公開

（事後評価・研究者用）

研究報告書

「頑強なハイブリッド深層学習モデルの自動探索システム」

研究期間： 2020 年 4 月～2022 年 3 月
研究者番号： 50243
研 究 者： ヴァルガス ダニロ

1． 研究のねらい

Current deep learning methods lack robustness to be employed in critical applications such as autonomous driving, medical aiding systems. The objective of this research is to investigate new paradigms and architectures that could lead to greater robustness in machine learning methods. Specifically, we investigated machine learning robustness from three completely different perspectives: (a) from a security perspective, (b) from a deep learning model perspective and (c) from a novel self-organizing paradigm that does not rely on optimization to learn (this novel paradigm also differs completely from self-organizing maps, using merely dynamical equations to learn patterns).

2． 研究成果

（1）概要

The research here discovered a novel paradigm for machine learning. This new paradigm called Self-Organizing Dynamical Equations does not rely on optimization or parametrized models to learn, instead it uses only dynamical equations. Experiments have shown that the new paradigm surpasses state-of-the-art unsupervised deep learning methods most of the time. In fact, the proposed paradigm is inherently adaptive and robust. Input fluctuations and changes in the problem structure does not affect the model learning capabilities which is something unheard of in machine learning research [Vargas, Asabuki, AAAI2021]. Beyond these results, investigation into the vulnerabilities of deep learning models revealed a clearer picture of the problems ahead for this research direction.

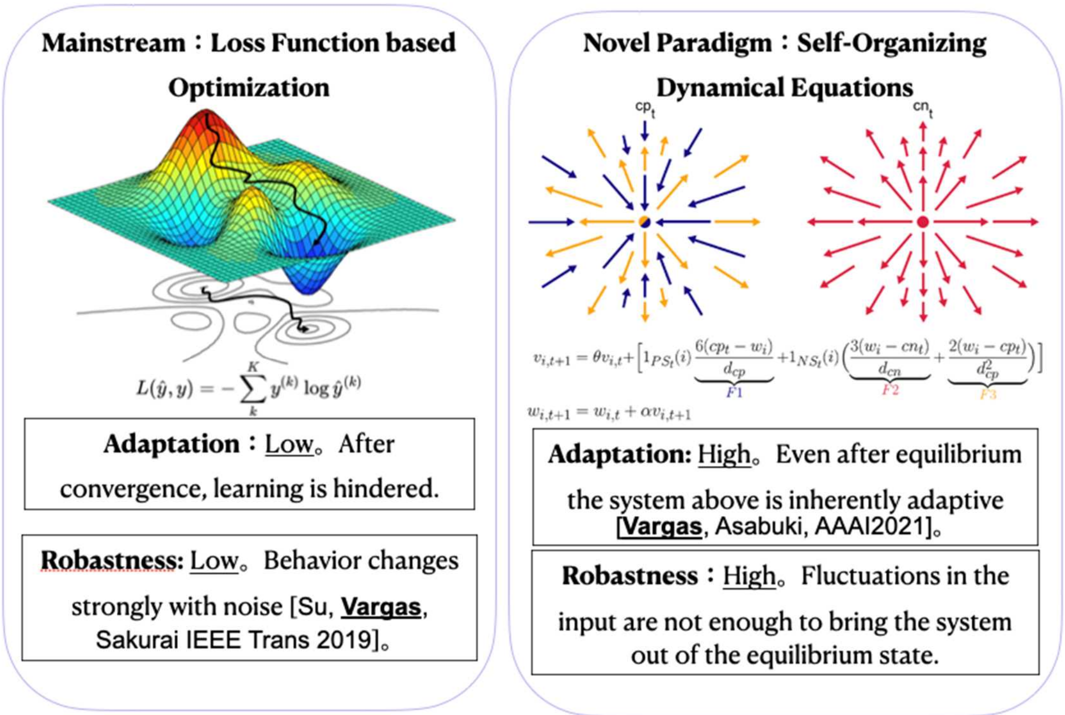
（2）詳細

Research Theme [Robust and Adaptive Machine Learning with Self-Organizing Dynamical Equations]

In the search to find a new foundation for machine learning that can be as robust as solutions found in Nature, I raised the question of “can learning happen without optimization?”. This is not a random guess, Nature itself does not seem to rely on complex optimization strategies. Evolution itself has a very simple logic that works both in ecosystems and immune systems alike. Albeit the simplicity, it seems that the complexity, adaptation, and robustness of Nature is orders of magnitude beyond our best artificial creations. Many of these systems, however, are self-organizing ones, which have many local interactions and complex emerging features.

