,$$

$$conf^{TC}_O(r) = \sum_{p \in O} \left(\frac{supp^{TC}_p(r) * conf^{TC}_p(r)}{supp^{TC}_O(r)}\right).$$

If the denominator in the formula for calculating confidence is equal to zero, then the confidence is equal to zero by definition.

5 Hierarchical object-based action rules

In previous sections, we discussed the advantages of limiting action rule extraction to the level of objects. The motivation was to extract more accurate action rules that better reflect real world cases, and to provide a more accurate actual representation of the observed information system. However, we also mentioned that specialization could cause over-fitting problems, and for that reason, it is the case that in some situations, generalization may be preferred; mainly when our number of observations within unique objects is limited. In this section, we provide a hybrid approach that combines both the specialized object-driven approach, and the general classical action rules approach. The hierarchy that we build will provide system users with tremendous convenience; allowing decision makers to take decision on how much generalization/specialization to be applied, by simply examining the dendrogram and specifying the proper level of specialization.

While there are many advantages for having a pure specialized object-driven approach to extract action rules, it is clear that by having more objects in our system, the number of extracted action rules, and their corresponding supports will automatically be negatively affected. Following the non-temporal pair-based action rule extraction (Subsection 4.3); for an object that consists of n number of instances, the maximum support (number of supporting pairs that could exist) is $(n/2)^2$, and that is only the case when half of our instances satisfy the preconditions of the action rule, and the other half satisfy the postcondition of the action rule; it would become apparent then, that by dividing our information system to multiple subsystems, the total support for our action rules will eventually decrease. For example, if we extract action rules from a system of 100 instances, the maximum support for a particular action rule would be 50^2, which is 2500; however, if we divide our 100-instance system into ten 10-element subsystems, the maximum number of support for a particular action rule would be $10 * 5^2$, which is 250, only one tenth the possible action rules that could be extracted by not using the object-driven approach; more accurately, the maximum support for a potential action rule is inversely proportional to the number of objects (or clusters) in our information system.

Table 3 shows the number of action rules, and the total support for our hypernasality disorder dataset described in the following section. The first column in Table 3 denotes the decision attribute of the action rules extracted from our system; the second column shows the total number of action rules extracted that trigger the corresponding decision shift from the first column, note that in the second column, we are treating each subsystem, identified by the object attribute, as an independent system. The third column on the other hand shows the number of action rules extracted, again, with respect to the corresponding decision shift from the first column, but after clustering objects of patients, forming 40 clusters; before clustering patients together, we are essentially extracting action rules from each of the 225 objects (or patients) independently; however, when we cluster objects together, we are combining objects (or patients), hence resulting in less subsystems, but with more instances in each subsystem (or cluster). Necessary details about our information system will be provided in following sections; however, we can still observe that when more objects (in this case patients) are combined, forming less independent subsystems, the more unique action rules get extracted. Combining patients (or objects) to form bigger groups of patients occurs through a clustering procedure where a similarity metric is proposed, as will be seen next.

Table 3 Total number of action rules with respect to number of clusters

decision shift	Number of Action Rules		Total Support	
	object-driven	40 clusters	object-driven	40 clusters
(2.5 → 2)	1608	35937	4456	354332
(2.5 → 1)	700	31467	1580	304808
(2 → 1.5)	112	12972	224	81072
(2 → 1)	480	25728	960	223753
(2 → 0)	14	91133	28	931985
(1.5 → 1)	388	10054	776	66874
(1 → 0.5)	96	59927	200	497465
(1 → 0)	954	85769	1996	755361

Also, the fourth column shows values for the total support for all action rules instead of the actual number of action rules; similarly, we can observe a pattern of changes, in which the more objects we combine, the higher support for action rules we get.

To address the limitations that may occur when extracting action rules from pure object-driven approach, we present a new hierarchical approach in which we cluster objects that react similarly to particular treatments using the minimum (or single-linkage) clustering criterion; this approach, which we will refer to as the hybrid hierarchical approach, can be regarded as the product of combining both the classical action rule extraction approach and the object-driven approach. The goal thus become to examine objects and find groups of objects that we would expect to react similarly when a set of actions were to be performed; in other words, react similarly to particular action rules. Our assumption that we base this work upon is that objects that share high similarities with respect to actions, tend to share same reactions with other unexamined actions. For example, if two patients react similarly to ten shared treatments, then one would be relatively confident that they share similar responses, hence are more likely to respond similarly to other actions as well; in this section, we provide a detailed explanation of the hybrid hierarchical action rule extraction approach. The idea proposed here is to use our existing action rule extraction system to learn similarities in behavior found in objects (or patients). As mentioned earlier, the goal and outcome of action rule extraction is to bring forth actionable patterns that describe reactions that will occur as a result of other performed actions; in other words, by performing some actions, others (in which we do not have direct control over) will be triggered as a result. Typically, we would be only interested in specific shifts; for example, in our hypernasality treatment case study, we were interested in transitions that shift the state of patient from more severe to less severe. It would be an act of bias however, if we cluster patients by only measuring similarities with respect to one direction of shift in our decision attribute. For that reason, to measure similarities, hence clusters, using an unbiased behavioral approach, it would only be fair to measure similarities in behavior in all directions. Next we provide steps to accomplish that.

Note that when objects are combined together forming clusters, it would be purposeless to consider the temporal aspect in further computations. The reason for that goes back to the assumption that objects are independent entities; the fact that one instance from one object occurred before another instance from another object has no significant meaning whatsoever; for example, by knowing that one patient was given a particular drug before another patient went into surgery will not help us to make a better prediction of the surgery outcome, which is clearly not the case when dealing with one entity (or patient).

