

# Riemannian Walk for Incremental Learning: Understanding Forgetting and Intransigence

Arslan Chaudhry , Puneet K. Dokania , Thalaiyasingam Ajanthan , Philip H. S. Torr

University of Oxford, United Kingdom  
{firstname.lastname}@eng.ox.ac.uk

**Abstract.** Incremental learning (IL) has received a lot of attention recently, however, the literature lacks a precise problem definition, proper evaluation settings, and metrics tailored specifically for the IL problem. One of the main objectives of this work is to fill these gaps so as to provide a common ground for better understanding of IL. The main challenge for an IL algorithm is to update the classifier whilst preserving existing knowledge. We observe that, in addition to *forgetting*, a known issue while preserving knowledge, IL also suffers from a problem we call *intransigence*, its inability to update knowledge. We introduce two metrics to quantify *forgetting* and *intransigence* that allow us to understand, analyse, and gain better insights into the behaviour of IL algorithms. Furthermore, we present RWalk, a generalization of EWC++ (our efficient version of EWC [6]) and Path Integral [25] with a theoretically grounded KL-divergence based perspective. We provide a thorough analysis of various IL algorithms on MNIST and CIFAR-100 datasets. In these experiments, RWalk obtains superior results in terms of accuracy, and also provides a better trade-off for forgetting and intransigence.

## 1 Introduction

Realizing human-level intelligence requires developing systems capable of learning new tasks continually while preserving *knowledge* about the old ones. This is precisely the objective underlying incremental learning (IL) algorithms. By definition, IL has ever-expanding output space, and no or limited access to data from the previous tasks while learning a new one. This makes it more challenging and fundamentally different from the classical learning paradigm where the entire dataset is available and the output space is fixed. Recently, there have been several works in IL [6,14,19,25] with varying evaluation settings and metrics making it difficult to establish fair comparisons. The first objective of this work is to rectify these issues by providing precise definitions, evaluation settings, and metrics for IL for the classification task.

Let us now discuss the key points to consider while designing IL algorithms. The first question is *‘how to define knowledge: factors that quantify what the model has learned’*. Usually, knowledge is defined either using the input-output behaviour of the network [4,19] or the network parameters [6,25]. Once the knowledge is defined, the objective then is to *preserve* and *update* it to counteract two inherent issues with IL algorithms: (1) *forgetting*: catastrophically forgetting knowledge of previous tasks; and

---

Joint first authors































