Open-world Machine Learning: Supplementary Material

JITENDRA PARMAR and SATYENDRA S. CHOUHAN, Malaviya National Institute of Technology Jaipur, INDIA
VASKAR RAYCHOUDHURY, Miami University, USA
SANTOSH S. RATHORE, ABV-IIITM Gwalior, INDIA

This file is created as supplementary material for the review paper: "Open-world Machine Learning: Applications, Challenges, and Opportunities". In this file, Section 1 discusses related areas such as Transfer Learning, Active Learning, Lifelong Machine Learning, and Multi-Task Learning.

ACM Reference Format:
Jitendra Parmar, Satyendra S. Chouhan, Vaskar Raychoudhury, and Santosh S. Rathore. 2022. Open-world Machine Learning: Supplementary Material. 1, 1 (July 2022), 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Related Areas

Some related areas closely associated with open-world machine learning are mentioned and discussed briefly in this section.

1.1 Transfer Learning

Traditional Machine Learning (ML) techniques predict the expected data by applying analytical principles trained on previously accumulated unlabeled or labelled training examples [3, 10, 26]. Analysis of transfer learning has brought attention since 1995 in several titles: knowledge transfer, learning to learn, multitask learning, knowledge consolidation, inductive transfer, knowledge-based inductive bias, context-sensitive learning, cumulative learning [23, 25], and multitask learning framework [14]. Transfer learning involves interpreting data for a reference task to provide a productive basis for a new task. Transfer learning is often applied to specific data sets with some labelled value. For example, an actual demonstrative prototype of one virus would have a significant advantage in developing a distinguishing prototype for another virus, for which fewer training samples are available. While all learning involves generalization across all queries, transfer learning illustrates the transfer of information across comparable but non-consistent fields, tasks, and distributions. In distinction, the unlabeled data does not require to be obtained from a similar task in the transfer learning framework. In the prior decade, there has been substantial development in improving cross-task transfer by utilizing both discriminative and generative strategies in a broad category of frames.

Authors' addresses: Jitendra Parmar, 2019rcp9044@mnit.ac.in; Satyendra S. Chouhan, sschouhan@mnit.ac.in, Malaviya National Institute of Technology Jaipur, JLN Marg, Jaipur, Rajsthan, INDIA, 302017; Vaskar Raychoudhury, Miami University, 510 E. High St., Oxford, Ohio, USA, raychov@miamioh.edu; Santosh S. Rathore, ABV-IIITM Gwalior, , Gwalior, MadhyaPradesh, INDIA.

XXXX-XXXX/2022/7-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1.2 Active Learning

Active Learning [24] is a discipline of machine learning where the algorithm is designed for learning and can choose the data for learning or learning strategy generated during learning (Figure 1). The active learning methodologies can play an essential role in domains that deal with real-time data such as speech recognition, information extraction [19], classification, and filtering. Moreover, active learning provides high accuracy with a small testing size of labelled data.

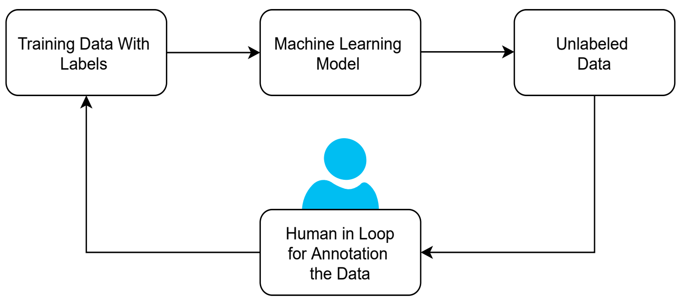

Fig. 1. Basic Framework of Active Learning [18]

There are different types of scenarios of active learning, such as membership query synthesis [1], stream-based selective sampling [8], and pool-based active learning [12]. The standard data mining methods learn models with isolated data and make a prediction based on static models [3, 10, 26]. It needs to use previous knowledge, or a learning model should transfer knowledge and be used to predict future learning. It is termed transfer learning [17]. The knowledge can be transferred in various forms such as transferring knowledge of instances [9], knowledge of feature representations [2] (for both supervised and unsupervised), knowledge of parameters [11] and relational knowledge [16].

1.3 Lifelong Learning/Continual Machine Learning

Lifelong machine learning is a system that can continuously learn from different domains, and this knowledge can be used effectively on future tasks in an efficient manner [21]. The selective knowledge is transferred when learning a novel task. Knowledge is retained from a different source and improves learning. The various techniques of lifelong learning as prior work in knowledge retention improves learning a new task. The significant tasks of lifelong machine learning are shown in Figure 2.

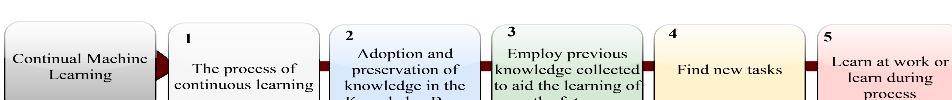

Fig. 2. Tasks of Continual Machine Learning

There are different names: constructive induction, incremental and continual learning, explanation-based learning, sequential task learning, and never-ending learning. These methods are further divided into different categories: lifelong machine learning is supervised learning, continual learning is reinforcement learning, and self-taught learning or never-ending learning is unsupervised learning.

