# Characterization of Learning Instances for Evolutionary Meta-Learning

William Raynaut[1], Chantal Soule-Dupuy[1], Nathalie Valles-Parlangeau[1], Cedric Dray[2], and Philippe Valet[2]

[1] IRIT UMR 5505, UT1, UT3
Universite de Toulouse, France
firstname.lastname@irit.fr
[2] Institut National de la Sante et de la Recherche Medicale (INSERM), U1048
Universite de Toulouse, France
firstname.lastname@inserm.fr

**Abstract.** Machine learning has proven to be a powerful tool in diverse fields, and is getting more and more widely used by non-experts. One of the foremost difficulties they encounter lies in the choice and calibration of the machine learning algorithm to use. Our objective is thus to provide assistance in the matter, using a meta-learning approach based on an evolutionary heuristic. We expand here previous work presenting the intended workflow of a modeling assistant by describing the characterization of learning instances we intend to use.

**Keywords:** Meta-Learning, Modeling, Prediction, Evolutionary Heuristics, Algorithm selection

## 1 Motivation

Over the last decades was produced an important variety of techniques and algorithms labeled as machine learning. But the performance of such techniques can vary a lot from a dataset to another, and the "no free lunch" theorems [21] showed that no algorithm could outperform all others on every possible problem. This led to many studies of algorithm's inner bias adequateness to diverse learning problem, such as [1] and [4] who used rule-generation machine learning techniques on the problem, describing the conditions under which the significant performance difference between algorithms holds. These applications of machine learning to the study of itself bore great significance over how this Meta-Learning problem would be addressed. Despite promising applications of such approaches over a limited range of learning tasks, like pairwise algorithm comparison [6], or recursion of adaptive learners [20], the Meta-Learning problem still carries many open perspectives. Another approach would be to address directly the question : "*Which learning algorithm will perform best on a given learning problem ?*", without having to comply to the limitation of the classic machine learning techniques employed at the meta-level.

Our own perspective view on the matter is that the meta-knowledge can be viewed as a population of meta-level learning instances (or meta-instances), each describing the evaluated application of a learning task to a given dataset, and that a good solution to a given meta-learning task can be obtained via the evolutionary exploration of this population. Such approach is giving interesting results among other classes of problems such as Boolean satisfiability (SAT) [22] or Instance selection [12], but, to our knowledge, has not yet been explored regarding Meta-Learning.

Our objective is to provide modeling assistance through an evolutionary algorithm-selection approach, which intended workflow is illustrated by figure 1 and was presented more thoroughly in [15].

