Meta-learning of Text Classification Tasks

Jorge G. Madrid[1(✉)] and Hugo Jair Escalante[1,2]

[1] Computer Science Department, Instituto Nacional de Astrofísica, Óptica y Electrónica, Tonantzintla, 72840 San Andrés Cholula, PUE, Mexico
{jgmadrid,hugojair}@inaoep.mx

[2] Computer Science Department, Centro de Investigación y Estudios Avanzados del IPN, Zacatenco, 07360 Mexico city, Mexico

Abstract. A text mining characterization is proposed consisting of a set of meta-features, unlike previous meta-learning approaches, some of them are extracted directly from raw text. Such novel description is useful for comparing text mining tasks and study their differences. The problem of determining the task associated to a text classification dataset is introduced and approached with our characterization. Experimental results on a set of 81 corpora show that the proposed meta-features indeed allow to recognize tasks with acceptable performance using only a few meta-features.

Keywords: Meta-learning · Text classification · Meta-features

1 Introduction

For humans, experiences from the past are usually helpful when learning a new skill or solving a new problem. Equivalently, in the context of machine learning, meta-learning takes advantage of prior experience acquired when solving related tasks for approaching new problems [12]. The main goals are to speed up the learning process and to improve the quantitative performance of models. Meta-learning has had an impact into several machine learning problems such as learning to design optimization algorithms [1], automatically suggesting supervised learning pipelines [4], learning architectures for deep neural networks [3] and few-shot learning [10].

Text classification is one of the most studied tasks in NLP, this is because of the number of problems and applications that can be approached as text classification tasks. Many techniques for pre-processing, feature extraction, feature selection and document representation have been developed over the last decades. Each of these being appropriate for different scenarios and types of tasks. However, despite the progress achieved by the NLP community, nowadays

This work was partially supported by CONACyT under grant A1-S-26314 *Integración de Visión y Lenguaje mediante Representaciones Multimodales Aprendidas para Clasificación y Recuperación de Imágenes y Videos.*

I. Nyström et al. (Eds.): CIARP 2019, LNCS 11896, pp. 107–119, 2019.
https://doi.org/10.1007/978-3-030-33904-3_10

it is still an NLP expert who determines the pipeline of text classification systems, including preprocessing methods, representation and classification models together with their hyperparameters.

This paper takes a first step towards the characterization of text classification problems with the ultimate goal of suggesting text classification pipelines for any type of problem, that is *Meta-learning of text classification tasks*. Earlier work in this direction (see Sect. 2) has defined straightforward meta-features and worked over a small number of datasets. What is more, previous work has focused exclusively on tabular data (i.e., they have extracted meta-features from a document-text matrix). Since natural language presents different characteristics from those of generic tabular data, herein we define a set of meta-features that are derived from the analysis of raw text and combine them with traditional meta-features. To the best of our knowledge this the first work on meta-learning extracting information from raw text directly.

As a first approximation, we approach the problem of learning to determine the type of task (e.g., topic-based vs. sentiment analysis) using the meta-features as predictive variables. We provide empirical evidence on the suitability of the proposed meta-features for characterizing text classification tasks. Additionally, we perform an analysis of the most important features for the approached meta-learning problem. Experimental results are encouraging and show that meta-learning of text classification is a promising research venue for NLP.

Our contributions are threefold: (1) introduction of the task-type prediction problem; (2) introduction of novel and effective meta-features that can be used for other meta-learning tasks; (3) experiments of larger scale than previous work (we proposed 73 meta-features, compared to 11 from previous references and report experiments on 81 corpora, compared to 9 from related work).

2 Background and Related Work

Meta-learning aims to learn from prior learning-experience in order to speed up the learning process when approaching a new task. A common way to learn from/across tasks is by characterizing them with a set of *meta-features* [13]. These attempt to describe a task (i.e., a dataset) by information readily available at a task/dataset level. In this way, each task is usually represented by a vector where dimensions are associated to meta-features. Meta-features can be as simple as the number of instances and features in a dataset and as complex as statistical measures from the data distribution. [11] provide a comprehensive description of the most commonly used meta-features.

