様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月 １２ 日現在

機関番号：８２４０１
研究種目：基盤研究(B)（一般）
研究期間：2020～2022
課題番号：２０Ｈ０４２４７

研究課題名（和文）Life-Long Deep Learning using Bayesian Principles

研究課題名（英文）Life-Long Deep Learning using Bayesian Principles

研究代表者
Ｋｈａｎ　Ｅｍｔｉｙａｚ（Khan, Emtiyaz）

国立研究開発法人理化学研究所・革新知能統合研究センター・チームリーダー

研究者番号：３０８５８０２２

交付決定額（研究期間全体）：（直接経費）　　14,000,000円

研究成果の概要（和文）：現在のディープラーニングは、継続的に学習することができず、過去に見た情報を簡単に忘れてしまう傾向にある。本研究では、忘却を抑制する新しい深層学習の手法を開発する。これは、過去の記憶の中から重要なデータを特定し、それを再利用することで実現する。提案手法は継続学習を普遍的に改善する手法であり、性能の良い継続学習方法の大半は提案方法の派生であることを示すことができた。また、提案方法はスケーラブルであり実用的なスケールの問題に適用することもできる。

研究成果の学術的意義や社会的意義
Deep-learning methods require a huge amount of computing resources and also a lot of data. Our work reduces the dependencies on such resources. We aim to design AI systems that continue to learn and improve throughout their lifetime.

研究成果の概要（英文）：Current deep learning method cannot learn continually, and can easily forget the past information seen a long time ago. We developed new methods for continual deep learning where we reduce the forgetting. We do so by identifying and reusing a memory of the past. We show that our methods are universal, that is, any method that work well must have similar properties to ours. Our method is scalable and can be applied in practical settings.

研究分野：Bayesian deep learning

キーワード：Continual learning　Bayesian deep learning　Lifelong learning

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

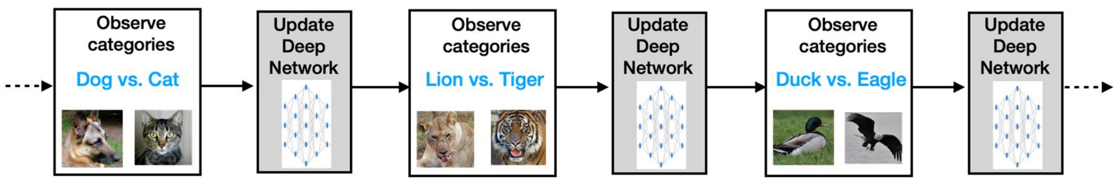
An important goal of data-driven AI is to design systems that continue to learn and improve throughout their lifetime. Deep-learning models are not trained this way currently. Rather, they rely on small, random subsets of data which are sampled from a large dataset (see Fig. 1 below). Unfortunately, such large datasets are unavailable in fields, such as medicine, robotics, education, climate research etc. which makes application of deep learning challenging. A life-long deep learning system would be ideal for such cases, but designing such system is one of the biggest challenges for data-driven AI.

Fig. 1: Deep learning requires a large dataset, (e.g., ImageNet), such that random subsets can be used to update a deep network.

２．研究の目的

How can we design deep-learning systems that can continue to learn and improve as it collects more and more data? A desirable example is shown below in Fig. 2 below, where, at a given time, a classifier only sees a subset of categories, e.g., images of certain animals. These categories are only seen once, and never again. Humans and animals can learn from such data streams, but deep networks forget old categories [Kirkpatrick et al., 2017]. Our main goal is to fix these type of problems and design more effective and practical methods for life-long deep learning.

Fig. 2: Life-long learning systems can learn continually, even when categories are seen sequentially and never stored together in a large dataset. This type of learning is called *Continual Learning*.



３．研究の方法

Our proposal departs significantly from traditional life-long learning approaches due to two key components. **The first component comprises of methods to "identify, memorize, and recall" useful past experiences** while collecting and analyzing new ones. Similar qualities are essential for humans and animals. During our lives, we identify and remember useful skills that can help us in the future. When learning a new skill, we can recall and use previously learned skills to improve quickly. Our goal is to incorporate similar qualities in a deep-learning system to enable effective life-long learning.

The second key component is a set of Bayesian principles to build the first component. Our recent work derives Bayesian principles for deep-learning (Khan et al. 2018; Osawa et al. 2019; Khan et al. 2019). Surprisingly, this also unravels the relationship between trained models and their training data, and can be used to identify important data which led to the discovery of the model (see Fig. 5, modified from Khan et al. 2019). This alternate view is completely missing in deep learning, where the models do not directly tell us which data to feed them, rather they simply ingest the random data given to them. Our Bayesian principle fix this issue, and in this proposal, we will use them to identify, memorize and recall useful data examples.

