Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/109/20241412

Impact of SVM-based Poisoning on the Semantic Recognition of Sounds

Shuobo Jiao

Huaxin Software College, Tianjin University of Technology, Tianjin, 300000, China

zuoweige@ldy.edu.rs

Abstract. Machine learning is a technique that enables computers to learn from data and make predictions or decisions, data poisoning is the process of machine learning training where malicious samples are put in to make the model predictions or classifications less accurate. Data poisoning attacks help to reveal security vulnerabilities in AI systems. In this paper, we study Support Vector Machine (SVM) poisoning for sound recognition techniques, using AISHELL-3 dataset training data, from which we find the most vulnerable features for SVM poisoning. In the field of speech recognition, SVM can be applied to speech feature extraction, speech classification and speech synthesis to find the best hyperplane by finding the maximum margin optimization for effective classification and recognition of speech signals. Experiments have resulted in biased semantic recognition of sounds, output of incorrect speech, reduced accuracy of model classification and generation of incorrect decision boundaries. The role of this research paper is to investigate whether SVM poisoning affects the semantic recognition of sounds, and the result of the research is that it does cause semantic bias.

Keywords: Machine learning, recognition of sounds, SVM, poisoning.

1. Introduction

A data poisoning attack is an attack that affects the accuracy of a model's prediction or classification by injecting malicious or misleading samples into the training data of a machine learning model. This attack aims to cause the model to produce incorrect results in future predictions or decisions by distorting the model's learning process and causing the model to deviate from the representation of real data. The significance of the existence of data poisoning attacks is to reveal the potential security risks of large language models, to advance the theory and to improve the safety and reliability of the models. Data poisoning attacks help to reveal the security vulnerabilities of artificial intelligence systems.

Attacks against machine learning can occur in the training phase, testing phase, and deployment phase, where data poisoning attacks mainly occur in the training phase of machine learning [1]. Data poisoning attacks can be aimed at affecting the classification boundaries of the model by modifying the labels of the training dataset or by directly changing some of the model inputs, thus reducing the correctness of the model. Machine learning is a technique that enables computers to learn from data and make predictions or decisions. Among them, Support Vector Machine (SVM) algorithms are supervised data mining algorithms that are mainly used in classification problems (both binary and multiclassification), but can also be used for regression analysis and outlier detection. For example, in a binary classification problem, by randomly modifying 40% of the training data labels, the model can

DOI: 10.54254/2755-2721/109/20241412

be rendered unable to classify properly. In data poisoning for SVM, by modifying the labels of the training data or directly changing part of the data, the decision boundary of the SVM model is made to change, thus affecting the classification accuracy of the model [2].

Sound Semantic Recognition is the process of recognizing and interpreting sounds by analyzing and understanding the semantic information in the sound signal. This improves the interaction between humans and machines, increases productivity, enhances the convenience of life, and brings innovative solutions to a wide range of industries. The significance of the application of sound semantic recognition is that it changes the way humans interact with computers, improves work efficiency, enhances social interaction as well as provides a wide range of applications in areas such as entertainment and healthcare. The significance of the application of speech recognition technology is mainly reflected in the application in the field of natural language processing, a huge leap in human-computer interaction, a powerful cross-industry enabler, contributing to social inclusiveness, and driving technological innovation. Thus, in recent years, for the poisoning attack of SVM in sound semantic recognition, the predecessors have proposed various defense measures, such as data cleaning, model retraining and adversarial training.

The purpose of this research paper is to observe the effect of bias on the semantic recognition of sounds through SVM poisoning. In this paper, experiments are conducted with a dataset, where the sound is extracted and then the most attackable features are found and an SVM attack is performed, and the final result is that the semantics of the sound is biased, indicating that the SVM poisoning attack has had an impact.

