Front. Comput. Sci., 2017, 11(3): 359–361
DOI 10.1007/s11704-016-6903-6

PERSPECTIVE

Lifelong machine learning: a paradigm for continuous learning

Bing LIU (✉)

Department of Computer Science, University of Illinois at Chicago, Chicago IL 60607, USA

Machine learning (ML) has been instrumental for the advances of both data analysis and artificial intelligence (AI). The recent success of deep learning brings it to a new height. ML algorithms have been successfully used in almost all areas of applications in industry, science, and engineering.

The current dominant paradigm for ML is to run an ML algorithm on a given dataset to generate a model. The model is then applied in real-life tasks. We call this paradigm *isolated learning* because it does not consider any other related information or the knowledge learned in the past. The fundamental problem with this isolated learning is that it has no *memory*. It does not retain knowledge learned in the past and use it to help future learning. As a result, a large number of training examples are needed to learn effectively. For supervised learning, training data labeling is often done manually, which is very labor-intensive and time-consuming. Since the world simply has too many possible tasks, it is almost impossible to label a large number of examples for each task in order for an ML algorithm to learn. To make matters worse, everything also changes constantly, and the labeling thus needs to be done continuously, which is a daunting task. The current isolated learning paradigm is probably not suitable for building a truly intelligent system, but only suitable for solving problems in very narrow domains.

We human beings seem to learn quite differently. We never learn in isolation. Instead, we always retain and accumulate the knowledge learned in the past and use it seamlessly in future learning. Over time we learn more and more and become more and more knowledgeable, and more and more effective at learning. *Lifelong machine learning* (LML) (or simply *lifelong learning*) aims to mimic this human learning process and capability. This type of learning is quite natural because things around us are closely related and interconnected. Concepts and their relationships learned in the past can help us understand and leam about a new subject better because a lot of things are shared across domains and tasks. For example, we humans do not need 1000 positive reviews and 1000 negative reviews of movies as an ML algorithm would need to build an accurate classifer to recognize positive and negative reviews about a movie. In fact, for this task, without a single training review, we can already perform the classification task. How can that be? The reason is simple. It is because we have already accumulated so much knowledge in the past about how people praise and criticize things, although none of those praises and criticisms may be in the form of online reviews. In fact, if there is no past knowledge, humans probably have great difficulty in manually building a good classifier with the 2000 positive and negative training reviews.

Definition of lifelong machine learning

Definition *Lifelong machine learning* (LML) is a continuous learning process where the learner has performed a sequence of N learning tasks, T_1, T_2, $\ldots$, T_N. When faced with the $(N+1)$th task T_{N+1} with its data D_{N+1}, the learner can leverage the prior knowledge in its knowledge base (KB) (the memory) to help learn T_{N+1}. KB stores and maintains the knowledge learned and accumulated in the past learning of the N tasks. After learning T_{N+1}, KB is updated with the learned (intermediate as well as the final) results from T_{N+1}.

This definition by Chen et al. [1] shows that the key characteristics of LML are 1) continuous learning, 2) knowledge accumulation in the knowledge base (KB), and 3) leveraging the knowledge in KB to help future learning. These charac-

E-mail: Liub@cs.uic.edu

teristics make it different from related learning tasks such as transfer learning [2] and multi-task learning [3].

Transfer learning (TL) uses a source domain to help a target domain learning. It assumes that the source domain S has a large amount of labeled training data, and the target domain T has few or no labeled training data but a large amount of unlabeled data. TL leverages the source labeled data to help learning in the target domain. TL is different from LML due to a few reasons. First of all, TL is not continuous. It only uses the source domain to help the target domain learning. Second, TL does not accumulate the learned knowledge. Third, TL is one-directional, using the source to help the target. LML can go in any directions. Fourth, TL assumes the source is very similar to the target. The similarity is determined by the human user. LML does not make such a strong assumption. Human users are usually not involved in determining the similarity of tasks.

Multi-task learning (MTL) aims to perform joint optimization of multiple similar learning tasks so that they can share each other's knowledge to achieve a better overall result. However, MTL still works in the traditional paradigm. Instead of optimizing a single task, it optimizes several tasks together. If we regard the several tasks as one bigger task, it reduces to the traditional optimization, which is actually the case in most optimization formulations of MTL. It does not accumulate any knowledge over time and it does not have the concept of continuous learning, which are the key characteristics of LML. Although one can argue that MTL can jointly optimize all tasks whenever a new task is added, it is quite difficult to optimize all tasks in the world simultaneously in a single process as the tasks are very different and too numerous.

History of lifelong machine learning

The concept of LML was proposed around 1995 by Thrun and Mitchell [4]. Since then it has been researched in four main directions.

