On Testing Membership to Maximal Consistent Extensions of Information Systems

Mikhail Moshkov[1], Andrzej Skowron[2], and Zbigniew Suraj[3,4]

[1] Institute of Computer Science, University of Silesia
Będzińska 39, 41-200 Sosnowiec, Poland
moshkov@us.edu.pl
[2] Institute of Mathematics, Warsaw University
Banacha 2, 02-097 Warsaw, Poland
skowron@mimuw.edu.pl
[3] Chair of Computer Science, Rzeszów University
Rejtana 16A, 35-310 Rzeszów, Poland
zsuraj@univ.rzeszow.pl
[4] Institute of Computer Science, State School of Higher Education in Jarosław
Czarnieckiego 16, 37-500 Jarosław, Poland

Abstract. This paper provides a new algorithm for testing membership to maximal consistent extensions of information systems. A maximal consistent extension of a given information system includes all objects corresponding to known attribute values which are consistent with all true and realizable rules extracted from the original information system. An algorithm presented here does not involve computing any rules, and has polynomial time complexity. This algorithm is based on a simpler criterion for membership testing than the algorithm described in [4]. The criterion under consideration is convenient for theoretical analysis of maximal consistent extensions of information systems.

Keywords: rough sets, information systems, maximal consistent extensions.

1 Introduction

Information systems can be used to represent knowledge about the behavior of concurrent systems. The idea of a concurrent system representation by information systems is due to Zdzisław Pawlak [3]. In this approach, an information system represented by a data table encodes the knowledge about global states of a given concurrent system. Columns of the table are labeled with names of attributes (interpreted as local processes of a given concurrent system). Each row labeled with an object (interpreted as a global state of a given concurrent system) includes a record of attribute values (interpreted as states of local processes). We assume that a given data table includes only a part of possible global states of a concurrent system, i.e., only those which have been observed by us

S. Greco et al. (Eds.): RSCTC 2006, LNAI 4259, pp. 85–90, 2006.

so far. In other words, it contains partial knowledge about a possible system behavior. Such an approach is called the Open World Assumption. This partial knowledge encoded in a data table can be represented by means of rules which can be extracted from the data table. Such knowledge is sufficient to construct a system model of the high quality. The remaining knowledge can be discovered from the constructed model. New knowledge derived from the model encompasses new global states of the system which have not been observed before. Such new states are consistent with the knowledge expressed by the rules extracted from the given data table.

The above approach has been motivation for introducing a notion of consistent extensions of information systems in [5]. A given information system S defines an extension S' of S created by adding to S all new objects including combinations of only known attribute-value pairs, i.e., those pairs which have occurred in S. Among all extensions of a given information system S, the so-called consistent extensions of S play a significant role. A consistent extension S' of S includes only objects, which satisfy all rules true and realizable in S. If the consistent extension S' with the above property is the largest extension of S (with respect to the number of objects) then S' is called a maximal consistent extension of S.

If an information system S describes a concurrent system, the maximal consistent extension of S represents the largest set of global states of the concurrent system consistent with all rules true and realizable in S. This set may include new global states not observed until now. It is – in a certain sense – new knowledge about the concurrent system behavior described by S which is discovered by us.

A crucial problem concerning maximal consistent extensions of information systems is computing such extensions. This problem has been considered in the literature, among others, in [4], [7] and [8]. In [1], [5], [6] some approaches have been presented, where maximal consistent extensions are generated by classical Petri net or colored Petri net models built on the basis of information systems describing concurrent systems. Majority of methods for determining maximal consistent extensions of information systems presented until now in the literature (with the exception of [4]) requires computing all minimal rules (or only a part of them) true and realizable in information systems. Such algorithms characterize exponential complexity. Therefore, elaborating efficient methods became an important research problem. In this paper, some theoretical background for the method of computing maximal consistent extensions of information systems not involving computing any rules in information systems is presented. The method considered in this paper is slightly different from the one considered in [4]. Our method is based on a simpler criterion than the method presented in [4] and is more appropriate for theoretical analysis.

The remaining part of the paper is organized as follows. Main notions are presented in Section 2. Section 3 describes a new algorithm for testing membership to maximal consistent extensions of information systems. Finally, Section 4 consists of some conclusions.

2 Main Notions

Let $S = (U, A)$ be an information system [2] where U is a finite set of objects and A is a set of attributes (functions defined on U). For any $a \in A$, by V_a we denote the set $\{a(u) : u \in U\}$ and we assume that $|V_a| \geq 2$.

We define an information system $S^* = (U^*, A^*)$, where U^* is equal to the Cartesian product $\times_{a \in A} V_a$ and $A^* = \{a^* : a \in A\}$, where $a^*(f) = f(a)$ for $f \in \times_{a \in A} V_a$.