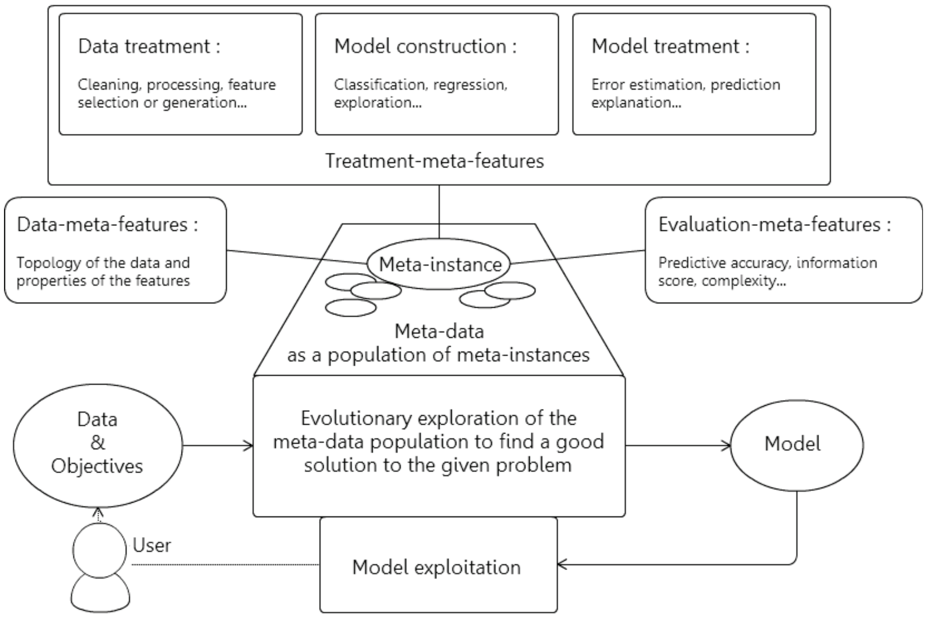


**Fig. 1.** Modeling assistant

One of the foremost issues we must address in order to complete our framework, and the main topic of this paper, will be the characterisation of the meta-instances. This problem can be viewed as an extended form of the dataset characterization problem faced by most meta-learning approaches, which consists in the definition of a subset of dataset properties (meta-level features of the dataset) that should allow a fine grain characterisation of datasets, while still complying to the requirements of the meta-level learner employed. It is typically solved through some kind of meta-level features selection [7], but to fit most learners requirements, dataset properties have to be aggregated into fixed-length feature vectors, which results into a important loss in information as stated in [6]. We intend to overcome this issue through the use of an evolutionary heuristic, whose fitness would rely on dissimilarity between meta-instances. Such approach would indeed allow the use of all available information to characterize the meta-instances. Relating in that way to the "anti-essentialist" representations such as discussed in [3], we believe that limitations in the representations of datasets are among the main obstacles to well performing algorithm selection, and are focusing our efforts toward the definition of such representation.

## 2 Characterising Learning Instances

We will here address the definition of the features that will describe our meta-instances, hence called meta-feature. Those sets of meta-features should be large enough to characterize well any modeling task, but a balance must be found to avoid the abundance of indecisive features and limit computational complexity. Furthermore, in order to discriminate between meta-features or meta-instances according to the user's need, the comparison of meta-features of a particular meta-instance - or of a given meta-feature over several meta-instances - should be possible and make sense.

As a meta-instance describes the evaluated application of a learning task to a given dataset, we can intuitively split those meta-features along three dimensions. First, meta-features describing the data (Fig.1 Data-meta-features), then, meta-features describing the applied treatments (Fig.1 Treatment-meta-features), and finally, meta-features evaluating the resulting model (Fig.1 Evaluation-meta-features).

### 2.1 Data Meta-features

The dataset characterization problem has been addressed along two main directions :

- In the first one, the dataset is described through a set of statistical or information theoretic measures. This approach, notably appearing in the STATLOG project [10], and in most studies afterwards [8, 20, 12], allows the use of many expressive measures, but its performance depends heavily on the adequateness of bias between the meta-level learner and the chosen measures. Experiments have been done with meta-level features selection [19] in order to understand the importance of different measures, but the elicited optimal sets of meta-feature to perform algorithm selection over two different pools of algorithms can be very different, revealing no significant tendencies among the measures themselves. This led [20] to the intuition that adapting the meta-learning process to specific tasks is in fact a meta-meta-learning problem, and so on, requiring an infinite recursion of adaptive learners to be properly solved.

- The second direction of approach to dataset characterization focuses, not on computed properties of the dataset, but on the performance of simple learners over the dataset. It was introduced as landmarking in [14], where the accuracies of a set of very simple learners are used as meta-features to feed a more complex meta-level learner. There again, the performance of the method relies heavily on the adequate choice of both the base and meta-level learner, with no absolute best combination. Further development introduced more complex measures than predictive accuracy over the models generated by the simple learners. For instance, [11] claims that using as meta-features different structural properties of a decision tree induced over the dataset by simple decision-tree learners can also result in well performing algorithm selection. [13] experiments with those approaches to algorithm selection, showing that all can result in good performance, but that no overall dominance between those methods or over the approaches relying on statistical measures can be found.

The dataset characterization problem has thus already received quite some attention in previous meta-learning studies, but, as stated before, the aggregation of meta-features into fixed-length vectors processable through the meta-level learner were source of an important information loss, even though it was partially limited in [8] with the use of histograms describing the distribution of meta-feature values. However, the paradigm shift between literal meta-learning and our approach will shift the issue to another : we are free to use varying-length meta-feature vectors, but have to design a sound way to compare them. This mostly comes as an issue when comparing meta-features computed over individual features of the dataset, as illustrated in the following example.

**Example** *We consider two datasets, $A$ and $B$ depicted in Fig.2. $A$ describes 12 features of 100 individuals, and $B$, 10 features of 200 individuals. Let us say we want to compare the results of a set of 5 statistical or information theoretic measures over each individual feature, like mean, variance, standard deviation, entropy, and kurtosis (as illustrated over the second feature of $A$ in Fig.2). The complete information we want to compare is then a 60-value vector for $A$, and a 50-values vector for $B$.*

