An enhanced prefix tree for mining Periodic Associated Sensor Patterns from Wireless Sensor Networks

Arun Pratap Singh Gurjar[1], and *Prof Vikas sejwar[2]
[1]Research Scholar, MITS Gwalior, India
[2]Faculty of CSE/IT, MITS Gwalior, India

Abstract

Wireless Sensor Networks produce a lot of information streams. Mining helpful data from these information streams is a testing assignment. Numerous calculations have been proposed to separate the helpful information from sensor information. Associated Sensor Pattern Mining from DataStream with a conservative tree structure, called Associated Sensor Pattern tree, is for the most part utilized calculation. All-confidence edge based ASP can decide co-events and worldly connection in sensor information. Be that as it may, number of produced examples might be huge if given all-confidence edge is low. In this manner time and space taken to mine these examples won't be proficient. To decrease such costs Periodic Associated Sensor Pattern (PASP) was presented, as often as possible happening Associated Sensor Pattern is called PASP. In this paper, we presented an Enhanced PASP Tree (EPASP) along with a sorted sensor order-list (SSO) for mining Periodic Associated Sensor Patterns from sensor database. We calculated support count of each sensor in sensor database and arranged them into SSO-list in descending order of their support count. Thereafter each epoch is scanned from database and inserted into EPASP tree according to SSO-list. After insertion of all epochs into EPASP tree we recursively mine the EPASP-tree of diminishing size to produced consistently visit patterns by making contingent pattern bases (PB) and relating conditional trees (CT) without extra database examine. Discovering periodic patterns by our methodology is time capable as showed up in a wide examination.

Keywords: *WSN, Data Mining, Associated Sensor Pattern, Periodicity, EPASP.*

I. Introduction

A wireless Sensor Network (WSN) is a remote system which stays of mostly scatter self-governing hardware that utilizes sensors to screen natural and genuine condition [1]. A WSN is an accumulation of vast sensor hubs which can be in hundreds or even a great many little, shabby hubs which are conveyed into a system at a certain area. These sensors have the capacity to detect, process and impart to its associate so as to cooperate in a helpful way. A sensor arrange sensed the gigantic majority of information and send it back to the Sink through passage for preparing reason by utilizing distinctive information mining methods. As of late information mining systems have got a lot of regard for separate captivating information from WSN [2]. These procedures have appeared to be a promising device to enhance WSN execution and nature of administrations [3]. Advances in remote correspondence and microelectronic gadgets prompted the improvement of low-control sensors and the arrangement of substantial scale sensor networks. With the abilities of obvious inspection, WSNs have pulled in huge consideration in numerous applications spaces, for example, living space observing, question following, condition observing, military, calamity administration, and also shrewd conditions. In these applications, real-time what's more, solid observing is the fundamental necessity. These applications yield the tremendous volume of dynamic, topographically conveyed and heterogeneous information. This crude information, assuming proficiently dissected and changed to usable data through information mining, can encourage computerized or human-induced strategic/vital choice. Accordingly, it is basic to create

* Correspondence author Prof Vikas Sejwar
Faculty of CSE/IT, MITS Gwalior
Vikassejwar@mitsgwalior.in

systems to dig the sensor information for designs in the request to settle on smart choices quickly. As of late, extricating information from sensor information has gotten a lot of consideration by the information mining network. Distinctive methodologies concentrating on clustering, association rules, visit designs, consecutive examples, and characterizations have been effectively utilized on sensor information. Be that as it may, the outline furthermore, the arrangement of sensor systems makes remarkable research challenges because of their expansive size (up to a huge number of sensor nodes), random and unsafe organization, lossy imparting condition, constrained power supply, and high disappointment rate. These difficulties make customary mining procedures inapplicable in light of the fact that customarily mining is concentrated and computationally costly, and it centers around plate occupant value-based information. Subsequently, new calculations have been made, and a portion of the information mining calculations have been adjusted to deal with the information created from sensor systems. A plenty of learning discovery methodologies, systems, and calculations have been proposed amid the most recent ten years. Some of them are listed below in II section. WSNs create a lot of information as streams. Accordingly, continuous information streams, locally available the constrained assets and the appropriated idea of sensor systems bring new difficulties for the information mining methods. Using information discovery in WSN one specific activity is to locate behavioral patterns of sensor nodes, which are developed from meta-data describing sensor behaviors. On the off chance that occasions from sensors s1 and s2 are accounted for, at that point there is a 80 percent shot of getting an occasion from sensor s3 and s4 inside λ units of time, the place 80 percent is the recurrence of the affiliation. Generating affiliation rules that have sure frequency wants to generate all the usual patterns present in the records that meet a precise frequency value. The procedure of producing the association rules is easy after identifying the regular patterns. The rule scheme is structured on a constraint termed minimum support threshold which specify minimal lower bound for the support of resulting association rules. It is feasible to extract excessive value information if the minimum support threshold is set high. By making the minimum support threshold low, an extraordinarily giant variety of affiliation rules are generated, most of which are non-informative. In this case, the legitimate correlation in the information objects receives a massive pile of pointless rules. Since WSN generates big amount of data, it is vital to use terrific interestingness measure to discover sensor behavioral patterns that has robust correlation among data. In light of this issue, another sort of sensor’s personal conduct standard, called ASP (associated sensor pattern)[11] was evolved. These behavioral patterns seize temporal correlations, affiliation like co-occurrences, which are linked with such co-occurrences in the sensor data.

