Learning How to Learn: Meta Learning Approach to Improve Deep Learning

Dr. Ashish Kr. Chakraverti[1]
Associate Professor, CSE, MIET, Gr. Noida UP

Sugandha Chakraverti[2]
Assistant Professor, CSE, RKGIT, Ghaziabad

Dr. Yashpal Singh[3]
Professor, MIET Gr. Noida UP

***Abstract*—Meta-Learning describes the abstraction to designing more elevated level components associated with preparing Deep Neural Networks. The expression "Meta-Learning" is tossed around in Deep Learning writing often referencing "AutoML", "Few-Shot Learning", or "Neural Architecture Search" when in reference to the robotized design of neural system architectures. Rising up out of entertainingly titled papers such as "Figuring out how to learn by inclination descent by slope descent", the success of OpenAI's rubik's solid shape mechanical hand demonstrates the development of the thought. Meta-Learning is the most promising worldview to propel the state-of-the-craft of Deep Learning and Artificial Intelligence.**

Meta-learning is one of the most dynamic regions of research in the profound learning space. A few ways of thinking inside the Artificial Intelligence(AI) people group buy in to the postulation that meta-learning is one of the venturing stones towards opening Artificial General Intelligence(AGI). As of late, we have seen a blast in innovative work of meta-learning systems. In any case, a portion of the essential thoughts behind meta-learning are still generally misconstrued by information researchers and designers. From that point of view, we figured it may be a smart thought to audit a portion of the crucial ideas and history of meta-learning just as a portion of the mainstream calculations in the space.

Keywords— Deep Learning; Meta learning;Artificial General Intelligence

I. INTRODUCTION

The term metalearning first happened in the territory of instructive brain research. One of the most refered to scientists right now, Biggs, portrayed met-alearning as monitoring and assuming responsibility for one's own learning [6]. Thus, metalearning is seen as a comprehension and adjustment of learning itself on a more elevated level than just securing subject information. In that manner, an individual mindful and equipped for metalearning can survey their learning approach and change it as per the prerequisites of a particular undertaking.

Metalearning as utilized in an AI setting has numerous likenesses to this portrayal. Subject information converts into base-realizing, where experience is gathered for one explicit learning task. Metalearning begins at a more significant level and is worried about collecting experience more than a few utilizations of a learning framework as per [9].

Over the most recent 20 years, AI look into was confronted with an expanding number of accessible calculations including a huge number of parametrisation, pre-preparing and postprocessing approaches just as a significantly stretched out scope of uses because of expanding registering power and more extensive availabil-ity of PC discernible informational collections. By advancing a superior comprehension of AI itself, metalearning can give a priceless assistance maintaining a strategic distance from broad experimentation techniques for calculation choice, and beast power looks for appropriate parametrisation. Seeing how to benefit from past ex-perience of a prescient model on specific undertakings can improve the presentation of a learning calculation and permit to all the more likely comprehend what causes an offered calculation to perform well on a given issue.

The possibility of metalearning isn't new, one of the first and original contri-butions having been given by [53]. In any case, the strict term just began showing up in AI writing during the 1990s, yet still numerous publi-cations manage issues identified with metalearning without utilizing the real word. This commitment attempts to get a handle on each perspective metalearning has been examined from, refering to books, research and survey papers of the most recent decade. We trust this review will give a valuable asset to the information mining and AI people group.

The rest of this paper is composed as follows. In Section 2 we re-see meanings of metalearning given in logical writing, concentrating on com-mon topics happening in every one of them. Segment 3 portrays various ideas of metalearning, connecting them to the definitions given in Section 2. In Section4 commonsense contemplations emerging when planning a metalearning framework are talked about, while open research headings are recorded in Section 5.

