ISSN 2515-8260 Volume 08, Issue 02, 2021

Slide Window Method Adapted for Privacy-Preserving: Transactional Data Streams

Jayendra Kumar
Assistant professor, CSE Department,Anurag Group of institutions.
Corresponding Author :jayendrakumarcse@cvsr.ac.in

Abstract:
Data streams mining on transactional data is very attractive area for researcher, identical about transactional data makes compromise the privacy of individual so identical information must be removed from transactional data. For publish static transactional streams many privacy- preserving techniques proposed ,these methods are not straightforwardly applied on transactional data streams because it have different characteristics .in sliding window addition and removal of transaction leads be unsuccessful to satisfy ρ-uncertainty . to maintain ρ-uncertainty remove the items of window, due this heavy information loss we proposed algorithm which dynamicallyselect items for remove to maintain satisfy ρ-uncertainty with less information loss and continuously make satisfy p-uncertainty of slide window with suppression of anonymize sliding window , experimental shows our method is more efficient than batch existing batch processing anonymization sliding window methods

I. Introduction

Maintain Individual privacy while publishing helpful information many tools and methods needs provide privacy-preserving. a moment ago, industry and academia are pay extensive attention and proposed methods are developed for data publishing situations[1,2,3,4,5,6,7].customer transactional data which consist of set of values haggard from list of items [8 ,9].With the beginning of big data, data coming continuously with uncontrolled and growing dataset ,such called data stream ,many data mining patterns can extracted from this this transactional data stream[10] ,such as frequent item set ,which mining from many applications such a online e-commerce ,retail chain, web server logs, click stream[11] and computer network . this raw transactional data disclosing person privacy, identical about transactional data makes compromise the privacy of individual so identical information must be removed from transactional data. For publish static transactional streams many privacy- preserving techniques proposed ,these methods are not straightforwardly applied on transactional data streams because it have different characteristics . its is more difficult than that for publishing traditional transactional static datasets.

A special data model is used for extraction of transaction from data streams, the data streams is continuously generated, this data model for Data stream mining techniques the damped window model(DW) .

ISSN 2515-8260 Volume 08, Issue 02, 2021

Many applications are use sliding window for data stream analysis,so we take privacy problem with sliding window model for data streams. This stream data used by data miners and attacker also,if attacker have knowledge about data in transactional of victim, attacker may find sensitive information of victim,here I takes one example a e-commerce company give this transactional data to a data mining company for finding patterns, let A person from data mining company and his know person B and his brought a item x,y,z from e-commerce company ,A find out only one transaction contain item x,y ,z.so A easily find its belongs to his friend B. along with items it have other sensitive data like AIDS test ,so B privacy is leaked