1. The first step would be to identify all possible transitions in our decision attribute. For example, if our decision attribute was d, where its possible values/states were listed in the following set: $\{d_1, d_2, d_3\}$, then the following set of atomic actions shows all possible transitions: $\{(d, d_1 \to d_2), (d, d_1 \to d_3), (d, d_2 \to d_1), (d, d_2 \to d_3), (d, d_3 \to d_1), (d, d_3 \to d_2)\}$
2. The second step would be to extract all action rules with respect to the transitions extracted in the first step; from every subsystem. We keep track of what subsystems support an action rule and what subsystems do not. In addition to that, we keep track of support and confidence for every action rule in every subsystem.
3. The third step would be to build a similarity matrix between every pair of objects, in which we measure the distance between the confidence and the relative support for every action rule in every subsystem. Since we are measuring similarities between objects that contain different number of observations; it is essential to use the relative support instead of the absolute support, defined as the support divided by the number of observation. The similarity between two objects o_1 and o_2 is defined as follows:

$$Similarity(o_1, o_2) = \sum_{r \in R} 2 - \left(|conf(o_1) - conf(o_2)| + \left| \frac{supp(o_1)}{E(o_1)} - \frac{supp(o_2)}{E(o_2)} \right| \right)$$

   Where R contains the set of all action rules shared by both o_1 and o_2, and where $E(o_1)$ and $E(o_2)$ denote the number of observations in object o_1 and o_2 respectively. The division by the number of observation will result in the relative support. For each two objects, all action rules similarities are computed then further aggregated; if the confidence and relative support are equal for one particular action rule, then the similarity will be 2 (for only that action rule), which is essentially the maximum similarity value. Note that in this formulation, we assume that the confidence and support have equal weights when calculating objects similarities; though we believe this is a fair assumption in our similarity formulation, adding a rate to set different weights is easily accomplished, and will not change the course of future computations.
4. Finally, we build a hierarchy of clusters using the minimum (or single-linkage) clustering criterion based on the similarity matrix built in step 3, hence producing a dendrogram.

Example to demonstrate the hybrid hierarchical object-driven approach Here, we provide an almost complete example by following the previous four steps to extract action rules using the hybrid hierarchical object-driven approach (Table 4).

The first step would be to identify all possible transitions in our decision attribute. In this example, and since we only have two states for our decision attribute d, the two decision transitions are $(d, d_1 \to d_2)$ and $(d, d_2 \to d_1)$.

Next, we extract all action rules with respect to all decision transitions from every subsystem; by applying the non-temporal pair-based approach. For the sake of simplicity, let us assume that attribute b is a stable attribute. The output of this second step would be:

The following action rules are extracted from *Object ID 1*:

$$(a, a_1 \to a_2) \Rightarrow (d, d_1 \to d_2);\ \text{support 2; confidence 50 \%}$$
$$(a, a_2 \to a_1) \Rightarrow (d, d_2 \to d_1);\ \text{support 2; confidence 100 \%}$$
$$(a, a_1 \to a_2) \wedge (b, b_2) \Rightarrow (d, d_1 \to d_2);\ \text{support 1; confidence 100 \%}$$
$$(a, a_2 \to a_1) \wedge (b, b_2) \Rightarrow (d, d_2 \to d_1);\ \text{support 1; confidence 100 \%}$$

Table 4 Information System S_3

	objectID	a	b	d
x_0	1	a_1	b_1	d_1
x_1	1	a_2	b_1	d_1
x_2	1	a_2	b_2	d_2
x_3	1	a_1	b_2	d_1
x_4	2	a_2	b_1	d_2
x_5	2	a_1	b_2	d_2
x_6	2	a_2	b_1	d_1
x_7	3	a_1	b_2	d_1
x_8	3	a_2	b_2	d_2
x_9	3	a_1	b_1	d_1
x_{10}	3	a_2	b_1	d_2
x_{11}	4	a_2	b_1	d_1
x_{12}	4	a_2	b_1	d_1
x_{13}	4	a_2	b_2	d_2
x_{14}	4	a_1	b_2	d_1
x_{15}	5	a_1	b_2	d_1
x_{16}	5	a_2	b_2	d_2
x_{17}	5	a_1	b_1	d_1
x_{18}	5	a_2	b_1	d_2

The following action rules are extracted from *Object ID 2*:

$$(a, a_1 \rightarrow a_2) \Rightarrow (d, d_2 \rightarrow d_1); \text{support } 1; \text{confidence } 100\ \%$$
$$(a, a_2 \rightarrow a_1) \Rightarrow (d, d_1 \rightarrow d_2); \text{support } 1; \text{confidence } 100\ \%$$

The following action rules are extracted from *Object ID 3*:

$$(a, a_1 \rightarrow a_2) \Rightarrow (d, d_1 \rightarrow d_2); \text{support } 4; \text{confidence } 100\ \%$$
$$(a, a_2 \rightarrow a_1) \Rightarrow (d, d_2 \rightarrow d_1); \text{support } 4; \text{confidence } 100\ \%$$
$$(a, a_1 \rightarrow a_2) \wedge (b, b_1) \Rightarrow (d, d_1 \rightarrow d_2); \text{support } 1; \text{confidence } 100\ \%$$
$$(a, a_1 \rightarrow a_2) \wedge (b, b_2) \Rightarrow (d, d_1 \rightarrow d_2); \text{support } 1; \text{confidence } 100\ \%$$
$$(a, a_2 \rightarrow a_1) \wedge (b, b_1) \Rightarrow (d, d_2 \rightarrow d_1); \text{support } 1; \text{confidence } 100\ \%$$
$$(a, a_2 \rightarrow a_1) \wedge (b, b_2) \Rightarrow (d, d_2 \rightarrow d_1); \text{support } 1; \text{confidence } 100\ \%$$

The following action rules are extracted from *Object ID 4*:

$$(a, a_1 \rightarrow a_2) \Rightarrow (d, d_1 \rightarrow d_2); \text{support } 1; \text{confidence } 100\ \%$$
$$(a, a_2 \rightarrow a_1) \Rightarrow (d, d_2 \rightarrow d_1); \text{support } 1; \text{confidence } 100\ \%$$
$$(a, a_1 \rightarrow a_2) \wedge (b, b_2) \Rightarrow (d, d_1 \rightarrow d_2); \text{support } 1; \text{confidence } 100\ \%$$
$$(a, a_2 \rightarrow a_1) \wedge (b, b_2) \Rightarrow (d, d_2 \rightarrow d_1); \text{support } 1; \text{confidence } 100\ \%$$