II. DEFINITION

First, In the 1990s, the term metalearning started to appear in machine learning re- search, although the concept itself dates back to the mid-1970s [53]. A number of definitions of metalearning have been given, the following list cites the main review papers and books from the last decade:

1. Metalearning studies how learning systems can increase in efficiency through experience; the goal is to

understand how learning itself can become flexible according to the domain or task under study. ([65])

2. The primary goal of metalearning is the understanding of the inter- action between the mechanism of learning and the concrete contexts in which that mechanism is applicable. ([25])

3. Metalearning is the study of principled methods that exploit meta- knowledge to obtain efficient models and solutions by adapting ma- chine learning and data mining processes. ([9])

4. Metalearning monitors the automatic learning process itself, in the context of the learning problems it encounters, and tries to adapt its behaviour to perform better. ([62])

Learning systems that adapt and improve by experience are a key concept of definitions 1, 3 and 4. This in itself however does not suffice as a descrip- tion, as it basically applies to all machine learning algorithms. Metalearning becomes metalearning by looking at different problems, domains, tasks or con- texts or simply past experience. This aspect is inherent in all of the definitions, although somewhat disguised in definition 3 using the term metaknowledge in- stead. Metaknowledge as described by the authors stands for knowledge to be exploited from past learning tasks, which may both mean past learning tasks on the same data or using data of another problem domain. Definition

2 differs in emphasising a better comprehension of the interaction between domains and learning mechanisms, which does not necessarily imply the goal of improved learning systems, but the pursuit of a better understanding of for which tasks individual learners succeed or fail.

Rephrasing, the common ground the above definitions share, we propose to define a metalearning system as follows:

Definition 1

1. A metalearning system must include a learning subsystem, which adapts with experience.

2. Experience is gained by exploiting metaknowledge extracted

(a) . . . in a previous learning episode on a single dataset, and/or

(b) . . . from different domains or problems.

Furthermore, a concept often used in metalearning is that of a bias, which, in this context, refers to a set of assumptions influencing the choice of hypotheses for explaining the data. [9] distinguishes declarative bias specifying the rep- resentation of the space of hypotheses (for example representing hypotheses using neural networks only) and procedural bias, which affects the ordering of the hypotheses (for example preferring hypothesis with smaller runtime). The bias in base-learning according to this theory is fixed, whereas metalearning tries to choose the right bias dynamically.

3 Notions of Metalearning

Metalearning can be employed in a variety of settings, with a certain disagree- ment in literature about what exactly constitutes a metalearning problem. Different notions will be presented in this section while keeping an eye on the question if they can be called metalearning approaches according to Def- inition 1. Figure 1 groups general machine and metalearning approaches in relation to Definition 1. Each of the three circles presents a cornerstone of the definition (1: adapt with experience, 2a: meta-knowledge on same data set, 2b: meta-knowledge from different domains), the approaches are arranged into the circles and their overlapping sections depending on which parts of the defini- tion applies to them. As an example, ensemble methods do generally work with experience gained with the same data set (definition 2a) and adapt with experience (definition 1), however, the only approach potentially applying all three parts of the definition is algorithm selection, which appears where all three circles overlap.

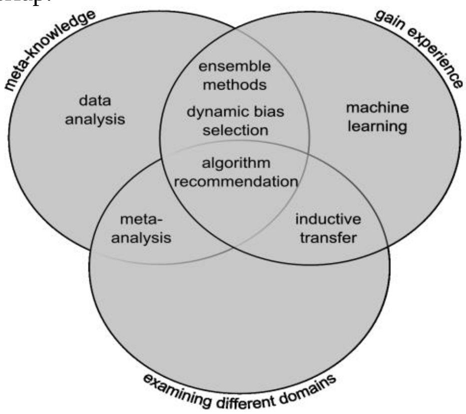

Fig. 1 Notions of metalearning vs. components of a metalearning system

3.1 Ensemble methods and combinations of base-learners

Model combination is often used when several applicable algorithms for a problem are available. Instead of selecting a single algorithm for a problem, the risk of choosing the wrong one can be reduced by combining all or a subset of the available outcomes. In machine learning, advanced model combination can be facilitated by ensemble learning according to [17] and [69], which comprises strategies for training and combining outputs of a number of machine learning algorithms. One often used approach of this type is resampling, leading to a number of ensemble generation techniques. Two very popular resampling- based ensemble building methods are:

– Bagging introduced in [12], which denotes repeated random sampling with replacement to produce a dataset of the same size as the original training set. The dataset is subsequently used for training of a base model and the

collection of models obtained in this way forms an ensemble with indi- vidual models' decisions combined typically using voting (in classification problems) or averaging (in regression problems).

