UtilitasMathematica

ISSN 0315-3681 Volume 122, 2025

NEW PREDICTION MODEL BASED ON CELLULAR GENETIC PROGRAM USING TWO-PHASE PRIVATE CLUSTERING

Kondeti Harika[1] Gudavalli Pratyusha[2] Sudula Lavanya[3]

Akula Kavya Sri[4] Devarasetti Prasad [5]

[1] M.Tech Scholor, CSE Department, DVR and Dr.HS MIC College of Technology, Kanchikacherla, NTR District, AP, India

[2] Assistant Professor, CSE Department, DVR and Dr.HS MIC College of Technology, Kanchikacherla, NTR District, AP, India

[3] Assistant Professor, IT Department, DVR and Dr.HS MIC College of Technology, Kanchikacherla, NTR District, AP, India

[4] Assistant Professor, IT Department, DVR and Dr.HS MIC College of Technology, Kanchikacherla, NTR District, AP, India

[5] Professor, CSE Department, DVR and Dr.HS MIC College of Technology, Kanchikacherla, NTR District, AP, India

Abstract: In recent years, forest fires have increased drastically due to global warming. Forest fire prediction is the best way to control the spread of fire. The model presented, for the first time, in this paper can predict the spreading of fire in both homogeneous and inhomogeneous forests and can easily incorporate weather conditions and land topography. We propose a cellular automaton (CA) that simulates the spread of wildfire. We embed the CA inside of a genetic program (GP) that learns the state transition rules from spatially registered synthetic wildfire data.. Several machine learning classification techniques, including logistic regression, support vector classifier, decision tree. In addition to natural role in ecosystem dynamics natural disasters that threaten human lives, property and ecosystems efficient algorithm to perform optimal label flipping poisoning attacks and reliable suspicious data points mitigating the effect of such poisoning attacks. A novel BDI-GIS model was then proposed intention was defined based on spatial or non-spatial data and GIS functions. The cluster based model was developed to determine the prediction of forest fires and implemented it on a real dataset

Index Terms: adversarial machine learning, poisoning attacks, label flipping attacks, Data Anonymization, K-means algorithm

.1. INTRODUCTION

Many modern services and applications rely on data driven schema use machine learning model to extract valuable

information from the data received, provide advantages to the users and allow the automation of many processes [1]. In addition to their natural role in ecosystem dynamics wildfires can morph into natural disasters that threaten human lives, property and ecosystems [2]. The adversary in this setting can only poison its own local data without observing the training data of other users. Moreover the poisoned data only influences the global model indirectly the masked features [3]. Although the training process becomes privacy preserving and cost efficient due to distributed computation as we highlight it remains susceptible to poisoning attacks [4]. K anonymity captures security of released data against identification of respondents to released data refers. K anonymity demands each topple in private table being released be indistinguishably related to k respondents [5]. As it seems impossible and limiting to assume as to which is a potential attacker and identify respondents k-anonymity requires that respondents indistinguishable in the released table itself regarding attributes set called quasi identifier which can be exploited for linking [6]. Implementing a sanitization process must consider expected threats. Briefly threats may be as simple as an attacker reading data with root access permissions complex as an attacker using laboratory equipment to read the storage media directly [7]. Guidelines for threats and appropriate sanitization levels have been published by several government agencies which require sanitization when purchasing storage [8]. The state-of-the-art privacy principle in the training of generative models is closure under post-processing property differential privacy ensures that the released model provides theoretically guaranteed privacy protection for the training data [9]. The use of generative models as the vehicles of data releasing enables the synthesized data to capture the rich semantics of the original data [10]. Modeling and simulation have been applied to fire fighting and management for many years, particularly in order to predict fire and spread in forests under various scenarios of weather conditions. [11].

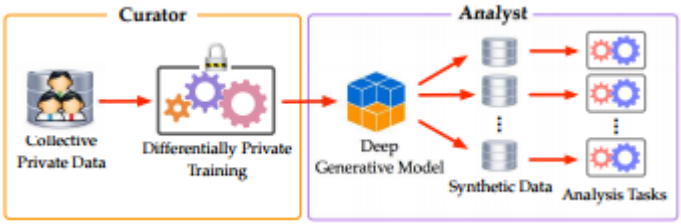

Figure 1: High-level design of releasing framework for semantic data.

