Semi-Supervised Lifelong Language Learning

Yingxiu Zhao[1*], **Yinhe Zheng**[2†], **Bowen Yu**[2], **Zhiliang Tian**[1], **Dongkyu Lee**[1], **Jian Sun**[2], **Yongbin Li**[2†], **Nevin L. Zhang**[1]

[1] The Hong Kong University of Science and Technology, Hong Kong SAR, China

[2] Alibaba Group, China

{yzhaocx,ztianac,dleear,lzhang}@connect.ust.hk, zhengyinhe1@163.com, {yubowen.ybw,shuide.lyb}@alibaba-inc.com, jiansun_china@hotmail.com

Abstract

Lifelong learning aims to accumulate knowledge and alleviate catastrophic forgetting when learning tasks sequentially. However, existing lifelong language learning methods only focus on the supervised learning setting. Unlabeled data, which can be easily accessed in real-world scenarios, are underexplored. In this paper, we explore a novel setting, semi-supervised lifelong language learning (SSLL), where a model learns sequentially arriving language tasks with both labeled and unlabeled data. We propose an unlabeled data enhanced lifelong learner to explore SSLL. Specially, we dedicate task-specific modules to alleviate catastrophic forgetting and design two modules to exploit unlabeled data: (1) a virtual supervision enhanced task solver is constructed on a teacher-student framework to mine the underlying knowledge from unlabeled data; and (2) a backward augmented learner is built to encourage knowledge transfer from newly arrived unlabeled data to previous tasks. Experimental results on various language tasks demonstrate our model's effectiveness and superiority over competitive baselines under the new setting SSLL. We will release the code and data [1].

1 Introduction

A remarkable ability of humans is to learn and accumulate knowledge continuously throughout their lifetime. Such *Lifelong Learning* ability is crucial for computational systems interacting with the real world and processing continuous streams of information (Parisi et al., 2019; Delange et al., 2021). However, most deep neural networks studies assume data distributions are stationary, which is not applicable in the real-world environments that dynamically evolve. In such real scenarios, models often suffer from *catastrophic forgetting* (McClelland et al., 1995; Parisi et al., 2019): a phenomenon where models forget previously learned knowledge when learning new tasks sequentially.

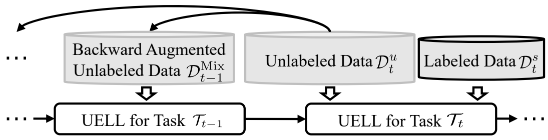

Figure 1: The Training process of our model UELL.

Various approaches have been proposed to alleviate catastrophic forgetting in lifelong scenarios. Attempts include constraining the variants of important weights with regularization (Schwarz et al., 2018; Mi et al., 2020), storing real samples or using pseudo samples for previous tasks to maintain the learned knowledge (d'Autume et al., 2019; Sun et al., 2019; Chuang et al., 2020), or dedicating task-specific modules to avoid the interference among tasks (Madotto et al., 2021; Qin and Joty, 2022; Zhang et al., 2022). Despite their reported effectiveness, these approaches are mostly designed to handle supervised learning tasks, where only labeled data are available. In real-world scenarios, labeled data are generally expensive and time-consuming to obtain, whereas unlabeled data are much easier to collect. These unlabeled data often carry rich information and have been successfully utilized to improve model performance in semi-supervised learning (Xie et al., 2020a; Chen et al., 2021).

In this paper, we investigate a novel setting: Semi-Supervised Lifelong Language learning (SSLL), where a model learns sequentially arriving language tasks with limited labeled data and adequate unlabeled data (see the training process in Fig. 1). The abundant information in the unlabeled data can not only facilitate learning the current task but also benefit learned tasks with similar data distributions. For example, sentiment analysis and topic classification tasks may only differ in their la-

[*] Work done while the author was interning at Alibaba.
[†] Corresponding author.
[1] https://github.com/AlibabaResearch/DAMO-ConvAI/tree/main/ssll

Findings of the Association for Computational Linguistics: EMNLP 2022, pages 3966 - 3980
December 7-11, 2022