The following action rules are extracted from *Object ID 5*:

$$(a, a_1 \rightarrow a_2) \Rightarrow (d, d_1 \rightarrow d_2); \text{support 4; confidence 100 \%}$$
$$(a, a_2 \rightarrow a_1) \Rightarrow (d, d_2 \rightarrow d_1); \text{support 4; confidence 100 \%}$$
$$(a, a_1 \rightarrow a_2) \wedge (b, b_1) \Rightarrow (d, d_1 \rightarrow d_2); \text{support 1; confidence 100 \%}$$
$$(a, a_1 \rightarrow a_2) \wedge (b, b_2) \Rightarrow (d, d_1 \rightarrow d_2); \text{support 1; confidence 100 \%}$$
$$(a, a_2 \rightarrow a_1) \wedge (b, b_1) \Rightarrow (d, d_2 \rightarrow d_1); \text{support 1; confidence 100 \%}$$
$$(a, a_2 \rightarrow a_1) \wedge (b, b_2) \Rightarrow (d, d_2 \rightarrow d_1); \text{support 1; confidence 100 \%}$$

After we generate all possible action rules from every subsystem, we calculate the similarity matrix, using the metric defined above. Next, we calculate the degrees of similarity between every pair of objects:

$$Similarity(o_1, o_2) = 0$$
$$Similarity(o_1, o_3) = \left(2 - |.5 - 1| - \left|\frac{2}{4} - \frac{4}{4}\right|\right) + \left(2 - |1 - 1| - \left|\frac{2}{4} - \frac{4}{4}\right|\right)$$
$$+ \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{1}{4}\right|\right) + \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{1}{4}\right|\right) = 1 + 1.5 + 2 + 2 = 6.5$$
$$Similarity(o_1, o_4) = \left(2 - |.5 - 1| - \left|\frac{2}{4} - \frac{1}{4}\right|\right) + \left(2 - |1 - 1| - \left|\frac{2}{4} - \frac{1}{4}\right|\right)$$
$$+ \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{1}{4}\right|\right) + \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{1}{4}\right|\right) = 1.25 + 1.75 + 2 + 2 = 7$$
$$Similarity(o_1, o_5) = \left(2 - |.5 - 1| - \left|\frac{2}{4} - \frac{4}{4}\right|\right) + \left(2 - |1 - 1| - \left|\frac{2}{4} - \frac{4}{4}\right|\right)$$
$$+ \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{1}{4}\right|\right) + \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{1}{4}\right|\right) = 1 + 1.5 + 2 + 2 = 6.5$$
$$Similarity(o_2, o_3) = 0; \; Similarity(o_2, o_4) = 0; \; Similarity(o_2, o_5) = 0$$
$$Similarity(o_3, o_4) = \left(2 - |1 - 1| - \left|\frac{4}{4} - \frac{1}{4}\right|\right) + \left(2 - |1 - 1| - \left|\frac{4}{4} - \frac{1}{4}\right|\right)$$
$$+ \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{1}{4}\right|\right) + \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{1}{4}\right|\right) = 1.25 + 1.25 + 2 + 2 = 6.5$$
$$Similarity(o_3, o_5) = \left(2 - |1 - 1| - \left|\frac{4}{4} - \frac{4}{4}\right|\right) + \left(2 - |1 - 1| - \left|\frac{4}{4} - \frac{4}{4}\right|\right)$$
$$+ \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{1}{4}\right|\right) + \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{1}{4}\right|\right) + \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{1}{4}\right|\right)$$
$$+ \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{1}{4}\right|\right) = 2 + 2 + 2 + 2 + 2 + 2 = 12$$
$$Similarity(o_4, o_5) = \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{4}{4}\right|\right) + \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{4}{4}\right|\right)$$
$$+ \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{1}{4}\right|\right) + \left(2 - |1 - 1| - \left|\frac{1}{4} - \frac{1}{4}\right|\right) = 1.25 + 1.25 + 2 + 2 = 6.5$$

Here, we show the corresponding similarity matrix (Table 5):

As shown in Fig. 1, only by visual examination of the dendrogram, system users are able to grasp a decent idea of the hierarchical structure of any information system. It is clear

Table 5 Similarity matrix for the hybrid hierarchical object-driven example

	Object ID 1	Object ID 2	Object ID 3	Object ID 4	Object ID 5
Object ID 1	∞	0	6.5	7	6.5
Object ID 2	0	∞	0	0	0
Object ID 3	6.5	0	∞	6.5	12
Object ID 4	7	0	6.5	∞	6.5
Object ID 5	6.5	0	12	6.5	∞

that the two most similar objects are Object 3 and Object 5, hence they would form the first cluster; Object 1 and Object 4 are then clustered together; next, the two clusters are joined together to form one combined cluster; and finally Object 2 joins that cluster.

By applying the proposed hybrid approach based on hierarchical clustering prior to rule extraction, we provide a substantial advantage to system users. Essentially, domain experts are given the convenience of choosing the level of generalization (or specialization) most appropriate to their application. As seen in Fig. 1, the degree of generalization can be specified by the level on which we decide to set the vertical threshold of the hierarchy; after predetermining the level of generalization, our objects are clustered accordingly, and finally action rule extraction is applied. At the lowest level, objects are not clustered together, and hence treated the way a typical object-driven rule-extracting approach. On the other hand however, at the highest level, all objects are clustered together, forming one big cluster, which will be identical to applying the classical action rule extraction approach. Next, we provide an explanation of our case-study: Hypernasality Speech Disorder.

6 Applications to hypernasality treatment

6.1 Overview of hypernasal speech disorder

Distortions of the velopharyngeal closure, resulting in speech hypernasality or hyponasality, may cause speech disorders in children (Gubrynowicz et al. 2007). The patient's nasopharynx disorders have been examined in the Children's Memorial Health Institute in Warsaw for many years. The gathered data also include general information on the patient's condition if it can be of importance, e.g., cerebral palsy, neurology, or myopathy. This way a rich collection of complex data describing hypernasality was gathered, in close cooperation with one of our collaborators, Prof. Ryszard Gubrynowicz, who is a speech scientist and expert in this area; the data were collected when he was working in the Children's Memorial Health Institute.

