

8-2019

An Introduction to Advanced Machine Learning : Meta Learning Algorithms, Applications and Promises

Farid Ghareh Mohammadi

M. Hadi Amini

Hamid R. Arabnia

Follow this and additional works at: https://digitalcommons.fiu.edu/cs_fac

This work is brought to you for free and open access by the College of Engineering and Computing at FIU Digital Commons. It has been accepted for inclusion in School of Computing and Information Sciences by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

An Introduction to Advanced Machine Learning : Meta Learning Algorithms, Applications and Promises

Farid Ghareh Mohammadi¹, M. Hadi Amini², and Hamid R. Arabnia¹

1 : Department of Computer Science, Franklin College of Arts and Sciences,
University of Georgia, Athens, Georgia, 30601

2 : School of Computing and Information Sciences, College of Engineering and Computing,
Florida International University, Miami, FL 33199

Emails : farid.ghm@uga.edu, amini@cs.fiu.edu, hra@cs.uga.edu

Abstract

In [1, 2], we have explored the theoretical aspects of feature extraction optimization processes for solving large-scale problems and overcoming machine learning limitations. Majority of optimization algorithms that have been introduced in [1, 2] guarantee the optimal performance of supervised learning, given offline and discrete data, to deal with curse of dimensionality (CoD) problem. These algorithms, however, are not tailored for solving emerging learning problems. One of the important issues caused by online data is lack of sufficient samples per class. Further, traditional machine learning algorithms cannot achieve accurate training based on limited distributed data, as data has proliferated and dispersed significantly. Machine learning employs a strict model or embedded engine to train and predict which still fails to learn unseen classes and sufficiently use online data. In this chapter, we introduce these challenges elaborately. We further investigate Meta-Learning (MTL) algorithm, and their application and promises to solve the emerging problems by answering *how autonomous agents can learn to learn?*.

Keywords : Meta learning, machine learning, online learning, online optimization, model-based learning, metric-based learning, gradient descent, low shot learning, few shot learning, one shot learning

1 Introduction

Machine learning algorithms enable researchers to learn from supervised / unsupervised data. Collected data is mainly offline and it is not evolving over time. Hence, total behavior of future data are vague enough for us to process. Conventionally, it is not possible to have the entire behaviour learned [3] using the traditional machine learning, evolutionary algorithms and optimization algorithms discussed earlier [1, 2] .

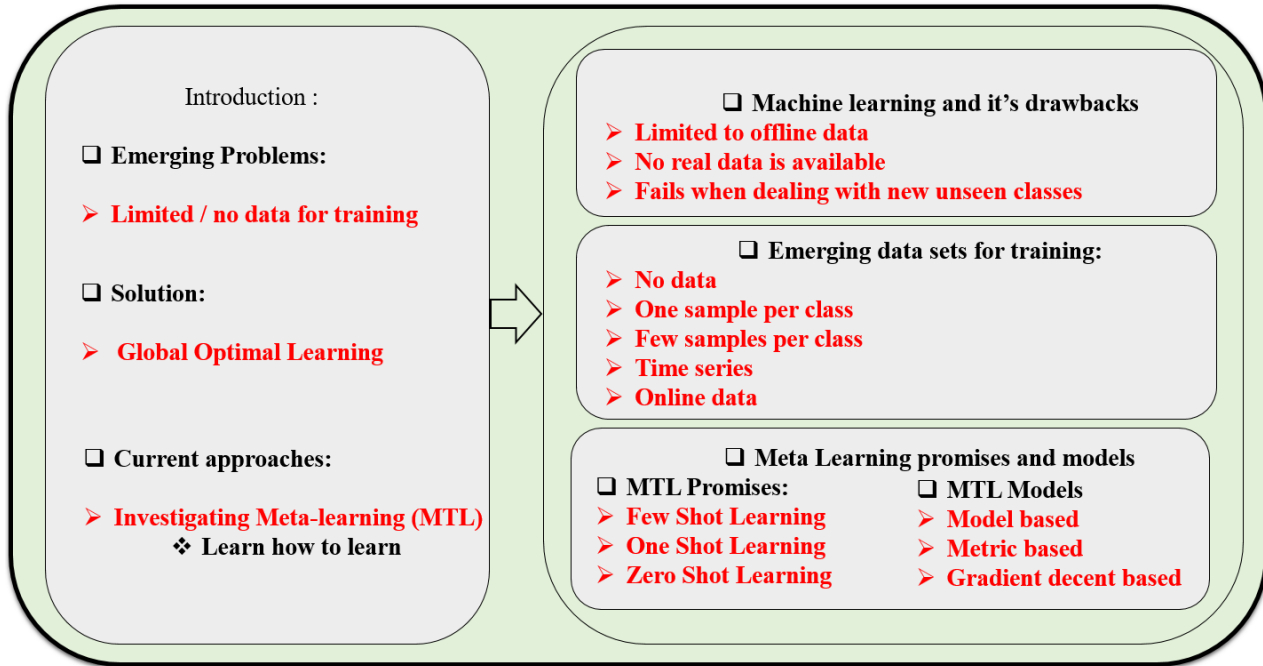


Figure 1: Overall structure of this study

Last decade, researchers have studied advanced research paradigms to solve learning process. They aim to learn using prior tasks or experiences and leverage them for future learning. One of the promising paradigm is Meta-learning (MTL). Prior studies investigated MTL methods that learn to update a function or learning rule [3, 4]. MTL differs from classic machine learning with respect to the level of adaptation [5]. MTL is the process of learning to learn. It leverages past experiences to ascertain a prior model's parameters and learning process i. e. algorithm. MTL investigates how to choose the right bias non-fixed, unlike base-learning where the bias is fixed a priori [5]. Concretely, MTL studies a setting where a set of tasks (\mathcal{T}_i) are made available together upfront. However, it cannot handle sequential and dynamic aspects of problems properly.

