International Journal of Approximate Reasoning 52 (2011) 881–893

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

A sequential pattern mining algorithm using rough set theory ☆

Ken Kaneiwa [a,*], Yasuo Kudo [b]

[a] Department of Electrical Engineering and Computer Science, Iwate University, 4-3-5 Ueda, Morioka, Iwate 020-8551, Japan
[b] College of Information and Systems, Muroran Institute of Technology, 27-1 Mizumoto, Muroran 050-8585, Japan

ARTICLE INFO

Article history:
Received 28 May 2010
Revised 4 January 2011
Accepted 14 March 2011
Available online 23 March 2011

Keywords:
Rough set theory
Sequential pattern mining
Local patterns

ABSTRACT

Sequential pattern mining is a crucial but challenging task in many applications, e.g., analyzing the behaviors of data in transactions and discovering frequent patterns in time series data. This task becomes difficult when valuable patterns are locally or implicitly involved in noisy data. In this paper, we propose a method for mining such local patterns from sequences. Using rough set theory, we describe an algorithm for generating decision rules that take into account local patterns for arriving at a particular decision. To apply sequential data to rough set theory, the size of local patterns is specified, allowing a set of sequences to be transformed into a sequential information system. We use the discernibility of decision classes to establish evaluation criteria for the decision rules in the sequential information system.

1. Introduction

Data mining algorithms have been developed as tools to discover valuable patterns and rules from large amounts of data. In the traditional algorithms, association rules are discovered from attributes found frequently in datasets. By using a more complex approach, sequential pattern mining algorithms [1,33,2] enable us to find frequent patterns in sequential datasets. Sequential pattern mining requires the analysis of an ordered list of itemsets (e.g., a list of actions or orders) that can be modeled by a sequence. In order to effectively carry out the task, we have to extract only valuable patterns included in sequences by skipping noisy and meaningless patterns. However, frequent data mining algorithms are not feasible when it comes to extracting local (or implicit) patterns from noisy data. This is because the algorithms may not work when valuable patterns do not appear frequently or when waste patterns appear frequently. In fact, the frequencies of such valuable patterns may be less than a user-specified threshold, but setting a lower threshold leads to the recovery of a number of meaningless patterns.

In order to solve the problem, we have to logically and combinationally analyze patterns in sequences by checking the occurrences of local patterns that consistently result in a decision. For such an analysis, rule generation in rough set theory [10,17,16,4,14,11,12] provides a data mining algorithm based on the notions of attribute reduction and reduced decision rules. One of the advantages of rough set data mining is that it can generate reduced and consistent decision rules by logically checking all combinations of condition and decision attributes in an information system. Thus, rough set theory can be used to generate essential attributes through attribute reduction of logical combinations. However, sequential pattern mining algorithms have not been well studied in the context of rough set theory. Extending this approach to sequential pattern mining entails a logical analysis of local patterns in granular computing, which differs from the frequency analysis of sequential patterns.

☆ This paper is an extended version of [13].

* Corresponding author.
E-mail addresses: kaneiwa@cis.iwate-u.ac.jp (K. Kaneiwa), kudo@csse.muroran-it.ac.jp (Y. Kudo).

0888-613X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2011.03.002

In this paper, we propose a sequential pattern mining algorithm using the rule generation from discernibility in rough set theory. This algorithm computes subsequences of a fixed size that are regarded as local patterns hidden inside sequences. A sequential information system consists of the subsequences obtained from a set of sequences so that we can apply sequential data to the rough set data mining. The decision rules generated from a sequential information system are said to be *sequential* decision rules. In each of the rules, the condition attributes represent the occurrences of local patterns in a sequence. In order to estimate the local patterns in the rules, we establish the evaluation of occurrence-based accuracy and coverage for sequential decision rules. This is because the accuracy and coverage measures in rough set theory [18,19,26,27,20] do not evaluate the occurring sequence patterns in each sequence.

Our algorithm for mining local sequence patterns has the following interesting features:

- *Occurrences of Local Patterns:* Given a set of sequences, a sequential information system is constructed from the attributes that denote the subsequences of a fixed size, where each attribute value represents the number of occurrences of a local pattern in a sequence.
- *Granularities of Sequences:* The different sizes of local sequence patterns determine the diversity of granularities in a sequential information system. In other words, longer subsequences correspond to smaller granularities because they contain more information.
- *Reduced and Consistent Decision Rules:* In rough set theory, attribute reduction generates *reduced* decision rules. In addition, the decision rules are *consistent*, and hence, they are significantly different from the frequent association rules in traditional data mining, because logically inconsistent rules are excluded due to the discernibility of decision classes.