• Lifelong supervised learning Thrun [5] first studied lifelong concept learning, where each past or new task is a class or concept. Several LML techniques were proposed in the contexts of memory-based learning and neutral networks. The neural network approach was improved in Ref. [6]. Fei et al. [7] extended this form of LML to cumulative learning, which, on encountering a new class, builds a new multi-class classifier that can classify all the past and the new classes. It also detects unseen classes in testing. This paves the way for self-learning because the ability to detect unseen classes allows it to learn new things. Ruvolo and Eaton [8] proposed an efficient LML algorithm (ELLA) to improve a multi-task learning method. Chen et al. [1] proposed an LML technique in the context of naive Bayesian classification. A theoretical study of LML was done by Pentina and Lampert [9].

• Lifelong unsupervised learning A lifelong topic model was first proposed by Chen and Liu [10]. Subsequently, they also reported several other models. The proposed techniques learn knowledge from topics produced from many past tasks and use the knowledge to help generate more coherent topics in the new task. Liu et al. [11] proposed an LML approach to information extraction and Shu et al. [12] presented a lifelong graph labeling method to separate two types of expressions in the context of opinion mining.

• Lifelong semi-supervised learning The work in this area is represented by the never-ending language learner (NELL) system [13]. NELL has been reading the Web continuously for information extraction since January 2010, and it has accumulated millions of entities and relations.

• Lifelong reinforcement learning Thrun and Mitchell [4] first studied lifelong reinforcement learning (LRL) for robot learning. Tanaka and Yamamura [14] proposed an LRL method that treats each environment as a task. Bou Ammar et al. [15] presented a policy gradient efficient LRL algorithm.

Summary

Although LML has been around for more than 20 years, not a great deal of research has been done so far. One reason could be that the ML research in the past 20 years focused on statistical and algorithmic approaches. LML typically needs systems approaches. However, as statistical machine learning becomes mature and researchers realize its limitations, LML will become more and more important. It is probably safe to say that without the LML capability to accumulate learned knowledge and learn new tasks with the help of the past knowledge in a self-motivated manner, we will not be able to build a truly intelligent system. We can only solve problems in very narrow domains.

Acknowledgements This work was supported in part by a grant from National Science Foundation (NSF) (IIS-1407927), a grant from NCI (R01CA192240), and a gift from Bosch.

References

1. Chen Z Y, Ma N Z, Liu B. Lifelong learning for sentiment classification. In: Proceedings of ACL Conference. 2015
2. Pan S J, Yang Q. A survey on transfer learning. IEEE Transaction on Knowledge and Data Engineering, 2010, 22(10): 1345–1359
3. Caruana R. Multitask learning. Machine Learning, 1997, 28(1)

4. Thrun S, Mitchell T M. Lifelong robot learning. In: Steels L, ed. The Biology and Technology of Intelligent Autonomous Agents. Berlin: Springer, 1995, 165–196
5. Thrun S. Is learning the *n*-th thing any easier than learning the first? Advances in Neural Information Processing Systems, 1996: 640–646
6. Silver D L, Mercer R E. The task rehearsal method of life-long learning: overcoming impoverished data. In: Proceedings of the 15th Conference of the Canadian Society for Computational Studies of Intelligence on Advances in Artificial Intelligence. 2002, 90–101
7. Fei G L, Wang S, Liu B. Learning cumulatively to become more knowledgeable. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016, 1565–1574
8. Ruvolo P, Eaton E. ELLA: an efficient lifelong learning algorithm. In: Proceedings of International Conference on Machine Learning. 2013, 507–515
9. Pentina A, Lampert C H. A PAC-Bayesian bound for lifelong learning. In: Proceedings of International Conference on Machine Learning. 2014, 991–999
10. Chen Z Y, Liu B. Topic modeling using topics from many domains, lifelong learning and big data. In: Proceedings of International Conference on Machine Learning. 2014
11. Liu Q, Liu B, Zhang Y L, Kim D S, Gao Z Q. Improving opinion aspect extraction using semantic similarity and aspect associations. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence. 2016
12. Shu L, Liu B, Xu H, Kim A. Separating entities and aspects in opinion targets using lifelong graph labeling. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, 2016
13. Mitchell T, Cohen W, Hruschka E, Talukdar P, Betteridge J, Carlson A, Dalvi B, Gardner M, Kisiel B, Krishnamurthy J, Lao N, Mazaitis K, Mohamed T, Nakashole N, Platanios E, Ritter A, Samadi M, Settles B, Wang R, Wijaya D, Gupta A, Chen X, Saparov A, Greaves M, Welling J. Never-ending learning. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence. 2015, 2302–2310
14. Tanaka F, Yamamura M. An approach to lifelong reinforcement learning through multiple environments. In: Proceedings of the 6th European Workshop on Learning Robots. 1997, 93–99
15. Bou Ammar H, Eaton E, Ruvolo P, Taylor M. Online multi-task learning for policy gradient methods. In: Proceedings of the 31st International Conference on Machine Learning. 2014, 1206–1214

Bing Liu is a professor of computer science at University of Illinois at Chicago, USA. His research interests include sentiment analysis and opinion mining, lifelong machine learning, data mining, machine learning, and natural language processing. He currently serves as the Chair of ACM SIGKDD. He is an ACM Fellow, AAAI Fellow, and IEEE Fellow.