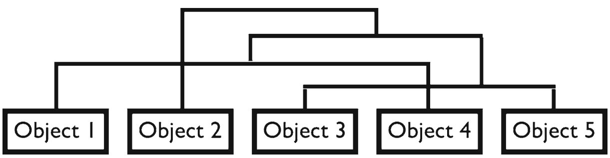

Fig. 1 Corresoponding dendrogram for Table 9; using the hybrid hierarchical object-drive action rule approach

Hypernasality can be examined by means of Czermak's mirror test of nasal air escape, see Fig. 2. The child is asked to repeat several times a syllable composed of a plosive consonant and an open vowel, e.g., /pa/-/pa/-/pa/, and the sizes of the fogging circles appearing on the mirror are rated on 4–point scale, from 0 (no hypernasality) to 3 (most severe hypernasality). Therefore, $Czermak's\ mirror\ test$ was used as a decision attribute in the nasality data set. All attributes, representing various medical conditions in the examined children, are listed in Table 6. More explanations about these attributes are given below.

6.2 Description of original attributes

Our dataset is composed of 225 patients; each patient was examined several times where each examination was performed on a separate visit, ranging from 2 to 11 visits for each patient. Personal data were recorded (first name and last name, sex, etc.), and for each examination the age of the child was marked. Personal data were removed before further processing, and replaced with ID data, representing the patient's ID combined with the sequential number of this patient's visit. During each visit, the articulation of selected vowels and consonants was recorded, and the recording date was marked ($recording\ date$ attribute). The data stored in columns marked as $diagnosis$ and $diagnosis2$ describe patient's condition related to nasality; only one diagnosis is stored in each of these columns, so $diagnosis2$ represents additional diagnosis, if there is more than one. The following diagnoses are described in these columns: R - cleft, RP - cleft palate, OR - after cleft palate surgery, WKP - congenital short velum, NO - hypernasality, NZ - hyponasality, BR - no diagnosis, PRP - submucous cleft palate, AT - after tonsillectomy, DKP - quite short palate, RJ - cleft uvula, III - hypertrophy of adenoids and possibly palatine tonsils, MP - hypertrophy of palatine tonsils, MPDz - cerebral palsy, AD - after adenotomy, ADT - after adenotonsillectomy, UK - larynx after injury/trauma, NS - hypoacusis, ORM - retarded speech development, NEU - neurology, ONR - after neurological surgery. If NO (hypernasality) is diagnosed and marked in the column $diagnosis$, it represents the most severe case of hypernasality. The numbers 0–3 in $diagnosis2$ refer to sleep apnoea, i.e. temporary cessation of respiration during sleep. 0 means no apnoea, 3 - very often. Sleep apnoea is also represented as a separate attribute, but the values assessed for the same patient may differ significantly, so they were kept in both columns. Generally, physicians may differ in their opinions, this is why we must be prepared to deal with some inconsistencies in the data. More of diagnostic details are given in the column $comments$, but these comments are not taken into account in the current version of our action rule software.

Other physical conditions recorded in the database include the degree of hypertrophy of adenoids and possibly palatine tonsils, and the degree of motility of the soft palate, represented as $tonsils$ and $motility$ attributes. The assessment of the patient's recorded speech

Fig. 2 Czermak's mirror fogging test, rating the degree of the patient's nasal air escape on a 4-point scale: none = 0; small = 1, medium = 2, large = 3 (Gubrynowicz et al. 2007)

Table 6 Attributes in the Hypernasality Data Set

Attribute	Description
ID	Patient's ID, with the sequential number of his/her visit
age	Age [years, months]
sex	Sex {M, F}
recording date	Recording Date [yyyy.mm.dd]
diagnosis	Diagnosis {AD, ADT, AT, BR, III, myopathy, MPDz, NEU, NO, ONR, OR, ORM, RJ, RP, UK, WKP}
comments	Comments, details of the diagnosis
diagnosis2	Diagnosis {0,1,2,3, DKP, RJ, WKP}
sleep apnoea	Sleep apnoea {0, 1, 2, 3}
tonsils	Hypertrophy of adenoids and possibly palatine tonsils {0, 1, 2, 3}
Czermak's mirror test	
- decision attribute	Mirror-fogging test {0, 1, 2, 3}
yeaoui	Measure of nasalization for vowels /I, e, a, o, u, i/ [0, 100]
i-long	Measure of nasalization for vowel /i/-long [0, 100]
bdg	Measure of nasalization for high pressure consonants /b, d, g/ [0, 100]
motility	Motility of the soft palate [0, 12]
difference level F1-F2	The difference level of 1^{st} & 2^{nd} formant measured for /i/-long [-14, 26]

Expansions of acronyms are given in the text

is represented in the following attributes: $yeaoui$ (vowels /I, e, a, o, u, i/ - a sequence of short vowel sounds spoken in isolation), $i-long$ (long vowel /i/ - vowel of sustained phonation), and bdg (high pressure consonants /b, d, g/); SAMPA coding of phonetic alphabet is used (SAMPA). These attributes describe the measure of nasalization (coefficient of nasalization), calculated from the analysis of mouth and nose signals (separately recorded), as the ratio of the nose signal level to the sum of the level of the nose and mouth signals for the phonemes indicated in each attribute. *difference level* $F1 - F2$ describes the vocal tract's first 2 resonances as the difference level of the 1^{st} and the 2^{nd} formant, measured for /i/-long.

The best diagnosis we are interested in is when the parameters' values are in normal ranges. Our decision attribute is Czermak's mirror test, so its values are most important in our research. The most desired value of our decision attribute is when it is equal to 0. The diagnosis is worse when Czermak's test value equals 2, next worse case is when Czermak's test value equals 3, and this is the most severe case. The lower the Czermak's test value, the better the diagnosis is. Therefore, we are interested in action rules indicating how to decrease the Czermak's test value. The goal of our system is to find action rules which purpose is to provide hints referring to doctor's interventions. They show how values of certain attributes need to be changed (through various medical procedures, according to the physician's order), so the patient's condition will get improved.