In relation to statistical data mining algorithms, these features are important in that they allow us to obtain implicitly local patterns, particularly when the patterns do not appear frequently. This is because each of the minimal subsets of the condition attributes calculated in rough set theory essentially discerns the decision classes among sequences without evaluating their frequencies.

This paper is arranged as follows: Section 2 briefly recalls the basic notions of rough sets. In Section 3, we describe the extension of rough set data mining to sequential pattern mining. We formalize a transformation from a set of sequences into a sequential information system. We then establish the occurrence-based accuracy and coverage of the sequential decision rules generated from the sequential information system. In Section 4, we present our algorithm for mining local sequence patterns from a set of sequences. The experimental results are reported in Section 5. Finally, we discuss related work in Section 6 and conclude this paper in Section 7.

2. Rough sets

An attribute a is a mapping $a\colon U \to V_a$ where U is a non-empty finite set of objects (called the universe) and V_a is the value set of a. An information system is a pair $T = (U, A)$ of the universe U and a non-empty finite set A of attributes. Let B be a subset of A. The B-indiscernibility relation is defined by an equivalence relation I_B on U such that $I_B = \{(x, y) \in U^2 \mid \forall a \in B\,.\, a(x) = a(y)\}$. The equivalence class of I_B for each object $x\,(\in U)$ is denoted by $[x]_B$. Let X be a subset of U. We define the lower and upper approximations of X by $\underline{B}(X) = \{x \in U \mid [x]_B \subseteq X\}$ and $\overline{B}(X) = \{x \in U \mid [x]_B \cap X \neq \emptyset\}$. A subset B of A is a reduct of T if $I_B = I_A$ and there is no subset B' of B with $I_{B'} = I_A$ (i.e., B is a minimal subset of the condition attributes without losing discernibility).

A decision table is an information system $T' = (U, A \cup \{d\})$ such that each $a \in A$ is a condition attribute and $d \notin A$ is a decision attribute. Let V_d be the value set $\{d_1, \ldots, d_u\}$ of the decision attribute d. For each value $d_i \in V_d$, we obtain a decision class $U_i = \{x \in U \mid d(x) = d_i\}$ where $U = U_1 \cup \cdots \cup U_{|V_d|}$ and for every $x, y \in U_i$, $d(x) = d(y)$. The B-positive region of d is defined by $P_B(d) = \underline{B}(U_1) \cup \cdots \cup \underline{B}(U_{|V_d|})$. A subset B of A is a relative reduct of T' if $P_B(d) = P_A(d)$ and there is no subset B' of B with $P_{B'}(d) = P_A(d)$.

We define a formula $(a_1 = v_1) \wedge \cdots \wedge (a_n = v_n)$ in T' (denoting the condition of a rule) where $a_j \in A$ and $v_j \in V_{a_j}$ $(1 \leq j \leq n)$. The semantics of the formula in T' is defined by $[\![(a_1 = v_1) \wedge \cdots \wedge (a_n = v_n)]\!]_{T'} = \{x \in U \mid a_1(x) = v_1, \ldots, a_n(x) = v_n\}$. Let φ be a formula $(a_1 = v_1) \wedge \cdots \wedge (a_n = v_n)$ in T'. A decision rule for T' is of the form $\varphi \to (d = d_i)$, and it is true if $[\![\varphi]\!]_{T'} \subseteq [\![(d = d_i)]\!]_{T'} (= U_i)$. The accuracy and coverage of a decision rule r of the form $\varphi \to (d = d_i)$ are respectively defined as follows:

$$accuracy(T', r, U_i) = \frac{|U_i \cap [\![\varphi]\!]_{T'}|}{|[\![\varphi]\!]_{T'}|}$$

$$coverage(T', r, U_i) = \frac{|U_i \cap [\![\varphi]\!]_{T'}|}{|U_i|}$$

In the evaluations, $|U_i|$ is the number of objects in a decision class U_i and $|[\![\varphi]\!]_{T'}|$ is the number of objects in the universe $U = U_1 \cup \cdots \cup U_{|V_d|}$ that satisfy condition φ of rule r. Therefore, $|U_i \cap [\![\varphi]\!]_{T'}|$ is the number of objects satisfying the condition φ restricted to a decision class U_i.