Another necessary criterion for figuring out the interestingness of related patterns may be the form of occurrence, i.e., whether they show up periodically, non-periodically, or by and large in precise time interval in the sensor database. An ASP that occurs after periodic intervals in WSNs called as PASP (periodic associated sensor patterns). Periodic associated sensor patterns can be used for predicting the source of future events. By understanding the origin of future event, we can become aware of the inaccurate nodes easily from the network. For example, we are expecting to get event from a particular node, and it does not occur. It also may be used to identify the source of the next event in the case of emergency preparedness class of applications. Periodic associated sensor patterns also can identify a set of temporally correlated sensors. This information can be beneficial to overcome the undesirable outcomes (e.g., neglected reading) of the unreliable wireless communications.

We proposed an enhanced prefix tree (EPASP) along with Sorted Sensor order-list (SSO) for mining Periodic Associated Sensor pattern from Wireless Sensor Network’s sensor data. A SSO-list maintains all sensors in descending order of their support count in SD. Our execution contemplate demonstrates that our proposed approach is time efficient in finding periodic associated sensor patterns. The paradigm of paper is as per the following: Section 2 bargains with the writing overview. Section 3 discusses the problem description. Section 4 clarifies about the proposed approach. Section 5 shows up the broad examination results of our approach and Section 6 portray the finish of the entire paper.

II. Literature Work

Y.K.Lee et al. [4] worn all-confidence to find out correlated patterns as it satisfies each null-invariance and descending closure belongings. Although CoMine algorithms are environment-friendly for transactional database, it is not appropriate for sensor information flow because of more than one scan requirement of the unchanged database.

Loo [5] and **Romer** [6] have concentrated on extricating design with respect to the event detected by the sensor hubs, in which the mining systems are connected to the detected information got from the sensors and put away in a Focal database.

Leung et al. [7] use a novel tree structure called Canonical-order tree (CanTree). However, for the frequency impartial canonical-order object insertion, CanTree achieves much less compactness and for this reason, results in terrible mining performance than FP-tree.

Boukerche et al. [8] proposed Sensor-Association rules and made use of a Positional Lexicographic Tree (PLT). Where patterns are extracted with respect to the sensor hubs instead of the region checked by the WSN and PLT is used to save a sensor's event detecting status. A case of sensor affiliation principles could be (s_1, $s_2 \rightarrow s_3$, 75%, λ) which implies that if sensor s_1 and s_2 identify occasions inside λ time interim, at that point there is 75% of chance that s_3 recognizes occasions inside same time interim. Notwithstanding, these guidelines regularly produce countless, the greater part of which are non-informative or failed to mirror the genuine relationship among sensors.

S. K. Tanbeer et al. [9] makes use of a unique tree structure, recognized as CP-tree (Compact Pattern tree) that captures information data with one scan within the insertion phase and offers the equal mining performance like FP-growth strategy within the restructuring section. However, it is not possible to locate related patterns it only mines frequent patterns.

R. U. Kiran et al. [10] in order to enhance the performance a novel notion referred to as items support interval and an algorithm CoMine++. To find out correlated patterns effectively CoMine++ is used. Using the prior information regarding the construction and mining of FP-tree, CoMine++ makes use of a novel pruning approach to construct the CPB of the suffix item well.