Motivated by such cues, I tried to create equations that mimic Hebbian learning (i.e., would pull weights close together when they activate together) together with an unusual anti-Hebbian rule (i.e., repel weights that do not activate together) that could reach equilibrium when used together (Fig. 1). Interestingly, I found out that for chunking variables in sequences, such weights create a space in which the distances are proportional to the correlation between input states. In fact, the space created is accurate enough, that clustering in this space surpasses other unsupervised algorithms of the state-of-the-art most of the time [Vargas, Asabuki, AAAI2021]. Beyond this, the equilibrium of the system is defined by the input and the dynamical equations, when the input structure changes, the previous equilibrium state also vanishes. This creates an inherently adaptive system that changes itself together with changes in the input.

When analyzed closely, the dynamical equations lead to the emergence of attractor-repeller points which shape the clusters and are in themselves the patterns learned. Changes in the structure of the problem also changes the place and existence of such attractor-repeller points [Tham, Vargas, 2021].



Mainstream：Loss Function based Optimization

$$L(\hat{y}, y) = -\sum_{k}^{K} y^{(k)} \log \hat{y}^{(k)}$$

Adaptation：Low。 After convergence, learning is hindered.

Robastness: Low。 Behavior changes strongly with noise [Su, **Vargas**, Sakurai IEEE Trans 2019]。

Novel Paradigm：Self-Organizing Dynamical Equations

$$v_{i,t+1} = \theta v_{i,t} + \left[1_{PS_t}(i) \underbrace{\frac{6(cp_t - w_i)}{d_{cp}}}_{F1} + 1_{NS_t}(i) \left(\underbrace{\frac{3(w_i - cn_t)}{d_{cn}}}_{F2} + \underbrace{\frac{2(w_i - cp_t)}{d_{cp}^2}}_{F3} \right) \right]$$

$$w_{i,t+1} = w_{i,t} + \alpha v_{i,t+1}$$

Adaptation: High。 Even after equilibrium the system above is inherently adaptive [**Vargas**, Asabuki, AAAI2021]。

Robastness：High。 Fluctuations in the input are not enough to bring the system out of the equilibrium state.

Figure 1 Overview of the Self-Organizing Dynamical Equations' paradigm

In fact, the investigation in the first phase of the ACT-I (before the acceleration stage) have identified a couple of features that are related to robustness in machine learning. The two main ones were (a) dynamical properties and (b) nonlinearity. This novel paradigm beyond adaptive has the two key features responsible for robustness and shows promising results towards a novel foundation for machine learning based on self-organization and dynamical equations rather than optimization and parametrized models.

Research Theme [Understanding and Evaluation of Robustness]

To further understand the problems of the mainstream of machine learning, especially current deep learning methods, I created some experiments to understand the reason for such a vulnerability [Kotyan, Vargas, AI Safety at IJCAI2021], [Vargas, Su, AI Safety at IJCAI2021]. One experiment investigated what happens with pixels modified in images and how change in such pixels propagate through the network. The result can be seeing in Fig. 2, in which the perturbation is shown to reach the deeper levels of a ResNet even when the modification is not originally from an adversarial sample (a sample in which the modification made the classification change). The reason for that is lack of non–linearity and dynamism in the model.

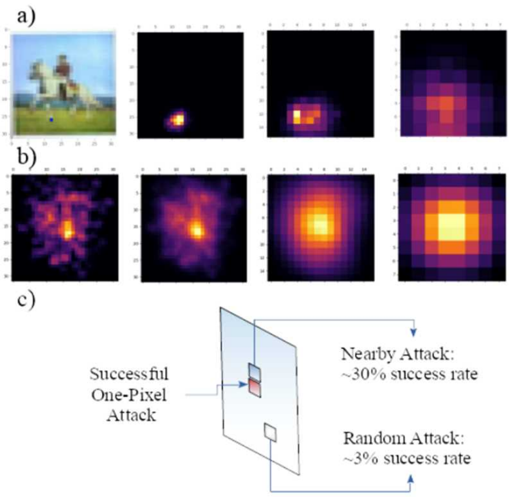


Figure 2 a) Propagation Maps of a successful one-pixel attack on Resnet shows how the influence of one pixel perturbation grows and spreads (bright colors show differences in feature map that are close to the maximum original layer output). b) Average Propagation Map over the entire set of propagation maps shows the overall distribution of attacks and their propagation. c) Illustration of locality analysis.