2. Semantic Recognition of Sound and Analysis of Poisoning Attacks

2.1. Semantic Recognition

Semantic recognition of sounds is an important part of speech recognition technology, which involves syntactic and semantic analysis of the recognition results to understand the meaning and purpose of the language in order to respond accordingly. Speech recognition technology is an exciting method of human-computer interaction that changes the way we interact with computers. Speech recognition technology has significantly improved people's productivity. In various office environments, employees no longer need to manually type text or operate a mouse but can accomplish a variety of tasks through voice commands alone. Speech recognition technology brings new possibilities for social interaction. Speech plays a vital role in human-to-human communication. Speech recognition technology enables machines to understand and respond to human language, enhancing the human-computer interaction experience. Speech recognition technology is also widely used in entertainment and healthcare. In entertainment, speech recognition technology adds a new element to games, allowing players to control games by voice. In the medical field, speech recognition technology provides convenience for healthcare professionals. For example, doctors can quickly access medical records, medication information and other critical medical data through voice commands.

2.2. Sound poisoning attacks

A sound poisoning attack is a technique that utilizes sound as a means of attack by creating sound waves of a specific frequency and intensity to cause damage to a target. This type of attack utilizes the physical properties of sound to cause damage to a human body or other target by creating specific sound waves. The principle of sound poisoning attacks is based on the way sound travels in pressure waves that can travel through solids, liquids, and even through the human body. Potential application scenarios for sound poisoning attacks include the use of sound waves of specific frequency and intensity to cause psychological or physiological effects on an enemy or hostile group for specific strategic or tactical purposes.

A sound poisoning attack algorithm is an attack that utilizes sound signals for data transmission or stealing information. Sound poisoning attack algorithms are implemented in various ways, including the Mosquito attack technique: this is a technique that utilizes sound waves for communication, which is

DOI: 10.54254/2755-2721/109/20241412

resistant to electromagnetic interference and does not require a microphone and achieves covert data exchange by converting the speaker function into a microphone function to record nearby audio and conversation information. Casper attack is an attack that utilizes the internal speakers of a computer as a data transmission channel of the attack. By encoding data as binary or Morse code and using frequency modulation to transmit undetectable ultrasonic waves through the internal speakers, the receiving end can pick up these signals through the microphone and decode the data.

3. SVM-based sound poisoning application

3.1. Methodology

The fundamentals of the poisoning attack include the following. First, SVMs introduce malicious samples that affect the localization of the hyperplane when they are trained on the data by finding the optimal classification hyperplane. Data contamination displaces the boundaries of the model by adding specific samples labelled as misclassified, leading to misclassification. Goal orientation is the selection of specific samples to be poisoned with the aim of attacking a specific business goal or degrading the recognition performance of a specific class [3]. The SVM algorithm performs data preprocessing on the noisy signals first so as to improve the classification efficiency, and then proceeds to the feature extraction, at which point the data poisoning reduces the recognition capability [4].

SVM poisoning attack methods mainly include label flipping based attacks, optimization based attack, gradient-based attack and clean label data poisoning attacks.

Relevant data is first collected, including audio samples of various timbres, accents, and noisy backgrounds. A Convolutional Neural Network (CNN) can be utilized to extract features from the image, followed by a fully connected network for classifying the extracted features [5]. The SVM is then analyzed in depth to determine what method of extracting sound is applied and to determine the most vulnerable features, at which point the poisoning samples can be generated, applying Fast Gradient Sign Method (FGSM) and Projected Gradient Descent (PGD) to apply perturbations each time so that the antagonistic samples are successfully generated At the same time, it is possible to customize the samples by inserting specific sounds or ambient noises to interfere with the model's decision (Figure 1).

The implementation of the attack is carried out at this time. Real-time injection attacks can be implemented by injecting adversarial samples while the microphone or speaker is playing, which is a more stealthy approach. It is also possible to use software to poison the audio stream, which can further increase stealth.

After that, an evaluation can be performed to compare the accuracy of the model recognition before and after the poisoning, to infer the success rate of the poisoning, as well as to perform the deletion of the parts that did not play a role in the poisoning.

It is important to enhance the robustness of the model to improve the resistance of the SVM model against sound poisoning, and at the same time, it should always be checked and monitored to block the attack in time. To improve the robustness of SVM against poisoning attacks, data cleaning can be done by anomaly detection methods, such as statistical-based anomaly detection algorithms, to clean up potential noise in the dataset. Enhancement modeling can be done by integration learning methods, such as integrating multiple different models, to improve the robustness against samples. Review mechanisms include a vetting process that increases sample labeling to ensure the authenticity and reliability of the data (Figure 2).

