IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.047 ⁜ Vol. 10, Issue 9, September 2022

DOI: 10.17148/IJIREEICE.2022.10907

Adversarial Attacks on Time Series

Bhoomika

PG Scholar, Dept. of MCA,VTU, Center of PG Studies, Mysuru, Karnataka, India

Abstract: Time series classification models have been garnering significant importance in the research community. However, not much research has been done on generating adversarial samples for these models. These adversarial samples can become a security concern. In this paper, we propose utilizing an adversarial transformation network (ATN) on a distilled model to attack various time series classification models. The proposed attack on the classification model utilizes a distilled model as a surrogate that mimics the behavior of the attacked classical time series classification models. Our proposed methodology is applied onto 1-Nearest Neighbor Dynamic Time Warping (1-NNDTW) and a Fully Convolutional Network (FCN), all of which are trained on 42 University of California Riverside (UCR) datasets. In this paper, we show both models were susceptible to attacks on all 42 datasets. When compared to Fast Gradient Sign Method, the proposed attack generates a larger faction of successful adversarial black-box attacks. A simple defense mechanism is successfully devised to reduce the fraction of successful adversarial samples. Finally, we recommend future researchers that develop time series classification models to incorporating adversarial data samples into their training data sets to improve resilience on adversarial samples.

I.INTRODUCTION

Since the turn of the century, the use of algorithms has been crucial in the progression of awide variety of diverse elements of society. Machine learning have a broad array of applications, some of which include wearables, online searches, and recommendation engines, to mention just a few of the many diverse fields in which they are used. As a result of developments in data collecting and storage at massive sizes, ease of data analytics, andpredictive modeling, it is now possible to examine time series data gathered from a varietyof sensors in hopes to identify recurring patterns that can be interpreted and exploited. This is made possible as a result of the fact that it is now possible to find recurring patterns that can be interpreted and exploited. These breakthroughs would not have been conceivable without the development of intelligent sensors, advancements in data gathering and storageat huge proportions, accessibility of data analytics, and predictive modeling.

The classification of this time series data is something that has captured the attention of a lot of scholars working in the academic sector. Data from electrocardiograms are used to identify patients who have considerable cognitive impairments, data from audio recordingsare used to categorize words into distinct phenomes, and data from show of affection recognition systems are used to categorize activities that are being carried out using time series classification models. The realm of medical care is home to each and every one of these possible applications.

On-device analytics may be able to assist in the prevention of severe issues that may arise during the regular functioning of a system by analyzing sensor data from significant applications such as manufacturing facilities, industrial engineering, and chemical compound synthesis. These applications include manufacturing facilities, industrial engineering, and chemical compound synthesis. It is important to evaluate how successful a time-series classification model was by determining whether or not it was able to properlycapture and generalize the pattern of time series signals in such a fashion that it was able to categorize data that was completely unseen.

The use of time series classification models has shown a significant surge in significance over the course of the last few years. Despite this, not a lot of research has been done on how to generate adversarial samples for these models yet. It is possible that these rogue samples pose a threat to the safety of the general population. [Here's a good example:] In this article, it is proposed that a n
model be used together in order to launch an attack against a number of different techniquesto time series categorization (ATN). This article presents a counterattack against the classification model, which makes use of a condensed version of the time series classification model. Using 42 datasets provided by the University of California Riverside(UCR), we trained 1-Nearest Neighbor Total Variation Warping (1-NN) DTW, Full Connected Networks (FCN), and Fully Convolutional Networks (FCN). All of the 42 datasets that were analyzed as part of this research contribute to the conclusion that both of the models under consideration are open to being attacked. A challenge of this kind to time series classification models does not seem to have ever been attempted, as far as our knowledge goes. We suggest that in the future, researchers who are building time series classification models should use the model's resilience as an assessment system of measurement and should include adversarial data samples in their training data sets. This is something that should be done in accordance with our referral. That's an idea that has been proposed by our group.

As a technique of deceiving different picture classification algorithms that rely on DNNs,a number of diverse methods for the production of asymmetrical samples have been presented (state-of-the-art models for computer vision). Due to the

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.047 ⁜ Vol. 10, Issue 9, September 2022