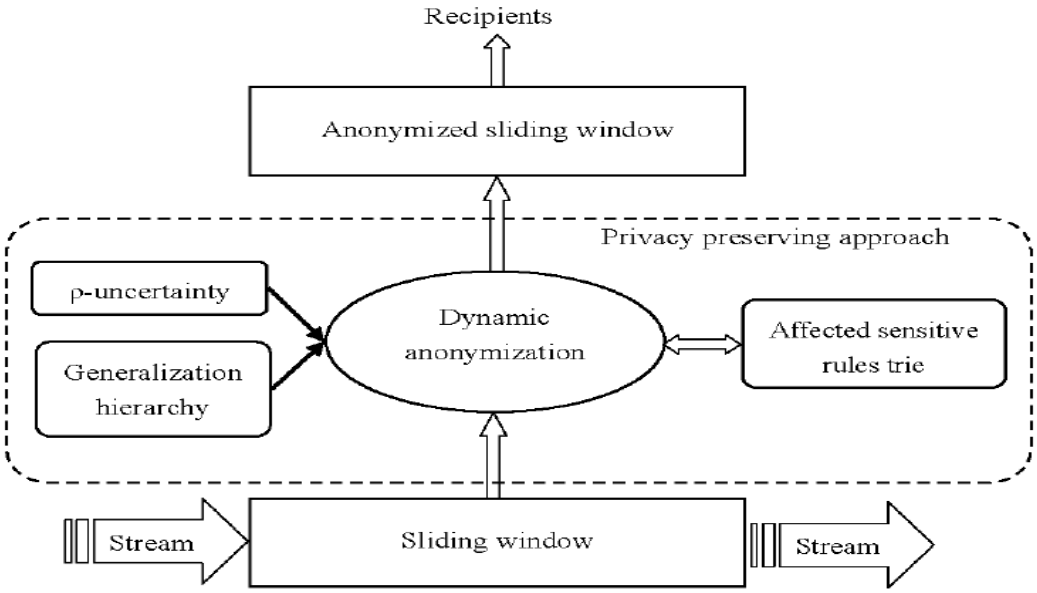


Fig 1: show data stream mining using sliding widow model

For existing transactional data mining technique are divided into two way that are they organizing data non-sensitive and sensitive [8] .data steam mining are extract pattern from newly arrived data only. So handling new date is very important in specially data stream environments. Sliding window approach is efficient for data streams mining, the traditionally method for anonymization data streams not suitable for preserving privacy ,because the sliding window is frequently updated .here item set S={ i1, i2, . . . , in} is limited set in which sensitive items are S_i and non-sensitive are S_n , TID={T_1,T_2,…T_M}… is infinite transactional,the sliding window represent with W_p the size is x ,where p represent it updated window with p size of data .we represent it following fig2.

ISSN 2515-8260 Volume 08, Issue 02, 2021

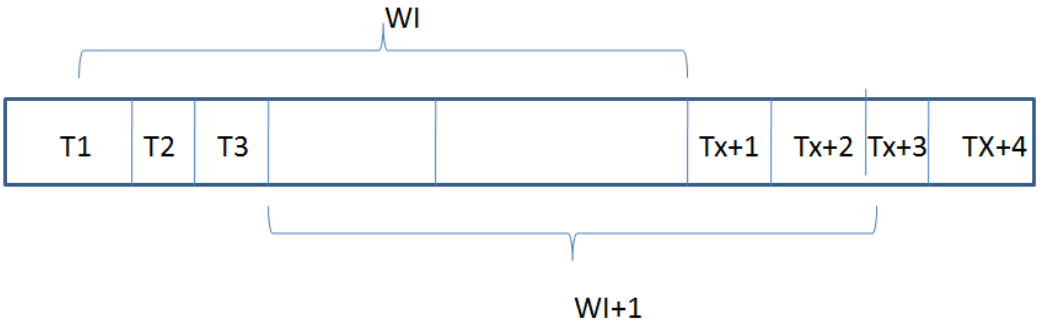


Fig 2: sliding window in data stream

In [18] is introduce on static transactional data privacy model proposed called as p-uncertainty, it is attacker knows some item sets *t* of transactional data which is monitor, so inferred sensitive data with this knowing data *t* with probability of $p>0$ and it *t->sn confidence also* $>p$, in data stream the sliding window data is deleted and add of p data, to avoid this we require anonymize sliding window transactional data confidence greater than threshold value.

In fig 2 sliding window step size is 3 and window size is *x which* contain non-sensitive data with window is updated mean T1,T2 and T3 are deleted and Tx+1,Tx+2 and Tx+3 are add into sliding window ,but in newly add transactions have sensitive data so data stream miner can inferred sensitive data by making relationship with knowing data so we need to compromised with privacy. Which dynamically and continuously make satisfy p-uncertainty of slide window with suppression and generalization

II. Related Work

Preserves privacy in generation of association rules in transactional data addressed by many researchers [2],but they methods based on perturbation [4,8,10,18,21,24] techniques, therefore data integrity damaging and false association rules generated [27],but they not stop sensitive rues(inferences),data stream data are continuous and unbounded[25],the Cao et al [25] done some research on relational data streams to privacy-preserving, through I-diversity and k- anonymization [26] cluster framework was presented with delay assurance with maximum between anonymized output and incoming data ,in [28]by time constraints on transactional data publication to reflect on k-anonymization move toward for data stream clustering and reuse the cluster, to decrease information loss and speed up the anonymization process in [29] SKY method proposed which constraints for privacy protection,in [30] create cluster on data stream data k-anonymizing approach,[31] delay free anonymization method on health data streams .