In the machine learning context, meta-learning has been studied for a while [12,13]. But it is only recently that it has become a mainstream topic, this mainly because of its successes in several tasks. For instance, Feurer et al. [5] successfully used a set of meta-features to warm-start a hyper-parameter optimization technique in the popular state-of-the-art AutoML solution *Autosklearn*. Likewise, the success of deep learning together with the difficulty in defining appropriate architectures and hyperparameters for users, has motivated a boom on neural architecture search, where meta-learning is common [3].

2.1 Meta-learning in Text Classification

In the context of text mining, meta-features from clustering text documents have been used directly for classification [2]. In the context of meta-learning these features have been used only in very specific domains [8]. Efforts dealing with generic datasets and closely related to the proposed research are reviewed in the remainder of this section.

Lam and Lai [7] introduced a meta-learning approach for text classification, they characterized subsets of the Reuters corpus with 8 *document-feature meta-characteristics* that were extracted from the document-term matrix representation. These consisted of simple meta-data such as the average document length or simple term statistics. These meta-features were later used to estimate the classification error of 6 classifiers and recommended a model depending on the prediction. Similarly, Gomez, et al. [6] proposed 11 meta-features which were also collected from a matrix representation of the documents, 9 different corpus were characterized with them. This method learned a set of rules that determine a suitable algorithm depending on the meta-feature values of the corpus.

Unlike previous approaches we do not assume a predefined representation of the documents, instead we derive meta-features from the raw text and combine these with traditional ones. This allows us to capture more language-relevant information. Also, we perform experiments of larger scale than previous work, considering 81 datasets (previous work used 6–9 collections) that have been characterized by 73 meta-features (in the past 8–11 meta-features have been considered).

3 Meta-learning Text Classification Tasks

We propose in this work a set of 73 meta-features with the aim of characterizing tasks (i.e., datasets), where the proposed meta-features comprise both, traditional and NLP-based ones. The ultimate goal of our work is to automatically suggest pipelines for solving text classification problems. As a first step in such direction, we show in this work that the proposed meta-features can be used as predictive variables to learn models able to recognize the type of task associated to a dataset. Different text classification tasks can be derived given the same dataset, our set of meta-features also acknowledges this since some of the proposed measures provide statistical information about the classes.

In NLP it is empirically known that certain methods work better according to the type of task that is aimed, for example, character-based n-grams are known to perform better than other representations in authorship attribution tasks because they determine better an author's style. Identifying correctly the type of task that is tackled is a fundamental step when modeling a text classification *pipeline*, thus we propose to automate this in pursuit of an automated recommendation system. In this work, we limit ourselves to learn to discriminate among types of tasks, and postpone to future work the problem of pipeline recommendation.

3.1 Proposed Meta-features

A common form of characterizing tasks are meta-features. Some sets of meta-features have proven to be useful for supervised machine learning problems, however we consider that these are not enough to characterize tasks in text classification; extracting them usually requires a tabular representation of the data, in the case of text documents some representation such as Bag-of-Words would be necessary. When a representation is selected some fundamental characteristics of language are lost, extracting *traditional* meta-features from it would result in a limited characterization of the task. We propose a set of 73 meta-features combining meta-learning traditional features with NLP ones. Below we organized them in groups.

- **General meta-features.** The *number of documents* and the *number of categories.*
- **Corpus hardness.** Most of these originally used in [9] to determine the hardness of short text-corpora.
  Domain broadness. Measures related to the thematic broadness/narrowness of words in documents. We included measures based on the vocabulary length and overlap: *Supervised Vocabulary Based (SVB)*, *Unsupervised Vocabulary Based (UVD)* and *Macro-averaged Relative Hardness (MRH).*
  Class imbalance. Class Imbalance (CI) ratio.
  Stylometry. Stylometric Evaluation Measure (SEM)
  Shortness. Vocabulary Length (VL), *Vocabulary Document Ratio (VDR)* and average *word length.*
- **Statistical and information theoretic.** We derive meta-features from a document-term matrix representation of the corpus.
  min, max, average, standard deviation, skewness, kurtosis, ratio average-standard deviation, and entropy of: vocabulary distribution, documents-per-category and words-per-document:
  Landmarking. 70% of the documents are used to train 4 simple classifiers and their performance on the remaining 30% was used based on the intuition that some aspects of the dataset can be inferred: *data sparsity - 1NN, data separability - Decision Tree, linear separability - Linear Discriminant Analysis, feature independence Naïve Bayes.* The *percentage of zeros* in the matrix was also added as a measure for sparsity.
  Principal Components (PC) statistics. Statistics derived from a PC analysis: *pcac* from Gomez, et al. [6]; for the first 100 components, the same statistics from documents per category and their *singular values sum, explained ratio and explained variance*, and for the first component its *explained variance.*
- **Lexical features.** We incorporated the distribution of parts of speech tags. We intuitively believe that the frequency of some lexical items will be higher depending on the task associated to a corpus, for instance a corpus for sentiment analysis may have more adjectives while a news corpus may have less. We tagged the words in the document and computed the average number of *adjectives, adpositions, adverbs, conjunctions, articles, nouns, numerals, particles, pronouns, verbs, punctuation marks* and *untagged words* in the corpus.