6.3 Description of the derived attributes

In this work, we derived a new set of attributes in accordance to Hajja et al. (2012). In addition to our attributes shown in Table 2, for each of the following four attributes: *yeaoui, i - long, bdg*, and *motility*, two new attributes were derived, resulting in eight new attributes. The two derived attributes are the difference, and the rate of change for every two consecutive instances, which we calculated as follows:

1. The difference of values for *yeaoui, i - long, bdg* and *motility* for every two consecutive visits is calculated, thus constituting the following new attributes: $yeaoui_1, i_1 - long, bdg_1$ and $motility_1$. For example, the value of bdg_1 equals to the value of bdg for the $(k+1)^{th}$ visit minus the value for the k^{th} visit.
2. The rate of change a_2 for every two consecutive visits is defined as:

$$a_2 = arctan\left(\frac{a_1}{\text{age difference in months}}\right)$$

where a_1 is the difference of values of the attribute a for the two visits.

Note that although by using the temporal-based approach, we are considering the direction of which instances transition before extracting action rule, it is worth noting here that the pace (and amount) of change is left unspecified; this is one of the main motivations for introducing the derived attribute, which would encode the amount (and pace of change) that occurs between instances to be used in the learning step. Additionally, and since the hierarchical object-driven approach need to discard the temporal-based approach, the derived attributes will be used to add a temporal significance to our instances. After calculating the derived attributes, we used the Rough Set Exploration System (Bazan and Szczuka 2005) to discretize our real-valued attributes wrt. our decision attribute. Next, our temporal object-driven action rule discovery system, presented in Section 2, was applied to the discretized data.

Our decision attribute *Czermak's mirror test* was not discretized. Moreover, when a physician could not decide between two neighboring Czermak's test values, an intermediate value was assigned. Therefore, the decision values are {0, .5, 1, 1.5, 2, 2.5, 3}.

7 Results

In all our following computations, we set the minimum support of our action rule extraction system to 1, and no minimum confidence. Also, all attributes (described in Table 6) for our hypernasality treatment dataset were set to *flexible attributes*.

7.1 Results of applying classical temporal-driven approach to hypernasality

In this subsection, a sample of the resulting action rules after running our classical object-driven approach discussed in SubSection 4.2 is presented:

Rule 1. $r_1 = [(\textit{palatine tonsils}, 0) \wedge (\text{i} - \text{long}, \geq 9.5) \wedge (i_2 - \text{long}, \geq 5.5 \rightarrow < 5.5)] \Rightarrow (\textit{Czermak's mirror test}, 2 \rightarrow 1.5)$; $supp(r_1) = 2$, $conf(r_1) = 66.7\,\%$.

Rule 1 means that if our patient has no hypertrophied palatine tonsils (value 0), and *i – long* (nasalization for /i/-long) is more than 9.5, then decreasing the rate of change of *i – long* (*i_2 – long)* from more than or equal to 5.5, to less than 5.5, would improve patient's *Czermak's mirror test* from 2 to 1.5.

Rule 2. $r_2 = [(\textit{palatine tonsils}, 1) \wedge (i-long, \geq 9.5) \wedge (i_2-long, \geq 5.5 \rightarrow < 5.5)] \Rightarrow (\textit{Czermak's mirror test}, 2 \rightarrow 1.5);\ supp(r_2) = 2,\ conf(r_2) = 66.7\ \%.$

Rule 2 means that if our patient has a bit hypertrophied palatine tonsils (value 1), and *i – long* is more than 9.5, then decreasing the rate of change of *i – long* (*i_2 – long)* from more than or equal to 5.5, to less than 5.5, would improve patients *Czermak's mirror test* from 2 to 1.5.

Rule 3. $r_3 = (i-long, \geq 9.5 \rightarrow [2.5, 7.5)) \Rightarrow (\textit{Czermak's mirror test}, 1 \rightarrow .5);\ supp(r_3) = 2,\ conf(r_3) = 66.7\ \%.$

Rule 3 means that if we decrease the value of *i – long* from greater than 9.5 to [2.5, 7.5), we would improve the patient's *Czermak's mirror test* from 1 to .5. Decreasing of *i – long*, i.e. the decrease of nasalization for /i/-long can be achieved, to some extend, through speech therapy, or (eventually) surgically through a nasopharynx surgery, moving posterior pharyngeal wall forward, towards soft palate. However, none of the examined patients underwent this surgery. These rules suggest decreasing palatine tonsils, and decreasing the nasalization coefficient for /i/-long. This confirms the importance of these attributes, which was expected, but not so obvious in the case of *i – long*. However, decreasing palatine tonsils causes only slight change of the Czermak's test, and this is also a confirmation for physicians that surgery on palatine tonsils might not be particularly beneficiary for patients. Palatine tonsils tend to regrow after surgery, and their influence on nasality is not so high, because of their location. At the same time, hypernasality can be treated, to some extent, through speech therapy aiming at decreasing nasalization for /i/-long.

Another interesting observation in our resulted action rules is that we may desire different transitions for the same attribute, depending on the current value of the *Czermak's mirror test*. For example, if our patient's *Czermak's mirror test* is 2, we would want for *i – long* to stay greater than 9.5 while changing *i_2 – long*. However, if our patient's *Czermak's mirror test* is 1, we would want to decrease the value of *i – long* from greater than 9.5 to [2.5 7.5).

Rule 4. $r_4 = [(\textit{palatine tonsils}, < 2) \wedge (i_2-long, < 5.5) \wedge (\textit{motility}, [4.5, 5.5)) \wedge (i-long, [1.5, 2.5) \rightarrow < 1.5)] \Rightarrow (\textit{Czermak's mirror test}, 0.5 \rightarrow 0);\ supp(r_4) = 3,\ conf(r_4) = 60\ \%.$

Rule 5. $r_5 = [(\textit{palatine tonsils}, < 2) \wedge (bdg_1, < 6.5) \wedge (\textit{motility}, [4.5, 5.5)) \wedge (i-long, [1.5, 2.5) \rightarrow < 1.5)] \Rightarrow (\textit{Czermak's mirror test}, 0.5 \rightarrow 0);\ supp(r_5) = 3,\ conf(r_5) = 60\ \%.$

The above two rules mean that very slight hypernasality might be completely removed by decreasing the nasalization coefficient for /i/-long, provided the indicated accompanying conditions are fulfilled.