M. M. Rashid et al. [11] uses a new type of behavioral pattern known as associated sensor patterns(A pattern is referred to as an associated pattern, if its all-confidence is higher than or equal to the given minimal all-confidence threshold), To seize this kind of patterns a mining algorithm called Associated Sensor Pattern and a compact tree structure, referred to as Associated Sensor Pattern tree is used. Associated Sensor Patterns seize association-like co-occurrences as well as temporal correlations which are linked with such co-occurrences. In ASP when information flow flows, historic information might also lose importance for the cutting-edge time. ASP-tree is in addition improved to SWASP-tree by using the capability of adopting sliding remark window to capture the significance of the latest information. The challenge of SWASP is that it takes plenty of time to mine associated patterns for a giant dataset.

Another necessary criterion for figuring out the interestingness of associated patterns would possibly be the structure of occurrence, i.e., whether or not they manifest periodically, non-periodically, or commonly in a precise time interval in the sensor database.

Tanbeer et al. [12] proposed an algorithm to mine periodic regular pattern from the transactional dataset. For periodicity calculation, they used maximum period (maxPrd) of an interval of the patterns. But in many real-world applications, it is hard for the patterns to show up commonly without any interruptions. Therefore in an inaccurate or noisy environment, maxPrd measure for regularity calculation is not effective.

Rashid et al. [13] have delivered a problem of discovering regularly frequent patterns that comply with a temporal regularity in their prevalence characteristics from sensor data. They used the variance of interval time between pattern occurrences in a database as a choice of maxPrd. In this model, at first, they mine the frequent patterns and then find out the regular patterns. Therefore, the massive amount of candidate patterns is generated and mining these large numbers of patterns may not environment-friendly in actual time.

Rashid et al. [14] recently developed a single-pass tree structure, referred to as the PASP-tree (periodic associated sensor patterns tree), that can seize vital know-how from the flow contents of sensor data in a very compact manner.

A pattern is referred to as a periodic associated sensor pattern if it satisfies each of the following three conditions:

(i) Its all-confidence value is no less than a user given min all confidence.
(ii) Its support count is not much less than a user given min sup.
(iii) its periodicity value is not higher than a user given max var.

PASP-tree can correctly mine the associated sensor patterns in sensor stream data for the user given min all conf, min sup and max_var thresholds. In this paper, we proposed an enhanced prefix tree (EPASP) along with Sorted Sensor order-list (SSO) for mine Periodic Associated Sensor pattern from Wireless Sensor Network's sensor data. A SSO-list maintains all sensors in descending order of their support count in SD. Our execution contemplate demonstrates that our proposed approach is time efficient in finding periodic associated sensor patterns.

III. Problem description

The PASP mining procedure comprises of two stages, insertion stage and Restructuring stage. For the insertion stage, PASP-tree organizes the sensors in agreement to the sensor's appearance arrange in the database. PASP-tree is built through embeddings each epoch in a database in a steady progression. However, an epoch may consist of the number of sensors whose support values are less than user precise min_sup value. This ends up by establishing a massive PASP-tree. As, in the restructuring phase, we have to deal with those sensors also who do not satisfies the user given min_sup value, it will take plenty of time to restructure that tree and to mine periodic associated sensor patterns. To overcome this deficiency EPASP-tree is introduced.

IV. Proposed enhance approach for PASP mining in WSN

Given a SD, min_sup, min_all_conf and max_var requirements, the objective is to gather a minimized prefix tree that discovers the whole arrangement of intriguing patterns in SD having support no less than min_sup and all confidence no less than min_all_conf and periodicity close to max_var. Presently we portray the mining procedure of our proposed calculation. Very first, we calculated the support hold of every sensor in SD. After that in addition stage, before the inclusion of an epoch into the PASP-tree, we arranged the sensors of every epoch into SSO-list as per their support esteems in SD. In the wake of arranging every one of the sensors in the sorted sensor order list (SSO), we dropped those sensors whose support esteem is not as much as the user given min_sup value. If a tail node fails to stand true against user given support count than the time slot value is shifted to the subsequent node and tail node dropped. Hence a node may be a tail node for one or more epochs in SD, such node contains multiple time slot values. Since the number of sensors in SD will be diminished now, our EPASP-tree will be conservative as it contains only those sensors whose support count is equal to or higher than user given min_sup values. Since sensors of each epoch are inserted into EPASP-tree according to sorted sensor order list (SSO), EPASP-tree is already sorted and does now not need restructuring anymore. Hence EPASP-tree will take less time to mine periodic associated sensor pattern. Thereafter like the PASP mining approach[14], we recursively mine the EPASP-tree of diminishing size to produced consistently visit patterns by making contingent pattern bases (PB) and relating conditional trees (CT) without extra database examine. At that point produce the successive pattern from the conditional tree. At last, we check the consistency of produced visit pattern to discover frequent sensor pattern.