Assuming that for any $u, u' \in U$ if $u \neq u'$ then $Inf_A(u) \neq Inf_A(u')$, where $Inf_A(u) = \{(a, a(u)) : a \in A\}$ we can identify any object $u \in U$ with the object $Inf_A(u) \in U^*$ and any attribute $a \in A$ with the attribute $a^* \in A^*$ defined by $a^*(Inf_A(u)) = a(u)$. Hence, the information system $S = (U, A)$ can be treated as a subsystem of the information system $S^* = (U^*, A^*)$. In the sequel we write a instead a^*.

For any information system $S = (U, A)$ it is defined the set of boolean combinations of descriptors over S [2]. Any descriptor over S is an expression $a = v$ where $a \in A$ and $v \in V_a$. Boolean combinations of descriptors are defined from descriptors using propositional connectives. For any boolean combination of descriptors α it is defined its semantics, i.e., the set $\|\alpha\|_S \subseteq U$ consisting of all objects satisfying α [2]. For example, if α is a formula $\bigwedge_{a \in B}(a = v_a)$, where $B \subseteq A$, $a \in B$, and $v_a \in V_a$ then $\|\alpha\|_S = \{u \in U : a(u) = v_a \text{ for any } a \in B\}$.

A rule r (over S) is any expression of the form

$$\bigwedge_{a \in B}(a = v_a) \longrightarrow a' = v_{a'}, \tag{1}$$

where $B \subseteq A$, $v_a \in V_a$ for $a \in B$, $a' \in A$, and $v_{a'} \in V_{a'}$.

The rule r (see (1)) is true for $u \in U^*$ if for some $a \in B$ we have $a(u) \neq v_a$ or $a'(u) = v_{a'}$. The rule r (see (1)) is S-true if $\| \bigwedge_{a \in B}(a = v_a)\|_S \subseteq \|a' = v_{a'}\|_S$ and it is S-realizable if $\| \bigwedge_{a \in B}(a = v_a)\|_S \cap \|a' = v_{a'}\|_S \neq \emptyset$.

The set of all S-true and S-realizable rules is denoted by $Rule(S)$.

Now, we can introduce the main concept of this paper, i.e., the maximal extension of S. The maximal extension of S, in symbols $Ext(S)$, is defined by

$$Ext(S) = \{u \in U^* : \text{ any rule from } Rule(S) \text{ is true in } u\}. \tag{2}$$

Let us consider an information system $S = (U, A)$ and $u, u' \in U$. The set of attributes on which u, u' are indiscernible in S is defined by

$$IND_A(u, u') = \{a \in A : a(u) = a(u')\}. \tag{3}$$

Such a set $IND_A(u, u')$ defines a pattern, i.e., the following boolean combination of descriptors over S:

$$T_A(u, u') = \bigwedge_{a \in IND_A(u,u')} (a = a(u)). \tag{4}$$

Now, for a given information system S and any $u^* \in U^* \setminus U$ we define an important for our considerations family of sets $\mathcal{F}(u^*, S)$ by

$$\mathcal{F}(u^*, S) = \{a(\|T_A(u, u^*)\|_S) : a \in A \setminus IND_A(u, u^*) \ \&\ u \in U\}, \tag{5}$$

where $a(\|T_A(u, u^*)\|_S) = \{a(x) : x \in \|T_A(u, u^*)\|_S\}$, i.e., $a(\|T_A(u, u^*)\|_S)$ is the image under a of the set $\|T_A(u, u^*)\|_S$.

3 Testing Membership to $\boldsymbol{Ext(S)}$

In [4] a polynomial algorithm has been considered, which for a given information system $S = (U, A)$ and a given object u from $U^* \setminus U$ recognizes whether this object belongs to $Ext(S)$ or not. This algorithm is based on a criterion of membership to maximal consistent extension of information system which uses comparison of sets of reducts of a special kind (reducts related to a fixed object and attribute) to a given set of objects and to its one-element extension.

We consider the following problem:

Membership Problem (MP)

INPUT: $S = (U, A)$ and $u^* \in U^* - U$
OUTPUT: 1 if $u* \in Ext(S)$
0 if $u^* \notin Ext(S)$.

We now present a polynomial algorithm $\mathcal{A}$ for solving the MP problem. Our algorithm is based on a simpler criterion than that presented in [4].

Algorithm $\mathcal{A}$

```
for all u ∈ U
    for all a ∈ A \ IND_A(u, u*)
        begin
            compute a(||T_A(u, u*)||_S);
            if |a(||T_A(u, u*)||_S)| = 1 then
                begin
                    return(0);
                    Stop
                end
        end
return(1)
```

The correctness of the algorithm $\mathcal{A}$ follows from the following proposition:

Proposition 1. *Let $S = (U, A)$ be an information system and let $u^* \in U^* \setminus U$. Then the following conditions are equivalent:*
(i) $u^* \notin Ext(S)$,
(ii) there exists $X \in \mathcal{F}(u^*, S)$ such that $|X| = 1$.

