Lifelong Machine Learning: Outlook and Direction

Xianbin Hong
Research Institute of Big Data Analytics, XJTLU
111 Ren' ai Rood
Suzhou Industrial Park
Xianbin.Hong@xjtlu.edu.cn

Prudence Wong
The University of Liverpool
Ashton Building, Ashton Street
Liverpool L69 3BX, UK
pwong@liverpool.ac.uk

Dawei Liu
Research Institute of Big Data Analytics, XJTLU
111 Ren' ai Rood
Suzhou Industrial Park
Dawei.Liu@xjtlu.edu.cn

Sheng-Uei Guan
Research Institute of Big Data Analytics, XJTLU
111 Ren' ai Rood
Suzhou Industrial Park
Steven.Guan@xjtlu.edu.cn

Ka Lok Man
Research Institute of Big Data Analytics, XJTLU
111 Ren' ai Rood
Suzhou Industrial Park
Ka.Man@xjtlu.edu.cn

Xin Huang
Research Institute of Big Data Analytics, XJTLU
111 Ren' ai Rood
Suzhou Industrial Park
Xin.Huang@xjtlu.edu.cn

ABSTRACT
The Lifelong machine learning is an advanced machine learning paradigm and also is the key to the stronger AI. In this paper, we review the development history of the lifelong machine learning and evaluate the current stage. The aim, definition and main components of it is introduced. In addition, the bottleneck and possible solution also is discussed and the further development waypoint is proposed.

CCS Concepts
• **Computing methodologies → Active learning settings**

Keywords
Lifelong Machine Learning, Classification, Big Data, Natural Language Processing

1. INTRODUCTION
Although the current machine learning had made a huge development and achieved a significant success, the intelligence level of learning algorithm still is very low. To make the learning machine more closes to human, an advanced learning paradigm was proposed. With more than 20-year development, the lifelong machine learning had become the key to the stronger AI. In this paper, we firstly review its' history and evaluate the current stage. And then an introduce to the lifelong machine learning is provided in details. Next the research direction and problems is discussed and finally some suggestions be proposed to advance the lifelong learning.

ICBDR 2018, October 27–29, 2018, Weihai, China

978-1-4503-6476-8/18/10…$15.00
DOI: http://dx.doi.org/10.1145/3291801.3291829

2. Why Lifelong Machine Learning

2.1 The ultimate goal of Machine Learning
The machine learning had been well developed over the past 65 years since it been raised. However, the mainstream still immerses in competing for isolate tasks [1]. Given a specific task, engineers and researchers only care about how to obtain the best model to such task. Also due to the boom of deep learning, it is much easier to get better and better performance on a task by feeding more and more data into more and more complex network. This kind of successes make the ML community tends to spend almost all resources on building the more complex models and collecting lager datasets, even ignores the goal of the machine learning.

Fortunately, there still are a few scientists who remember the ultimate goal of the machine learning and artificial intelligence – to make the machine can think, act like human being. They aware of the gap between the machine learning with the human, and devote themselves to reduce it.

2.2 Problems of Current Algorithms

2.2.1 Lacking of Human-Level Intelligence
Although human also often is trained for a specific task, the difference with the machine is that human can accumulate knowledge and use them for the future tasks. The current algorithms is under a setting that all tasks is not related with others [1]. Human is not only taught for a single task, but also is expected to learn the skills well to analysis and figure out different tasks. Once obtained the basic knowledge and skills, human can learn other tasks during the practices by themselves. This ability is exactly what called as "learning to learn" [2].

The reason why current learning algorithm lacks the human level intelligence is that the missing of the system for the knowledge accumulation and reuse [1]. When design the algorithms, we didn't consider about which kind knowledge could be obtained from a task and how to use them in the future. Without such kind of accumulation, an algorithm is never possible to become a "general learning algorithm". A set of different learning algorithms for different tasks without a high-level organization is never the real AI we want.

2.2.2 Wasting and Lacking of Data

As said above, the current way of mainstream deviates from our ultimate goal. In addition, the lacking of data is a practical problem need to face. Due to current learning algorithm only build model upon the single tasks[1], the collection of data turns to be essential. However, without accumulation from previous tasks, the archived data is hard to utilize. Such kind of isolated learning algorithms [1] lead to the “data island” problem. The valuable data collected only for a single task, it’s really a huge wasting.

A significant character of the isolated learning algorithm is that it may can obtained better performance if has more data. Especially in deep learning filed, data annotation becomes the core of whole processing. The volume of the dataset means the competitiveness. Vast resource is needed but is hard to be afforded by the individual or small organization. The algorithm which can learn from small dataset to achieve the art of state performance is needed.