– Boosting proposed in [21], which manipulates the probability with which samples are drawn from the original training data, to sequentially train classifiers focusing on the 'difficult' parts of the training set. Hence each consecutive ensemble member focuses on the training examples that can- not be successfully handled by the ensemble developed up to that point. The ensemble is usually built until a specified number of ensemble mem- bers is generated (although other stopping criteria are possible) and their decisions are combined using a weighted voting mechanism. Although the ensemble members can be 'weak' learners (i.e. models only slightly better than chance), this property must hold in the context of an increasingly difficult resampled dataset. As a result at some stage the 'weak' learner may in fact need to be quite complex and powerful.

The above approaches exploit variation in the data and are referred to as met- alearning methods in [9] and [62]. Bagging however does not satisfy point 2 of Definition 1, as consecutive random samples from the original dataset are independent from each other, so there is no experience from previous learn- ing episodes involved. In the case of boosting however, the ensemble is built sequentially and it is the performance of previous ensemble members (i.e. ex- perience gained while trying to solve the problem) that influences the sampling process.

More often, the following two approaches are considered as metalearning techniques:

– Stacked generalisation (or stacking) as introduced in [68], where a number of base learners is trained on the same dataset. Their outputs are subse- quently being used for a higher level learning problem, building a model linking the outcomes of the base learners to the target value. The meta-model then produces the final target outcome.

– Cascade generalisation [23], which works sequentially. When building a model, the output of the first base learner is appended to the original feature set and passed on to the next learner with the original target values. This process can then be repeated.

Although in these cases the information about base-learning is drawn in the sense of point 2a of Definition 1, these algorithms are limited to a single problem domain with a bias that is fixed a priori, so that they, using the definition above, do not undoubtedly qualify as metalearning methods.

3.2 Algorithm recommendation

A considerable amount of metalearning research has been devoted to the area of algorithm recommendation. In this special case of metalearning, the aspect of interest is the relationship between data characteristics 1 and algorithm per-formance, with the final goal of predicting an algorithm or a set of algorithms suitable for a specific problem under study. As a motivation, the fact that it is infeasible to examine all possible alternatives of algorithms in a trial and error procedure is often given along with the experts necessary if pre-selection of algorithms is to take place. This application of metalearning can thus be both useful for providing a recommendation to an end-user or automatically selecting or weighting algorithms that are most promising.

[62] points out another aspect: it is not only the algorithms themselves, but different parameter settings that will naturally let performance of the same algorithm vary on different datasets. It would be possible to regard versions of the same algorithm with different parameter settings as different learning algo- rithms altogether, but the author advocates treating the subject and studying its effects differently. Such an approach has for example been taken in [26] and [41], where the authors discuss a hybrid metalearning and search based tech- nique to facilitate the choice of optimal parameter values of a Support Vector Machine (SVM). In this approach, the candidate parameter settings recom- mended by a metalearning algorithm are used a starting point for further optimization using Tabu Search or Particle Swarm Optimization techniques, with great success. [51] investigate increasing the accuracy and decreasing runtime of a genetic algorithm for selecting learning parameters for a Support Vector Machine and a Random Forests classifier. Based on past experience on other datasets and corresponding dataset characteristics, metalearning is used to select a promising initial population for the genetic algorithm, reducing the number of iterations needed to find accurate solutions.

An interesting treatment of the above problem can also be found in [31], where the authors propose to take into account not only the expected per- formance of the algorithm but also its estimated training time. In this way the algorithms can be ordered according to the estimated training complexity, which allows to produce relatively well-performing models very quickly and then look for better solutions, while the ones already trained are producing predictions. These ideas are further extended in [30], where some modifications of the complexity measures used are introduced.

The classic application area of algorithm selection in machine learning is classification. [56] however tries to generalise the concepts to other areas including regression, sorting, constraint satisfaction and optimisation. Met-alearning for algorithm selection has also been investigated in the area of time series forecasting, where the term was first used in [48]. A comprehensive and recent treatment of the subject can be found in [66] and [37], where time series are clustered according to their characteristics and recommendation rules or combination weights derived with machine learning algorithms. Maintaining the Integrity of the Specifications

III. CONSIDERATIONS FOR USING METALEARNING

Before applying metalearning to any problem, certain practical choices have to be made. This includes the choice of a metalearning algorithm, which can even constitute a meta-metalearning problem itself. Selection of appropriate metaknowledge and the problem of setting up and maintaining metadatabases have to be tackled, research efforts of which will be summarised in this section.

