Formal Proofs of Orthogonality for Class-Incremental Learning for Wireless Device Identification in IoT

Yongxin Liu [1], Jian Wang [2], Jianqiang Li [2], Shuteng Niu [2], and Houbing Song [2]

[1]Embry-Riddle Aeronautical University
[2]Affiliation not available

October 30, 2023

Posted on 19 May 2020 — CC-BY 4.0 — https://doi.org/10.36227/techrxiv.14559765.v1 — e-Prints posted on TechRxiv are preliminary reports that are not peer reviewed. They should not b...

Abstract

This document provides a formal proof and supple- mentary information of the paper: Class-Incremental Learning for Wireless Device Identification in IoT. The original paper focuses on providing a novel and efficient incremental learning algorithm. In this document, we explicitly explain why the mem- ory representations (latent device fingerprints in our application) in Artificial Neural Networks approximate orthogonality with insights for the invention of our Channel Separation Incremental Learning algorithm.

Formal Proofs of Orthogonality for Class-Incremental Learning for Wireless Device Identification in IoT

Yongxin Liu, Jian Wang, Jianqiang Li, Shuteng Niu, and Houbing Song, *Senior Member, IEEE*

***Abstract*—This document provides a formal proof and supplementary information of the paper: Class-Incremental Learning for Wireless Device Identification in IoT [1]. The original paper focuses on providing a novel and efficient incremental learning algorithm. In this document, we explicitly explain why the memory representations (latent device fingerprints in our application) in Artificial Neural Networks approximate orthogonality with insights for the invention of our Channel Separation Incremental Learning algorithm.**

***Index Terms*—Internet of Things, Cybersecurity, Big Data Analytics, Non-cryptographic identification, Zero-bias Neural Network, Deep Learning, Memory orthogonality.**

We reused the existing proofs and formulas in the original class-incremental learning paper but with a slightly modified expression to be more generalizable and explicit.

We use the term memory representations to replace the specific term device fingerprints [2]. The decisional memory representations usually exist within the last dense layer of neural networks. And in this document, we do not consider the bias neurons and amplificative attentions, because we have proved that such a simplification will not impair the performance of neural networks [3], [4].

I. Separation of fingerprints at converging point

Intuitively, if the memory representations (the devices' fingerprints), are distantly separated in the latent space, we will have less chance to confuse different concepts (wireless devices). To quantify the separation, the sum of the mutual cosine distances of all memory representations (devices' fingerprints) in a classification model can be defined as:

$$\begin{aligned} TD(\boldsymbol{f_1},\cdots,\boldsymbol{f_C}) &= \sum_{i=1,j<i}^{C} CosineDistance(\boldsymbol{f_i},\boldsymbol{f_j}) \\ &= \sum_{i=1,j<i}^{C} x_i^{(1)}x_j^{(1)} + x_i^{(2)}x_j^{(2)} + \cdots + x_i^{(N1)}x_j^{(N1)} \end{aligned} \tag{1}$$

where $\boldsymbol{f_i} = (x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(N1)})$ and $\boldsymbol{f_j} = (x_j^{(1)}, x_j^{(2)}, \cdots, x_j^{(N1)})$ are devices' fingerprint vectors. Suppose we have C devices with N_1-D fingerprint vectors. Noted that the fingerprints have been normalized into unit vectors. Therefore, if we need to find the optimal value of $TD(\cdot)$, we need to incorporate the constraints:

$$\forall i,\ g(\boldsymbol{f_i}) = \textstyle\sum_{d=1}^{N1}(x_i^{(d)})^2 - 1 = 0 \tag{2}$$

Equation 1 has now become a constrained optimization problem. We solve this constrained optimization problem with the Lagrange Multiplier as:

$$\begin{aligned} &L(\boldsymbol{f_1},\cdots,\boldsymbol{f_C},\lambda_1,\cdots,\lambda_C) \\ &\qquad = TD(\boldsymbol{f_1},\cdots,\boldsymbol{f_C}) - \textstyle\sum_{i=1}^{C}\lambda_i g(\boldsymbol{f}_i) \end{aligned} \tag{3}$$

And we need to solve:

$$\nabla_{x_1^{(1)}\cdots x_1^{(N1)},\cdots,x_C^{1}\cdots x_C^{N1},\lambda_1\cdots\lambda_i} L(\boldsymbol{f_1}\cdots\boldsymbol{f_C},\lambda_1\cdots\lambda_C) = 0 \tag{4}$$