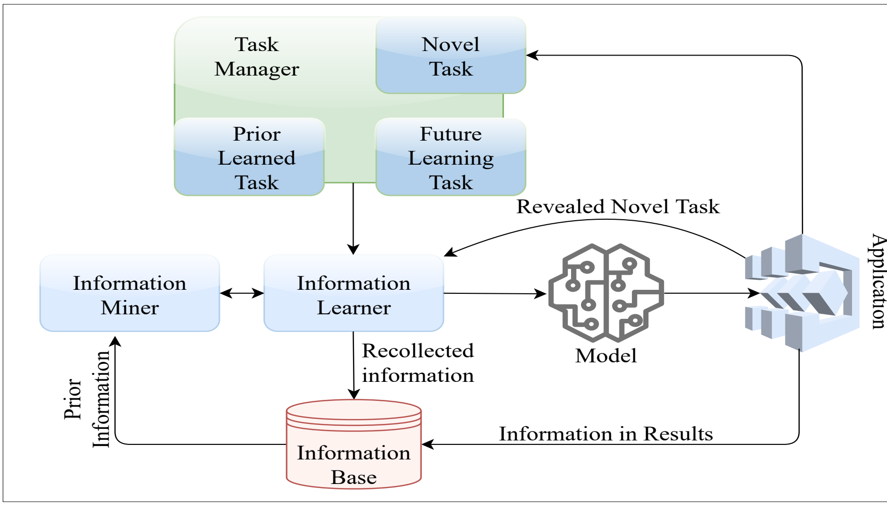

Fig. 3. Basic Framework of Lifelong Machine Learning [7]

Supervised lifelong learning uses an explanation-based Neural Network (EBNN) with a back propagation gradient. Whenever new learning tasks occur, EBNN utilises prior domain information of the task. EBNN gives more accurate outcomes even with fewer amounts of data. In [20], authors suggested knowledge-based cascade correlation neural networks. This method utilises prior trained networks and concealed units to set the new bias for a novel task. Unsupervised Lifelong learning is used to increase the system's scalability, and adaptive resonance theory has been used to map the bottom and top nodes of clusters. It uses a threshold for a new example node, a map with an attention parameter (below threshold).

In [22], authors proposed a novel approach to ensemble clusters from the primary partition of objects; it uses labels of the cluster deprived of accessing the original features. The self-taught learning models build high-level features using unlabeled data and test such models for various image, web, and song genre classification applications. Lifelong learning goals can be achieved by another popular method: Never-ending Language Learner (NELL) [4]. NELL extracts data or reads information from the web and increases its knowledge, then learns how to perform a new task better than the same task done the earlier day. Rather than focus on traditional machine learning, the system should retain knowledge and transfer this to the system as the learning agent. The system should learn sequential tasks and increase their magnitude.

1.3.1 Challenges and Benefits of Lifelong Learning Models [21].

- **Input /output Type, Complexity and Cardinality:** The real-time environment has a variety of data from different domains; it can differ in nature. The attributes of each input may vary according to their source and required task.
- **Training Examples Vs Prior Knowledge :** In life-long learning systems, prior knowledge is a crucial part of the end-to-end system to achieve accuracy while performing a new task. There is a need to retain valid data from the knowledge base that must have information deception as a training example.
- **Effective and Efficient knowledge Retention:** The system must retain efficient information that must not be erroneous. Furthermore, it must use finite memory to store knowledge

with limited computational capacity. The system must be capable of handling duplicate data and increasing the accuracy of prior knowledge.

- **Effective and Efficient knowledge Transfer :** Prior knowledge should not increase computational time and effort. Moreover, the knowledge transfer does not generate less accurate inputs/models for new tasks. There are three major components of lifelong learning.

(1) Retention of learned task knowledge,
(2) Selective transfer of prior while learning a new task, and
(3) The system must ensure that retention and transfer of knowledge must be efficient.

- **Scalability:** Scalability is one of the most challenging and essential aspects of almost all fields of computer science. The system must be able to adapt increments in volumes of input data. The lifelong learning systems must be able to address the space and time complexity of all these factors.
- **Heterogeneous Domain of Task:**The lifelong learning systems must handle data from different domains by establishing relations between the origin and targeted domains. There are many standard features, but diversity in transferred knowledge data also exists. The system must have the ability to map features in transferred knowledge.

1.4 Multi-task Learning

Multi-task learning (MTL) [5, 6, 13] acquires various associated tasks concurrently, beaming at delivering a more reliable representation by using the associated knowledge yielded by various jobs. Introducing inductive bias in Multi-task learning is to create a joint hypothesis space for every job by utilizing the task-relatedness building. It inhibits over-fitting in the specific job and therefore has a more immeasurable generalization capability. Unlike transfer learning, it mainly does various jobs preferably than various areas as much of the area's existing research is based on several comparable jobs of the identical application area. Multi-task learning allows those jobs are strictly associated with each other. There are several hypotheses regarding job-relatedness, which drive another modelling strategy. Many researchers hypothesize that all job data come from the same sources and correlate to the standard or global models. According to this hypothesis, they created the association among jobs employing a task-coupling parameter, including regularization.