3.1 Prerequisites

As also elaborated on in [9], metalearning can not be seen as a magic cure to machine learning problems for a variety of reasons. First of all, the extracted metafeatures need to be representative of their problem domain, otherwise, an algorithm will fail to identify similar domains. On the same note, if a problem has not been seen before, metalearning will be unable to exploit past knowledge to improve prediction performance. Performance estimation may be unreliable because of the natural limitations of estimating the true performance of the dataset. Different metafeatures might be applicable to each dataset. These is- sues emphasise the importance of being critical when designing a metalearning system.

3.2 Metalearning algorithms

[62] gives a survey on efforts to describe properties of algorithms. The au- thor distinguishes qualitative properties (for example type of data that can be handled, learning strategy, incrementality) and quantitative properties (bias-variance profile, runtime properties like scalability and

resilience). In an effort to find an implementation and vendor-independent method for representing machine learning models, the XML-based standard PMML has been devel- oped and gained some recognition in the last years. A detailed description of PMML can be found in [27].

The choice of a metalearning algorithm naturally depends on the prob- lem and the task to be solved. Generally, traditional classification algorithms are very successful in metalearning algorithm selection and can include meta- decision trees [60], neural networks, Support Vector Machines or any other classification algorithms, with the k -Nearest Neighbours being another pop- ular choice [9]. Applying regression algorithms is less popular, even smaller is the number of available algorithms to learn rankings. One of the simplest ranking method involves dividing the problem space using clustering of avail- able datasets according to a distance measure (usually k -Nearest Neighbour) of the metafeatures and using average performance ranks of the cluster into which a new problem falls [11]. [10] also look at the magnitude and significance of the differences in performance. The NOEMON approach introduced by [35] builds classifiers for each pair of base forecasting methods with a ranking be- ing generated using the classifiers' outputs. [58] build decision trees using the positions in a ranking as target values.

3.3 Extracting metaknowledge

According to [9], metaknowledge is derived in the course of employing a learn- ing system. A very common form of metaknowledge is the performance of algorithms in certain problem domains, which is to be linked with charac- teristics of the task. Several possibilities for characterising a problem domain exist.

The most straightforward form of metaknowledge extracted from the data include statistical or information-theoretic features. For classification prob- lems, [9] mention the number of classes and features, ratio of examples to fea- tures, degree of correlation between features and target concept and average class entropy. For other application areas, features can look completely differ- ent, as for example summarised in [38] for the area of time series forecasting, where features can include, for example, length, seasonality, autocorrelation, standard deviation and trends of the series.

[64] propose measures for the difficulty of a classification problem that can be used as an input for metalearning. They include class variation, denoting the probability that, by means of a distance measure, any two neighbouring data records have a different class value and example cohesiveness, measuring the density of the example distribution in the training set. In a similar approach, [36] also suggest comparing observations with each other and extract 'case base properties', which assess the quality of a dataset using measures such as redundancy, for example induced by data records that are exactly the same, or incoherency, which, for example occurs if data records have the same features but different class labels.

Alternatively to looking at the data only, information of individual algo- rithms and how they solved the problem can be considered, for example their predicted confidence intervals. This can be achieved by using a model that is fast to build and train and investigating its properties. In this spirit, [4] sug- gest building a decision tree for a classification problem and using properties of the tree such as nodes per feature, tree depth or shape to characterise it. Another approach is landmarking as proposed in [47], using the performance of simple algorithms to describe a problem and correlating this information with the performance of more advanced learning algorithms. A list of land- marking algorithms can be found in [62]. Landmarking algorithms can also be run on only a small sample of the data available, reducing the training time required. Performance information of different algorithms and learning curves generated when more data is added to the training set can then be used to select an algorithm according to [22].

Empirical evaluation of different categories of metafeatures in the con- text of their suitability for predicting classification accuracies of a number of standard classifiers can be found in [52]. The authors distinguish 5 such categories of features i.e. simple, statistical, information-theoretic, landmark- ing and model-based, which corresponds to the general categorization evident from the literature.