Extracting patterns from recent arrived data is very important like stream data, in data mining researcher focus on these tasks. Generated patterns from such kind of data a common approach was sliding window [11], above anonymization methods are applicable to data streams and not satisfies privacy requirements and privacy leakage,in sliding window approach data is updated rapidly so maintain protection of privacy is big problem in [32] wang et al proposed SWAF framework to solve this problem by constantly facilitating k-anonymity on sliding window. In

[33] incrementally anonymize a sliding window under the LKCprivacy requirements[34] ,with low data loss consider privacy preserving publishing data continuously in sliding window.

III. MinLossSupression Method

With global suppression method void the generation false association rules [37],delete item from all transactions which contain item , in [18] suppress control initially with sliding window method, due to addition and removal data from sliding window it not satisfy theρuncertainty,we proposed algorithm to continuously maintain theρ uncertainty below threshold value of SAR association rules, SAR association are have the sensitive items in either sides rules$SAR = \{A \rightarrow \alpha | \alpha \in I_S \}$.

In sliding window methods transaction added and some are deleted ,the delete the transaction may give information about newly added transaction so before delete the transaction find the minimum information loss due to suppress the item ,for that find which item are have maximum payoff ,this payoff$\left\{\frac{Count(b,SAR^1)}{Sup(b)inTSW_{i+1}}\right\}$With respective currently window with new added transactions.Maintain a global suppression item, its frequently updated $i_{sup}$with newly find maximum payoff items.

Algorithm: MinLossSupression

Input:TDS stream data,TSW_i, *current window transactions,*$\rho thresholdvalue$

Output:anonymization windownTSW_{i+1}

1. *slide* TSW_i, *over TDS by add* T_{add}
2. $TSW_{i+1} = TSW_i + TSW_{add}$
3. *Delete* i_{sup} *from* TSW_{i+1}
4. $SAR = \{A \rightarrow \alpha | \alpha \in I_S \}$
5. $SAR^1 = \{r | r, r \in SAR, conf(r) > \rho\}$
6. *While* $SAR^1 \neq \varphi$
7. *Find max* $\left\{\frac{Count(b,SAR^1)}{Sup(b)inTSW_{i+1}}\right\} forallb \in r$
8. *Delete b from* TSW_I,*add b to* i_{sup}
9. *Update* SAR^I
10. *Return* $TSW_{i+1} = TSW_{i+1} - TSW_{del}$

IV. Results And Discussion

To evaluate our proposed algorithm with SuppressionControlin [1] in terms of efficiency and data quality.Our evaluation with static anonymization with batch processing .we implement our algorithm in Matlab Environment and run on windows7 with 8 GB primary memory(RAM) and four-core 3.2.GHz central processor Unit(CPU).

TABLE 1. Two data set Description

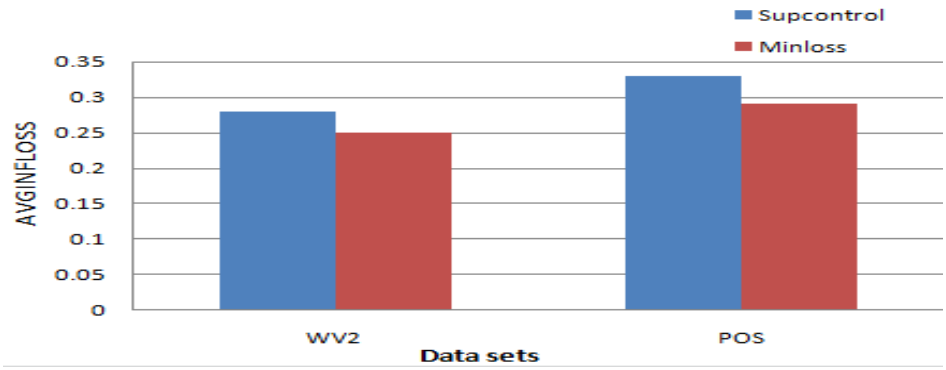
Data sets	Size |D|	Items|I|	Max|t|	Averg|t|