2.3 Mission for Lifelong Machine Learning

In the basic setting, the lifelong machine learning (or named as lifelong learning) algorithm, as a learner, an artificial system will learn a series of or even infinite tasks. In this situation, how to learn efficiently and effectively which means how to avoid repeat learning on the same things and how to learn knowledge from a task as much as possible. In other words, the lifelong learning should learn knowledge not only for the current task, but also for the future tasks. Once meeting with the new task, all the knowledge already learnt from previous tasks should not been repeat learnt.

When face to a new problem, we only need to study a few examples due to we already have a basic concept for the task and only need some fine-tune and consolidation [1]. With such ability, there are two advances: the learning will be effective because the previous knowledge will reinforce our current tasks to achieve better performance; the learning will be efficient due to we can easier to accomplish task because we have a higher baseline acquired by archive tasks.

3. Lifelong Machine Learning

As the lifelong learning has not finish its’ mission, it’s too early to give a final definition. So, in this paper, the development history and related learning paradigm will be reviewed first. Then, a current definition will be presented and discussed.

3.1 History

The concept of lifelong machine learning (LML) started from the representation transfer from a task to another and we only introduce it within supervised learning. Michalski [3] invented an approach to transfer old representation to new task to achieve better performance. Solomonoff [4] introduce a great idea of incremental learning which propose to learn from simplest tasks and then go on with more difficult. This principle provided a possible way for the continuous learning. Thrun [5] used gradients of previous tasks to help the training process of the new task. It involved the early idea of the transfer learning. Lifelong machine learning also was named as constructive induction, increment and continual learning, explanation-based learning, sequential task learning, never ending learning and etc [6].

It was firstly called as lifelong machine learning since 1995 by Thrun [7, 8]. Efficient Lifelong Machine Learning (ELLA) [9] raised by Ruvolo and Eaton. Comparing with the multi-task learning [10], ELLA is much more efficient. Zhiyuan and Bing [11] improved the sentiment classification by involving knowledge. The object function was modified with two penalty terms which corresponding with previous tasks.

3.2 Related Paradigm

Although lifelong learning is an advanced learning paradigm, some current algorithms already achieved a lot. These algorithms include transfer learning (domain adaption), multi-task learning, online learning and reinforcement learning. From these paradigms, we can obtain some useful tools and inspires.

3.2.1 Transfer Learning

The paradigm most related to the lifelong learning is the transfer learning (or called domain adaption). Transfer learning attempts to transfer knowledge from source task/domain to the target task/domain. Normally, the source task/domain contain large volume data and the target task/domain could benefit from it [1]. However, transfer learning generally only considers two tasks and without explicit knowledge base. It’s not a continuous learning, and the enhance direction only from source to the target [1].

3.2.2 Multi-Task Learning

Multi-task learning (MTL) learns a group of tasks simultaneously, and optimize over all tasks. A group means the tasks are very related and generally within the same domain. In another view, multi-task learning integrate a series of isolate tasks to a large task [1]. The reason why the MTL works due to the related tasks normally sharing the same representation, so the integrate of task equals to increase the data number on a single task, and the performance of course will be better. Meanwhile, the restrict setting to the task relatedness also causes that MTL will meet serious problem when tasks are not very related. Once the tasks have different distribution and representation, the MTL always worse than the single task learning. Comparing with MTL, lifelong learning should can leverage knowledge from different domain and be more flexible.

3.2.3 Online Learning

Online learning is a flexible algorithm for the situation that data is incoming in the whole time. Hence, the online learning just adapts to the new data rather re-train the whole dataset. In this view, online learning is similar with lifelong learning. The difference is, for lifelong learning, the new tasks are incoming, rather than data within the same task.

3.2.4 Reinforcement Learning

Reinforcement Learning simulate an environment for the agent and make it is able to learn by the rewards (feedbacks). It also only suit for a single task, but is similar with the learning approach of human being. Inspire from it, LML also could collect rewards at the backend during learning and fine-tine the whole model.

3.3 Referential Definition

The earlier definition of LML by Thrun [8] means the new task should be improved by the knowledge gained from previous tasks. Zhiyuan [1] extend it with new components mainly include the explicit Knowledge Base (KB). In summary, LML is a kind of continuous learning, the key of it is to accumulate knowledge from previous tasks and maintain in the knowledge base and then to use for new task. It will handle a series of tasks and more like a kind of never-ending learning. With more knowledge, LML will have better performance.