2. RELATED WORK

To reduce computation time, increase accuracy and leverage the advances in satellite imagery recent work has modeled wildfire dynamics with machine learning and evolutionary strategies [12]. As of late clustering techniques has been enhanced to accomplish a protection safeguarding in neighborhood recoding anonymization[13]. From the utility security conservation viewpoint the nearby recoding likewise utilizes the best down partner and a base up new approach are as one pit-forward in view of the bunch measure the agglomerative grouping procedure and disruptive bunching systems get enhanced [14]. These model leverages the technique of differential privacy to hide the information about the training data simple aggregation generates classifiers with very high accuracy. Recently use differential privacy to design a deep learning model that supports collaboration among users while preserving the privacy of their training data [15]. A new PPDM framework of multi-dimensional data proposed developed a new and flexible PPDM approach without needing new problem-specific algorithms as it mapped original data set to new anonym zed data set [16]. Most previous sanitization research discusses erasing the entire storage system, while our customers wish to sanitize individual files addressed the issue of finding regions of solid state drives that hold previous versions of a file by implementing a full scan of the pages to find physical-to-logical mappings [17].

3. SYSTEM MODEL

This model can help us to be panic-free from the last-minute chaos. Several machine learning approaches like Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), Support Vector Classifier (SVC), K Nearest Neighbors Classifier (KNN) were applied on two separate datasets [18]. The key insight in designing the poisoning of training data strongly influences the share of the masked features learnt in the system. Since each masked feature corresponds to different information in the training data the main challenge lies in identifying which set of masked features are affected due to poisoning of the dataset [19]. We propose a naive spreading function and

examine how well the regression can reproduce the spread patterns generated. synthetic burns [20].

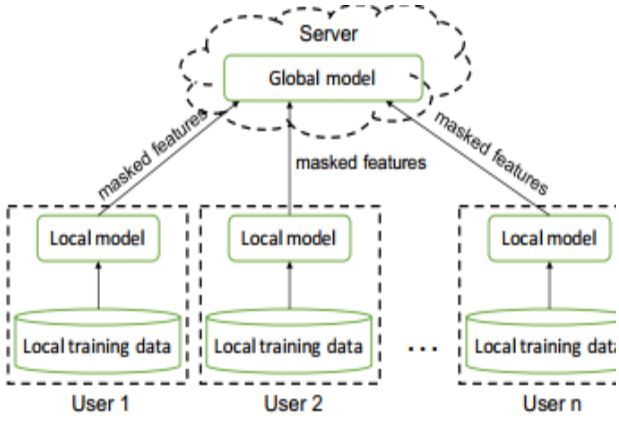

Figure 2: The server computes a global model

4. PROPOSED SUPPRESSION TECHNIQUE:

The algorithm optimizes feature suppression process as yielding best optimal features cannot be removed during anonymization without affecting classification accuracy. The neighborhood of a cell is taken to be the cell itself and some or all of the immediately adjacent cells. The artificial agents are classified as employed bee onlooker bee and scout. Each plays a different role: employed bee stays at a food source and provides the neighborhood of the source in its memory the onlooker gets food source information from employed bees in the hive and selects one to gather nectar the scout is responsible to find new nectar sources. If value at corresponding position is 1, it indicates that a feature is part of subset needing evaluation.

1. Initialize the food source positions.
2. Evaluate the food sources.
3. Produce new food sources (solutions) for the employed bees .
4. Apply greedy selection .
5. Calculate the fitness and probability values
6. Produce new food sources for onlookers.
7. Apply greedy selection .
8. Determine the food source to be abandoned and allocate .
9. Memorize the best food source found 10
10. Repeat steps 3-9 for a pre determined number of iterations

Partial Differential Equations (PDEs), contain much more information than is usually needed, because variables may take an infinite number of values