*Our stance on the matter is to compare those features by most similar pairs, while comparing $A$'s two extra features with empty features (features with no value at all). The assumption taken here is that a feature with absolutely no value is equivalent to no feature at all. To get back to our example, we end up comparing the 5 measures taken on the two closest (according to these very measures) features in $A$ and $B$, then of the second closest, and so on, to finish on comparing the measures taken over the two extra features of $A$ with measures taken over an artificial empty feature. These different comparisons sum up to an accurate description of how different $A$ and $B$ are, according to our set of measures. These pairwise comparisons would allow to ignore the presentation order of the features (which holds no meaningful information), focusing on the actual topology of the dataset.*

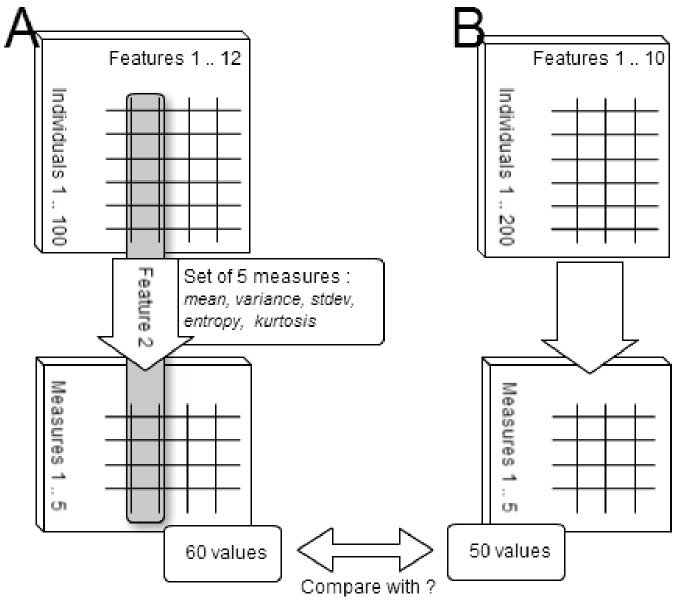


**Fig. 2.** Measures over individual features

Assuming that a very expressive comparison will result in a better performing fitness, this only emphasizes the need for an extensive set of meta-features. We intend to use most of the classic statistical and information theoretic measures, from the number of instances to features entropy, considering also measures of feature correlation. As the various landmarking approaches showed interesting results, we additionally consider using such measures as meta-features, but further studies might be required to limit overlapping information between the two kinds of measures.

### 2.2 Evaluation and Treatment Meta-features

The meta-features describing the evaluation of the resulting model should consider a wide range of criteria and allow some flexibility in its comparison to the user's need. Among many usual criteria, we are giving a particular attention to meaningful information-based criteria such as described in [9]. We also wish to investigate the definition of some explainability criteria following [17] prediction explanations, as the ability of the model to explain its predictions has been shown to be a very important factor in allowing non-experts to understand and accept them [18].

The meta-features describing the modeling treatments should consider all potential treatments producing a model of the given dataset. The characterization of treatments has notably been addressed by the algorithm profiles presented in [5], where sets of algorithm properties are learned from modeling tasks on arbitrary chosen datasets. We intend to describe a modeling algorithm not from a-priori learned properties, but from aggregated properties of the meta-instances of our population presenting the use of this particular algorithm. For instance, the current *predictive accuracy* property of a given algorithm could be defined as the mean of the *predictive accuracy* evaluation-meta-feature among the meta-instances in our current base featuring that particular algorithm. We also consider relative aggregations, such as rank over known algorithms, as no absolute value is required for comparison.

## 3 Conclusion and perspectives

The set of all meta-features presented above should allow fine grain description of evaluated modeling experiments, and will thus define the structure of the meta-instances over which the evolutionary heuristic will be applied. In other terms, those meta-features will be the genome of the meta-instances, along which evolution will take place, to find a modeling treatment answering the user's need.

However, in order to complete and thus evaluate this framework, several important tasks are yet to be addressed. First, a representation of the user's modeling need that would allow its automatic or semi-automatic elicitation will be required. Indeed, as the target user is a non-expert, he should be walked through the definition of his modeling need, that will define the goal of the evolution. Also, such representation could allow to lessen the computational complexity of the heuristic, by considering only instances that could answer the user's need.

Then, meta-instances comparison metrics shall be formalized in order to define the evolutionary fitness as a similarity with the evolution goal that was elicited from the user's need.

Finally two of the important challenges to address will be the definition and calibration of the evolutionary heuristic employed, and the creation of predatory mechanisms limiting the population of meta-instances. We intend to use the framework of genetic algorithms [2] and memetic algorithms [16], which present desirable properties such as unconstrained individuals and native parallelism, the later being required to deal with the important computational complexity of the intended workflow.