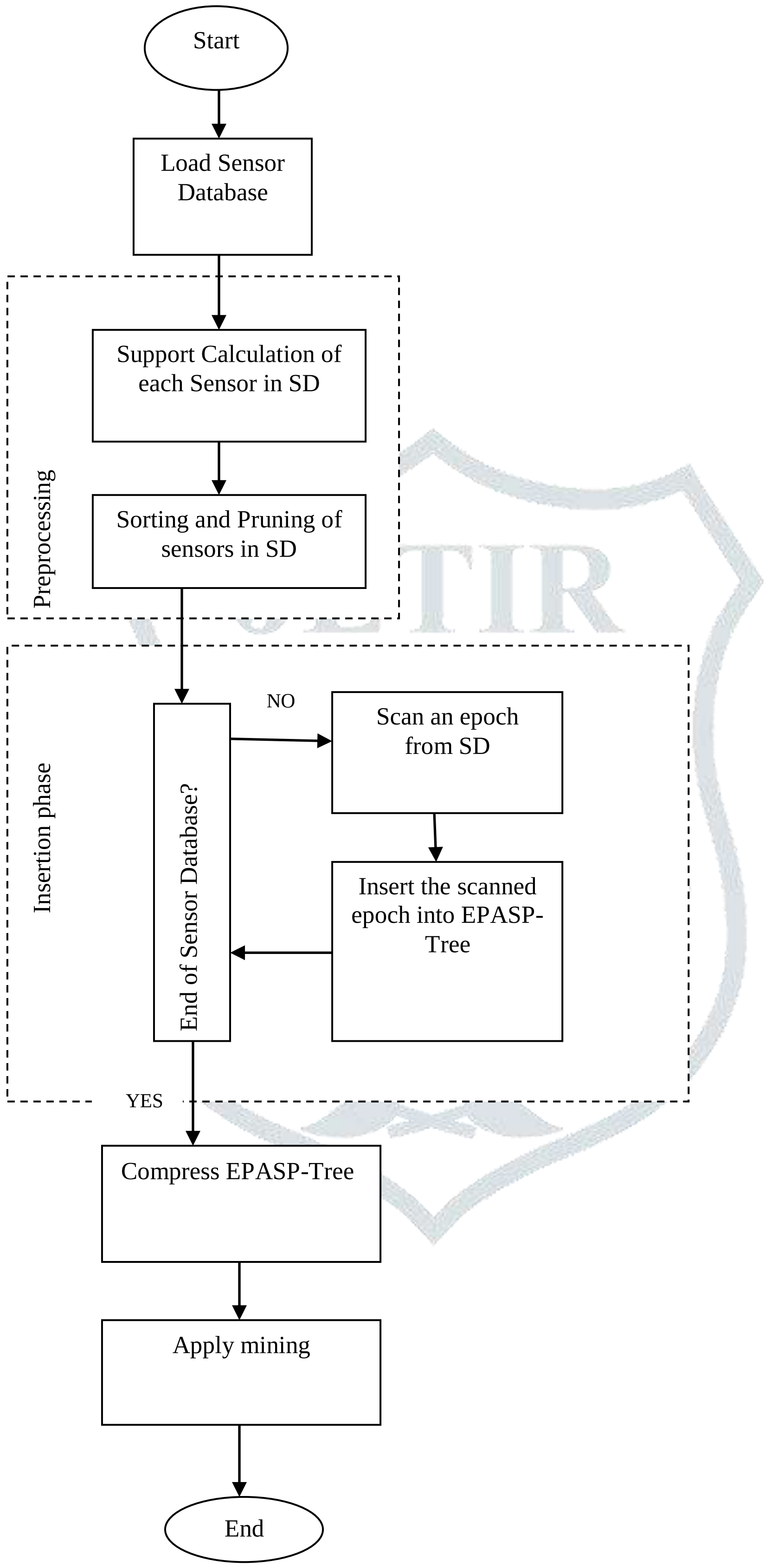


Fig.1 Proposed Flow Diagram

Algorithm: Enhanced PASP algorithm

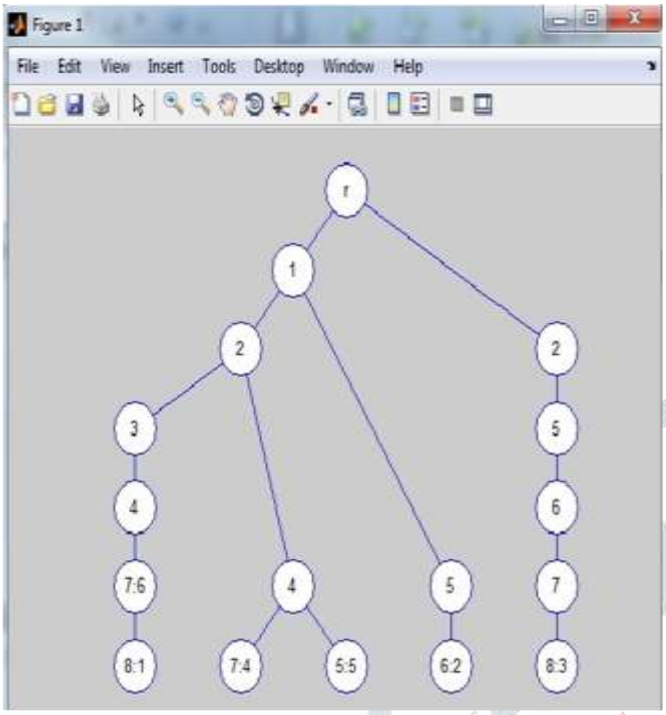
Input: Sensor database (SD), min_sup, min_all_confidence, max_var.
Output: Complete set of periodic associated sensor patterns
1: **Begin**
2: Select the Sensor database SD
3: Calculate the support count of each sensor in SD
4: Arrange all sensor into SSO in descending order of their support count
5: Remove sensors whose support count is less than min_sup
6: PASP-tree a prefix-tree with null initialization
7: **while** (Not end of SD) **do**
8: Scan an epoch from the current location in SD;
9: Insert the scanned epoch into PASP-tree according to SO by following the tree construction method
10: **end while**
11: **for** each branch in PASP-tree **do**
12: Identify the equal support node in each branch and merge them to a single node