In contrast, online learning is the process of learning sequentially, however, it does not leverage past experiences like MTL, i.e., it may not consider how past experience can help to enhance the adaptation to a new task. The Earliest research studies introduced sequential learning [6, 7] where tasks are revealed one after another repeatedly. The aim of learning is to learn as independent as possible to attain zero-shot learning with non task-specific adaptation. We argue that neither setting is ideal for studying continual lifelong learning. MTL deals with learning to learn, but neglects the sequential and non-stationary aspects of the problem. Online learning offers an appealing theoretical framework, but does not generally consider how past experience can accelerate adaptation to a new task. In this work, we motivate and present the online MTL problem setting, where the agent simultaneously uses past experiences in a sequential setting to learn good priors, and also adapt quickly to the current task at hand.

The rest of this chapter is organized as follows. In section 2 emerging challenges in machine learning are

discussed . After that, applications of MTL using transfer learning are covered. finally we have promises of MTL. Figure 1 represents the overall structure of this study.

2 Machine learning : challenges and drawbacks

Prior works on learning process, regression and optimization problems, have attempted to learn the behavior of input data, analyze and categorize it to attain a high performance algorithms. Machine learning (ML) has been strongly applied to solve supervised and unsupervised problems. ML deploys different algorithms, such as online learning, multi-task learning and supervised algorithms, including rule based [8, 9], function based [10, 11], lazy [12], and bootstrap [13]. Some of them are used to transform data, special example would be dimension reduction for optimization, some to build classifiers like supervised algorithms, others for prediction like regression, etc. Machine learning still yields subtle drawbacks for time-varying input data which restrict it to consider properly future and unseen classes to provide general idea and knowledge from data.

Traditionally, machine learning is a machine learns only input data and predict new data which follow the rule of the equation $\mathcal{P}_i \times \mathcal{D} \rightarrow \mathcal{M}$, where \mathcal{P}_i stands for the specific supervised algorithm parameters, \mathcal{D} represents the space of training data distribution and \mathcal{M} defines the space of generated models which will be applied on test data to evaluate the supervised algorithm performance.

Figure 2 presents few machine learning approaches and algorithms which provide different applications with respect to the wide variety of data such as offline data vs. online data, labeled data vs. unlabeled data, multi-model data vs. single model data, and multi-domain data vs. single domain data. As it shows, machine learning has critical drawbacks which cannot handle whole data once. Moreover, it just considers each data as a new model and each model is separate from previous ones.

Furthermore, figure 2 depicts the relationship between traditional machine learning and advance machine learning. In traditional machine learning we have to deal with offline and limited amount of data and the number ground-truth. However, in the world of technology, where data growth have proliferated significantly and are coming from wherever technology exists, it is very critical to get to know the pattern and rules that govern whole data and learn the trend of the generated data for a specific domain. For that end, we need to classify data into three big categories, time series data, offline data and online data. These all categories are shown in three different aspects : supervised and unsupervised; multi-model data and multi-domain data.

Machine learning involves transfer learning and online learning, which is compatible to learn tasks and classes, which are consequential. Transfer learning is the theory of transferring knowledge from one task to another and learning from non-randomness. Meta learner also is one of the bootstrap algorithms which learn data by sampling given data set and generating different data sets. Meta learner deploys different supervised algorithms and then select a meta learner to give a vote to make a decision about the class of current instance.

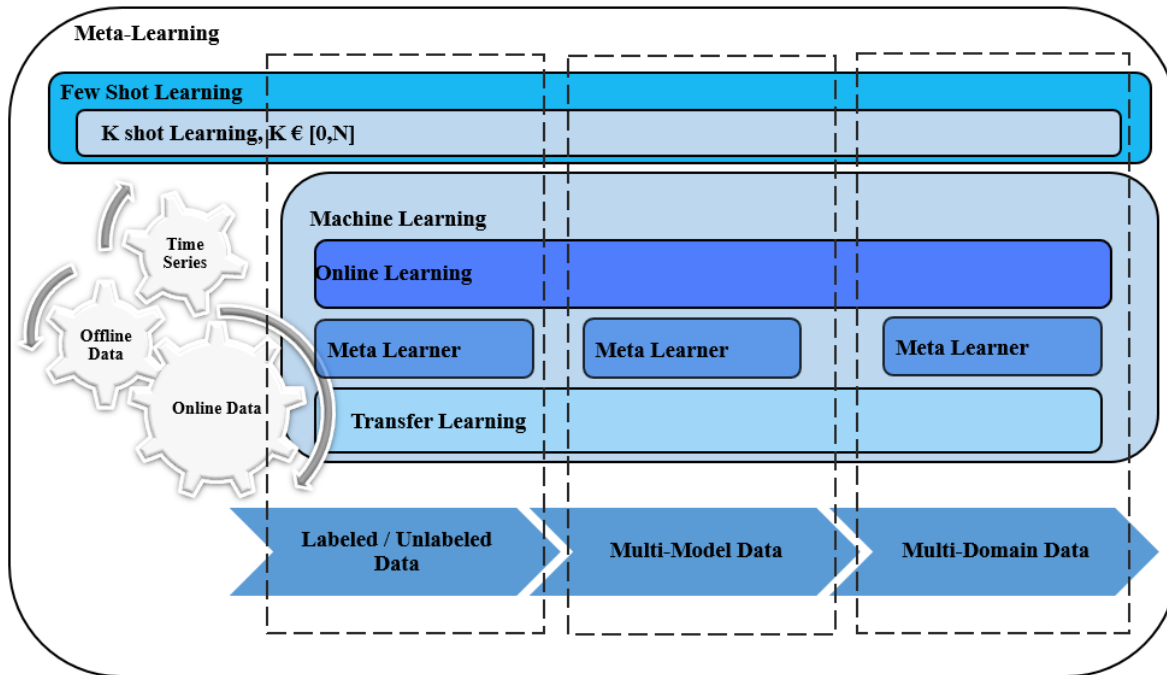


Figure 2: The general overview of Learning against emerging data

3 Meta-learning algorithms

Meta-learning (MTL) is firstly presented in [3] and [14]. After a decade gap, lately research studies have tried to deploy MTL again. MTL is a machine which learns the variety of input data. ML methods need to learn new tasks faster by leveraging previous experiences. MTL does not consider past experiences separately. MTL is the process of learning how to learn. MTL is an emerging learning algorithms with new challenges and research questions. It is an extension of transfer learning, which is one of the multi-task learning algorithms. MTL has covers three different aspects as illustrated in Figure 5. Few shot learning (FSL), one shot learning (OSL) and zero shot learning (ZSL). FSL and OSL yield highly accurate results as compared with traditional machine learning algorithms. However, they still have a critical challenge which limits them from converging to optimal results. Limits of ZSL have been addressed using domain semantic space, where includes all information system as presented in Figure 5.