Proof.
(ii) $\Rightarrow$ (i)

Let us assume that for some one element set X we have $X \in \mathcal{F}(u^*, S)$. Then the following equality holds: $X = a(\|T_A(u, u^*)\|_S)$ for some $a \in A \setminus IND_A(u, u^*)$ and $u \in U$. Hence, the rule r defined by

$$T_A(u, u^*) \longrightarrow a = a(u)$$

is S-true, because $|X| = 1$. We also have $a(u) \neq a(u^*)$ because $a \notin IND_A(u, u^*)$. Hence, r is not true for u^*, so $u^* \notin Ext(S)$.
(i) $\Rightarrow$ (ii)

Let us assume $u^* \notin Ext(S)$. It means that there exists a rule r of the form $\alpha \longrightarrow a = v$, where α is a boolean combination of descriptors over S, which is not true for u^* but is S-true and S-realizable. Hence, $u^* \in \|\alpha\|_{S_*}$ and $a(u^*) \neq v$. The rule r is S-realizable. Hence, for some $u \in U$ we have $u \in \|\alpha\|_S$ and $a(u) = v$. From the definition of $T_A(u, u^*)$ we obtain that $T_A(u, u^*)$ consists of all descriptors from α. Hence, $\|T_A(u, u^*)\|_S \subseteq \|\alpha\|_S$ and in S is true the following rule:

$$T_A(u, u^*) \longrightarrow a = a(u).$$

Let us now consider the set $a(\|T_A(u, u^*)\|_S)$. Since $\|T_A(u, u^*)\|_S \subseteq \|\alpha\|_S$ and $|a(\|\alpha\|_S)| = 1$ we also have $|a(\|T_A(u, u^*)\|_S)| = 1$. The last equality follows from the fact that $|a(\|T_A(u, u^*)\|_S)| \geq 1$ if the set $\|T_A(u, u^*)\|_S$ is non-empty.

Let us consider an example.

Example 1. Let $S = (U, A)$, $A = \{a_1, a_2\}$ and $U = \{(0, 1), (1, 0), (0, 2), (2, 0)\}$. The application of the considered algorithm to each object u from $\{0, 1, 2\}^2 \setminus U$ allows to find the set $Ext(S)$ which is equal to $\{(0, 1), (1, 0), (0, 2), (2, 0), (0, 0)\}$.

4 Conclusions

In this paper, a new method for testing membership to maximal consistent extensions of information systems is proposed. This method significantly differs from the majority of methods presented in the literature, as it does not involve computing any rules. Moreover, the presented method is useful for theoretical analysis of maximal consistent extensions of information systems.

We also plan to extend the presented approach to the case of nondeterministic or probabilistic rules used in the definition of the extension of a given information system. Moreover, filtration methods can be used for selecting relevant rules in constructing models, analogously to methods used for constructing of rule based classifiers.

One of the problem we would like to study is a decision problem for checking if a given information system has consistent extension consisting of at least k new states, where k is a given positive integer.

Acknowledgments

The authors are greatly indebted to anonymous reviewers for useful suggestions. This paper has been partially supported by the Ministry of Scientific Research and Information Technology of the Republic of Poland research grants No. 3 T11C 005 28 and 3 T11C 002 26.

References

1. Pancerz, K., Suraj, Z.: Synthesis of Petri net models: a rough set approach. Fundamenta Informaticae **55**(2) (2003) 149–165
2. Pawlak, Z.: Rough Sets – Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers, Dordrecht, 1991
3. Pawlak, Z.: Concurrent versus sequential the rough sets perspective. Bulletin of the EATCS **48** (1992) 178–190
4. Rząsa, W., Suraj, Z.: A new method for determining of extensions and restrictions of information systems. Proceedings of the 3rd International Conference on Rough Sets and Current Trends in Computing. Lecture Notes in Artificial Intelligence **2475**, Springer-Verlag (2002) 197–204
5. Skowron, A., Suraj, Z.: Rough sets and concurrency. Bulletin of the Polish Academy of Sciences **41**(3) (1993) 237–254
6. Suraj, Z.: Rough set methods for the synthesis and analysis of concurrent processes. Rough Set Methods and Applications (Studies in Fuzziness and Soft Computing **56**). Edited by L. Polkowski, S. Tsumoto, and T.Y. Lin. Physica-Verlag (2000) 379–488
7. Suraj, Z.: Some remarks on extensions and restrictions of information systems. Rough Sets and Current Trends in Computing. Lecture Notes in Artificial Intelligence **2005**, Springer-Verlag (2001) 204–211
8. Suraj, Z., Pancerz, K.: A new method for computing partially consistent extensions of information systems: a rough set approach. Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. Paris, France, July 2006, Editions EDK (2006), Vol. III 2618-2625