13: **end for**
14: **while** any mining request from the user **do**
15: **for** sensor S from the bottom of SSO-list **do**
16: Call Mining (ASP, max_var, periodicity)
17: **end for**
18: **end while**
20: **End**

For a given sensor database SD (Table 1.), min_sup=3, min_all_conf=0.55, and max_var=1, we examine the structure of PASP and EPASP tree after the Insertion phase.

Table 1. Sensor Database

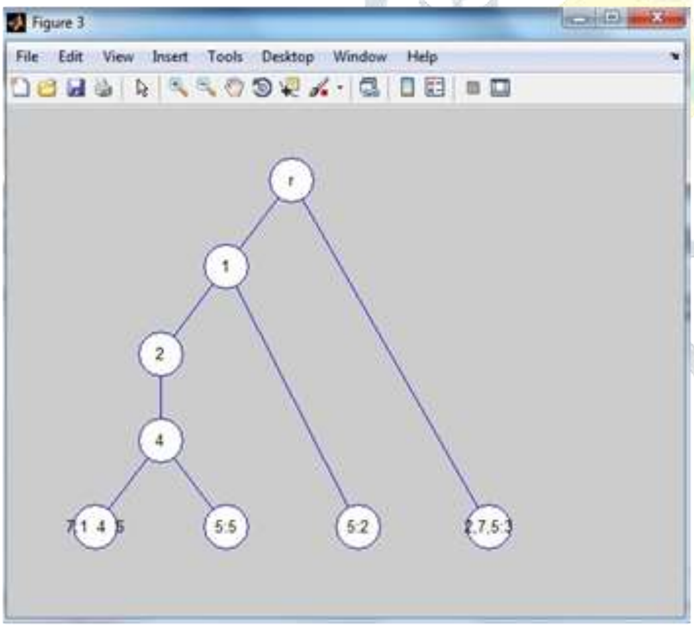
TS	Epoch
1	S_1 S_2 S_3 S_4 S_7 S_8
2	S_1 S_5 S_6
3	S_2 S_5 S_6 S_7 S_8
4	S_1 S_2 S_4 S_7
5	S_1 S_2 S_4 S_5
6	S_1 S_2 S_3 S_4 S_7

However the sensors S_2, S_3, and S_8's support counts are less than min_sup esteem, figure 2 clearly shows that PASP-tree still considers sensors S_2, S_3, and S_8. This leads us to have a big size prefix tree and increased complexity at the restructuring phase. Thus tree sorting becomes an essentiality of PASP-tree.