As with any learning problem, metalearning is subject to the 'curse of dimensionality' [7] and other issues, which can traditionally be solved by se- lecting a subset of relevant features. Although to the best of our knowledge, in the context of metalearning this issue has only been addressed in relatively few publications (e.g. [59, 34, 52]), we assume that the reason for this is quite simple – meta-feature selection does not differ from feature selection at the base-level, and the machine learning literature is very rich in this regard (a comprehensive review of various feature selection techniques can be found in [28]).

3.4 Metadatabases

As metalearning profits from knowledge obtained while looking at data from other problem domains, having sufficient datasets at one's disposal is impor- tant. [57] propose transforming existing datasets ('datasetoids') to obtain a larger number of them and show success of the approach on a metalearn- ing post-processing problem. [62] states that there is no lack of experiments being done, but datasets and information obtained often remain in 'people's heads and labs'. He proposes a framework to export experiments to specifically designed experiment databases based on an ontology for experimentation in machine learning. The resulting database can then, for example, give informa- tion on rankings of learning algorithms, the behaviour of ensemble methods, learning curve analyses and the bias-variance behaviour of algorithms. One example of such database can be The Open Experiment Database 4 . An analysis of this database together with a critical review can be found in [19].

An alternative approach to the problem of scarcity metadatabases has been presented in [50], where the authors describe a dataset generator able to pro- duce synthetic datasets with specified values of some metafeatures (like kur- tosis and skewness). Although the proposed generator appears to be at a very early stage of development, the idea is definitely very promising, also from the point of view of performing controlled experiments on datasets with spec- ified properties. Similarly to feature selection, synthetic data generation has received a considerable attention in the recent generic machine learning and data mining literature, especially in the context of data streams and concept drift (please see [3] and references therein).

IV. CONCLUSIONS AND RESEARCH CHALLENGES

Research in the area of metalearning is continuing in several directions. One area is the identification of metafeatures. As mentioned before, the vast majority of publications investigates extracting features from the dataset, mostly in the form of statistical or information theoretic measures. Landmarking is a different approach using simple base learning algorithms and their performance to describe the dataset at hand. However, [9] argue that characteristics of learning algorithms and gaining a better understanding of their behaviour would be a valuable research avenue with very few publications, for example [63], that exist in this area to date.

A lot of publications on metalearning focus on selecting the base-learning method that is most likely to perform well for a specific problem. Fewer pub- lications like [11] and [49] consider ranking algorithms, which can be used to guide combination weights and to increase robustness of a metalearning system.

Regarding adaptivity and continuous monitoring, many approaches go fur- ther than the static traditional metalearning approaches, for example by using architectures that support life-long learning such as in [33]. However, research in this area can still go a long way further investigating continuous adjust- ment, rebuilding or discarding of base-learners with the help of metalearning approaches.

Users of predictive systems are faced with a difficult choice of an ever in- creasing number of models and techniques. Metalearning can help to reduce the amount of experimentation by providing dynamic advice in form of assistants, decrease the time that has to be spent on introducing, tuning and maintaining models and help to promote machine learning outside of an academic environment.

REFERENCES

[1] Abbasi, A., Albrecht, C., Vance, A.O., Hansen, J.V.: Metafraud: a meta-learning frame- work for detecting financial fraud. Management Information Systems Quarterly 36(4), 1293–1327 (2012)

[2] Aiolli, F.: Transfer learning by kernel meta-learning. Journal of Machine Learning Research-Proceedings Track 27, 81–95 (2012)

[3] Albert Bifet Geoff Holmes, R.K., Pfahringer, B.: Data stream mining a practical approach. Tech. rep., The University of Waikato (2011)

[4] Bensusan, H., Giraud-Carrier, C., Kennedy, C.: A higher-order approach to metalearning. Proceedings of the ECML'2000 workshop on Meta-Learing: Building auto- matic advice strategies for Model Selection and Method Combination (2000)

[5] Bernstein, A., Provost, F., Hill, S.: Toward intelligent assistance for a data mining process: an ontology-based approach for cost-sensitive classification. IEEE Transactions on Knowledge and Data Engineering 17, 503–518 (2005)

[6] Biggs, J.B.: The role of meta-learning in study process. British Journal of Educational Psychology 55, 185–212 (1985)