Which results in a linear system of equations. For each kth ($k = 1\cdots N_1$) dimension of memory representation vectors $x_1^{(k)},\cdots,x_C^{(k)}$, we have:

$$\begin{matrix} \frac{\partial L}{x_1^{(k)}} = -2\lambda_1 x_1^{(k)} + \sum_{i=1,i\neq 1}^{C} x_1^{(i)} = 0 \\ \vdots \qquad\qquad \cdots \qquad\qquad \vdots \\ \frac{\partial L}{x_C^{(k)}} = -2\lambda_C x_C^{(k)} + \sum_{i=1,i\neq C}^{C} x_C^{(i)} = 0 \end{matrix} \tag{5}$$

This is a homogeneous system of equations, and it is unlikely that it only has a trivial solution (zeros). Hence, $\lambda_1 = \lambda_2 = \cdots = \lambda_C = -0.5$ and Equation 5 can be converted into one equation:

$$\textstyle\sum_{i=1}^{C} x_i^{(k)} = 0 \tag{6}$$

We square Equation 6 and expand it. According to Multinomial Theorem [5] we have:

$$\textstyle\sum_{i=1}^{C}(x_i^{(k)})^2 + 2\sum_{n=1,m<n}^{C} x_n^{(k)}x_m^{(k)} = 0 \tag{7}$$

Given that $k = 1\cdots N_1$, we have N_1 Equations with an identical form of Equation 7. By summing them up, we have:

$$\sum_{k=1}^{N1}\sum_{i=1}^{C}(x_i^{(k)})^2 + 2\sum_{k=1}^{N1}\sum_{n=1,m<n}^{C} x_n^{(k)}x_m^{(k)} = 0 \tag{8}$$

On the left of Equation 8, the first part is the sum of the magnitude of fingerprint vectors. And its value is C. The second part is exactly two times $TD(\boldsymbol{f_1},\cdots,\boldsymbol{f_C})$ in Equation 1. Now, we have:

Yongxin Liu, Jian Wang, Shuteng Niu, and Houbing Song are with the Security and Optimization for Networked Globe Laboratory (SONG Lab), Embry-Riddle Aeronautical University, Daytona Beach, FL 32114 USA

Jianqiang Li is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, Guangdong 518060 China

Corresponding author: Houbing Song

Manuscript received March 22, 2021; revised XXX.

Remark 1. *The sum of the mutual cosine distances of memory representations (device fingerprints) of DNN at a converging point is a predictable constant:*

$$TD(\boldsymbol{f_1}, \cdots, \boldsymbol{f_C}) = -\frac{C}{2} \tag{9}$$

When such a value is reached, the separation of memory representations are maximized in the latent space, indicating the lowest degree of conflict. Here, conflict can be expressed as interference in neuroscience. We will use the term *Degree of Conflict (DoC)* to describe the characteristic of the zero-bias DNN. Noted that the range of DoC is from $-\frac{C}{2}$ to $\frac{C(C-1)}{2}$. The maximum value is reached when all fingerprints collide into one single vector.

II. Orthogonality Approximation

We define that the averaged cosine distance between N_1 classes is $\bar{D}_0$, according to Remark 1, after initial training we have:

$$\frac{N_1(N_1-1)}{2}\overline{D_0} = -\frac{N_1}{2} \text{ and } \overline{D_0} = -\frac{1}{N_1-1} \tag{10}$$

If N_1 becomes larger, we will have:

$$\overline{D_0} \approx -0 \tag{11}$$

And the averaged cosine distance between device fingerprints or memory representations approximates 90 degrees, thus orthogonal. Apparently, if all memory representations (device fingerprints) are orthogonally distributed, then $\overline{D_0}$ will directly approximate zero.

III. Insights to The Invention of New Incremental Learning Algorithms

If the newly added memory representations are orthogonal to the existing ones, there will not be any conflicts or interference introduced. This is the most essential finding that motivates the invention of Channel Separation Incremental Learning, in which memories of different learning stages are organized into orthogonally separated spaces. And the biological evidence of our work has been revealed in the most recent advancement of neuroscience [2], but with a totally different roadmap.

Acknowledgment

This research was partially supported by the National Science Foundation under Grant No. 1956193.