DOI: 10.17148/IJIREEICE.2022.10907

fact that DNNs includeinformation about gradients, they are vulnerable to the attacks outlined in the previous paragraphs. Despite the fact that doing so may provide a substantial risk to the confidentiality of data, the creation of adversarial samples for use in time series classification models is not even a priority. These models are used to identify patterns in time series data. This same process of converting text to speech using voice recognition poses a number of extremely significant problems about individuals' right to privacy and the integrity of their personal information. Carlini and Wagner provide a tactic that may beused to test the accuracy of text-to-speech classifiers. [Carolini with Wagner] [Commentary] This same text-to-speech classifier known as Deep Speech is unable to accurately recognize the speech that is included inside a number of audio recordings that were provided by the researchers. Any use of time series classification algorithms in medical equipment raises additional concerns regarding patient safety. This is due to the fact that these algorithms can be manipulated purpose of providing inaccurate diagnoses ofpatients, which can have a negative effect on the therapy that these patients receive. It is feasible that the algorithms that are used to identify and monitor seismic activity may be adjusted in such a manner as to provoke terror and hysteria in our society. This is a possibility since it is possible that these algorithms could be hacked. Wearables that categorise the behaviors of their users based on time-series data run the risk of fooling theusers into thinking they are engaged in activities other than those that the garment is really

tracking them for. The most complex algorithms for moment series classification that are available at the moment include traditional approaches such as 1 Nearest Neighbor - Dynamic Time Warping (1NNDTW), Kernel - Based virtual Machines (SVMs), and Fast-Shapelet.

II.EXISTING SYSTEM

The development of hostile examples for time series classification methods has not been the subject of nearly as much research as has been done on the issue of time series classification models, despite the fact that hostile sampling provide a potentially significant security risk. [Here's a good example:] [Here's a good example:] [Here's a good example:][Operations that use voice recognition, specifically those that convert speech to text, present a significant risk of security being compromised. This is especially true of operations that convert speech to text. Carlini and Wagner provide an example of how vulnerability in speech-to-text classifiers might be used for nefarious reasons. cannot determine with any degree of accuracy the words that are being said.

There is a possibility that medical devices that employ time series classification techniques might be deceived into erroneously diagnosing patients, which could result in a modification to the original diagnosis of the patient's condition. This may start happening since it is hypothetically possible for these gadgets to be tricked in some way. This vulnerability is an additional possible security problem that may be uncovered in these devices. The devices contain this weakness. It is feasible to make changes to the algorithmsthat are used to classify time series and which are used to identify and monitor seismic activity. It is possible that this weakness will be used in order to incite panic and disseminate fear across our society.

When conducting black-box assaults, the only piece of information that may be known insome instances is the total amount of time that the input time series data was collected over.It's possible that this will be the case in certain circumstances. In this body of work, we give both a black-box and a white-box strategy to approaching the problem at hand. Both of these approaches are able to provide a challenge to the most trying to cut models that are presently in use for classical nor deep learning time series classification. These modelsare used to categorize time series data. To be more particular, the black-box attack is moredirect than the white-box strategy.

III. PROPOSED SYSTEM

Time series classification techniques are considered to as "teachers," while neural networkscreated from those models are referred to as "students." Time series classification methodsare used to train neural networks generated from those models.

The adversarial transformation network that we provide may be programmed to initiate attacks on the student model after it has been trained to attack the student model. The FullyConnected Network and the Convolutional Neural Network that we use for training were both trained using data sets provided from the University of California, Riverside. Both ofthese networking are used for training purposes.

The adversarial black-box attacks produced by the Fast Gradient Symbol Method, which is the baseline aggressive methodology, are less powerful than just the attack that we havediscussed here.

In addition, we explain how a basic defensive mechanism may decrease the percentage ofsuccessful antagonistic samples across a wide range of GATN assaults, as well as how thismay be achieved. When researchers in the future design time series classification techniques, they should have used model resilience as an evaluation factor and should include adversarial data samples in training data sets.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.047 ⁜ Vol. 10, Issue 9, September 2022

DOI: 10.17148/IJIREEICE.2022.10907

Disadvantage:

The use of educated GATN models for assaults that are viable on hands - on practical, including black box attacks, is the one-of-a-kind consequence that emerges purely as a result of this generalization and is caused by the fact that it has been generalized. It is possible to create a considerable quantity of adversarial samples at a cost that is very closeto being constant over the course of time when a trained GATN is coupled with a paired student model. This makes it viable to produce a significant proportion of adversarial data.The use of a deep convolutional training network makes this success conceivable (GATN).
These attacks can be constructed on small devices that are portable without the need for a significant amount of computation because with a forward pass of the GATN only uses a small amount of time and resources, and indeed the student model only needs to be a certainsize to be able to calculate the input gradient in a reasonable amount of time. Both of thesefacts are due to the fact that the GATN only uses a small amount of resources. This is feasible thanks to the fact that the GATN requires a minuscule number of available resources.

It is concerning that certain classifiers that have been trained on some datasets are susceptible to attack even without the need of any further on-device training being carriedout on the device itself.

In every one of our experiments, the perpetrator was well aware of the scale of the time series data, which brings us to our last and most important point. In the domain of constructing time series adversaries, more research may be done, especially in circumstances in which the attacker is uninformed of the breadth of the time series data.

Advantages:

Since traditional statistical classification models are black boxes with non-differentiable storage and processing, they are more hard to challenge than their more modern counterparts. As a direct result of this, it is unable to utilize any information on gradients in any way. Because of the easiness with which a pale skinned attack potentially exploit the gradient information that is included in DNN models, these models are more vulnerableto attack.