3.1 Model-based MTL

Model-based MTL depends on a model and no conditional probabilistic method which enable it to be the best match for fast learning model where it updates it's hyper-parameters so fast by training just few examples. The process of updating their hyper-parameters is done either internal architecture or external meta-learner. The concept of model-based MTL is having one neural network interact with sequential neural networks to

accelerate the learning process. In other words, it tries to learn a model per each label using pixel by pixel value, according to the figure 3. In other words, this model's algorithms try to train a recurrent model like the work presented [4], which proposed long short term memory (LSTM). Hochreiter and Schmidhuber [15] proposed for the first time in 1997 the theory of LSTM. Model-based algorithms take the data set sequentially and analyze instances one by one. Since Model-based algorithms leverage RNN for learning, they become the most least efficient model in comparison with other models. Nagabandi *et al* [16] proposed online deep learning using MTL towards Continual adaptation for model-based reinforcement learning. Santoro *et al* [17] proposed memory-augmented neural network using MTL.

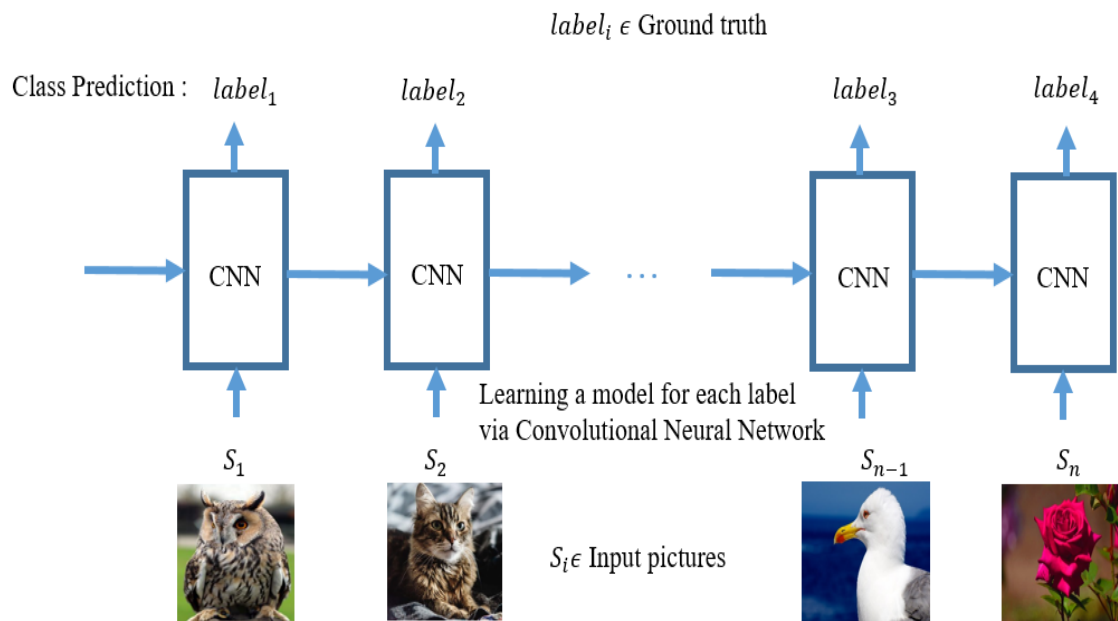


Figure 3: Model-based MTL

3.2 Metric-based learning

Metric-based learning leverages metric space learning, which leads to efficient data processing and is suitable for few-shot learning. Let's consider that our goal is image classification. As model-based learning tries to learn each image pixel by pixel which takes long time and time consuming, metric-based learning overcomes this limitation by leveraging comparing given two images to the network. The output per each input yields a vector, comparing these two vector states that whether they are similar or not. Figure 4. One of the most common application of metric-based learning is Siamese network presented in [18]. Koch *et al* presented Siamese neural network (SNN) for one-shot learning which achieved strong and better results. The idea behind SNN is that it tries to use twin or half-twin network to compare the input images. Note that one of the input is already computed and we only need to take the second image and try to go through the layers and compute the vector.

Then, SNN tries to compute the distance between them, if the result is small they similar otherwise they are different. Another application of metric-based learning is [19], where Vinyals *et al* proposed a matching network(MN) for one-shot learning.