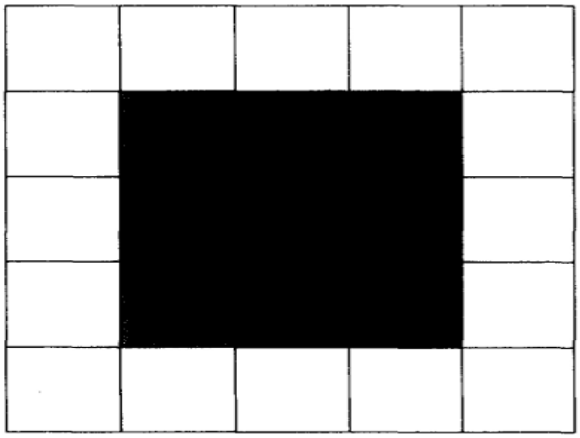

Fig. 3. The neighbourhood of the (i,j) cell comprises the nine grey cells

5. TWO-PHASE PRIVATE CLUSTERING USING MAPREDUCE

The agent's intention is also important in generating its desire. An optimization operator such as a genetic algorithm (GA) can be used to generate such an intention. For versatility viewpoint, point-task strategies are perfect for neighborhood recoding in Map Reduce. Point bunches are utilized to pick an arrangement of information records to shape a group from that whatever is left of the records will dole out into these bunches. Be that as it may, for the extensive arrangement of information records under perceptions, the size will be 1/k of a unique informational collection [21]

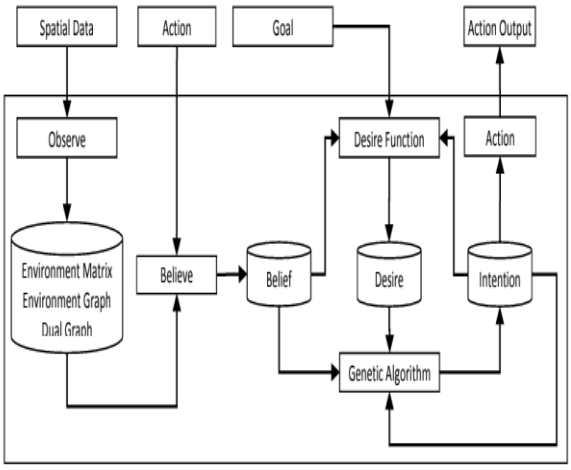

Figure 4. The proposed architecture for handling spatial problems

Algorithm1: Design of Two-Phase Clustering

Input: Data set B, anonymity parameter k

Output: Anonymous data set B*

1. Run the t-ancestor clustering algorithm on B, get a set of α-clusters: Cα = {C1α ,……..,Ct α }. 2.
2. For each α-cluster Ci α ϵ Cα; 1≤i≤t; run ε- differential privacy algorithm Let Sε() be an ε-differentially private sanitizer ȳ← Partitioned data set TA(Y)for R=1 to n doyε← Sε(Qr(ȳ)) End for Return Yε 3.
3. For each cluster Cjϵ C , where C=Ui=1l Ci ,generalizeCjtoCj* by

replacing each attribute value with a general one. 4.]

4. Generate B*= Uj=1 mj C*j, where mj=∑ m1

Technique for producing the differentially private informational index X. let X is an informational index with m numerical traits..

A. Hash Algorithm

The algorithm is based on the hash, displace and compress approach [22]. The perfect hash function data structure consists of two levels. we use a pair from the sequence {(0, 0),(0, 1), ...,(0, m − 1),(1, 0),(1, 1), ...,(1, m − 1), ...,(m − 1, m − 1)}. Instead of storing a pair (d0, d1) for each bucket Bi, we store the index of the first pair in that sequence that successfully places all keys in Bi,, d(i). The data structure only has to store the sequence {d(i)|0 ≤ i.

Algorithm 2 Hash, displace, and compress

1. Split S into buckets Bi = {x ∈ S|g(x) = i}, 0 ≤ i
2. Sort buckets Bi in falling order according to size |Bi|;
3. Initialize array T[0 ...m − 1] with 0's;
4. for all i ∈ [r], in the order from (2), do
5. for ℓ = 0, 1,... repeat forming Ki = {hi(x) | x ∈ Bi}
6. until |Ki| = |Bi| and Ki ∩ {j | T[j] = 1} = ∅;
7. let d(i) = the successful ℓ;
8. for all j ∈ Ki let T[j] = 1;
9. Compress (d(i))0≤i

B. Weight_ Clustering

While it is natural to group the bias parameters together as many of them are close to zero, the grouping of the weight parameters is much less obvious. We propose a simple yet effective strategy to stratify and cluster the weight parameters [23]. Assuming that we have the optimal parameter-specific clipping bound {c(дi)}i for each weight's gradient {дi } we then cluster these parameters into a predefined number of groups using a hierarchical clustering procedure..