In what is known as a "fair skinned assault," sometimes known as a "backdoor attack," anadversary has the opportunity to get access to the training dataset, the development of employees, the hyperparameters and weights, as well as the model architecture itself.

It is feasible that the computers that are being used to identify and control seismic activitymay be adjusted in such a manner as to provoke terror and hysteria in our society. This is a possibility since it is possible that these algorithms could be hacked.
Wearable device users who use the devices to categorize their activities on the basis of datasets face the threat of developing the gadgets misrepresent them.

It is possible to judge what successful a time-series classification model was by determining whether or not it was able to properly capture and make assumptions the pattern of time series signals in such a manner that it was able to summarize data that waspreviously unknown. This is also true for classification methods that are used in computervision. These algorithms take advantage of the innate spatial organization that is present inpictures in addition to do particular tasks.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.047 ⁜ Vol. 10, Issue 9, September 2022

DOI: 10.17148/IJIREEICE.2022.10907

IV.SYSTEM ARCHITECTURE

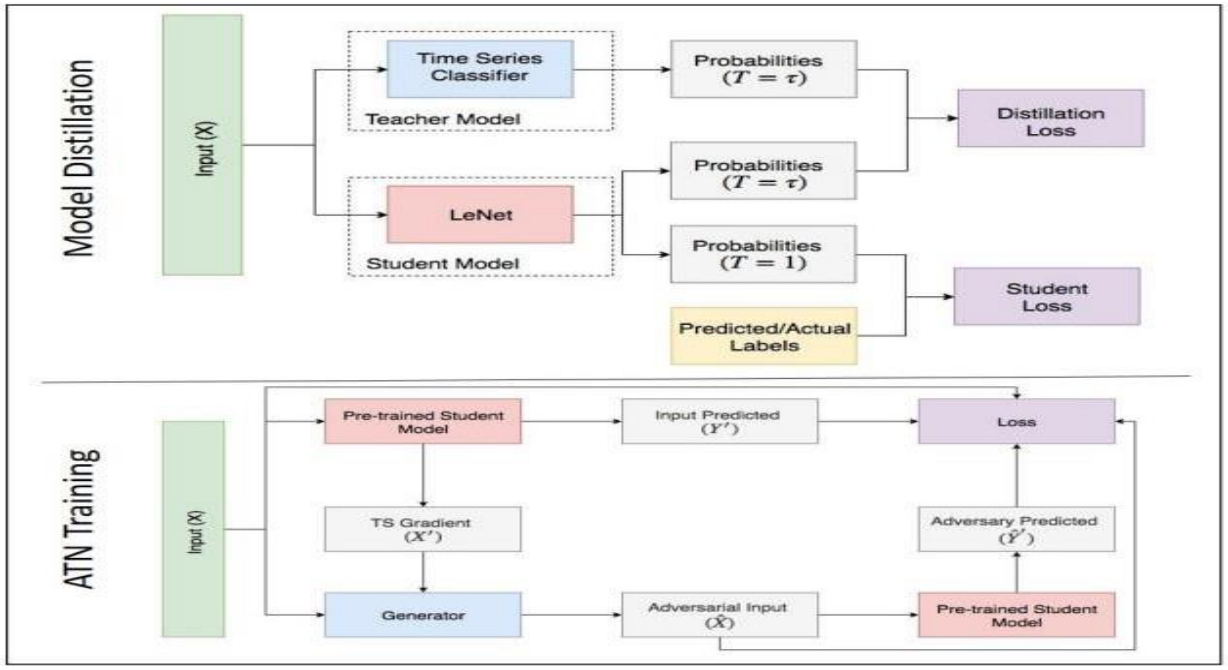

Fig 1 SYSTEM DESIGN

V.CONCLUSION AND FUTURE ENHANCEMENT

We present a model extraction strategy to simulate classical time - series data classificationmodels and an adversarial transformations network to attack time series datasets. Both additions increase scope of the project and usefulness. We also explore non-differentiabletarget models, which are common in time series, and present the student-teacher structure as a general proxy attack solution.

Future time - series data classification model programmers might utilize the model's resistance as an evaluation metric and incorporate adversarial data samples to boost modelaccuracy.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” nature, vol. 521, no. 7553, p. 436, 2015.

[2] J. Boyan, D. Freitag, and T. Joachims, “A machine learning architecture for optimizingweb search engines,” in AAAI Workshop on Internet Based Information Systems, 1996, pp. 1–8.

[3] M. J. Pazzani and D. Billsus, “Content-based recommendation systems,” in The adaptive web. Springer, 2007, pp. 325–341.

[4] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, “A deep learning approach to on-node sensordata analytics for mobile or wearable devices,” IEEE journal of biomedical and health informatics, vol. 21, no. 1, pp. 56–64, 2017.