References

[1] Y. Liu, J. Wang, J. Li, S. Niu, and H. Song, “Class-incremental learning for wireless device identification in iot,” *IEEE Internet of Things Journal*, pp. 1–1, 2021.

[2] A. Libby and T. J. Buschman, “Rotational dynamics reduce interference between sensory and memory representations,” *Nature Neuroscience*, pp. 1–12, 2021.

[3] Y. Liu, J. Wang, S. Niu, and H. Song, “Deep learning enabled reliable identity verification and spoofing detection,” in *International Conference on Wireless Algorithms, Systems, and Applications*. Springer, 2020, pp. 333–345.

[4] Y. Liu, J. Wang, J. Li, H. Song, T. Yang, S. Niu, and Z. Ming, “Zero-bias deep learning for accurate identification of internet of things (iot) devices,” *IEEE Internet of Things Journal*, 2020.

[5] “Multinomial theorem — brilliant math & science wiki,” https://brilliant.org/wiki/multinomial-theorem/, (Accessed on 03/18/2021).

Yongxin Liu (LIU11@my.erau.edu) received his first Ph.D. from South China University of Technology in 2018. He is a Ph.D. student in Electrical Engineering and Computer Science at Embry-Riddle Aeronautical University, Daytona Beach, Florida, and a graduate research assistant in the Security and Optimization for Networked Globe Laboratory (SONG Lab, www.SONGLab.us). His major research interests include data mining, wireless networks, IoT, and unmanned aerial vehicles.

Jian Wang (wangj14@my.erau.edu) is a Ph.D. student in the Department of Electrical Engineering and Computer Science, Embry-Riddle Aeronautical University, Daytona Beach, Florida, and a graduate research assistant in the Security and Optimization for Networked Globe Laboratory (SONG Lab, www.SONGLab.us). He received his M.S. from South China Agricultural University in 2017. His research interests include wireless networks, unmanned aerial systems, and machine learning.

Jianqiang Li (lijq@szu.edu.cn) received his B.S. and Ph.D. degrees from the South China University of Technology in 2003 and 2008, respectively. He is a Professor with the College of Computer and Software Engineering, Shenzhen University, Shenzhen, China. His major research interests include Internet of Things, robotic, hybrid systems, and embedded systems.

Shuteng Niu (shutengn@my.erau.edu) is a Ph.D. student in the Department of Electrical Engineering and Computer Science, Embry-Riddle Aeronautical University (ERAU), Daytona Beach, Florida, and a graduate research assistant in the Security and Optimization for Networked Globe Laboratory (SONG Lab, www.SONGLab.us). He received his M.S. from ERAU in 2018. His research interests include machine learning, data mining, and signal processing.

Houbing Song (M’12–SM’14) received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, in August 2012.

In August 2017, he joined the Department of Electrical Engineering and Computer Science, Embry-Riddle Aeronautical University, Daytona Beach, FL, where he is currently an Assistant Professor and the Director of the Security and Optimization for Networked Globe Laboratory (SONG Lab, www.SONGLab.us). He has served as an Associate Technical Editor for IEEE Communications Magazine (2017-present), an Associate Editor for IEEE Internet of Things Journal (2020-present), IEEE Transactions on Intelligent Transportation Systems (2021-present) and IEEE Journal on Miniaturization for Air and Space Systems (J-MASS) (2020-present). He is the editor of seven books, including Big Data Analytics for Cyber-Physical Systems: Machine Learning for the Internet of Things, Elsevier, 2019, Smart Cities: Foundations, Principles and Applications, Hoboken, NJ: Wiley, 2017, and Industrial Internet of Things, Cham, Switzerland: Springer, 2016. He is the author of more than 100 articles. His research interests include cyber-physical systems/Internet of things, cybersecurity and privacy, AI/machine learning/big data analytics, and unmanned aircraft systems. His research has been featured by popular news media outlets, including IEEE GlobalSpec’s Engineering360, Association for Unmanned Vehicle Systems International (AUVSI), Fox News, USA Today, U.S. News & World Report, The Washington Times, and New Atlas.

Dr. Song is a senior member of ACM and an ACM Distinguished Speaker. Dr. Song was a recipient of 5 Best Paper Awards (CPSCom-2019, ICII 2019, ICNS 2019, CBDCom 2020, WASA 2020).