Mathematically, a poisoning attack can be realized by modifying the data points in the training set. Specifically, assume that the original training set is

$$D=(X1,y1),(X2,y2), \ldots ,(Xn,yn) \quad (1)$$

where x; is the feature vector and y is the corresponding category label. By modifying some x; or y, the attacker makes the modified training set D' learn a different decision boundary when used to train the SVM model. At this point the sound recognition is performed with bias and wrong semantics are output.

DOI: 10.54254/2755-2721/109/20241412

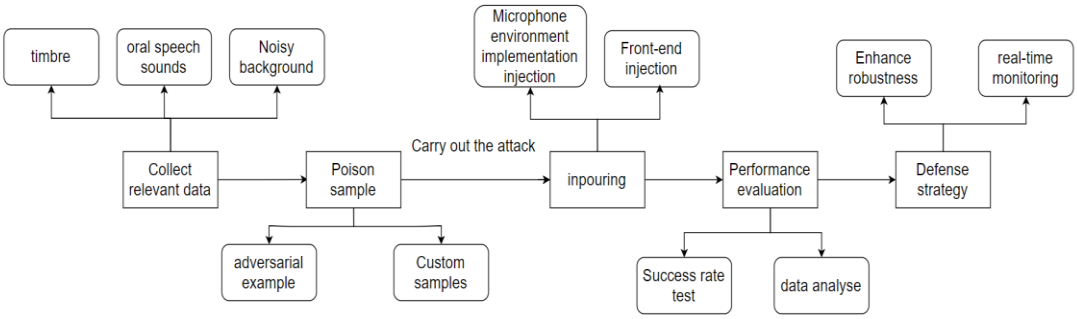

Figure 1. Processes carried out with the collected data

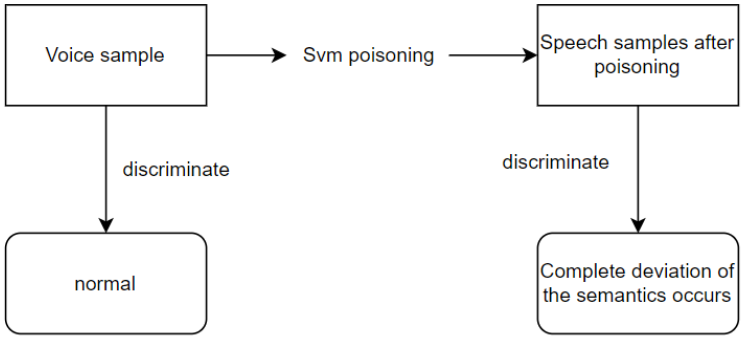

Figure 2. Results of poisoning

3.2. Experimental procedure

The data used for semantic recognition of sounds include LibriSpeech, TED-LIUM, TIMIT, CHiME, AISHELL, and Switchboard. The AISHELL-3 dataset is used as an example in this process. There are a total of 88,035 utterances, 40,000 in the training set and 48,035 in the test set. The corpus contains about 85 hours of emotionally neutral recordings spoken by 218 native speakers of Chinese. Their auxiliary attributes such as gender, age group and native accent are explicitly labelled and provided in the corpus. A new dataset training with 70% of the original dataset and 30% of the confrontation samples is utilized for 30 rounds of training. At this point, backdoor or hidden features are implanted in the model training and perturbation attacks are applied using FGSM and PGD.

3.3. Case Study

In this paper, we analyze bird sound recognition by taking the research of T. Y. Wu as an example, creating Bark sound spectrogram to extract the sound features, extracting the Spectral Descriptors (SSD) and Rhythmic Patterns (RP), and the Mel Frequency Cepstrum Coefficients (MFCC) features, which are input into the SVM classifier and then trained and tested to find the feature method to get the optimal recognition rate of the SSD which is 82. 1%.

Then it is classified by visual features. The audio image algorithm is used to analyze the bird song audio frequency signal and convert it into a visual image from which the artificial features are extracted. Most of the artificial visual features are found to perform better than the acoustic feature methods. The highest recognition rate of 88. 0% was found for multi-scale local phase quantization (MLPQ).