Rule 6. $r_6 = [(\textit{sleep apnoea}, < 2 \rightarrow > 2) \wedge (bdg, > 8.5 \rightarrow [6.5, 8.5)] \Rightarrow (\textit{Czermak's mirror test}, 1.5 \rightarrow 0);\ supp(r_6) = 2,\ conf(r_6) = 100\ \%.$

This rule is very interesting because it shows that the increase of sleep apnoea (along with decreasing the nasality of /bdg/) cures light-medium hypernasality.

Rule 7. $r_7 = [(\textit{palatine tonsils}, < 2) \wedge (\textit{motility}, < 3.5 \rightarrow [4.5, 5.5)) \wedge (\textit{difference level F1-F2}, [5.5, 6.5) \rightarrow < 4.5)] \Rightarrow (\textit{Czermak's mirror test}, 1.5 \rightarrow 0);\ supp(r_7) = 2, conf(r_7) = 100\ \%.$

Rules 6 and 7 are especially interesting because their confidence is 100 %, and they both induce a significant shift in Czermak's test, from 1.5 to 0, i.e.

curing light-medium hypernasality completely, through procedures increasing the motility of the soft palate, and decreasing the difference for the first two formants of the vocal tract for /i/-long.

7.2 Results of applying pair-based object-driven approach to hypernasality

In this section, we show a sample of results after running our pair-based approach to extract object-driven action rules from temporal systems. We show that by using the pair-based approach, not only we were able to extract a larger set of action rules, but also we were able to extract action rules that provide more dramatic decrease of patient severity than the rules extracted in Hajja et al. (2012). For an action rule to be eligibly used on a patient, the preconditions of the action rule and the patient's current condition have to match, meaning that only a subset of our patients will benefit from each particular action rule. Having said that, using our pair-based approach to extract action rules will generate a significant amount of action rules that can be appropriately used for various sets of patients.

Rule 1. $r_1 = (\textit{difference level F1-F2}, \geq 9.5 \rightarrow [6.5, 9.5)) \Rightarrow (\textit{Czermak's mirror test}, 3 \rightarrow 2);\ supp(r_1) = 2,\ conf(r_1) = 100\ \%$.

This rule means that by decreasing the difference between the first two formants of the vocal tract for /i/ - long, we would notice a decent shift of the Czermak's mirror test, decreasing from 3 to 2. In Hajja et al. (2012), we extracted a similar action rule that also indicated the importance of *difference level F1-F2* attribute. However, this action rule is exclusive to the work described in this paper.

Rule 2. $r_2 = (i_2 - long, \geq 5.5 \rightarrow < 5.5) \Rightarrow (\textit{Czermak's mirror test}, 3 \rightarrow 2);\ supp(r_2) = 3,\ conf(r_2) = 66.7\ \%$.

This rule means that decreasing the value of $i - long$ in a short period of time, since $i_2 - long$ is defined as the rate of change, will result in a similar decrease of the Czermak's mirror test from 3 to 2. Again, this rule affirms the importance of the attribute $i - long$.

Rule 3. $r_3 = (i_2 - long, \geq 5.5 \rightarrow < 5.5) \wedge (bdg, \geq 8.5) \Rightarrow (\textit{Czermak's mirror test}, 2.5 \rightarrow 2);\ supp(r_3) = 2,\ conf(r_3) = 100\ \%$.

This rule is similar to Rule 1. It confirms the effect of decreasing the rate of change of the nasalization measured for /i/ - long, but also adds an additional condition concerning the nasality of /bdg/, that is, this rule only applies to patients suffering from high nasality for /bdg/ (≥ 8.5).

Rule 4. $r_4 = (tonsils, < 2) \wedge (i_2 - long, \geq 5.5 \rightarrow < 5.5) \wedge (motility, [4.5, 5.5)) \Rightarrow (\textit{Czermak's mirror test}, 2 \rightarrow 1.5);\ supp(r_4) = 2,\ conf(r_4) = 100\ \%$.

This rule states that when a patient is experiencing a little hypertrophied adenoids and possibly palatine tonsils ($tonsils < 2$), we can slightly improve his condition from Czermak's mirror test 2 to 1.5 by decreasing the rate of change in /i/ - long, and if the motility of the soft palate does not change.

Rule 5. $r_5 = (bdg, \geq 8.5 \rightarrow [6.5, 8.5)) \Rightarrow (\textit{Czermak's mirror test}, 1 \rightarrow .5);\ supp(r_5) = 3,\ conf(r_5) = 66.7\ \%$.

This rule states that by only decreasing the nasality of /bdg/, we would be able to shift the patients' Czermak's mirror test state from 1 to .5.

Rule 6. $r_6 = (i - long, \geq 9.5 \rightarrow [2.5, 7.5)) \Rightarrow (\textit{Czermak's mirror test}, 1 \rightarrow .5);\ supp(r_6) = 2,\ conf(r_6) = 100\ \%$.

Although the support of this action rule is not high, the rule is rather interesting. It states that by decreasing only one attribute; /i/ - long, there is a 100 % chance that the Czermak's mirror test will shift from 1 to .5.

Rule 7. $r_7 = (motility, < 3.5 \rightarrow [4.5, 5.5)) \wedge (diagnosis, OR) \Rightarrow (\textit{Czermak's mirror test}, 1 \rightarrow 0)$; $supp(r_7) = 3,\ conf(r_7) = 100\ \%$.

This rule states that if a patient has gone through a cleft palate surgery (OR), then increasing the motility of the soft palate would significantly improve the patient's condition, to the level where the patient is entirely cured, which will result in shifting the Czermak's mirror test value from 1 to 0.

Rule 8. $r_8 = (i_2 - long, \geq 5.5 \rightarrow < 5.5) \Rightarrow (\textit{Czermak's mirror test}, .5 \rightarrow 0)$; $supp(r_8) = 7,\ conf(r_8) = 71\ \%$.

