DeepSign: Digital Rights Management in Deep Learning Models

Tech ID: 29386 / UC Case 2018-218-0

CONTACT

University of California, San Diego
Office of Innovation and Commercialization
innovation@ucsd.edu
tel: 858.534.5815.

OTHER INFORMATION

KEYWORDS

Artificial intelligence, deep learning, watermarking, intellectual property protection, neural network

CATEGORIZED AS

- **Computer**
  - Security
  - Software
- **Security and Defense**
  - Cyber security

RELATED CASES

2018-218-0

BACKGROUND

As society becomes more and more complicated, we have also developed ways to analyze and solve some of these complexities via the convergence of the fields of artificial intelligence, cognitive science and neuroscience. What has emerged is the development of machine learning, which allows computers to improve automatically through experience. Thus, developers working on artificial intelligence (AI) systems have come forth to align AI with machine-learning algorithms to cover a wide variety of machine-learning problems. The most advanced of these are called supervised learning methods which form their predictions via learned mapping, which can include decision trees, logistic regression, support vector machines, neural networks and Bayesian classifiers. More recently, deep networks have emerged as multilayer networks involved in a number of applications, such as computer vision and speech recognition.

A practical concern in the rush to adopt AI as a service is the capability to perform model protection: AI models are usually trained by allocating significant computational resources to process massive amounts of training data. The built models are therefore considered as the owner’s intellectual property (IP) and need to be protected to preserve the competitive advantage.

TECHNOLOGY DESCRIPTION

Researchers at UC San Diego have developed DeepSign, the first generic Deep Learning watermarking framework that is applicable in both black-box and white-box settings. DeepSign works by embedding the watermark information in the probability density distribution of the activation sets corresponding to different layers of a neural network. A digital watermark is a type of marker covertly embedded in a signal or IP including audio, video image, or functional design. It is commonly adopted to identify ownership of the copyright of such a signal or function.

The performance of the proposed framework is evaluated on MNIST and CIFAR-10 datasets using three different topologies. The results demonstrate that DeepSign satisfies all the criteria for effective watermarking including fidelity, robustness, generalizability, and integrity. DeepSign attains comparable accuracy to the baseline neural network after embedding the watermark and resists potential attacks such as parameter pruning, model fine-tuning, and watermark overwriting.

APPLICATIONS

The invention works by iteratively learning and adjusting the corresponding probability density function of data abstractions to incorporate the desired watermarking information within each layer of the neural network. The watermarking information can later be detected and leveraged to claim the ownership of the neural network or detect IP infringement

ADVANTAGES

DeepSign, for the first time, introduces a generic watermarking methodology that enables IP protection in both and black-box settings, where the adversary may or may not know the internal details of the model.

STATE OF DEVELOPMENT

A working prototype has been designed

INTELLECTUAL PROPERTY INFO

A provisional patent has been submitted and the technology is available for licensing.

RELATED MATERIALS

- Bita Darvish Rohani, Huili Chen, and Farinaz Koushanfar. DeepSign: A Generic Framework for Watermarking and IP Protection of Deep Learning Models. arXiv. 2018 - 04/17/2018

PATENT STATUS

Country	Type	Number	Dated	Case
Patent Cooperation Treaty	Published Application	2019/190886	10/03/2019	2018-218

University of California, San Diego
Office of Innovation and Commercialization
9500 Gilman Drive, MC 0910, ,
La Jolla,CA 92093-0910

Tel: 858.534.5815
innovation@ucsd.edu
https://innovation.ucsd.edu
Fax: 858.534.7345

© 2018, The Regents of the University of California
Terms of use
Privacy Notice