There are also deep learning methods for classification. The converted audio images are fed into a convolutional neural network, and the trained CNN model is migrated through the ImageNet dataset to the classification task. Different CNN architectures are tested and compared. It is found that robust

DOI: 10.54254/2755-2721/109/20241412

classification can be achieved by using transfer learning. The best classification result was achieved by Xception network reaching 90. 6%.

Classification method based on multi-model fusion. Using the output of different visual and acoustic descriptors to improve the recognition rate, it is found that the fusion accuracy is significantly improved. Secondly combining different fine-tuned CNNs is found to improve the model recognition ability. Finally combining the CNN ensemble with artificial and acoustic features gives the best recognition accuracy of 98. 0%.

It was found that the late fusion of acoustic feature methods, visual feature methods and deep learning best-achieved classification performance [6].

4. Challenges

The implementation of SVM sound poisoning can be used to target the sound in the audio recognition system for poisoning attacks [7]. Currently, SVM has security risks, such as adding malicious data in the attack, which leads to recognition errors and increases the proportion of misclassification, is sensitive to noise, is not applicable to multi-classification problems, requires the selection of an appropriate kernel function, and there is a reduction in the adaptability and robustness [8]. SVM faces many challenges in sound poisoning, including the high requirement of linear divisibility of the data, the large amount of training data, the low accuracy of the feature extraction, interference from the external environment, increasing poisoning detection mechanisms and legal consequences [9]. SVM in voice poisoning provides methods but in practical problems, implementation is difficult. The challenges in speech recognition technology are: noise often interferes with the performance of speech recognition systems, accent and speech rate differences, privacy and security [10].

5. Conclusion

This paper focuses on exploring the impact of SVM poisoning on sound recognition systems. This paper concludes that SVM has a significant impact on sound recognition system poisoning, in the speech recognition stage, finding the weakest angle of attack for SVM poisoning, predicting a high impact of poisoning, poisoning results in reduced robustness, increased misclassification, shifted decision boundaries, antagonistic sample generation to further affect the performance of the model, and an increase in the security risks may be exploited by malicious use for spoofing.

In the field of speech recognition, SVM can be applied to speech feature extraction, speech classification and speech synthesis to find the best hyperplane by searching for the maximum margin optimization to achieve effective classification and recognition of speech signals. Future research can explore how to combine the advantages of SVM and deep learning to improve the accuracy and efficiency of speech recognition. Through offline evaluation and arithmetic evaluation, we comprehensively explore the methods and strategies of performance evaluation to provide a basis for model optimization.

Future research can be centred on the design of test queries, the writing of test reports, and the application of test questions in order to comprehensively analyze the effectiveness and reliability of AI applications. Meanwhile, exploring how to combine SVM with other technologies to improve the accuracy and practicality of speech recognition is also an important future research direction.

References

[1] Wen Y and Fan Y 2024 SVM machine learning based malicious domain name detection method CN201910971102. 0 CN110866611A.

[2] Xiang 2022 International Conference on Augmented Intelligent and Sustainable Systems (ICAISS) Trichy, India pp 1244-1247.

[3] Wang Q and Wu 2022 International Conference on Big Data Analytics (ICBDA) Guangzhou, China pp 248-252.

[4] Zhu J, Ju Y and Xia M 2021 International Symposium on Artificial Intelligence, Networking and Information Technology (AINIT) Shanghai, China pp 294-297.

DOI: 10.54254/2755-2721/109/20241412

[5] Kanth R and Saraswathi S 2015 IEEE International Conference on Computational Intelligence and Computational Research (ICCIC) Madurai, India pp 1-6.
[6] Wu T Y 2022 Research on bird sound recognition method based on multi-feature fusion Xi'an University of Electronic Science and Technology.
[7] Gupta S, Mehra A and Vinay 2015 International Conference on Signal Processing and Integrated Networks (SPIN) Noida, India pp 570-574.
[8] Li P, Zhao W T, Liu Q et al. 2018 Computer Science and Exploration 12(2):171-184.
[9] Raj K V, Patil A, Roopashree N, Gudaje S and Kavya G B 2020 International Conference on Computing, Communications and Networking Technologies (ICCCNT) Kharagpur, India pp 1-6.
[10] Carlini N and Wagner D 2018 Proceedings of the 2018 IEEE European Symposium on Security and Privacy (EuroS&P).