[7] Bishop, C.: Neural Networks for Pattern Recognition. Oxford University Press, New York, USA (1995)

[8] Bonissone, P.P.: Lazy meta-learning: creating customized model ensembles on demand. In: Advances in Computational Intelligence, pp. 1–23. Springer (2012)

[9] Brazdil, P., Giraud-Carrier, C., Soares, C., Vilalta, R.: Metalearning: Applications to Data Mining. Springer (2009)

[10] Brazdil, P., Soares, C.: A comparison of ranking methods for classification algorithm selection. In: R. de Mantaras, E. Plaza (eds.) Machine Learning: Proceedings of the 11th European Conference on Machine Learning ECML2000, pp. 63–74. Springer (2000)

[11] Brazdil, P., Soares, C., de Costa, P.: Ranking learning algorithms: Using IBL and metalearning on accuracy and time results. Machine Learning 50(3), 251–277 (2003)

[12] Breiman, L.: Bagging predictors. Machine Learning 24(2), 123–140 (1996)

[13] Bruha, I., Famili, A.: Postprocessing in machine learning and data mining. ACM SIGKDD Explorations Newsletter 2, 110–114 (2000)

[14] Budka, M., Gabrys, B.: Ridge regression ensemble for toxicity prediction. Procedia Computer Science 1(1), 193–201 (2010). DOI 10.1016/j.procs.2010.04.022. URL http://www.sciencedirect.com/science/article/pii/S1877050910000232

[15] Budka, M., Gabrys, B., Ravagnan, E.: Robust predictive modelling of water pollution using biomarker data. Water Research 44(10), 3294–3308 (2010). DOI 10.1016/j.watres.2010.03.006. URL Http://www.sciencedirect.com/science/article/pii/S004313541000179X

[16] Cao, L.: Domain-driven data mining: Challenges and prospects. IEEE Transactions on Knowledge and Data Engineering 22, 755–769 (2010)

[17] Dietterich, T.: Ensemble methods in machine learning. In: Proceedings of the First International Workshop on Multiple Classifier Systems, pp. 1–15 (2000)

[18] Domingos, P., Hulten, G.: Mining high-speed data streams. In: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 71–80 (2000)

[19] Driessens, K., Vanwinckelen, G., Blockeel, H.: Meta-learning from an experiment database. In: Proceedings of the Workshop on Teaching Machine Learning at the 29th International Conference on Machine Learning, Edinburgh, UK (2012)

[20] Evgeniou, T., Micchelli, C., Pontil, M.: Learning multiple tasks with kernel methods. Journal of Machine Learning Research 6, 615–637 (2005)

[21] Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55(1), 119–139 (1997). DOI http://dx.doi.org/10.1006/jcss.1997.1504

[22] Fu¨rnkranz, J., Petrak, J., Brazdil, P., Soares, C.: On the use of fast subsampling esti- mates for algorithm recommendation. Tech. rep., sterreichisches Forschungsinstitut fr Artificial Intelligence (2002)

[23] Gama, J., Brazdil, P.: Cascade generalisation. Machine Learning 41(3), 315–343 (2000)

[24] Giraud-Carrier, C.: The data mining advisor: Meta-learning at the service of practi- tioners. In: Proceedings of the Fourth International Conference on Machine Learning and Applications, ICMLA '05, pp. 113–119. IEEE Computer Society, Washington, DC, USA (2005)

[25] Giraud-Carrier, C.: Metalearning - a tutorial. Tutorial at the 7th International Con- ference on Machine Learning and Applications (ICMLA), San Diego, California, USA (2008)

[26] Gomes, T.A., Prudˆencio, R.B., Soares, C., Rossi, A.L., Carvalho, A.: Combining meta- learning and search techniques to select parameters for support vector machines. Neu- rocomputing 75(1), 3–13 (2012)

[27] Guazzelli, A., Zeller, M., Lin, W.C., Williams, G.: PMML: An open standard for sharing models. The R Journal 1(1), 60–65 (2009)

[28] Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. The Journal of Machine Learning Research 3, 1157–1182 (2003)

[29] Hernansaez, J.M., Bot´ıa, J.A., Go´mez-Skarmeta, A.F.: METALA: a J2EE technology based framework for web mining. Revista Colombiana de Computacio´n 5(1) (2004)