In [15], authors proposed multi-task learning for the deep neural network. They classify multi-tasking tasks into two categories for deep learning. The first is classification, and the second is ranking; in classification, the model identifies the queried domain, whereas the ranking model finds the relevant queries.

References

[1] Dana Angluin. 1988. Queries and concept learning. *Machine learning* 2, 4 (1988), 319–342.
[2] Andreas Argyriou, Theodoros Evgeniou, and Massimiliano Pontil. 2008. Convex multi-task feature learning. *Machine learning* 73, 3 (2008), 243–272.
[3] Elena Baralis, Silvia Chiusano, and Paolo Garza. 2007. A lazy approach to associative classification. *IEEE Transactions on Knowledge and Data Engineering* 20, 2 (2007), 156–171.
[4] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R Hruschka, and Tom M Mitchell. 2010. Toward an architecture for never-ending language learning. In *Twenty-Fourth AAAI conference on artificial intelligence.*
[5] Rich Caruana. 1997. Multitask learning. *Machine learning* 28, 1 (1997), 41–75.
[6] Jianhui Chen, Lei Tang, Jun Liu, and Jieping Ye. 2009. A convex formulation for learning shared structures from multiple tasks. In *Proceedings of the 26th Annual International Conference on Machine Learning*. 137–144.
[7] Zhiyuan Chen and Bing Liu. 2018. Lifelong machine learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 12, 3 (2018), 1–207.
[8] David Cohn, Les Atlas, and Richard Ladner. 1994. Improving generalization with active learning. *Machine learning* 15, 2 (1994), 201–221.

[9] Wenyuan Dai, Ou Jin, Gui-Rong Xue, Qiang Yang, and Yong Yu. 2009. Eigentransfer: a unified framework for transfer learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*. 193–200.
[10] Ludmila I Kuncheva and Juan J Rodriguez. 2007. Classifier ensembles with a random linear oracle. *IEEE Transactions on Knowledge and Data Engineering* 19, 4 (2007), 500–508.
[11] Neil D Lawrence and John C Platt. 2004. Learning to learn with the informative vector machine. In *Proceedings of the twenty-first international conference on Machine learning*. 65.
[12] David D Lewis and William A Gale. 1994. A sequential algorithm for training text classifiers. In *SIGIR'94*. Springer, 3–12.
[13] Hui Li, Xuejun Liao, and Lawrence Carin. 2009. Multi-task Reinforcement Learning in Partially Observable Stochastic Environments. *Journal of Machine Learning Research* 10, 5 (2009).
[14] Qiuhua Liu, Xuejun Liao, Hui Li, Jason R Stack, and Lawrence Carin. 2008. Semisupervised multitask learning. *IEEE transactions on pattern analysis and machine intelligence* 31, 6 (2008), 1074–1086.
[15] Xiaodong Liu, Jianfeng Gao, Xiaodong He, Li Deng, Kevin Duh, and Ye-Yi Wang. 2015. Representation learning using multi-task deep neural networks for semantic classification and information retrieval. (2015).
[16] Lilyana Mihalkova, Tuyen Huynh, and Raymond J Mooney. 2007. Mapping and revising markov logic networks for transfer learning. In *Aaai*, Vol. 7. 608–614.
[17] Sinno Jialin Pan and Qiang Yang. 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2009), 1345–1359.
[18] Burr Settles. 2009. Active learning literature survey. (2009).
[19] Burr Settles, Mark Craven, and Lewis Friedland. 2008. Active learning with real annotation costs. In *Proceedings of the NIPS workshop on cost-sensitive learning*. Vancouver, CA:, 1–10.
[20] Thomas R Shultz and Francois Rivest. 2001. Knowledge-based cascade-correlation: Using knowledge to speed learning. *Connection Science* 13, 1 (2001), 43–72.
[21] Daniel L Silver, Qiang Yang, and Lianghao Li. 2013. Lifelong machine learning systems: Beyond learning algorithms. In *2013 AAAI spring symposium series*.
[22] Alexander Strehl and Joydeep Ghosh. 2002. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of machine learning research* 3, Dec (2002), 583–617.
[23] Kristinn R Thórisson, Jordi Bieger, Xiang Li, and Pei Wang. 2019. Cumulative learning. In *International Conference on Artificial General Intelligence*. Springer, 198–208.
[24] Sebastian Thrun. 1995. Exploration in active learning. *Handbook of Brain Science and Neural Networks* (1995), 381–384.
[25] Sebastian Thrun and Lorien Pratt. 1998. Learning to learn: Introduction and overview. In *Learning to learn*. Springer, 3–17.
[26] Xiaoxin Yin, Jiawei Han, Jiong Yang, and Philip S Yu. 2006. Efficient classification across multiple database relations: A crossmine approach. *IEEE Transactions on Knowledge and Data Engineering* 18, 6 (2006), 770–783.