BMS-POS	306984	1178	5	2.64
BMS-WebView	77513	3341	161	5.0

In [38] introduce two data set ,those are BMS-POS and BMSwebVew-2 these are very point of reference in data mining community , For our experiment we use these data sets.Description of these two data sets characteristics show in table 2,the characteristics are avg|t| and max|t| represents average and maximum size, correspondingly. Items in data sets are not differentiated as non-sensitive and sensitive. We consider 40% of items as sensitive and The default values of the window size w, step size p and privacy requirement are 10000, 50 and 0.5, respectively

By varying window size w, privacy requirement ρ ,step size P and two data observe average information losses equation (1) show in Fig 4 To 7

$$AvgInfoLoss = \frac{\sum_{i=1}^{w} infoLoss(TSW)}{w} \qquad (1)$$

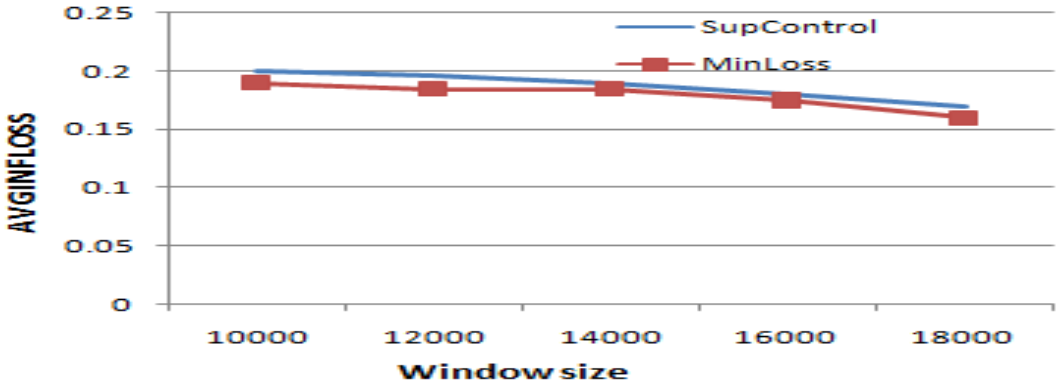


a)Information Loss

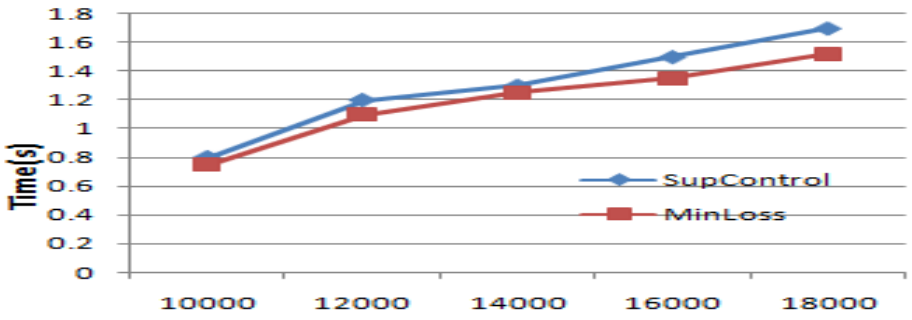
b) Execution Time

Fig 3 :The performance by varying datasets

ISSN 2515-8260 Volume 08, Issue 02, 2021

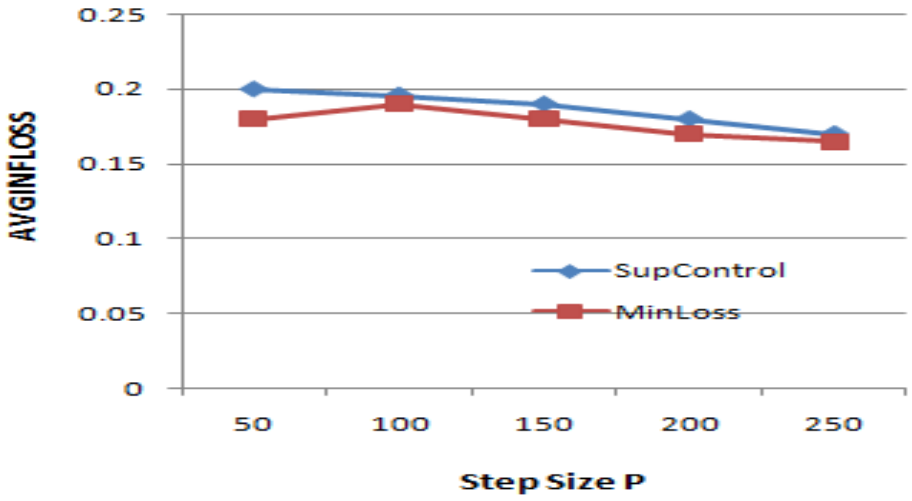