– **Corpus readability.** Statistics from text that determine readability, complexity and grade from textstat library[1]: *Flesch reading ease:*

$$206.835 - 1.015\left(\frac{total\_words}{total\_sentences}\right) - 84.6\left(\frac{total\_syllables}{total\_words}\right)$$

SMOG grade:

$$1.043\sqrt{polysyllables \times \frac{30}{total\_sentences}} + 3.1291$$

Flesch-Kincaid grade level:

$$0.39\left(\frac{total\_words}{total\_sentences}\right) + 11.8\left(\frac{total\_syllables}{total\_words}\right) - 15.59$$

Coleman-Liau index:

$$0.0588L - 0.296S - 15.8$$

where L is the average number of letters per 100 words and S the average number of sentences per 100 words, *automated readability index:*

$$4.71\left(\frac{total\_chars}{total\_words}\right) + 0.5\left(\frac{total\_words}{total\_sentences}\right) - 21.43$$

Dale-Chall readability score:

$$0.1579\left(\frac{difficult\_words}{total\_words}\right) + 0.0496\left(\frac{total\_words}{total\_sentences}\right)$$

the number of difficult words, Linsear Write formula:

$$\frac{3(complex\_words) + (non\_complex\_words)}{2(total\_sentences)}$$

where complex words are those with more than 3 syllables *Gunning fog scale:*

$$0.4\left(\frac{total\_words}{total\_sentences}\right) + 40\left(\frac{complex\_words}{total\_words}\right)$$

and the *estimated school level to understand the text* that considers all the above tests.

Apart from general, statistical and PC based, the rest of the listed features have not been used in a meta-learning context.

[1] https://github.com/shivam5992/textstat.

Table 1. Meta-features identified as relevant after feature selection. We show the ranked features for each problem, in bold we show the features used for obtaining the results from Table 5.

Hate	Irony	Sentiment	Topics	Author	All 5 TASKS
number of categories	**number of categories**	**dpc min**	**adverbs**	**dpc min**	**number of categories**
dpc min	**dpc kurtosis**	**numerals**	**MRH**	**100pca kewness**	**dpc kurtosis**
Flesch reading ease	**adpositions**	**SMOG**	**pronouns**	**dpc max**	**dpc min**
dpc kurtosis	**wpd average**	**unmarked**	nouns	feature independence NB	**dpc entropy**
zeros in matrix	**Flesch reading ease**	**pca singular sum**	punctuation marks	number of documents	**MRH**
voc skewness	**zeros in matrix**	**pca explained variance**	dpc entropy	pca kurtosis	**adverbs**
dpc entropy	**readability index**	**adpositions**	number of categories	pca explained ratio	**adjectives**
pca explained variance	**Kincaid grade**	**pca max**	scholar grade	dpc entropy	**wpd average**
imbalance degree	**dpc min**	**number of categories**	SMOG	pcac	**Flesch reading ease**
voc kurtosis	**dpc skewness**	**wpd average**	data separability DT	pca explained variance	**pca explained variance**
	Linsear	**Gunning**			**zeros in matrix**
	MRH	**Linsear**			**SMOG**
	pca singular sum	**zeros in matrix**			
	voc kurtosis	**Articles**			
	pca min	**Flesch reading ease**			
		dpc skewness			
		MRH			
		adverbs			
		Coleman Liau			
		number of documents			
		nouns			
		wpd entropy			
		pca explained variance			
		conjuctions			
		dpc entropy			

3.2 Datasets

For the extraction of the meta-features and the experimental evaluation we collected 81 text corpora associated to different problems. We associated each corpus with a task-type-label according to the associated classification problem, where the considered labels were: *authorship analysis, sentiment analysis,* *topic/thematic tasks, irony* and *hate speech detection*. Table 2 illustrates the distribution of the datasets as labeled by their task.