SO-list
S_1:5
S_2:5
S_3:2
S_4:4
S_7:4
S_8:2
S_5:3
S_6:2

Fig.2 PASP-tree after insertion of all epochs and Unsorted Sensor Order list



SSO-list
S_1:5
S_2:5
S_4:4
S_7:4
S_5:3

Fig.3 EPASP-tree after insertion of all epochs and Sorted Sensor Order list

While in EPASP-tree as shown in figure 3 is already sorted, as sensors are arranged in sorted sensor order list in descending order of their respective support counts, prior to insertion of epochs into prefix tree. EPASP-tree doesn't contain the sensors whose support count is not as much as given min_sup value.

So the compactness of the prefix tree is increased and also it doesn't need sorting & restructuring anymore.

V. Experimental Results analysis

In this segment, We analyzed our experimental results to check the change in runtime EPASP-tree accomplishes over the best in class calculations[14] proposed for PASP mining from WSNs. The examinations were held utilizing the datasets regularly utilized in PASP mining executions. We thought about both IBM artificial information (T10I4D100K) and genuine dataset (mushroom) acquired from[15]. All projects are composed in MATLAB R2013a and keep running with Windows 7 (32-bit) on a 2.80 GHz CPU and 2 GB main memory. Circumstance and stuff in these datasets are like the epochs and sensors in the lingo of this paper. Here, we demonstrate the viability of EPASP-tree in mining frequent sensor patterns regarding execution time. To examination the execution time, tests were led with a digging demand for the given datasets by fluctuating the min sup and max var esteems. In our examination, for both PASP-tree and EPASP-tree we settled min all conf = 10% and max var = 0.4% and fluctuated min sup esteems. The x-pivot in each graph below demonstrates the difference in min sup in the form of level of database estimate and the y-pivot shows the general execution time. The general execution time is the aggregate of the development time, tree rebuilding time (for PASP-tree just) and the mining time.