Algorithm 2: Weight-Clustering

Input: k - targeted number of groups; {c(дi)}i - parameter-specific gradient clipping bounds

Output: G - grouping of parameters

1 G ← {(дi : c(дi))}i ;

2 while |G| > k do

3 G,G ′ ← arg minG,G′ ∈G max c(G) c(G′) , c(G ′) c(G) ;

4 merge G and G ′ with clipping bound as p c(G) 2 + c(G′) 2 ;

5 return G

6. EXPERIMENT RESULTS

The logistic model shows an upward trend; individuals resulting from crossover or mutation thus improved in mean fitness. The genetic programming model demonstrates poor fitness in early function evaluations but very fast improvement. For the learning algorithm we set the learning rate to 0.01 and the number of epochs to 100. For the defensive algorithm, we set the confidence parameter η to 0.5 and selected the number of neighbours k according to the performance of the algorithm evaluated in the validation dataset. We assume that the attacker has not access to the validation data.

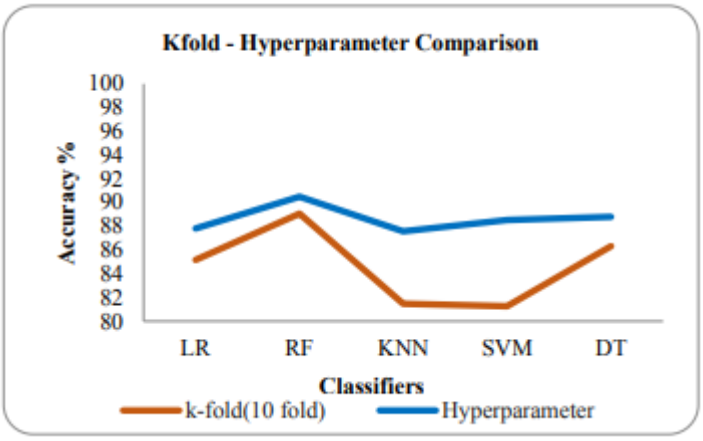

Figure 4. K-fold and hyper parameter Accuracy-Dataset 2

7. CONCLUSION AND FUTURE WORK

Forest fires or wildfires are the major threats created by mankind or natural disasters. This leads to measure losses to our nature as well as the humans. Predicting these forest fires may help the environment and humans to protect themselves from the uncontrollable fires turning each and every bit into ashes A label flipping poisoning attack strategy that is effective to compromise machine learning classifiers defense mechanism based on k-NN to achieve label sanitization, aiming to detect malicious poisoning points. To achieve this, dp-GAN integrates the generative adversarial network framework with differential privacy mechanisms, provides refined

analysis of privacy loss within this framework, and employs a suite of optimization strategies to address the training stability and scalability challenges. Further work on increasing the accuracy and speed of our model is required. Live predictions or on-site predictions using satellite images can also be performed. Ensemble approach may be used for prediction of forest fire

8. REFERENCES

[1] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97–107, 2014.

[2] Arjovsky, M., Chintala, S., and Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017 (2017), D. Precup and Y. W. Teh, Eds., vol. 70 of Proceedings of Machine Learning Research, PMLR, pp. 214–223.

[3] Beaulieu-Jones, B. K., Wu, Z. S., Williams, C., and Greene, C. S. Privacypreserving generative deep neural networks support clinical data sharing. bioRxiv (2017).

[4] Beimel, A., Brenner, H., Kasiviswanathan, S. P., and Nissim, K. Bounds on the sample complexity for private learning and private data release. Machine Learning 94, 3 (Mar 2014), 401–437.

[5] Chaudhuri, K., Monteleoni, C., and Sarwate, A. D. Differentially private empirical risk minimization. Journal of Machine Learning Research 12, Mar (2011), 1069–1109.

[6] Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., and Abbeel, P. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In Advances in Neural Information Processing Systems (2016), pp. 2172–2180.

[7] Donahue, J., Krähenbühl, P., and Darrell, T. Adversarial feature learning. CoRR abs/1605.09782 (2016).