Table 2. Tasks by their type.

Type-task	Frequency	Avg. documents	Avg. classes
Topic	16	93,797(±191,833)	15.81(±20)
Author	13	10,490(±15,790)	12.31(±14)
Irony	7	13,579(±10,372)	2.00(±0)
Sentiment	39	362,660(±905156)	2.95(±2)
Hate	6	14,969(±9134)	2.33(±1)

The full list of datasets is available in the appendix material. Some of the considered datasets are well known benchmarks (e.g. Yelp) while the rest can be found in competition sites like Kaggle and SemEval. After pre-processing each corpus to share the same format and codification, we extracted the 73 meta-features for each of the 81 collections and we assigned a task type label to each dataset according to the associated classification problem. To accelerate the feature extraction process we limited the number of documents to 90,000 for each collection, where these were randomly sampled from the categories of the corpus. The resultant matrix of size 81×73 comprises our *knowledge base* characterizing multiple corpora.

3.3 Meta-learning of Task Labels

We approached the problem of recognizing the classification task of a dataset by using the proposed meta-features. We studied the prediction problem as both multiclass (predicting one of the 5 task labels) and binary (distinguishing one label from the rest at a time) classification problems. The following classifiers were considered for the evaluation: Random Forest (RF), XGBoost (XG), Support Vector Machines and 1NN.

4 Experiments

For the evaluation we adopted a leave-one-out cross-validation: 80 tasks were used for training and 1 for testing, repeating this process 81 times, each time changing the test task; the average performance over the 81 folds is reported. As evaluation measures we report accuracy and f_1 measure for the positive class; in the case of the multiclass problem average accuracy and Macro-f_1 are reported.

Table 3. Task prediction results with 73 meta-features

	Accuracy		f_1	
Task/model	XG	RF	XG	RF
Hate	0.94	0.94	0.29	0.29
Irony	**0.95**	0.93	**0.67**	0.25
Sentiment	**0.89**	0.85	**0.83**	0.77
Topics	0.86	**0.89**	0.62	**0.64**
Author	**0.90**	0.89	**0.60**	0.52
All 5-tasks	**0.77**	0.75	**0.64**	0.59

Table 3 shows the results obtained by the 2-best performing classifiers (XG and RF). Table 4 shows the results of experiments with different classifiers. It can be seen that performance for all of the tasks is greater than random guessing. The high accuracy contrasted by moderate f_1 values reveals the models are favouring the majority class. In fact, high imbalance makes prediction quite difficult, specially for the *hate* and *irony detection* tasks where there are 6 and 7 positive examples, respectively.

Table 4. Task prediction f1 score for different classification models.

	f_1			
Task/model	XG	1NN	SVM	RF
Hate	0.29	**0.36**	0.23	0.29
Irony	**0.67**	0.50	0.35	0.25
Sentiment	**0.83**	0.60	0.39	0.77
Topics	0.62	0.30	0.53	**0.64**
Author	**0.60**	0.38	0.33	0.53
All 5-tasks	**0.64**	0.43	0.32	0.59

An additional experiment involved a feature selection process prior to the classification stage. Mutual information was used to select the top K features and used for training and predicting. Table 5 shows the best performance obtained when performing feature selection together with the number of meta-features selected. It can be observed that there is a performance improvement after the selection of meta-features in all binary cases. Improvements are dramatic in terms of the f_1 measure in some cases (e.g., *Hate, Topics, Author*). Surprisingly, for some problems only few meta-features were required to achieve better performance, see, e.g., *Hate*. For the multiclass problem meta-feature selection did not improve the initial results on either evaluation measures.