3.4 Components of LML

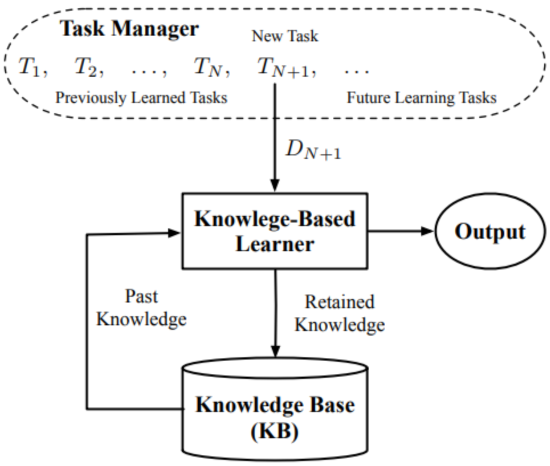

Figure 1. Knowledge System in the Lifelong Machine Learning [1]

3.4.1 Knowledge Base (KB)

The knowledge Base [1] mainly used to maintain the previous knowledge. Based on the type of knowledge, it could be divided as following:

3.4.1.1 Past Information Store (PIS)

The PIS [1] mainly focus some explicit information of previous task: raw data, results and the parameters of the model. This kind of information is normal in the machine learning and the PIS just need to simply store them.

3.4.1.2 Meta-Knowledge Miner (MKM)

Except the explicit information, meta-knowledge as the high-level information will be more helpful. Hence, a meta-knowledge miner [1] is necessary.

3.4.1.3 Meta-Knowledge Store (MKS)

Meta-knowledge likes the summary or conclusion generated from the raw data. As the useless and harmful information was eliminated During the processing of extracting it, meta-knowledge is able to improve our model. The same reason, human also wish to memory the meta-knowledge rather than the original data, and the storage cost also could be reduced.

3.4.1.4 Knowledge Reasoner (KR)

Inference based on the achieved knowledge involves the logic induction to generate more data. This requires a strict system design and the most of the LML algorithms assumes this as optional.

3.4.2 Knowledge-Base Learner (KBL)

The Knowledge-Based Learner [1] aims to searcher and transfer achieved knowledge to current task. Hence, it includes two parts: task knowledge miner and leaner. The miner searches and determines which knowledge could be used, and the learner applies such knowledge to the current task.

4. Current State of LML

In the recent years, more researchers start to pay more attention to the lifelong learning. Most researches focus on the basic knowledge related issues, like knowledge representation, retention and transfer. The knowledge could be global knowledge generated from all task like which in ELLA [9], or also local knowledge only extracted from each task like LSC[11]. In common, the knowledge still quite simple, without complicate structure. There also isn't a powerful backend to organize the learning process. We just implemented the very basic level lifelong learning which can retain and transfer knowledge among tasks. The correctness of knowledge is unable to evaluate and the knowledge representation also is hard. We need to invent a more general knowledge representation and to consider how to assess knowledge. Furthermore, knowledge reason also is valuable research problem.

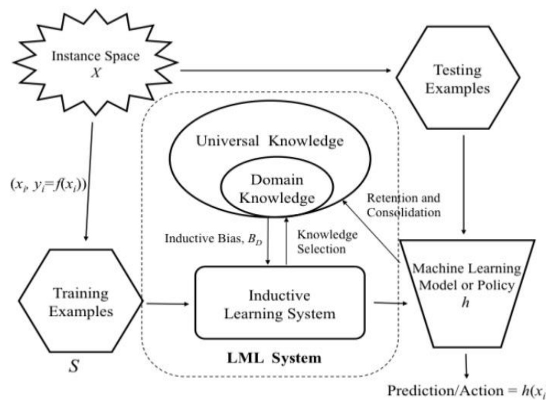

Figure 2. Framework for the Lifelong Machine Learning [6]

5. Role of Big Data in lifelong learning

With Big data, we gain a reliable source of knowledge and which is the foundation of lifelong learning. The success of the LML need the knowledge retention, more knowledge more powerful. Before the big data era, the collection and storage of data are too expensive to provide an environment for the LML. In another hand, the LML will really discover the potentiality of big data. Under the single task learning setting, one dataset only contributes for one task. In contrast, numerous tasks could be benefit from the same dataset and the efficiency of the knowledge reusing will significantly increase.

6. Research Directions for LML

In current stage, the researchers are working on the two research directions: to make the LML more efficient and effective [6].

6.1 Efficiency

Efficiency is important for the LML due to there are a series of tasks, improper approach will lead to the unacceptable processing time. Therefore, the researchers are working hard to make the algorithm efficient, especially comparing with the multi-task learning. ELLA (Efficient Algorithm for Lifelong Learning) [9] is a typical example. This algorithm divides the parameters into two parts: global knowledge and task-specific knowledge. Rather than training all tasks together like multi-task learning, ELLA only train the task specific parameters individually because the global knowledge is same and sharing with all task. With this way, the training workload is significantly reduced and the training speed up 1000x times comparing with multi-task learning.