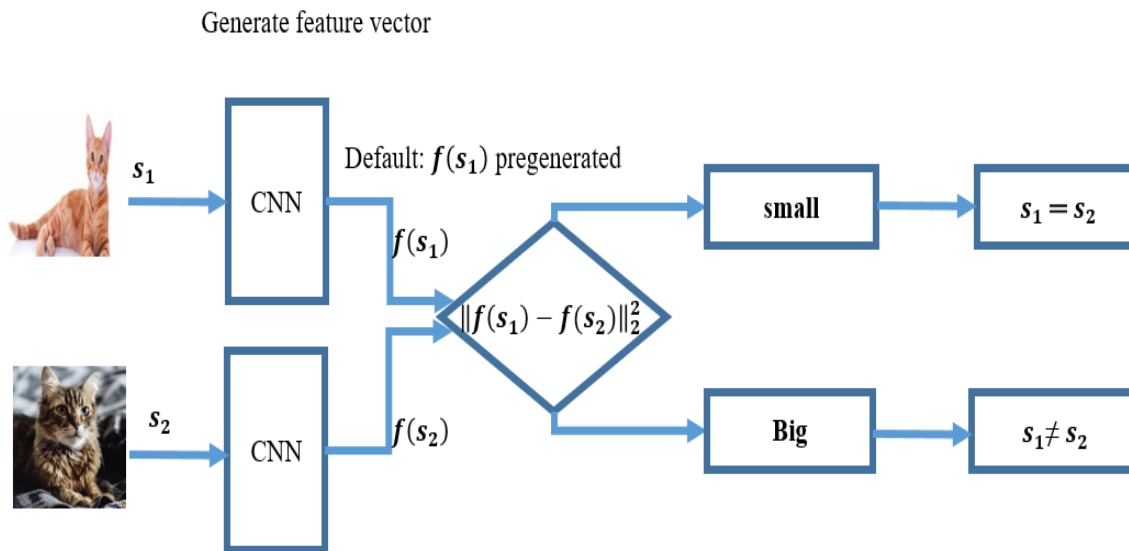


Figure 4: Metric based MTL

3.3 Gradient decent-based learning

This model of MTL is also known as optimization-based model for tuning the parameter (θ). The idea here is to leverage stochastic gradient-decent (SGD) and for new given sample, it updates the parameters to be a universal learner. It may not converge to a local optimal since it does not rely on a small number of samples.

Although gradient-based learning models work well, they still have some drawbacks. Ravi and Larochelle [4] addressed these problems carefully and provided LSTM-based MTL to overcome those problems. Finn [20] presented MAML to improve the accuracy of LSTM-based MTL.

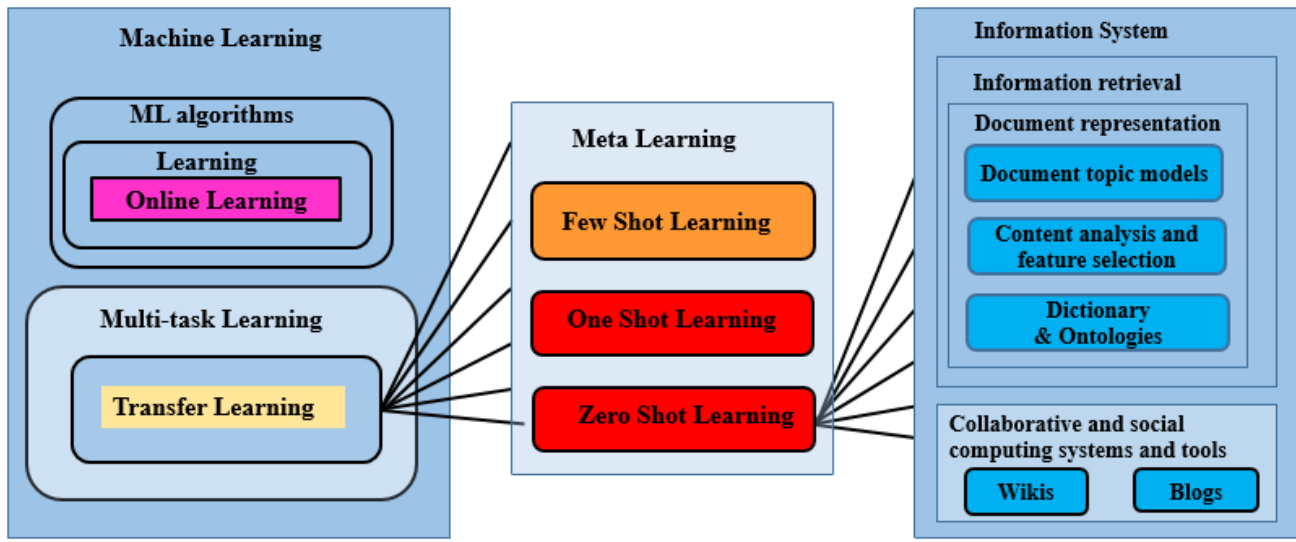


Figure 5: The relation among Machine Learning, Meta Learning, and Information system

| Abb | Definition |
|----------|--|
| CNN | Convolutional neural network |
| EGNN | Edge-labeling graph neural network |
| fSL | Few-shot learning |
| GSP | Goal-conditioned skill policy |
| LSL | Low-shot learning |
| MAML | Model-agnostic MTL |
| MANN | Memory-augmented neural network |
| MIL | Meta Imitation Learning |
| MTL | Meta-learning |
| ML | Machine learning |
| MN | Matching Network |
| OSL | One-shot learning |
| PN | Prototypical Network |
| RN | Relation network |
| RNN | Recurrent neural network |
| SAE | Semantic AutoEncoder |
| SNN | Siamese neural network |
| TL | Transfer learning |
| ZSL | Zero-shot learning |
| ZSL-FGVD | ZSL Fine-Grained Visual Descriptions |
| ZSL-H | Zero-shot learning by mitigating the Hubness problem |
| ZSL-KT | Zero-shot learning and knowledge transfer |