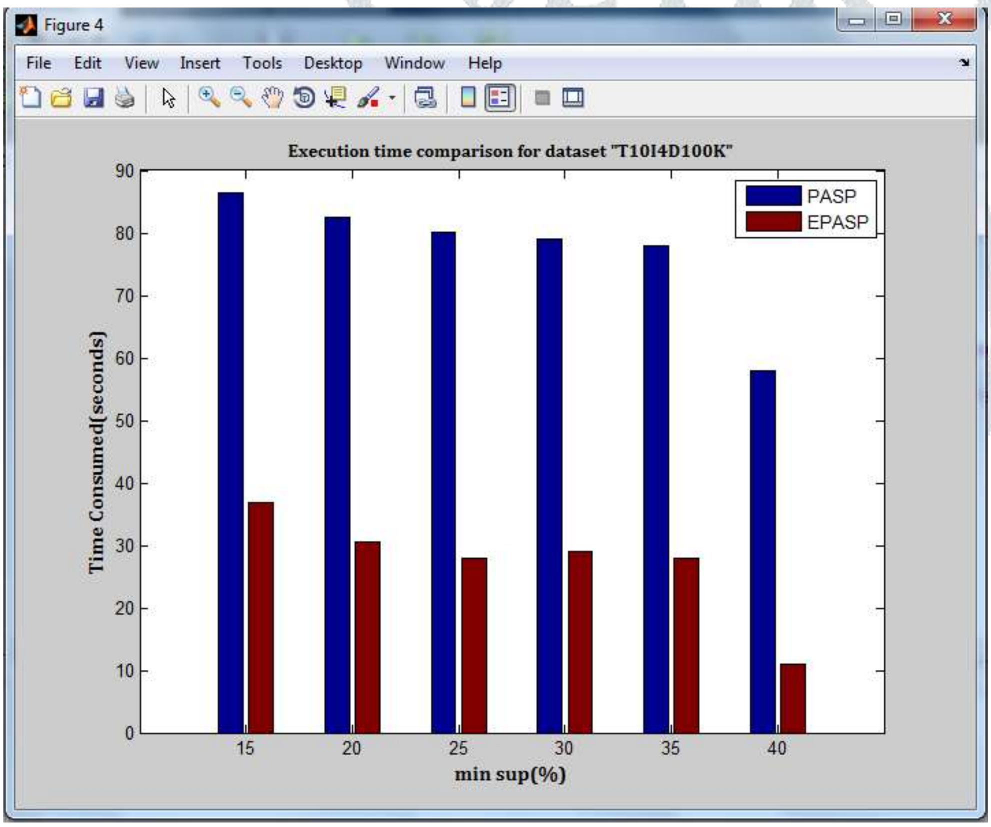


Fig.4 T10I4D100K

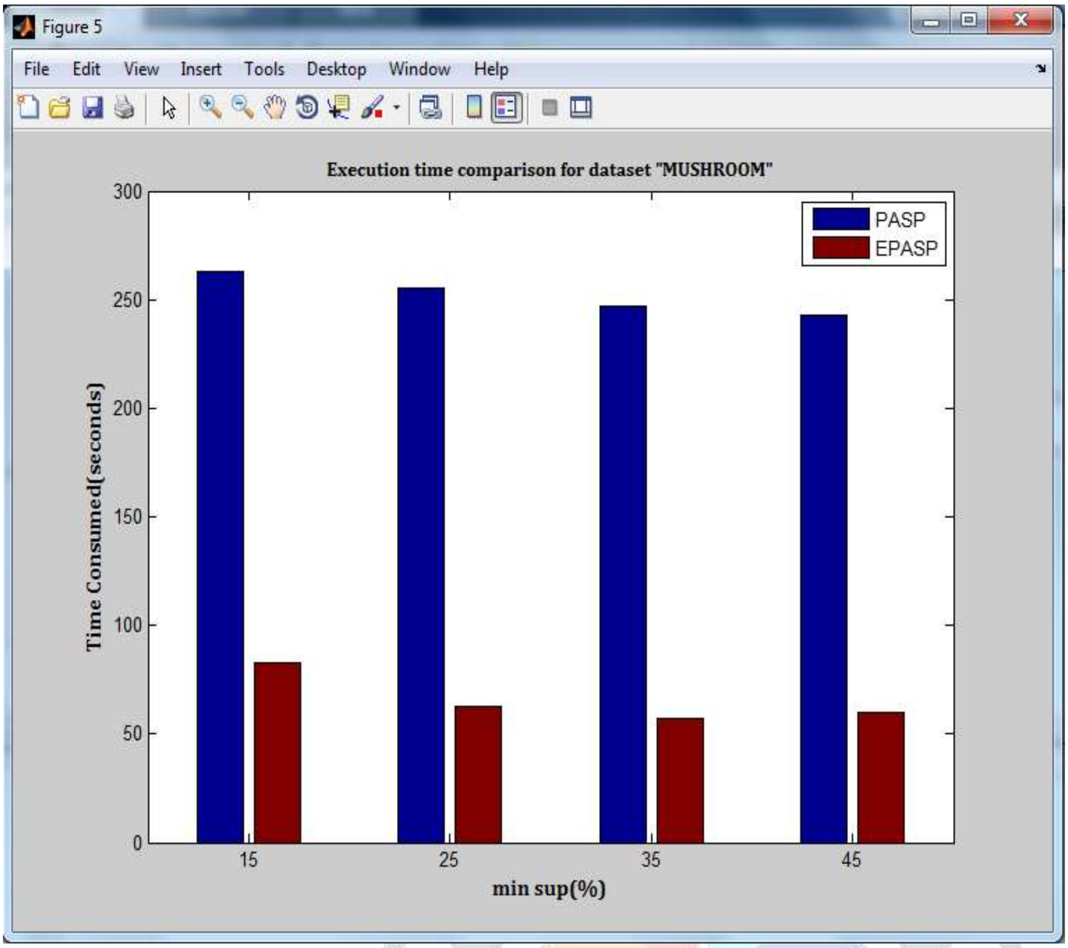


Fig.5 MUSHROOM

Fig 4 & Fig 5 demonstrate that, when the min sup esteem is expanded the digging time for the two trees diminished, however, EPASP-tree performs better than the PASP-tree in the execution time all of the cases. Since, EPASP-tree deals with less number of sensors; in this way, its general runtime is better than PASP-tree in every case.

VI. Conclusion

The key commitment of this paper is to give a productive strategy to mining periodic associated sensor patterns from WSNs information utilizing a prefix tree called EPASP-tree. We assessed the execution of EPASP-tree over various datasets. Broad execution examinations demonstrate that EPASP-tee is exceptionally effective for periodic associated sensor pattern mining and essentially superior to the current strategy. The proposed procedure is appropriate for application in numerous other spaces.

References

[1] Y. Jennifer, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," Computer Networks, pp. 2292-2330, 2008. doi>10.1016/j.comnet.2008.04.002

[2] Mahmood, A., Shi, K., Khatoon, S., Xiao, M.: Data mining techniques for wireless sensor networks: a survey. Int. J. Distrib. Sens. Netw. 2013 (2013). doi:10.1155/2013/406316

[3] Tan, P.-N.: Knowledge discovery from sensor data. Sensors 23, 14–19 (2006).