## References

1. Aha D.W., Generalising from case studies: a case study, Proceedings Ninth International Conference on Machine Learning, Morgan Kaufmann, San Mateo, CA (1992), pp. 110, 1992
2. Jaume Bacardit and Xavier Llor. Large scale data mining using genetics-based machine learning. In Proceedings of the 15th annual conference companion on Genetic and evolutionary computation (GECCO '13 Companion), 2013
3. RPW Duin, The Dissimilarity Representation for finding Universals from Particulars by an anti-essentialist Approach, Pattern Recognition Letters (2015).
4. Gama J and Brazdil P, Proceedings of the seventh Portuguese Conference on Artificial Intelligence, Characterization of Classification Algorithms, 189-200, 1995
5. M. Hilario and A. Kalousis. Building algorithm profiles for prior model selection in knowledge discovery systems. In Proceedings of the IEEE SMC'99, International Conference on Systems, Man and Cybernetics. IEEE press, October 1999.
6. Kalousis A., Hilario M. Model Selection via Meta-learning : a Comparative Study. In Proceedings of the 12th IEEE International Conference on Tools with AI, Vancouver, November 2000
7. Alexandros Kalousis, Melanie Hilario, Feature Selection for Meta-learning, Advances in Knowledge Discovery and Data Mining Lecture Notes in Computer Science Volume 2035, 2001, pp 222-233, 2001
8. Kalousis, A. Algorithm Selection via Meta-Learning. PHD Thesis, Thesis Number : 3337. University of Geneve, Department of Computer Science, 2002
9. Kononenko I, Bratko I, Information-Based Evaluation Criterion for Classifier's Performance, Machine Learning, January 1991, Volume 6, Issue 1, pp 67-80, 1991
10. D. Michie, D.J. Spiegelhalter, C.C. Taylor. Machine Learning, Neural and Statistical Classification. Ellis Horwood Series in Artificial Intelligence, 1994
11. Yonghong Peng , Peter A Flach , Pavel Brazdil , Carlos Soares, Decision Tree-Based Data Characterization for Meta- Learning, 2002
12. Raul Perez, Antonio Gonzalez, Enrique Leyva, "A Set of Complexity Measures Designed for Applying Meta-Learning to Instance Selection," IEEE Transactions on Knowledge and Data Engineering, 2015
13. Johannes Furnkranz, Johann Petrak, Extended data characteristics, A MetaLearning Assistant for Providing User Support in Machine Learning and Data Mining, METAL, 2002
14. Bernhard Pfahringer, Hilan Bensusan, Christophe Giraud-Carrier, Metalearning by landmarking various learning algorithms. Proceedings of the Seventeenth International Conference on Machine Learning, ICML'2000, pp. 743750. June 2000
15. William Raynaut, Chantal Soule-Dupuy, Nathalie Valles-Parlangeau, Cedric Dray and Philippe Valet. Addressing the Meta-Learning problem with Metaheuristics (Extended Abstract), Metaheuristics International Conference, 2015

16. Smith J, Coevolving memetic algorithms: A review and progress report, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 6–17, 2007
17. Erik Strumbelj, Igor Kononenko, Towards a Model Independent Method for Explaining Classification for Individual Instances, Data Warehousing and Knowledge Discovery, Lecture Notes in Computer Science Volume 5182, pp 273-282, 2008
18. Erik Strumbelj, Zoran Bosni, Igor Kononenko, Branko Zakotnik, Cvetka Grai Kuhar, Explanation and reliability of prediction models: the case of breast cancer recurrence, Knowledge and Information Systems, 2010, Volume 24, Number 2, Page 305
19. L. Todorovski, P. Brazdil, and C. Soares. Report on the experiments with feature selection in meta-level learning. In P. Brazdil and A. Jorge, editors, Proceedings of the Data Mining, Decision Support, Meta-Learning and ILP Workshop at PKDD2000, pages 27-39, 2000
20. Ricardo Vilalta, Youssef Drissi, A Perspective View and Survey of Meta-Learning, Artificial Intelligence Review 2002, Volume 18, Issue 2, pp 77-95, 2002
21. Wolpert David, The Lack of A Priori Distinctions between Learning Algorithms, Neural Computation, pp. 1341-1390, 1996
22. J. Shen H. Hoos-K. Leyton-Brown L. Xu, F. Hutter. Satzilla2012: Improved algorithm selection based on cost-sensitive classification models. SAT Challenge 2012.