Table 1: Abbreviation of words

| Paper | Meta-learning | Proposed method | Meta-learning Models | Conference / journal | domain | year |
|-------|--------------------|--------------------|-----------------------|----------------------|------------------------------------|------|
| [20] | Few-shot learning | MAML | Gradient decent based | ICML | Image classification | 2017 |
| [21] | Few-shot learning | MAML++ | Gradient decent based | ICRL | Image classification | 2019 |
| [22] | Few-shot learning | probabilistic MAML | Gradient decent based | NIPS | Image classification | 2018 |
| [23] | Few-shot learning | PN | Metric based | NIPS | Image classification | 2017 |
| [24] | Few-shot learning | HML | Model based | arXiv | Image classification | 2019 |
| [25] | Few-shot learning | RN | Metric based | CVPR | Image classification | 2018 |
| [4] | Few-shot learning | LSTM-Metalearner | Gradient decent based | ICLR | Image classification | 2017 |
| [26] | Few-shot learning | EGNN | Model based | CVPR | Image classification | 2019 |
| [27] | Few-shot learning | LSL | Metric based | CVPR | Image classification | 2019 |
| [24] | One-shot learning | HML | Model based | arXiv | Image classification | 2019 |
| [17] | One-shot learning | MANN | Model based | ICML | Image classification | 2016 |
| [19] | One-shot learning | MN | Metric based | NIPS | Image classification | 2016 |
| [20] | One-shot learning | MAML | Gradient decent based | ICML | Image classification | 2017 |
| [28] | One-shot learning | MIL | Gradient decent based | CoRL | Visual imitation | 2017 |
| [18] | One-shot learning | MIL | Metric based | ICML | Image recognition | 2015 |
| [29] | Zero-shot learning | ZSL-FGVD | Metric based | CVPR | Image classification and retrieval | 2016 |
| [30] | Zero-shot learning | ZSL-H | Metric based | ICLR-workshop | Hubness problem | 2015 |
| [25] | Zero-shot learning | RN | Metric based | CVPR | Image classification | 2018 |
| [31] | Zero-shot learning | SAE | Metric based | CVPR | Image classification | 2017 |
| [25] | Zero-shot learning | ZSL-RN | Metric based | CVPR | Benchmark classification | 2018 |
| [32] | Zero-shot learning | ZSL-KT | Metric based | arXiv | Music classification | 2019 |
| [33] | Zero-shot learning | GSP | Metric based | CVPR-workshop | Visual imitation | 2018 |
| [34] | Zero-shot learning | GSP | Metric based | ICLR | Visual imitation | 2018 |

Table 2: An overview of previous studies on Meta-learning.

4 Promises of meta-learning

Learning to learn is an advance process which provides three promises : one few-shot learning(FSL), one one-shot learning (OSL), one zero-shot learning(ZSL). Figure 6 presents a general view of each promises. we have three layers : input data, meta-training, and meta testing. Input data for FZL and OSL are the same type, particularly images for particular image classification aims. Further, ZSL becomes an independent learning MTL algorithm which evaluates input data based on domain semantic space and visual information of that domain.

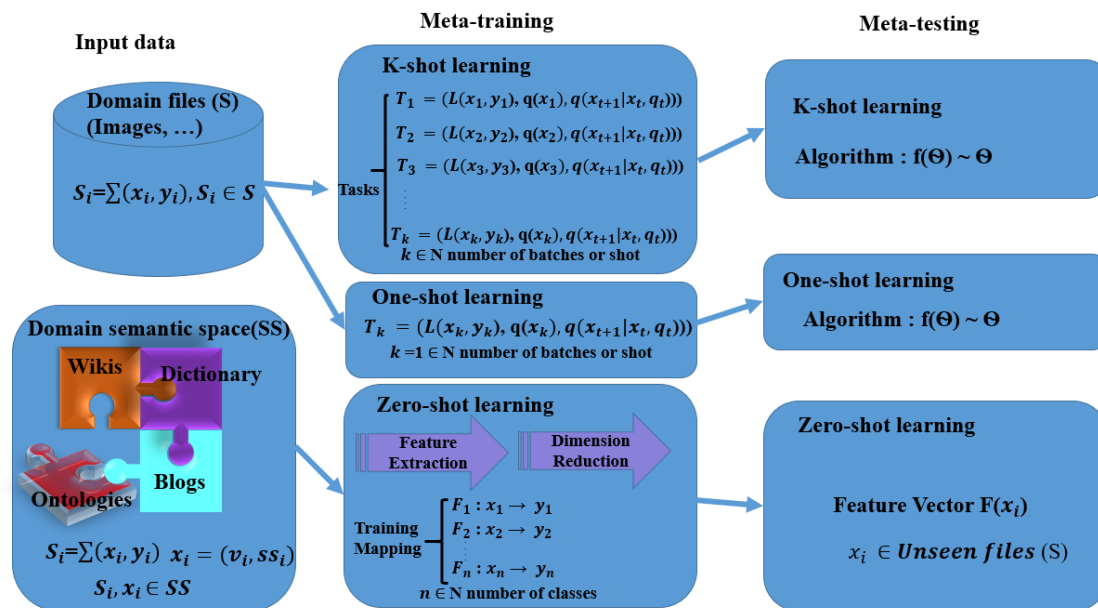


Figure 6: Structure of meta-learning models

In second layer, FSL tries to learn k-shot tasks, which means MTL is training by leveraging k different training data set. K-shots had generated in advance before learning process have started. Thus, MTL is known as a certain type of bootstrap algorithms, however, in k-shot data set we only have specific number of K instances per class. However, the bootstrap algorithm tries to split given data set with different rate and would be keeping the ratio of number of classes. MTL attempts to calculate the loss of each shot using loss function. Furthermore, OSL also ties to learn the task based on k=1 shot learning which means that OSL only has one shot at a moment. In other words, when it start bootstrapping, it only select one sample per class as a training set. Note that OSL represents a special kind of K-shot or few shot learning. Both FSL and OSL use the equitation below from [20]. $\mathcal{T}_i = \sum_{i=1}^k (\mathcal{L}(x_i, y_i), q(x_i), q(x_{t+1}|x_t, q_t))$

Further, in second lay, ZSL unlike FSL and OSL algorithms ties to work with domain semantic space rather than domain files like images. The goal here is to find an optimal mapping from semantic space to vector space. ZSL tries to map given extracted features to a new space called vector space.

Finally, last layer stands for the meta-testing which is responsible to predict the given test data and analyze them. First two algorithms try to predict unseen data using $f(\theta)$, however, zsl attempts to solve the problem by mapping the unseen data to the new vector space.

4.1 Few-shot learning

The first and one of the most common promises of MTL is few-shot learning (FSL). Few-shot classification is a specific extension of MTL in supervised learning. Lake *et al* [35] challenged traditional machine learning algorithms by enabling them to learn every concept from one or few shot of that data set. The idea behind that, MTL tries to re-sample the given input data set for training using only K samples per each class. In other words, meta-training process is accomplished by learning k shot meta sets which are selected by replacement. Although few-shot learning outperforms traditional machine learning algorithms, it has an explanatory challenge, called task ambiguity. This problem happens when a small task, which is generated from large input data set, to learn via few-shot learning. After taking a new task, which appears too ambiguous to ascertain a single model for that task that covers quite a large number of samples.