(a) Information loss



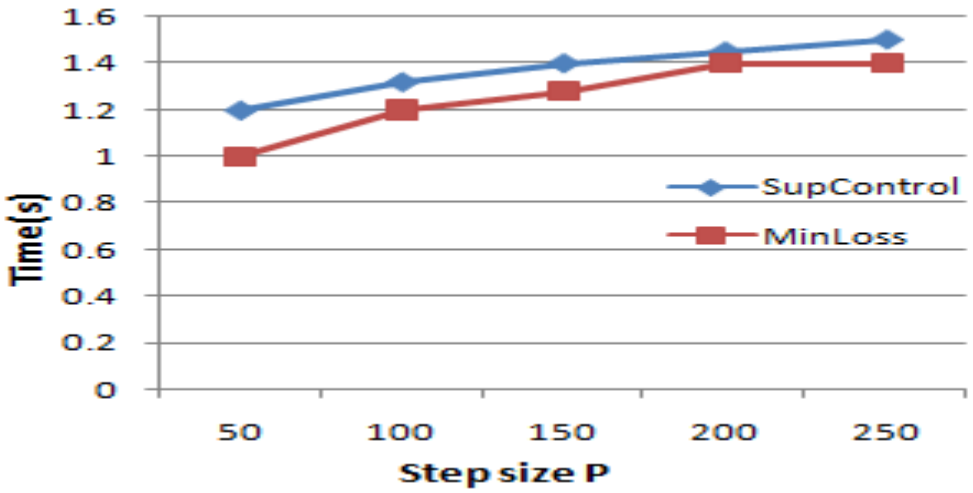
(b) Execution time

Fig 4 The performance by varying window size w



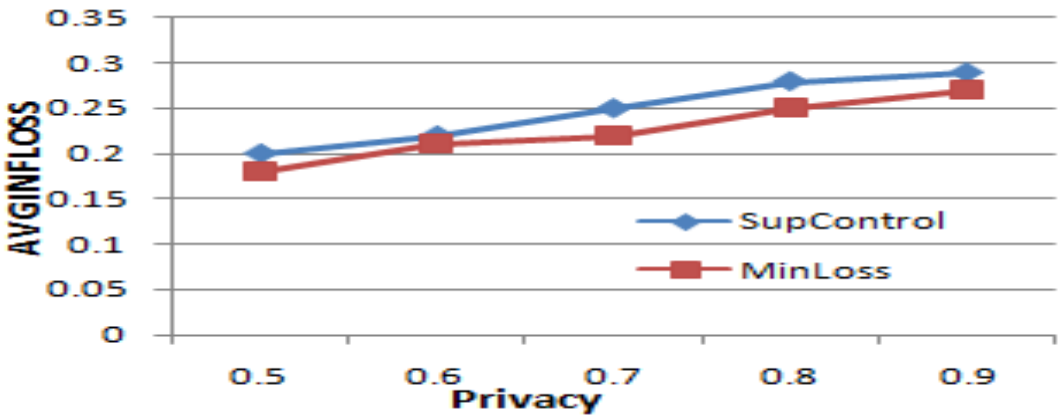
(a) Information loss

ISSN 2515-8260 Volume 08, Issue 02, 2021



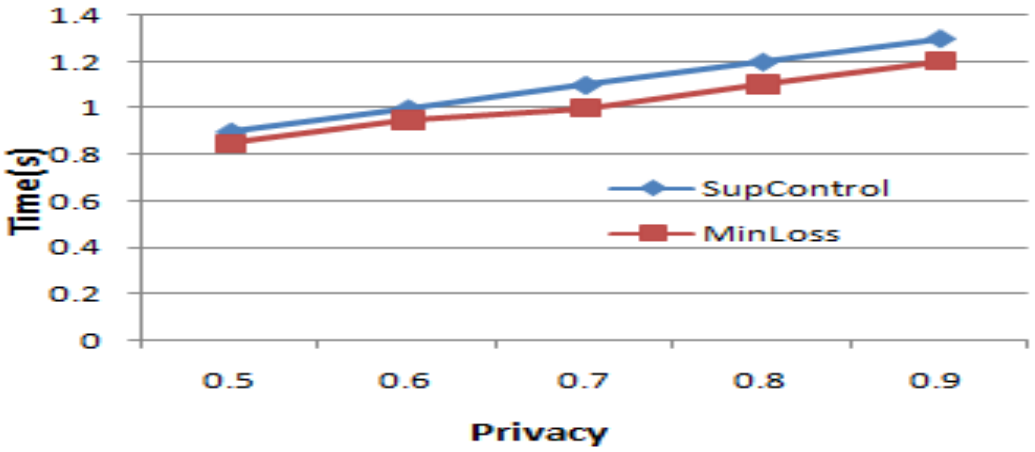
(b) Execution time

Fig 5 The performance by varying Step Size p



(a) Information loss

ISSN 2515-8260 Volume 08, Issue 02, 2021



(b) Execution time

Fig6The performance by varying privacy requiremenρ

in fig 3(a) shows our MinLoss Method is less average information compared to existing high performanceSuppressControl Method because payoff of items are calculated after the add new item,with two data sets also. In fig 3(b) execution of algorithm shown ,our proposed algorithm give higher performance due to consider the whole sliding window, while our algorithms scan TSW_{del} and TSW_{add} to anonymize the sliding window.

We investing the our algorithm in decrease size of window in fig4(a) and (b) average information loss and execution time respectively the average information loss is decreased with window size increased because the probability the lower the probability of a SAR violating ρuncertainty is , in fig 5 (a) and (b)increase step size p both algorithms average loss also decreased ,compared to existing algorithm our minloss method have less loss because suppression will influence the current sliding window with new add itemsIn fig 6(a) and (b) shows the performance of algorithms with vary privacy requirement, average information loss is decrease with ρ increase ,and execution time is minimum effect by ρ.

V. CONCLUSION

publish static transactional streams many privacy- preserving techniques methods are not straightforwardly applied on transactional data streams because it have different characteristics. In sliding window addition and removal of transaction leads be unsuccessful to satisfy ρ-uncertainty which dynamically and continuously make satisfy p-uncertainty of slide window with suppression anonymize sliding window by selecting optimal select items by use of maximal payoff and attacker not able to find knowledge about data and sensitive information of victim.