４．研究成果

During the three years of the projects, we published 3 papers where we build a few comopnents of our approach. In 2020, we proposed a new “continual” deep learning where we regularize the “memory” of the past data to ensure that the model does not forget them. This new regularization method improves over the traditional EWC“weight-regularization” method by Kirkpartick et al. (2017). Our paper was accepted as an oral presentation at NeurIPS 2020 (105 out of 9454 submissions)

- Continual Deep Learning by Functional Regularisation of Memorable Past (NeurIPS 2020) P. Pan*, S. Swaroop*, A. Immer, R. Eschenhagen, R. E. Turner, M.E. Khan

(a) Most memorable (b) Least memorable (a) Most memorable (b) Least memorable

Figure 3: Most memorable and least memorable datapoints on MNIST (left) and CIFAR-10 (right). Memorable points are difficult to classify and lie on the decision boundary. Additionally, our method for choosing memorable points (Step B in FROMP) is computationally cheap.

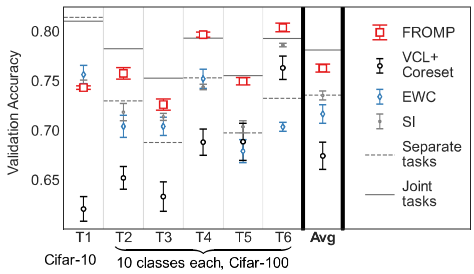


Figure 2: FROMP outperforms baselines on Split CIFAR (see ‘Avg’ column). ‘Separate tasks’: different networks are trained on each task separately. ‘Joint tasks’: a single network is trained jointly on all task data (upper-bound to continual learning performance).

In 2021, we worked on improving our understanding of this method. Current implementation uses the Gaussian-Process viewpoint but this leads to difficult computations. We wanted to go beyond this and understand fundamental principles behind the working of this method; our hope was to not only simplify the computations involved but to show that our method is universal and that any adaptation task will require a regularization of the form we used in this work. This lead to the creation of Knowledge-Adaptation Priors or K-priors. Our functional regularizer is a special type of K-priors. We show that, by carefully choosing the past memory, K-priors can reconstruct the gradients of the past faithfully. This principle is present in all sorts of methods, ranging from simple linear regression, to kernel methods, and also deep neural networks (including knowledge distillation). Our paper was published in NeurIPS 2021.

Khan, M. E. and Swaroop, S. "Knowledge-adaptation priors" Advances in Neural Information Processing Systems 34, pp. 19757-19770, (2021).

In the last year, we aimed to apply our method to large-scale problem, such as ImageNet. Using the principle of gradient reconstruction, we proposed an improvement to combine the EWC method with our memory-based approach. We applied this method to Tiny-ImageNet. We could not finish the ImageNet work, but we have made an important step towards it. The paper is finished and we are submitting it to a journal.

- (Under submission) Improving Continual Learning by Accurate Gradient Reconstructions of the Past, Erik Daxberger, Siddharth Swaroop, Kazuki Osawa, Rio Yokota, Richard E Turner, Jose; Miguel Hernandez-Lobato, Mohammad Emtiyaz Khan

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　2件／うち国際共著　2件／うちオープンアクセス　2件）

１．著者名 Pan, P., Swaroop, S., Immer, A., Eschenhagen, R., Turner, R. and Khan, M. E.	４．巻 374
２．論文標題 Continual deep learning by functional regularization of the memorable past	５．発行年 2020年
３．雑誌名 Advances in Neural Information Processing Systems	６．最初と最後の頁 4453-4464
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Khan, M. E. and Swaroop, S.	４．巻 34
２．論文標題 Knowledge-Adaptation Priors	５．発行年 2021年
３．雑誌名 Advances in Neural Information Processing Systems	６．最初と最後の頁 19757--19770
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕　計0件

〔図書〕　計0件

〔産業財産権〕

〔その他〕

－

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	Ａｌｑｕｉｅｒ　Ｐｉｅｒｒｅ (Alquier Pierre) (10865645)	国立研究開発法人理化学研究所・革新知能統合研究センター・研究員 (82401)	
研究分担者	横田　理央 (Yokota Rio) (20760573)	東京工業大学・学術国際情報センター・准教授 (12608)	

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関