The majority of MTL algorithms leverage few-shot learning. FSL has decent important extension : one Finn *et al* proposed model-agnostic MTL (MAML) [20], which adapts to new tasks via gradient descent-based MTL. In [22], Finn *et al* re-sampled models for a new task using a model distribution. This paper extends MAML to conduct a parameter distribution that is trained through different lower bound. In [22] Finn et al addressed the ambiguity problem by proposing probabilistic MAML.

Second important extension is Online learning which is learning process of training data sequentially and continuously. The next one is online MTL. Finn *et al* [36] proposed online MTL based on the regret-based meta-learner. Kim *et al* [26] proposed EGNN, which applied a deep neural network on a certain model, edge-labeling graph. Furthermore, Sun *et al* [37] proposed an advanced Meta-transfer learning for few-shot learning. Zou and Feng [24] introduced new type of MTL which works based on hierarchy structure, called Hierarchical MTL(HML). HML overcomes previous MTL limitation which are limited to the tasks where training sets and identical output structure. HML enables MTL to optimize adaptability of meta-model to tasks, that are similar. Figure7 provides a general view of few shot learning, one-shot learning and 2-shot learning and generalized k-shot learning.

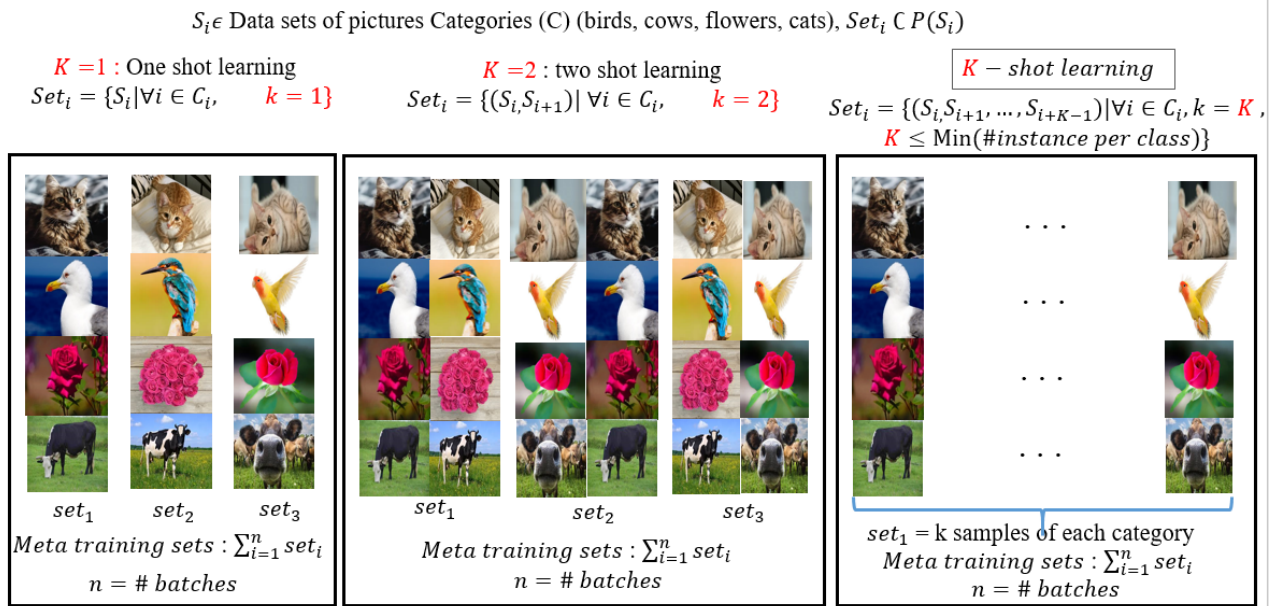


Figure 7: Few-shot learning structure

4.2 One shot learning

One-shot learning (OSL) is a critical challenge in the applications of deep neural networks. OSL is special type of few-shot learning or k-shot learning in which it choose k=1 shot for training section. In other words, when the algorithm starts training, they only leverage from one instance per class at a time with different batches. The research studies have been done for one-shot learning are listed as following : Matching networks [19] which is a metric based MTL.

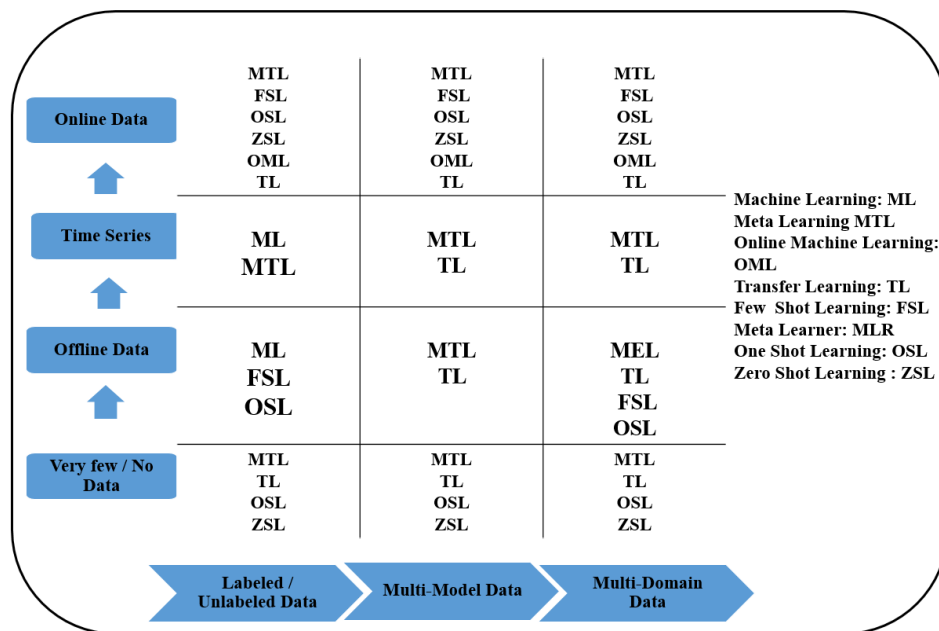


Figure 8: Machine Learning : ML, Meta-Learning: MTL, Online Machine Learning: OML, Transfer Learning: TL, Few Shot Learning: FSL, Meta Learner: MLR, One Shot Learning: OSL