[8] Dwork, C. Differential privacy. In Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II (2006), ICALP'06, pp. 1–12.

[9] Dwork, C. The differential privacy frontier (extended abstract). In Proceedings of the 6th Theory of Cryptography Conference on Theory of Cryptography (2009), TCC '09, pp. 496–502.

[10] Dwork, C., and Roth, A. The algorithmic foundations of differential privacy. Found. Trends Theor. Comput. Sci. 9, 3–4 (2014), 211–407

[11] A. Adya, W. Bolosky, M. Castro, G. Cermak, R. Chaiken, J. Douceur, J. Howell, J. Lorch, M. Theimer, and R. Wattenhofer. Farsite: Federated, available, and reliable storage for an incompletely trusted environment. ACM SIGOPS Operating Systems Review, 36(SI):1–14, 2002.

[12] D. Belazzougui, F. C. Botelho, and M. Dietzfelbinger. Hash, displace, and compress. In Proceedings of the 17th Annual European Symposium on Algorithms, ESA'09, pages 682–693, 2009.

[13] M. A. Bender, M. Farach-Colton, R. Johnson, R. Kraner, B. C. Kuszmaul, D. Medjedovic, P. Montes, P. Shetty, R. P. Spillane, and E. Zadok. Don't thrash: How to cache your hash on flash. In Proceedings of the 38th International Conference on Very Large Data Bases, 2012

[14] M. Barreno, B. Nelson, R. Sears, A. D. Joseph, and J. D. Tygar. Can machine learning be secure? In Symposium on Information, computer and communications security, pages 16–25. ACM, 2006.

[15] B. Biggio, I. Corona, G. Fumera, G. Giacinto, and F. Roli. Bagging classifiers for fighting poisoning attacks in adversarial classification tasks. In Multiple Classifier Systems, pages 350–359. Springer, 2011.

[16] B. Biggio, I. Corona, D. Maiorca, B. Nelson, N. Šrndic,´ P. Laskov, G. Giacinto, and F. Roli. Evasion attacks against machine learning at test time. In Machine Learning and Knowledge Discovery in Databases, pages 387–402. Springer, 2013.

[17] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu,"Privacypreserving data publishing: A survey of recentdevelopments," ACM compute. Survey, vol. 42,no. 4, pp. 1–53, 2010.

[18] J. Li, R. C.-W. Wong, A. W.-C. Fu, and J. Pei,"Anonymization by local recoding in

data with attributehierarchical taxonomies," IEEE Trans. Knowl. Data Eng.,vol. 20, no. 9, pp. 1181–1194, Sep. 2008.

[19] G. Aggarwal, R. Panigrahy, T. Feder, D. Thomas, K.Kenthapadi, S. Khuller, and A. Zhu, "Achievinganonymity via clustering," ACM Trans. Algorithms,vol. 6, no. 3, 2010.

[20] R. C.-W. Wong, J. Li, A. W.-C. Fu, and K. Wang, "(a, k)- anonymity: An enhanced k-anonymity model for privacypreserving data publishing," in Proc. 12th ACM SIGKDDInt. Conf. Knowl. Discovery Data Mining, 2006, pp.754–759.

[21] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding Deep Learning Requires Rethinking Generalization," arXiv preprint arXiv:1611.03530, 2016.

[22] B. Nelson, M. Barreno, F. J. Chi, A. D. Joseph, B. I. P. Rubinstein, U. Saini, C. A. Sutton, J. D. Tygar, and K. Xia, "Exploiting Machine Learning to Subvert Your Spam Filter," LEET, vol. 8, pp. 1–9, 2008.

[23] A. Paudice, L. Mu ˜noz-Gonz´alez, A. Gyorgy, and E. C. Lupu, "Detection of Adversarial Training Examples in Poisoning Attacks through Anomaly Detection," in arXiv pre-print arXiv:1802.03041, 2018.

[24] S. Mei and X. Zhu, "Using Machine Teaching to Identify Optimal Training-Set Attacks on Machine Learners," in AAAI, 2015, pp. 2871–2877.

[25] P. W. Koh and P. Liang, "Understanding Black-box Predictions via Influence Functions," in Int. Conf. on Machine Learning, 2017, pp. 1885–1894.