6.2 Effectivity

To enhance effectivity is the main reason we researching the LML. With more data, the better performance is expected. However, this goal isn't easy to achieve due to knowledge may include noises and will influence the learning. As the result, the performance of LML even is worse than the single task learning on a few of tasks. To solve this problem, LSC (Lifelong Learning for Sentiment Classification) [11] balances the usage of the domain knowledge and global knowledge to achieve a better result. In the experiment, LSC is much better than the multi-task and single task learning.

6.3 NLP

As discussed above, lifelong learning needs the big data. Currently, the two largest sources of big data are texts and images. The neural networks work well in the vision filed, even perform better than the human being. However, the huge demand of data annotation and computing capability make the training and deployment cost unacceptable. Therefore, the vision filed very needs the LML, but it's also quite difficult. As is well-known, the feature extraction and representation are complicated. Why neural network is popular because that it avoids the manual feature engineering. However, this also makes the knowledge discovering difficult and hard to explain. Without the fully understanding to the model parameters and meta-knowledge, it is nearly impossible to transfer useful information to the new tasks.

In contrast, this becomes friendlier in NLP filed [1]. The language is invented by human so the semantic level information is easy to understand for human. The word meaning and grammar is same among different tasks even different domain. And some traditional approaches like Bayesian methods are working well and have good interpretability, which makes the knowledge representation and transfer is possible.

7. Missing Part for Current LML

As lifelong learning is in an initial stage, the definition of LML is incomplete, a few components still are missing.

Firstly, The LML system lacks a knowledge validation system. In previous definition, only knowledge mining and retention were mentioned, but the knowledge assessment is missing. The knowledge could be wrong and all knowledge have an application range, hence it's irresponsible that just directly stores all knowledge to the knowledge base.

Secondly, while the knowledge learning, the learning skill also should be trained. The importance of the knowledge was discussed enough, but the learning skill development still is ignored. We human being not only gain knowledge while learning but also are developing learning skills. Therefore, for the real lifelong machine learning, the learning skills also should be considered and even a skill base is necessary. Thirdly, a self-reflection and self-improving system is needed. People always use their experiences and knowledge during life, but it they get unexpected feedback, they will doubt previous thoughts and then modify their knowledge system or even the learning skill. The LML system also needs this ability to close to human.

8. The Blueprint for LML

In order to advance the lifelong learning, some action need to be taken. NLP is the first choice for the research and it will bring us a breakthrough. Knowledge validation is needed and must be put emphasis on. Skill learning and retention is a new design proposed by this paper and should be added to the design of the lifelong learning.

9. ACKNOWLEDGMENTS

This research is supported by the Research Institute of Big Data Analytics, Xi'an Jiaotong – Liverpool University and supported by the CERNET Innovation Project under Grant NGII20161010.

10. REFERENCES

[1] Z. Chen, and B. Liu, "Lifelong machine learning," Synthesis Lectures on Artificial Intelligence and Machine Learning, vol. 10, no. 3, pp. 1-145, 2016.

[2] S. Thrun, "Lifelong learning algorithms", Learning to Learn, Sebastian Thrun and Lorien Y. Pratt," Kluwer Academic Publisher.

[3] R. S. Michalski, "Learning= inferencing+ memorizing," Foundations of Knowledge Acquisition, pp. 1-41: Springer, 1993.

[4] R. J. Solomonoff, "A system for incremental learning based on algorithmic probability."

[5] S. Thrun, "Explanation-based neural network learning," Explanation-Based Neural Network Learning, pp. 19-48: Springer, 1996.

[6] D. L. Silver, Q. Yang, and L. Li, "Lifelong Machine Learning Systems: Beyond Learning Algorithms." p. 05.

[7] S. Thrun, and T. M. Mitchell, "Lifelong robot learning," The biology and technology of intelligent autonomous agents, pp. 165-196: Springer, 1995.

[8] S. Thrun, "Is learning the n-th thing any easier than learning the first?." pp. 640-646.

[9] P. Ruvolo, and E. Eaton, "ELLA: An efficient lifelong learning algorithm." pp. 507-515.

[10] R. Caruana, "Multitask learning," Machine learning, vol. 28, no. 1, pp. 41-75, 1997.

[11] Z. Chen, N. Ma, and B. Liu, "Lifelong learning for sentiment classification," arXiv preprint arXiv:1801.02808, 2018.