4.3 zero shot learning

Zero-shot learning (ZSL) is an emerging paradigm of machine learning which is recently proposed [29, 25, 25, 32] to yield a better result than supervised learning algorithms by covering their critical limitations, which work only with a fixed number of classes. Zero shot learning is as a joint embedding problem of domain specific and side information, which includes ontology, wikis, dictionary and blogs. To be certain, ZSL overcome few-shot learning and one shot learning limitation and promised to yield a result better than FSL and OSL. The goal is to classify unseen samples of different classes without having a training data set. This is possible once you have proper information about the domain and classes, properties and most importantly the functionality of the problem. The ZSL process is a journey from feature space to a vector space in which it leverages feature extraction and dimension reduction algorithms technically. The feature vector describes shared features among classes. Reed *et al* [29] applied neural language model to overcome supervised learning limitation. ZSL has been accomplished for visual recognition [29], music classification [32], and image classification [25]. More recent methods have been proposed by Kodirov *et al.* [31] using auto-encoders for ZSL, Nagabandi *et al.* [16] to deploy MTL for online Learning and by Finn *et al* [36] for online MTL.

5 Discussion

Choosing the appropriate type of data for machine learning algorithms is an important yet challenging task. According to [38], it is crucial to select an optimal algorithm to solve each specific problem with to ensure optimal decision making. They combined experimental result and interviewed with domain experts. It is essential to know where we are, what are the challenges and what kind of data we have now. Further, what is the relationship among emerging data with respect to traditional and modern machine learning algorithms. Figure 8 presents the information to choose which algorithm are suitable, compatible and applicable given the specific type of data.

In figure 9, we have identified some of the publications in top venues. According to our investigation, few-shot learning is one of the most promising areas.

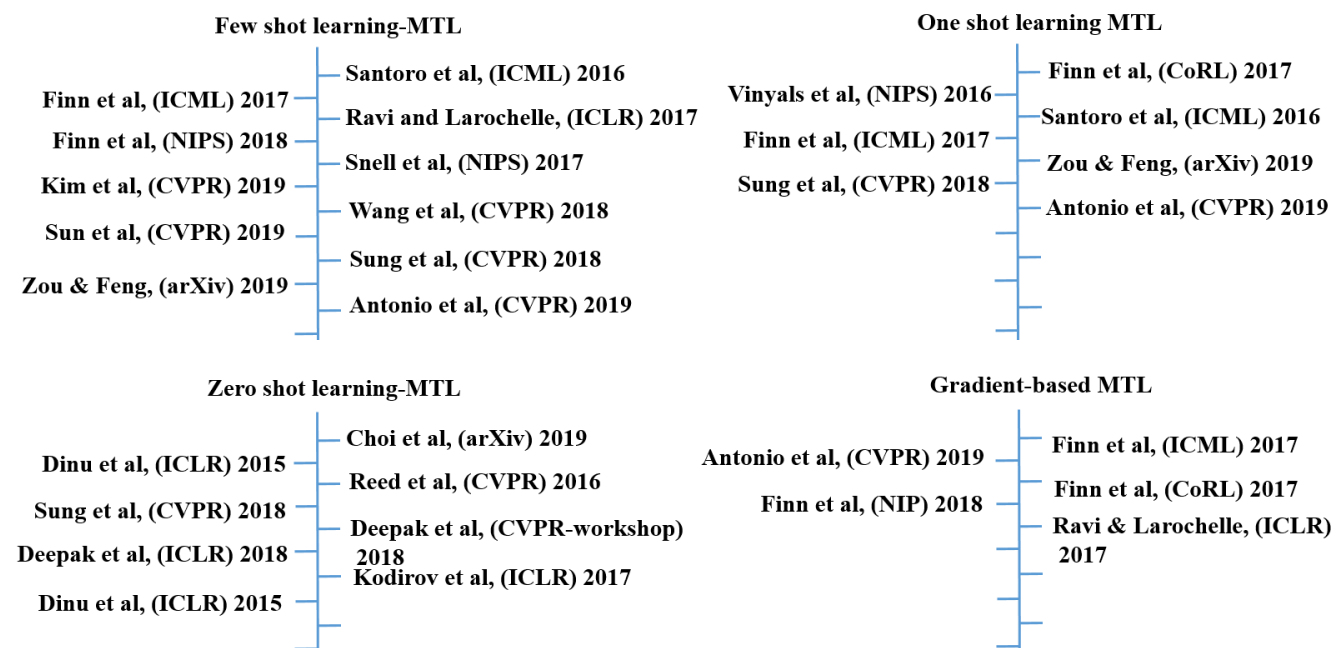


Figure 9: A brief studies over promises of Meta-learning

6 Conclusion

Optimizing algorithms to work with offline data is almost ubiquitous in each domain, such as engineering applications. The majority of studies have determined an optimal way to deal with large-scale problems. Advancing technologies have people provided data available wherever they have access to internet. Thus, it is critical to process continues data which is online and introduce an advance learning algorithm to help scientists to predict future properly. In this chapter, we addressed this problems and investigated an advanced machine

learning algorithm to solve them optimally using MTL. MTL has three important categories which covers whole research studies have accomplished in this area. One model based, one metric based, one gradient decent based, which also known as optimization method. Further, MTL has three critical extension for emerging data and large-scale problems. The first one, few-shot learning which is practically worked on k-shots of training classes. The second extension is special type of few-shot learning which here we have only one-shot for each training classes. The last one but not the least one is zero-shot learning. Although decent work have been done using FSL and OSL, but ZSL is the promising extension of meta-learning where researchers have no idea about the new classes and no enough data available.

