Finding Context Association Rules over Sensor-Actuator Data Streams

Ho Jin Woo[1], Se Jung Shin[1], Kil Hong Joo[2*] and Won Suk Lee[1],

[1] Yonsei University, 50 Yonsei-ro, Seodaemun-gu, 120-749 Seoul, Korea
[2] Gyeongin National University of Education, San 6-8 Seoksu-dong, Manan-gu,
430-040 Anyang-si, Gyeonggi
{judas, starofu}@database.yonsei.ac.kr, khjoo@ginue.ac.kr, leewo@database.yonsei.ac.kr

Abstract. In a ubiquitous computing environment of sensor networks, monitoring the frequent contexts of a user occurred by those sensors is very important to provide a proactive service to the user. Such frequent contexts of the sensors can be recognized by associations between the set of sensor values and a set of actuator operations. This paper proposes a method of generating a new type of an association rule, called a context association rule, over an online sensor/actuator transactional data stream in order to invoke proper operations of actuators relevant to values of the sensors. To enumerate context association rules, a new prefix tree structure is introduced. It maintains all frequent context itemsets over the current data steam of sensor networks, so that a set of frequently co-occurred sensors and actuators items can be captured efficiently.

Keywords: Sensor/actuator data streams, data mining, association rules

1   Introduction

In a ubiquitous sensor network environment, numerous diverse sensors and actuators are distributed to monitor the inherent behavior of a user on the actuators with respect to the ongoing values of the sensors. A sensor is a device that can capture a particular state of the environment while a context of the environment is described by a set of co-occurred sensor values. An actuator is a device that can convey a service to a user. The behavior of a user on a specific actuator with respect to a context can be modeled by associations between a set of sensors values and a set of actuator operations.

For a set of items I, an association rule is represented by a form of $X \rightarrow Y$, where $X \subseteq I$ and $Y \subseteq I$. The intuitive implication of an association rule is that the presence of an itemset X in a transaction also indicates a high possibility of the presence of an itemset Y in the same transaction. Either a distinct sensor value range or a distinct actuator operation can be regarded as an item. Given a minimum confidence C_{min}, for a frequent itemset e and one of its non-empty subsets q, an association rule $q \rightarrow (e-q)$ is

---

* Corresponding Author

generated only if $S(e)/S(q) \geq C_{min}$ where $S(x)$ is the support of x. Association rules mining has been applied to sensor networks [1, 2]. A sensor association rule [1] is a behavioral pattern proposed to solve the problem of finding temporal associations among sensors in a particular sensor network. A coverage-based rule [2] discovers the correlation of the same sensor event over a number of disjoint network areas by detecting the event during profiling a behavior of the networks as a sensor association rule. However, these two approaches regard a sensor as a minimum unit of information, so that only one temporal association can be identified at most for a sensor.

This paper introduces a new type of an association rule, called a *context association rule*, and it proposes a method of generating context association rules over an online sensor/actuator transactional data stream. A particular context is represented by a set of sensor items and a set of actuator items captured at the same time. Context association rules can associate the sets of temporally correlated sensor values with relevant actuator operations. Since the proposed approach concentrates on a dedicated form of an association rule for a sensor/actuator data stream, the number of rules is much less than that of general association rules. Consequently, a succinct set can be traced more efficiently. In addition, context association rules can improve the mining quality of sensor networks because the set of similar rules is not redundantly included. The rest of this paper is organized as follows: Section 2 describes how a complete set of context association rules is traced over an online sensor/actuator data stream. In Section 3, the performance of the proposed method is evaluated by a series of experiments to identify its characteristics. Finally, Section 4 concludes this paper.

2   Context Association Rules over Sensor/Actuator Data Streams

Given a ubiquitous computing environment U with a set of sensors S and a set of actuators A, the ongoing value of each sensor is sampled and discretized to be an item at a fixed rate and the operations of each actuator for a particular set of discretized sensor items is also captured. An atomic context of a user in the environment of sensor networks is represented by a context transaction which contains actuator operations for a set of sensor items captured at the same time. A current context data stream CD_k is an infinite set of context transactions generated continuously. The context of a particular situation for a user in a ubiquitous computing environment of sensor networks is represented by how the user operates a set of the actuators for a particular situation of sensor values. Therefore, a particular set of sensor values can be regarded as a precondition for a set of relevant actuator operations. To represent this type of a context, a *context itemset* is considered as the union of the values of sensors and actuators. In other words, given the current context data stream CD_k for a ubiquitous environment U with a set of sensor items I_S and a set of actuator items I_A, an itemset e in CD_k is called as a context itemset if and only if $e=e_S \cup e_A$ such that $e_S \subseteq I_S$, $e_A \subseteq I_A$, $e_S \neq \emptyset$ and $e_A \neq \emptyset$. An context itemset e is a *frequent* if the current support $S_k(e) \geq S_{min}$ where S_{min} is a predefined minimum support. For a frequent context itemset, the number of association rules can be reduced greatly. This is because only a proper

subset of sensor items needs to be placed in the left-hand side of an association rule. To address this distinction, a new form of an association rule is shown in Definition 1.