This rule has a relatively high support. It states that decreasing the rate of change of $i - long$ from greater than or equal to 5.5, to less than 5.5, will result in curing a light hypernasality.

Rule 9. $r_9 = (i_2 - long, \geq 5.5 \rightarrow < 5.5) \wedge (\textit{sleep apnoea}, < 2) \Rightarrow (\textit{Czermak's mirror test}, .5 \rightarrow 0)$; $supp(r_9) = 6,\ conf(r_9) = 83\ \%$.

This rule is a similar, but more specific than rule 8. By expanding the condition side of action rules, we are able to generate action rules with higher confidence. Rule 8 states that by only decreasing the rate of change of $i - long$, we would have a 71 % chance of shifting the Czermak's mirror test from .5 to 0. However, rule 9 states that by decreasing the rate of change of $i - long$ and maintaining a low value of sleep apnoea, we would have an 83 % chance of shifting Czermak's mirror test from .5 to 0.

Rule 10. $r_{10} = (tonsils, \geq 2 \rightarrow < 2) \Rightarrow (\textit{Czermak's mirror test}, .5 \rightarrow 0)$; $supp(r_{10}) = 5$, $conf(r_{10}) = 100\ \%$.

This rule states, with absolute certainty (confidence 100 %), that by decreasing the hypertrophied adenoids and possibly palatine tonsils that the patient is experiencing, the Czermak's mirror test will shift from .5 to 0. Although the improvement does not appear to be significant, the high support and high confidence make this rule highly valuable. In our hypernasality dataset, most of the patients were experiencing slight to no hypernasality speech (Czermak's mirror test .5 or 0). As a consequence, the last three action rules had a much higher support compared to the others.

7.3 Results of applying the hybrid hierarchical object-driven approach to hypernasality

In this subsection, we provide sample results of action rules extracted by applying the hybrid hierarchical object-driven approach, while using three different levels of generalization. As we have seen in Table 3, the more patients (or objects) we tend to cluster together, the more potential action rules are to be extracted, with relatively higher support (Fig. 3 shows a more detailed chart of how the number of action rules increases as we decrease the number of clusters); however, by combining (or clustering) patients (or objects) into super-groups, generalization effect will be more apparent; until it reaches a point where generalization may cause negative effects; for that reason, using multiple levels of generalization is usually advised. Next, we provide action rules extracted from the hybrid hierarchical object-driven approach proposed in this paper using various levels of generalization, ranging from 10 clusters to 40.

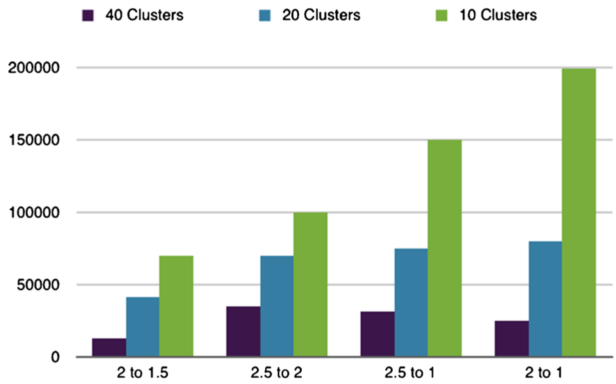

Fig. 3 This figure shows how the number of action rules increases as the number of clusters decreases for some desired transitions in the Czermak's mirror test

Rule 1. $r_1 = (\textit{palatine tonsils}, \leq 2) \wedge (\textit{difference level F1-F2}, \leq 4.5 \rightarrow [6.5, 9.5]) \Rightarrow (\textit{Czermak's mirror test}, 1 \rightarrow 0.5);\ \textit{supp}(r_1) = 18,\ \textit{conf}(r_1) = 100\ \%$.

This rule has a relatively high support; it states that for patients who are experiencing low levels of hypertrophied, by increasing the difference between the first two formants of the vocal tract for */i/ - long*, we expect for the patient's condition to improve from 1 to .5.

Rule 2. $r_2 = (\textit{palatine tonsils}, \leq 2) \wedge (\text{motility}, \geq 5.5) \wedge (\textit{yeaou}, \geq 7.5 \rightarrow [3.5, 4.5]) \Rightarrow (\textit{Czermak's mirror test}, 2 \rightarrow 0);\ \textit{supp}(r_2) = 12,\ \textit{conf}(r_2) = 66.6\ \%$.

This rule states that for patients who are experiencing low levels of hypertrophied, and that have high motility level for their soft palate, by decreasing the levels of nasalization for vowels /I, e, a, o, u, i/, we expect the patient's condition to improve substantially from 2 to 1.

Rule 3. $r_3 = (\textit{palatine tonsils}, \leq 2) \wedge (\text{motility}, \geq 5.5 \rightarrow \leq 3.5) \wedge (\textit{bdg}, \geq 8.5 \rightarrow \leq 6.5) \Rightarrow (\textit{Czermak's mirror test}, 2 \rightarrow 0);\ \textit{supp}(r_3) = 18,\ \textit{conf}(r_3) = 75\ \%$.

Similar to rule 1, this rule has a relatively high support; it states that for patients who are experiencing low levels of hypertrophied, by both changing the motility level from greater than 5.5 to less than 3.5, and by decreasing the nasalization of /b, d, g/, we expect the patient's condition to improve substantially from 2 to 1.

Rule 4. $r_4 = (\textit{palatine tonsils}, \leq 2) \wedge (\text{diagnosis, III}) \wedge (\textit{yeaou}, \geq 7.5 \rightarrow [3.5, 4.5]) \Rightarrow (\textit{Czermak's mirror test}, 1 \rightarrow 0);\ \textit{supp}(r_4) = 3,\ \textit{conf}(r_4) = 100\ \%$.

The above rule states that for patients who are experiencing low levels of hypertrophied, and that have been diagnosed with hypertrophy of adenoids and possibly palatine tonsils, by decreasing the level of nasalization for vowels /I, e, a, o, u, i/, we expect the patient's condition to improve from 1 to 0, which would completely remove any hypernasality.