References

- [1] F Ghareh Mohammadi, M Hadi Amini, and Hamid R. Arabnia. Evolutionary computation, optimization and learning algorithms for data science. *arXiv preprint arXiv: 1908.08006*, 2019.
- [2] F Ghareh Mohammadi, M Hadi Amini, and Hamid R. Arabnia. Applications of nature-inspired algorithms for dimension Reduction: Enabling efficient data analytics. *arXiv preprint arXiv: 1908.1908.08563*, 2019.
- [3] Jurgen Schmidhuber. Evolutionary principles in self-referential learning. *Diploma thesis*, 1987.
- [4] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *International Conference on Learning Representations (ICLR)*, pages 281–288, 2017.
- [5] Ricardo Vilalta and Youssef Drissi. A perspective view and survey of meta-learning. *Artificial intelligence review*, 18(2):77–95, 2002.
- [6] James Hannan. Approximation to bayes risk in repeated play. *Contributions to the Theory of Games*, 3:97–139, 1957.
- [7] Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.
- [8] Sholom M Weiss and Nitin Indurkha. Rule-based machine learning methods for functional prediction. *Journal of Artificial Intelligence Research*, 3:383–403, 1995.
- [9] Juan M Banda, Martin Seneviratne, Tina Hernandez-Boussard, and Nigam H Shah. Advances in electronic phenotyping: from rule-based definitions to machine learning models. *Annual review of biomedical data science*, 1:53–68, 2018.
- [10] Mingzhe Chen, Ursula Challita, Walid Saad, Changchuan Yin, and Mérouane Debbah. Artificial neural networks-based machine learning for wireless networks: A tutorial. *IEEE Communications Surveys & Tutorials*, 2019.
- [11] Arya Iranmehr, Hamed Masnadi-Shirazi, and Nuno Vasconcelos. Cost-sensitive support vector machines. *Neurocomputing*, 343:50–64, 2019.

- [12] Rashmi Agrawal. Integrated parallel k-nearest neighbor algorithm. In *Smart Intelligent Computing and Applications*, pages 479–486. Springer, 2019.
- [13] David Poland, Slava Rychkov, and Alessandro Vichi. The conformal bootstrap: Theory, numerical techniques, and applications. *Reviews of Modern Physics*, 91(1):015002, 2019.
- [14] Yoshua Bengio, Samy Bengio, and Jocelyn Cloutier. *Learning a synaptic learning rule*. Université de Montréal, Département d’informatique et de recherche . . . , 1990.
- [15] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [16] Anusha Nagabandi, Chelsea Finn, and Sergey Levine. Deep online learning via meta-learning: Continual adaptation for model-based rl. *arXiv preprint arXiv:1812.07671*, 2018.
- [17] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. Meta-learning with memory-augmented neural networks. In *International conference on machine learning*, pages 1842–1850, 2016.
- [18] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2, 2015.
- [19] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In *Advances in neural information processing systems*, pages 3630–3638, 2016.
- [20] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1126–1135. JMLR. org, 2017.
- [21] Antreas Antoniou, Harrison Edwards, and Amos Storkey. How to train your maml. *arXiv preprint arXiv:1810.09502*, 2018.
- [22] Chelsea Finn, Kelvin Xu, and Sergey Levine. Probabilistic model-agnostic meta-learning. In *Advances in Neural Information Processing Systems*, pages 9516–9527, 2018.
- [23] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, pages 4077–4087, 2017.
- [24] Yingtian Zou and Jiashi Feng. Hierarchical meta learning. *arXiv preprint arXiv:1904.09081*, 2019.
- [25] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1199–1208, 2018.

- [26] Jongmin Kim, Taesup Kim, Sungwoong Kim, and Chang D Yoo. Edge-labeling graph neural network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11–20, 2019.
- [27] Yu-Xiong Wang, Ross Girshick, Martial Hebert, and Bharath Hariharan. Low-shot learning from imaginary data. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [28] Chelsea Finn, Tianhe Yu, Tianhao Zhang, Pieter Abbeel, and Sergey Levine. One-shot visual imitation learning via meta-learning. In *1st Conference on Robot Learning (CoRL)*, 2017.
- [29] Scott Reed, Zeynep Akata, Honglak Lee, and Bernt Schiele. Learning deep representations of fine-grained visual descriptions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 49–58, 2016.
- [30] Angeliki Lazaridou Georgiana Dinu and Marco Baroni. Improving zero-shot learning by mitigating the hubness problem. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*. workshop track, 2015.
- [31] Elyor Kodirov, Tao Xiang, and Shaogang Gong. Semantic autoencoder for zero-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3174–3183, 2017.
- [32] Jeong Choi, Jongpil Lee, Jiyoung Park, and Juhan Nam. Zero-shot learning and knowledge transfer in music classification and tagging. *arXiv preprint arXiv:1906.08615*, 2019.
- [33] Deepak Pathak, Parsa Mahmoudieh, Guanghao Luo, Pulkit Agrawal, Dian Chen, Yide Shentu, Evan Shelhamer, Jitendra Malik, Alexei A Efros, and Trevor Darrell. Zero-shot visual imitation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 2050–2053, 2018.
- [34] Deepak Pathak, Parsa Mahmoudieh, Guanghao Luo, Pulkit Agrawal, Dian Chen, Yide Shentu, Evan Shelhamer, Jitendra Malik, Alexei A. Efros, and Trevor Darrell. Zero-shot visual imitation. In *International Conference on Learning Representations (ICLR)*, 2018.
- [35] Brenden M Lake, Ruslan Salakhutdinov, and Joshua B Tenenbaum. Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338, 2015.
- [36] Chelsea Finn, Aravind Rajeswaran, Sham Kakade, and Sergey Levine. Online meta-learning. *arXiv preprint arXiv:1902.08438*, 2019.
- [37] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 403–412, 2019.
- [38] Qianwen Wang, Yao Ming, Zhihua Jin, Qiaomu Shen, Dongyu Liu, Micah J Smith, Kalyan Veeramachaneni, and Huamin Qu. Atmseer: Increasing transparency and controllability in automated machine learning. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 681. ACM, 2019.