[30] Jankowski, N.: Complexity measures for meta-learning and their optimality. Solomonoff 85th Memorial. Lecture Notes in Computer Science. Springer-Verlag (2011)

[31] Jankowski, N., Grabczewski, K.: Universal meta-learning architecture and algorithms. In: W. Duch, K. Grabczewski, N. Jankowski (eds.) Meta-learning in Computational Intelligence. Springer (2009)

[32] Kadlec, P., Gabrys, B.: Learnt topology gating artificial neural networks. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN 2008) as part of the 2008 IEEE World

Congress on Computational Intelligence (WCCI'2008), pp. 2605–2612 (2008)
[33] Kadlec, P., Gabrys, B.: Architecture for development of adaptive on-line prediction models. Memetic Computing 4(1), 241–269 (2009)
[34] Kalousis, A., Hilario, M.: Feature selection for meta-learning. In: D. Cheung, G. Williams, Q. Li (eds.) Advances in Knowledge Discovery and Data Mining, Lec- ture Notes in Computer Science, vol. 2035, pp. 222–233. Springer Berlin Heidelberg (2001)
[35] Kalousis, A., Theoharis, T.: NOEMON: design, implementaion and performance results of an intelligent assistant for classifier selection. Intelligent Data Analysis 5(3), 319–337 (1999)
[36] Ko¨pf, C., Iglezakis, I.: Combination of task description strategies and case base proper- ties for meta-learning. In: Proceedings of the 2nd international workshop on Integration and Collaboration Aspects of Data Mining, Decision Support and Meta-Learning, pp. 65–76 (2002)
[37] Lemke, C., Gabrys, B.: Meta-learning for time series forecasting and forecast combina- tion. Neurocomputing 73(10), 2006–2016 (2010)
[38] Lemke, C., Riedel, S., Gabrys, B.: Dynamic combination of forecasts generated by diversification procedures applied to forecasting of airline cancellations. In: Proceedings of the IEEE Symposium Series on Computational Intelligence, pp. 85–91 (2009)
[39] Matijaˇs, M., Suykens, J.A., Krajcar, S.: Load forecasting using a multivariate meta- learning system. Expert Systems with Applications 40(11), 4427–4437 (2013)
[40] Metal: Meta-learning assistant for providing user support in machine learning and data mining. http://www.metal-kdd.org/ (2002)
[41] de Miranda, P., Prudˆencio, R., de Carvalho, A., Soares, C.: An experimental study of the combination of meta-learning with particle swarm algorithms for svm parameter selection. Computational Science and Its Applications–ICCSA 2012 pp. 562–575 (2012)
[42] Molina, M.D.M., Romero, C., Ventura, S., Luna, J.M.: Meta-learning approach for au- tomatic parameter tuning: A case of study with educational datasets. In: EDM, pp. 180–183 (2012)
[43] Morik, K., Scholz, M.: The miningmart approach to knowledge discovery in databases. In: Intelligent Technologies for Information Analysis, pp. 47–65. Springer (2004)
[44] Nguyen, P., Kalousis, A., Hilario, M.: A meta-mining infrastructure to support kd work- flow optimization. eCML PKDD 2011 p. 1 (2011)
[45] Nguyen, P., Kalousis, A., Hilario, M.: Experimental evaluation of the e-lico meta-miner. 5th Planning to learn workshop WS28 at ECAI 2012 p. 18 (2012)
[46] 46. Pan, S., Yang, Q.: A survey on transfer learning. Knowledge and Data Engineering, IEEE Transactions on 22(10), 1345–1359 (2010)
[47] Pfahringer, B., Bensusan, H., Giraud-Carrier, C.: Meta-learning by landmarking various learning algorithms. In: In Proceedings of the Seventeenth International Conference on Machine Learning, pp. 743–750. Morgan Kaufmann (2000)
[48] Prudencio, R., Ludermir, T.: Using machine learning techniques to combine forecast- ing methods. In: Proceedings of the 17th Australian Joint Conference on Artificial Intelligence, pp. 1122–1127 (2004)