Research Theme [A Multidisciplinary View of Intelligence]

Robust and adaptive intelligence is barely understood not only in machine learning but also in psychology, neuroscience, decision making, among other areas. A multidisciplinary view of intelligence might be the key to understand its most intricate mechanisms. Such multidisciplinary understandings can also ignite new methods in machine learning creating a mutual benefit cycle.

Based on this, I also investigated decision making [Vargas, Lauwereyns, Cognitive Neurodynamics2021] and started still unpublished research on psychology and neuroscience. Such an unusual multidisciplinary approach might seem unnecessary at first glance; however, they were the basis of the main result of this research project (namely the Self-Organizing Dynamical Equations paradigm). This reveals the importance of such a holistic multidisciplinary view.

3．　今後の展開

Motivated by the strong results from this research proposal, I will continue to investigate (a) self-organization adaptive and robust paradigms that could transform the field of artificial intelligence, (b) architectures for deep learning that could increase robustness substantially and (c) study intelligence from a multi-disciplinary perspective using psychology and neuroscience, aiming at understanding how intelligence is developed in live beings as well as utilize this information to improve methods in machine learning. The advances in robustness for machine learning would allow applications to critical systems such as autonomous driving while understanding intelligence from a multi-disciplinary perspective can open borders for applications, collaborations, among other social economical benefits.

4．　自己評価

The objective of this research proposal was to investigate paths to robustness with different paradigms.

・Achievement – The main objective was achieved with an unexpected strong success! A primer into a possible new foundation for machine learning which is robust and adaptive right from its first smaller element was created, i.e., Self-Organizing Dynamical Equations.

・Research Progress – Progress and planning for new foundations rarely follow a linear path, because they need some key ideas or developments to ignite. Despite the expected nonlinearity, the progress was unusually fast and steep for such foundational research.

・Future Prospects – Further developments on the novel paradigm proposed here should allow for inherently adaptive and robust machine learning systems to be employed in real world applications. This is especially important for tasks in which deep learning do not convince, i.e., robotics, autonomous driving, and other critical applications.

・Novelty – In the research here, a new foundation for machine learning has been shown to be possible.

5．　主な研究成果リスト

（1）論文（原著論文）発表

1. Danilo Vasconcellos Vargas, Toshitake Asabuki (2021), ″Continual General Chunking Problem and SyncMap″, Accepted, AAAI 2021. (Acceptance rate 19.8%)
2. Y. F. Tham and Danilo Vasconcellos Vargas, "Towards learning Hierarchical Structures with SyncMap", CYBCONF 2021.
3. Danilo Vasconcellos Vargas, Johan Lauwereyns, "Setting the space for deliberation in decision-making", Cognitive Neurodynamics, 1-13, 2021. (**Impact Factor: 3.9**)
4. Shashank Kotyan and Danilo Vasconcellos Vargas, "Deep neural network loses attention to adversarial images", AI Safety Workshop at IJCAI, 2021.

5. Danilo Vasconcellos Vargas, and Su, J., "Understanding the one-pixel attack: Propagation maps and locality analysis", AI Safety Workshop at IJCAI2020.

（2）特許出願

（3）その他の成果（主要な学会発表、受賞、著作物、プレスリリース等）

受賞：

2022 - IEEE Transactions on Evolutionary Computation Outstanding Paper Award 2022

本：

Van Uytsel, S., & Vargas, D. V. (2021) *Autonomous Vehicles: Business, Technology and Law*. Springer. ISBN: 978-981-15-9254-6.

招待公演：

2021 - "On the Deeper Secrets of Deep Neural Networks and a Path Forward", Beyond AI – Summer School, Virtual Vehicle Research GmbH, Austria (Online)

2021 - "SAN (ノベルティに基づくサブポピュレーションアルゴリズム)による多目的最適化の最先端" IMI 研究集会「進化計算の数理」, Kyushu University, Japan

2020 - "1ピクセルで誤魔化される人工知能が人間を超えた？" 第7回 AI Optics 研究会, Online