ISSN 2515-8260 Volume 08, Issue 02, 2021

REFERENCES

1. Jinyan Wang , Chaoji Deng, and Xianxian Li, "Two Privacy-Preserving Approaches for Publishing Transactional Data Streams", Special Section On Recent Computational Methods in Knowledge Engineering and Intelligence Computation, IEEE Access, Vol. 6, pp- 23648- 23658, 2018.
2. S.Wang,L.Minku,andX.Yao,"Alearningframeworkforonlineclass imbalance learning," in Proc. IEEE Symp. Comput. Intell. Ensemble Learn., Apr. 2013, pp.36–45.
3. S. Wang, L. L. Minku, and X. Yao, "Online class imbalance learning and its applications in fault search," Int. J. Comput. Intell. Appl., vol. 12, no. 4, pp. 1340001(19 pages),2013.
4. J. Kivinen, A. Smola, and R. Williamson, "Online learning with kernels," IEEE Trans. Transaction Process., vol. 52, no. 8, pp. 2165–2176, Aug.2004.
5. N. Japkowicz, "Concept-learning in the presence of between-class and within-class imbalances," in Proc. 14th Biennial Conf. Can. Soc. Comput. Stud. Intell.: Adv. Artif. Intell., 2001, pp.67–77.
6. T. Jo and N. Japkowicz, "Class imbalances versus small disjuncts," SIGKDD Explor. Newsl., vol. 6, no. 1, pp. 40–49, Jun.2004.
7. P. Mallapragada, R. Jin, and A. Jain, "Non-parametric mixture models for clustering," in Proc. Int. Conf. Struct., Syntactic, and Statistical Pattern Recog., 2010, vol. 6218, pp.334–343.
8. K. Bache and M. Lichman. (2013). UCI machine learning repository [Online]. Past: http://archive.ics.uci.edu/ml
9. R. Li, S. Wang, H. Deng, R. Wang, and K. C.-C. Chang, "Towards social user profiling: Unified and discriminative influence model for inferring home locations," in Proc. 18th ACM SIGKDD Int. Conf. Know. Discovery Data Mining, 2012, pp.1023–1031.
10. H. He and E. Garcia, "Learning from imbalanced data," IEEE Trans. Know. Data Eng., vol. 21, no. 9, pp. 1263–1284, Sep.2009.
11. S.Wang,L.Minku,andX.Yao,"Resampling-basedensemblemethods foronlineclassimbalancelearning,"IEEETrans.Know.DataEng,vol. 27, no. 5, pp. 1356–1368, May2015.
12. I. Ozalp, M. E. Gursoy, M. E. Nergiz, and Y.Saygin,``Privacy-preserving publishing of hierarchical data," *ACM Trans. Privacy Secur.*, vol. 19, no. 3, Sep. 2016, Art. no. 7.
13. Y. Xin, Z.-Q. Xie, and J. Yang, ``The privacy preserving method for dynamictrajectoryreleasingbasedonadaptiveclustering,"*Inf.Sci.*,vol. 378, pp. 131_143, Feb.2017.
14. H. Zakerzadeh, C. C. Aggarwal, and K. Barker, ``Managing dimensionality in data privacy anonymization," *Knowl. Inf. Syst.*, vol. 49, no. 1, pp. 341_373, Oct.2016.
15. R.Chen,N.Mohammed,B.C.Fung,B.C.Desai,andL.Xiong``Publishing set-valued data via differential privacy," in *Proc. VLDB*, Seattle, WA, USA, 2011, pp. 1087_1098.
16. J. Liu and K. Wang, ``Anonymizing transaction data by integrating suppression and generalization," in *Proc. PAKDD*, Hyderabad, India, 2010, pp.171_180.