Rule 5. $r_5 = (\textit{sleep apnoea}, \leq 2) \wedge (\text{bdg}, \leq 6.5 \rightarrow \geq 8.5) \wedge (\textit{motility}, \leq 3.5 \rightarrow [3.5, 4.5]) \Rightarrow (\textit{Czermak's mirror test}, 1 \rightarrow 0);\ \textit{supp}(r_5) = 2,\ \textit{conf}(r_5) = 100\ \%$.

The above rule states that for patients who are experiencing low levels of sleep apnoea, by both increasing the level of nasalization for /b, d, g/, and by increasing the level of motility for the soft palate, we are expected to observe an improvement in the patient's condition from 1 to 0.

Rule 6. r_6 = (difference level F1-F2, $\leq$ 4.5) $\wedge$ (*motility*, $[4.5, 5.5] \rightarrow [3.5, 4.5]$) $\Rightarrow$ (*Czermak's mirror test*, $1 \rightarrow 0$); $supp(r_6) = 6$, $conf(r_6) = 100\ \%$.

This rule states that for patients who have difference between the first two formants of the vocal tract for */i/ - long* less than 4.5, by decreasing the motility level of the soft palate, we would expect the patient's condition to slightly improve from 1.5 to 1.

Rule 7. r_7 = (diagnosis, AD) $\wedge$ (*motility*, $[4.5, 5.5] \rightarrow [3.5, 4.5]$ $\Rightarrow$ (*Czermak's mirror test*, $1 \rightarrow 0$); $supp(r_7) = 6$, $conf(r_7) = 100\ \%$.

This rule states that for patients who have had their adenoids removed by adenotomy surgery, we can improve their condition from 1 to .5 by decreasing the motility level of the soft palate.

8 Conclusion

In this paper, we presented a new hierarchical hybrid approach to extract action rules from information systems of object-driven nature. We proposed a novel hierarchical clustering approach based on a behavior response similarity metric that takes advantage of our existing action rule functionalities. The motivation behind this work is to expand on the work proposed by Hajja et al. (2012, 2013), and to propose a system that will overcome the limitation that often occurs when unique objects do not have many observations.

We have shown that by creating a combined system that utilizes the advantages of both the classical action rule approach, and the object-driven approach, we are able to present system users with a convenient and interactive action rule extraction system that allows them to extract action rules with respect to their unique requirements; by controlling intrinsically trivial threshold value accordingly.

Acknowledgments The authors would like to thank Dr. Danuta Chojnacka-Wądołowska and Dr. Cecylia Konopka, both from Children's Memorial Health Institute in Warsaw; and Dr. Ryszard Gubrynowicz from PJIIT, for their help with data collection and providing medical diagnoses.

This project was partially supported by the Research Center of PJIIT, supported by the Polish Ministry of Science and Higher Education. It is also based upon work supported by the National Science Foundation under Grant No. OISE-0730065.

References

Agrawal, R., Imieliski, T., Swami, A. (1993). Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD international conference on Management of data (SIGMOD '93), P. Buneman and S. Jajodia (Eds.), ACM, New York, USA, 207–216. doi:10.1145/170035.170072.

Bazan, J., & Szczuka, M. (2005). The rough set exploration system. In *Transactions on Rough Sets III*, pp 25–42.

Gubrynowicz, R., Chojnacka-Wądołowska, D., Konopka, C. (2007). Assessment of velum malfunction in children through simultaneous nasal and oral acoustic signals measurements. *Archives of Acoustics*, *32*(1), 165–175.

Hajja, A., Wieczorkowska, A., Raś, Z.W., Gubrynowicz, R. (2012). Object-driven action rules and their application to hypernasality treatment. In *Proceedings of ECML-PKDD workshop on new frontier in mining complex patterns*, Bristol, September 24–28, 2012, 104–115.

Hajja, A., Wieczorkowska, A., Raś, Z.W., Gubrynowicz, R. (2013). Pair-based object-driven action rules. *Post-proceedings of the first ECML/PKDD 2012 international workshop on new frontiers in mining complex patterns*. LNAI 7765, Springer.

Pawlak, Z. (1981). Information systems - theoretical foundations. *Information Systems Journal*, *6*, 205–218.
Raś, Z.W., & Wieczorkowska, A. (2010). Advances in music information retrieval, studies in computational intelligence. (Vol. 274). Springer.
Raś, Z.W., & Wieczorkowska, A. (2000). Action-Rules: How to increase profit of a company. In Zighed, D.A., Komorowski, J., Zytkow, J. (Eds.)*, PKDD 2000. LNAI* (Vol. 1910, pp. 587–592). Heidelberg: Springer.
Raś, Z.W., & Dardzinska, A. (2011). From Data to Classification Rules and Actions. In the Special Issue on Rough Sets. Theory and Applications. In C.-C. Chan et al. (Eds.), *International Journal of Intelligent Systems* (Vol. 26(6), pp. 572–590). Wiley.
Raś, Z.W., Tzacheva, A., Tsay, L.S., Gürdal, O. (2005). Mining for interesting action rules. In *Proceedings of IEEE/WIC/ACM international conference on intelligent agent technology (IAT 2005)*, IEEE Computer Society, 187–193.
Raś, Z.W., Dardzińska, A., Tsay, L.S., Wasyluk, H. (2008). Association action rules. In *IEEE International conference on data mining workshops*, 283–290.
Raś, Z.W., Wyrzykowska, E., Wasyluk, H. (2008). Action rules discovery based on agglomerative strategy. In *Mining complex data, post-proceedings of 2007 ECML/PKDD third international workshop (MCD 2007)*, LNAI, Vol. 4944, 196–208, Springer.
SAMPA. - computer readable phonetic alphabet. http://www.phon.ucl.ac.uk/home/sampa/.
Tsay, L.S., & Raś, Z.W. (2006). Action rules discovery system DEAR3. In *Foundations of intelligent systems, Proceeding of ISMIS 2006*, LNAI, Vol. 4203, 483–492, Springer.
Zhang, X., Raś, Z.W., Jastreboff, P.J., Thompson, P.L. (2010). From Tinnitus data to action rules and tinnitus treatment. In *Proceedings of 2010 IEEE conference on granular computing*. Silicon Valley, IEEE Computer Society, 620–625.