Definition 1. Context association rules. Given the current context data stream CD_k for a ubiquitous environment U with a set of sensor items I_S and a set of actuator items I_A, for a frequent context itemset $e=e_S\cup e_A$, let $p\subseteq e_S$, $q\subseteq e_A$. An association rule $p\rightarrow q$ is a *context association rule* if $S_k(p\cup q)\geq S_{min}$ and $cnt_k(p\cup q)/cnt_k(p)\geq C_{min}$ where S_{min} and C_{min} are a minimum support and minimum confidence, respectively. ∎

To generate all the currently valid context association rules, a new prefix tree called *CAR*(context association rule)-tree is introduced. The basic structure of a *CAR*-tree is the same as that of a prefix tree used in the *estDec* method [3]. Each node of a *CAR*-tree consists of three fileds: *item-name*, *count* and *sensor/actuator type*. The item-name field holds the item of the node and the count field maintains the current occurrence count of the itemset represented by the node. The sensor/actuator type field is additionally included to indicate whether the item is a sensor item or an actuator item. Every node of the *CAR*-tree is traversed in a preorder traversal manner. Two pointers, namely *sensor-actuator boundary pointer P_{sa}* and *consequent pointer P_c* are employed to enumerate every possible context association rule that can be derivable from the tree. At first, starting from the root node, all the nodes of the left-most path are visited by these two pointers P_{sa} and P_c.

When a path from the root to a node corresponding to a frequent context itemset e is visited, e should contain both sensor and actuator items, i.e., $e=e_S\cup e_A= s_1s_2...s_i a_1a_2...a_j$ ($1\leq i\leq m$, $1\leq j\leq n$). Therefore, all the nodes which are in the 1-level with actuator items have no use for generating context association rules. Upon visiting a child node, its item is checked first whether the item is a sensor item or not. If so, the next child node is visited. Only when the item of the currently visiting node is the first actuator item in the path from the root, the context association rule generation is started. When a node v corresponding to a frequent context itemset $e=e_S\cup e_A$ that contain an actuator item $a_1(\in I_A)$ is visited, the pointer P_{sa} is set to pointer at the parent node of v. The current confidence of $s_1...s_i\rightarrow a_1...a_n$ ($s_i\in I_S$, $1<i\leq|e_S|$, $a_j\in I_A$, $n=|e_A|$) is examined by computing $cnt_k(s_1...s_i\cup a_1...a_n)/cnt_k(s_1...s_i)$. Subsequently, all the nodes of the subtree rooted at the node corresponding to s_i are examined. After all the nodes of a subtree rooted at a 1-item s_1 are examined, the subtree rooted at each of its sibling nodes are examined by the same manner. This procedure is terminated when all the 1-level sensor nodes of frequent itemsets are traversed. These nodes are included the context valid range which has items of both I_S and I_A.

3   Performance Evaluations

In this section, the performance of the proposed method is analyzed for the data set *T10.I4.D1000K*. The numbers used in the name denote the average transaction size (*T*), the average maximal frequent itemset size (*I*) and the total number of transactions (*D*) respectively. The transactions of the data set are looked up one by one in

sequence to simulate the environment of an online sensor/actuator data stream. Fig. 1 shows the performance of the proposed method by comparing with the *Assoc-all-naive* method [4]. Fig. 1-(a) illustrates the effects of C_{min} on the average processing time of the proposed method. When the value of C_{min} is set to be higher, fewer itemsets are maintained, so that the processing time is shortened. For the *Assoc-all-naive* method, an additional filtering operation is required to select context association rules from the resulting set. Therefore, the processing time of the proposed method is much faster than that of the *Assoc-all-naive* method. Fig. 1-(b) shows the effects of sensor and actuator items distribution per transaction. When the ratio of the number of sensor items over the number of actuator items is closed to 1, more context association rules are generated, so that the average processing time is enlarged.
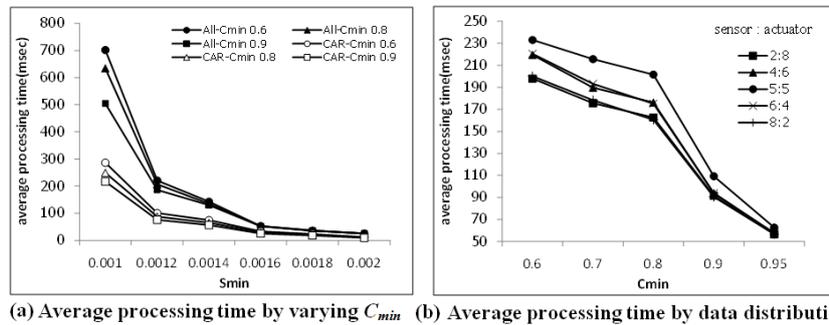

(a) Average processing time by varying C_{min} (b) Average processing time by data distribution

Fig. 1. Performance evaluations

4   Concluding Remarks

This paper proposes an association rule mining technique for a sensor/actuator data stream. A context association rule can capture a set of frequently co-occurred sensors and actuators items for a ubiquitous computing environment over a sensor/actuator transactional data stream. For this purpose, a *CAR*-tree is proposed to enumerate all the context association rules of every frequent itemset efficiently. Experiments results verify that the proposed method can find a set of context association rules efficiently.

References

1. Boukerche, A., Samarah, S.: A Novel Algorithm for Mining Association Rules in Wireless Ad Hoc Sensor Networks. IEEE Transactions on Parallel and Distributed Systems, vol. 19, no. 7, pp. 865--877 (2008)
2. Samarah, S., Boukerche, A., Yonglin, R.: Coverage-Based Sensor Association Rules for Wireless Vehicular Ad Hoc and Sensor Networks. In: IEEE GLOBECOM, pp. 1--5 (2008)
3. Chang, J.H., Lee, W.S.: Finding Recent Frequent Itemsets Adaptively over Online Data Streams. In: 9th ACM SIGKDD, pp. 487--492 (2003)
4. Shin, S.J., Lee, W.S.: On-line Generation of Association Rules over Online Data Streams. Information and Software Technology, vol. 50, issue. 6, pp.569--578 (2008)