Table 5. Results with meta-feature selection

Task	Model	K	Accuracy	f_1
Hate	RF	2	0.94 (+0%)	0.55 (+89.6%)
Irony	XG	15	0.96 (+1%)	0.73 (+8.9%)
Sentiment	XG	24	0.90 (+1%)	0.85 (+2.4%)
Topics	RF	3	0.90 (+1%)	0.75 (+17.1%)
Author	RF	3	0.91 (+1%)	0.70 (+16.6%)
5 tasks	XG	12	0.70 (−7%)	0.64 (+0%)

Table 1 shows the complete subsets of features considered for obtaining the results from Table 5. Meta-features are ordered by their mutual information values. It is hard to find a common pattern but we found that some features are part of almost every subset: the *percentage of adverbs,* the *number of categories,* vocabulary overlapping in classes: *MRH,* and some statistic of *documents per category.* Hence showing the importance of the novel meta-features extracted from raw text. For hate detection and authorship analysis simple statistical measures appear to be better to describe the corpora, for the rest of the tasks the subsets that improved the original performance include a wide variety of meta-features from the groups presented in Sect. 3 (Fig. 1).

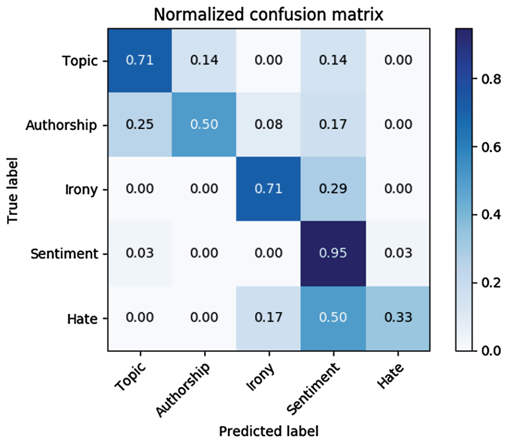

Fig. 1. Normalized confusion matrix of predicting all 5-tasks with XG.

5 Conclusions

We introduced the problem of automatically predicting the type of text classification tasks from meta-features derived from text. A set of 73 meta-features have been proposed and evaluated in 81 data sets associated to 5 types of tasks. Experimental results demonstrate that the proposed meta-features entail discriminative information that could be useful for other meta-learning tasks. Results of a meta-feature selection analysis showed that *traditional* meta-features are not good enough to characterize datasets by themselves, proving the effectiveness of the newly introduced ones. This paper comprises the first steps in trying to meta-learn from raw text directly, we foresee our work will pave the way for the establishment of meta-learning in NLP.

A Appendix

See Tables 6 and 7

Table 6. List of datasets.

Name	Task	# of docs	Voc size	# of classes
20 Newsgroups	Topics	18828	229710	20
Women's reviews	Author	23473	15153	8
Amazon cellphones	Sentiment	999	2241	2
Every song	Author	20779	48752	40
authorship_poetry	Author	200	9141	6
SouthPark episodes	Author	11953	14068	5
Spanish songs	Author	3947	35571	23
Bias Politics	Sentiment	5000	21328	2
Brown	Topics	500	48778	15
Progressive tweets	Topics	1159	5491	4
ccat	Author	1000	20416	10
Classic	Topics	7095	29518	4
Cyber trolls	Hate	20001	21193	2
Davidson hate	Hate	24678	24289	2
BBC News	Topics	2225	33771	5
BBC Summaries	Topics	2225	22921	5
Doctor deception	Sentiment	556	4453	2
Op_spam-	Sentiment	800	8819	2
Op_spam+	Sentiment	800	6548	2
Restauran reviews	Sentiment	400	5353	2
Deflate	Sentiment	11786	25616	5
Gender-microblog	Author	781	2439	2
Gender-twitter	Author	19953	50910	4
Imperium	Hate	6593	28031	2

(*continued*)

Table 6. (*continued*)

Name	Task	# of docs	Voc size	# of classes
Hate tweets	Hate	24783	41639	3
Iro-eduReyes	Irony	20000	32714	2
Iro-humReyes	Irony	19870	30485	2
iro-mohammad	Irony	1929	6040	2
Iro-polReyes	Irony	20000	31882	2
iro-riloff	Irony	2080	6132	2
Iro-semeval18	Irony	4466	10906	2
Kaggle hate	Hate	6594	25646	2
Machado	Topics	246	79461	8
Hate-Malmasi	Hate	7162	14456	3

Table 7. List of datasets.