[4] Y.K.Lee, W.Y.Kim, Y.D.Cai, and J.Han, "CoMine: Efficient mining of correlated patterns, "Proc.Int.Conf.Data Mining, Melbourne, FL, USA, 2003, pp.581-584.

[5] Loo, K.K., Tong, I., Kao, B.: Online algorithms for mining inter-stream associations from large sensor networks. In: Ho, T.B., Cheung, D., Liu, H. (eds.) PAKDD 2005. LNCS, vol. 3518, pp. 143–149. Springer, Heidelberg (2005). doi:10.1007/11430919 18

[6] Romer, K.: Distributed mining of spatio-temporal event patterns in sensor networks. In: EAWMS/DCOSS, pp. 103–116 (2006).

[7] C. K. S. Leung, Q. I. Khan, Z. Li, and T. Hoque, "CanTree: A canonical-order tree for incremental frequent-pattern mining," Knowl.e Inf. Sys., vol. 11, pp. 287–311, 2007. doi:10.1007/s10115-006-0032-8

[8] Boukerche, A., Samarah, S.A.: Novel algorithm for mining association rules in wireless ad-hoc sensor networks. IEEE Trans. P & D. Syst. 865–877 (2008). doi**:** 10.1109/TPDS.2007.70789

[9] S. K. Tanbeer, C. F. Ahmed and B. S. Jeong, "An efficient single-pass algorithm for mining association rules from sensor networks wireless," IETE Tech. Rev., vol. 26, p. 280, 2009. doi: 10.4103/0256-4602.52997

[10] R. U. Kiran and M. Kitsuregawa, "Efficient discovery of correlated patterns in transactional databases using items' support intervals," In Proc. 23rd Int. Conf. Database Expert Syst. Appl., 2012, pp.234–248. doi:10.1007/978-3-642-32600-4_18

[11] M. M. Rashid, I. Gondal, and J. Kamruzzaman, "Mining Associated Patterns from Wireless Sensor Networks," in IEEE Transactions on Computers, vol. 64, no. 7, pp. 1998-2011, 1 July 2015. doi: 10.1109/TC.2014.2349515

[12] Tanbeer, S.K., Ahmed, C.F., Jeong, B. S., and Lee, Y. K.: Discovering periodic frequent patterns in transactional databases. In: Theeramunkong, T., Kijsirikul, B., Cercone, N., Ho, T.-B. (eds.) PAKDD 2009. LNCS (LNAI), vol. 5476, pp. 242–253. Springer, Heidelberg (2009). doi:10.1007/978-3-642-01307-2 24

[13] Rashid, M.M., Gondal, I., Kamruzzaman, J.: Regularly frequent patterns mining from sensor data stream. In: Lee, M., Hirose, A., Hou, Z.-G., Kil, R.M. (eds.) ICONIP 2013. LNCS, vol. 8227, pp. 417–424. Springer, Heidelberg (2013). doi:10. 1007/978-3-642-42042-9 52

[14] Rashid M.M., Kamruzzaman J., Gondal I., Hassan R. (2017) Periodic Associated Sensor Patterns Mining from Wireless Sensor Networks. In: Liu D., Xie S., Li Y., Zhao D., El-Alfy ES. (eds) Neural Information Processing. ICONIP 2017. Lecture Notes in Computer Science, vol 10638. Springer, Cham. doi:10.1007/978-3-319-70139-4_25

[15] Frequent itemset mining repository. http://fimi.cs.helsinki.fi/data/.

[16] Goethals B. (2005) Frequent Set Mining. In: Maimon O., Rokach L. (eds) "Data Mining and Knowledge Discovery" Springer, Boston, MA. doi: 10.1007/0-387-25465-X_17

[17] Robert Nisbet, John Elder and Gary Miner "Statistical Analysis and Data Mining Applications", Elsevier Inc., 2009, ISBN: 978-0-12-374765-5. doi:10.1016/B978-0-12-374765-5.X0001-0

[18] Oded Maimon Lior Rokach "Data Mining and Knowledge Discovery Handbook" Springer, Boston, MA, 2010, ISBN: 978-0-387-09823-4, doi: 10.1007/978-0-387-09823-4

[19] Narendra Kumar Kamila "Wireless Sensor Network Trends, Technologies, and Applications" IGI Global, 2016, ISBN:9781522505013.

[20] Data Mining: Concepts and Techniques, Third Edition by Han, Kamber & Pei (2013)

[21] Data Mining and Analysis Fundamental Concepts and Algorithms, by Zaki & Meira (2014)

[22] Data Mining: The Textbook, by Aggarwal (2015)