[49] Prudencio, R.B., Ludermir, T.B.: Meta-learning approaches to selecting time series mod- els. Neurocomputing 61, 121–137 (2004)
[50] Reif, M., Shafait, F., Dengel, A.: Dataset generation for Meta-Learning. In: KI-2012: Poster and Demo Track, pp. 69–73 (2012)
[51] Reif, M., Shafait, F., Dengel, A.: Meta-learning for evolutionary parameter optimization of classifiers. Machine Learning 87, 357–380 (2012)
[52] Reif, M., Shafait, F., Goldstein, M., Breuel, T., Dengel, A.: Automatic classifier selection for non-experts. Pattern Analysis and Applications pp. 1–14 (2012). DOI 10.1007/s10044-012-0280-z
[53] Rice, J.: The algorithm selection problem. In: M. Rubinov, M.C. Yovits (eds.) Advances in Computers, vol. 15. Academic Press, Inc. (1976)
[54] Silver, D., Bennett, K.: Guest editor's introduction: special issue on inductive transfer learning. Machine Learning 73, 215–220 (2008)
[55] Silver, D.L., Poirier, R., Currie, D.: Inductive transfer with context-sensitive neural networks. Machine Learning 73(3), 313–336 (2008)
[56] Smith-Miles, K.: Cross-disciplinary perspectives on meta-learning for algorithm selec- tion. ACM Computing Surveys 41(6), 1–25 (2008)
[57] Soares, C.: UCI++: Improved support for algorithm selection using datasetoids. In: T. Theeramunkong, B. Kijsirikul, N. Cercone, T.B. Ho (eds.) Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science, vol. 5476, pp. 499–506. Springer Berlin Heidelberg (2009)
[58] Todorovski, L., Blockeel, H., Dzeroski, S.: Ranking with predictive clustering trees. In: T. Elomaa, H. Mannila, H. Toivonen (eds.) Proceedings of the 13th European Confer- ence on Machine Learning, pp. 444–455. Springer (2002)
[59] Todorovski, L., Brazdil, P., Soares, C.: Report on the experiments with feature selection in meta-level learning. In: Proceedings of the PKDD-00 Workshop on Data Mining, Decision Support, Meta-Learning and ILP: Forum for Practical Problem Presentation and Prospective Solutions. Citeseer (2000)
[60] Todorovski, L., Dˇzeroski, S.: Combining classifiers with meta decision trees. Machine learning 50(3), 223–249 (2003)
[61] Tsai, C.F., Hsu, Y.F.: A meta-learning framework for bankruptcy prediction. Journal of Forecasting 32(2), 167–179 (2013)
[62] Vanschoren, J.: Understanding machine learning performance with experiment databases. Ph.D. thesis, Arenberg Doctoral School of Science, Engineering & Tech- nology, Katholieke Universiteit Leuven (2010)
[63] Vanschoren, J., Blockeel, H.: Towards understanding learning behavior. In: In Proceed- ings of the Annual Machine Learning Conference of Belgium and the Netherlands, pp. 89–96 (2006)
[64] Vilalta, R., Drissi, Y.: A characterization of difficult problems in classification. In: Proceedings of the 6th European Conference on Principles and Practice of Knowledge Discovery in Databases, Helsinki, Finland (2002)
[65] Vilalta, R., Drissi, Y.: A perspective view and survey of meta-learning. Artificial Intel- ligence Review 18, 77–95 (2002)
[66] Wang, X., Smith-Miles, K., Hyndman, R.: Rule induction for forecasting method selec- tion: Meta-learning the characteristics of univariate time series. Neurocomputing 72, 2581–2594 (2009)
[67] Wirth, R., Shearer, C., Grimmer, U., Reinartz, T., Schloesser, J., Breitner, C., En- gels, R., Lindner, G.: Towards process-oriented tool support for kdd. Proceedings of the 1st European Symposium on Principles of Data Mining and Knowledge Discovery, Trondheim, Norway (1997) 68. Wolpert, D.: Stacked generalization. Neural Networks 5, 241–259 (1992)
[68] Yao, X., Islam, M.: Evolving artificial neural network ensembles. IEEE Computational Intelligence Magazine 3, 31–42 (2008)
[69] Zhang, J., Ghahramani, Z., Yang, Y.: Flexible latent variable models for multi-task learning. Machine Learning 37, 221–242 (2008)