17. S. Ramírez-Gallego, B. Krawczyk, S. García, M.Wo¹niak, and F. Herrera, ``A survey on data preprocessing for data stream mining: Current status and future directions," *Neurocomputing*, vol. 239, pp. 39_57, May2017.
18. S. K. Tanbeer, C. F. Ahmed, B.-S. Jeong, and Y.-K. Lee, ``Sliding windowbased frequent pattern mining over data streams," *Inf. Sci.*, vol. 179, no. 22, pp. 3843_3865, Nov.2009.
19. Y.ZhuandD.Shasha,``StatStream:Statisticalmonitoringofthousands of data streams in real time," in *Proc. VLDB*, Hong Kong, 2002, pp. 358_369.
20. Z. Farzanyar, M. Kangavari, and N. Cercone, ``Max-FISM: Mining (recently)maximalfrequentitemsetsoverdatastreamsusingthesliding window model," *Comput. Math. Appl.*, vol. 64, no. 6, pp. 1706_1718, Sep.2012.
21. J. Kim and B. Hwang, ``Real-time stream data mining based on CanTree and Gtree," *Inf. Sci.*, vols. 367_368, pp. 512_528, Nov.2016.
22. F. Nori, M. Deypir, and M. H. Sadreddini, ``A sliding window based algorithmforfrequentcloseditemsetminingoverdatastreams,"*J.Syst. Softw.*, vol. 86, no. 3, pp. 615_623, Mar.2013.
23. H. Chen, L. Shu, J. Xia, and Q. Deng, ``Mining frequent patterns in a varying-size sliding window of online transactional data streams," *Inf. Sci.*, vol. 215, pp. 15_36, Dec.2012.
24. H. Ryang and U. Yun, ``High utility pattern mining over data streams with sliding window technique," Expert Syst. Appl., vol.57, pp. 214_231, Sep. 2016.
25. 25. Z. Zheng, R. Kohavi, and L. Mason, ``Real world performance of association rule algorithms," in Proc. KDD, San Francisco, CA, USA, 2001, pp. 401-406.
26. J. Cao, B. Carminati, E. Ferrari, and K. Tan, "CASTLE: A delayconstrained scheme for ks-anonymizing data streams," in Proc. ICDE,Cancun, Mexico, 2008, pp. 1376-1378.
27. A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "ldiversity: Privacy beyond k-anonymity," in Proc. ICDE, Atlanta, Georgia,USA, 2006, article 24, 12 pages.
28. K. Guo, and Q. Zhang, "Fast clustering-based anonymization approaches with time constraints for data streams," Knowl.-Based Syst., vol. 46, pp.95-108, Jul. 2013.
29. J. Li, B. C. Ooi, and W. Wang, "Anonymizing streaming data for privacy protection," in Proc. ICDE, Cancun, Mexico, 2008, pp. 1367-1369.
30. B. Zhou, Y. Han, J. Pei et al, "Continuous privacy preserving publishing of data streams," in Proc. EDBT, Saint-Petersburg, Russia, 2009, pp. 648-659.
31. S. Kim, M. K. Sung, and Y. D. Chung, "A framework to preserve the privacy of electronic health data streams," J. Biomed. Inform., vol. 50, pp. 95-106, Aug. 2014.
32. W. Wang, J. Li, C. Ai, and Y. Li, "Privacy protection on sliding window of data streams," in Proc. CollaborateCom, White Plains, New York, USA, 2007, pp. 213-221.
33. K. Al-Hussaeni, B. C. M. Fung, and W. K. Cheung, "Privacy-preserving trajectory stream publishing," Data Knowl. Eng., vol. 94, part A, pp. 89- 109, Nov. 2014.

34. N. Mohammed, B. C. M. Fung, and M. Debbabi, "Walking in the crowd: anonymizing trajectory data for pattern analysis," in Proc. CIKM, Hong Kong, China, 2009, pp. 1441-1444.
35. D. Molodtsov, "Soft set theory–First results," Comput. Math. Appl., vol.37, no. 4-5, pp. 19-31, Feb.-Mar. 1999.
36. T. Herawan, and M. M. Deris, "A soft set approach for association rules mining," Knowl.-Based Syst., vol. 24, no. 1, pp. 186-195, Feb. 2011.
37. V. S. Verykios, A. K. Elmagarmid, E. Bertino et al, "Association rule hiding," IEEE T. Knowl. Data En., vol. 16, no. 4, pp. 434-447, Apr. 2004.
38. Z. Zheng, R. Kohavi, and L. Mason, "Real world performance of association rule algorithms," in Proc. KDD, San Francisco, CA, USA, 2001, pp. 401-406.
39. Ravi, P., Haritha, D."Computing iceberg queries on map reduce framework" International Journal of Advanced Trends in Computer Science and Engineering, 2020, 9(5), pp. 8325–8329

40. Ravi, P., Haritha, D."Average iceberg queries computation with state buckets counter"SSRG International Journal of Engineering Trends and Technology, 2020, 68(8), pp. 53–57

PROFILE


Author-1 Photo

Mr. Jayendrakumar is currently working at Anurag Group of Institutions as Assistant professor in Computer Science and Engineering Department. He obtained M.Tech CSE from JNTU Hyderabad. His research interest are Data Mining, Internet of Things and Machine Learning . He is Life Member of Computer Society of India.