Name	Task	# of docs	Voc size	# of classes
masc_tagged	Topics	389	43234	20
Medium papers	Topics	185	530	3
Movie reviews	Sentiment	2000	39768	2
polarity	Sentiment	1386	36614	2
Politic	Topics	5000	21328	9
Pros cons	Sentiment	45875	14015	2
Women's clothing	Sentiment	23486	15160	5
rawdata_cric	Author	158	13787	4
rawdata_fin	Author	175	15517	6
rawdata_nfl	Author	97	8940	3
rawdata_travel	Author	172	15560	4
Recommendations	Sentiment	23486	15160	2
Relevance economic news	Sentiment	8000	53162	3
Relevance short news	Sentiment	5007	20111	3
Reuters	Topics	13328	41600	84
Sarcasm Headlines	Irony	26709	25437	2
IMDB short	Sentiment	748	3401	2
Sent-semeval16	Sentiment	30631	36451	3
sent-semevalSA	Sentiment	6999	18042	3
Twitter-airline	Sentiment	14640	18614	3

(*continued*)

Table 7. (*continued*)

Name	Task	# of docs	Voc size	# of classes
Twitter-self-dirve	Sentiment	7156	18017	6
Short yelp	Sentiment	1000	2379	2
Sharktank	Sentiment	706	5175	2
smsspam	Sentiment	310	1610	2
Socialmedia disaster	Sentiment	10860	33768	2
Starter test	Sentiment	10876	33606	3
subjectivity	Sentiment	10000	21001	2
Tripadvisor reviews	Sentiment	17223	32423	5
Sentences polarity	Sentiment	10662	18408	2
Yahoo answers	Sentiment	1459998	180241	10
YouTube	Sentiment	1956	5929	2
Yelp	Sentiment	699998	125757	5
Ag News	Topics	127598	64504	4
Kickstarter	Sentiment	215513	81252	2
News_Categories	Topics	124989	37183	30
Ohsumed	Topics	56984	79479	23
Short Amazon	Sentiment	568454	68831	5
Amazon	Sentiment	3649998	139289	5
sarcasm	Sentiment	1010826	62765	2
Amazon B	Sentiment	3999998	138968	2
Sentiment140	Sentiment	1600000	93115	2
Semeval17	Sentiment	62618	62304	3
Yelp B	Sentiment	597998	113897	2
Sogou news	Topics	509998	42991	5
Dbpedia	Topics	629998	199912	14
Victorian authorship	Author	53678	9977	45
Stanford	Sentiment	25000	95550	2

References

1. Andrychowicz, M., et al.: Learning to learn by gradient descent by gradient descent. In: Advances in Neural Information Processing Systems, pp. 3981–3989 (2016)
2. Canuto, S., Sousa, D.X., Gonçalves, M.A., Rosa, T.C.: A thorough evaluation of distance-based meta-features for automated text classification. IEEE Trans. Knowl. Data Eng. **30**(12), 2242–2256 (2018)
3. Elsken, T., Metzen, J.H., Hutter, F.: Neural architecture search: a survey. arXiv preprint arXiv:1808.05377 (2018)

4. Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., Hutter, F.: Efficient and robust automated machine learning. In: Advances in Neural Information Processing Systems, pp. 2962–2970 (2015)
5. Feurer, M., Springenberg, J.T., Hutter, F.: Initializing bayesian hyperparameter optimization via meta-learning. In: Twenty-Ninth AAAI Conference on Artificial Intelligence (2015)
6. Gomez, J.C., Hoskens, S., Moens, M.F.: Evolutionary learning of meta-rules for text classification. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion, pp. 131–132. ACM (2017)
7. Lam, W., Lai, K.Y.: A meta-learning approach for text categorization. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 303–309. ACM (2001)
8. Lee, S.I., Chatalbashev, V., Vickrey, D., Koller, D.: Learning a meta-level prior for feature relevance from multiple related tasks. In: Proceedings of the 24th International Conference on Machine Learning, pp. 489–496. ACM (2007)
9. Pinto, D.: On clustering and evaluation of narrow domain short-text corpora. Ph.D. UPV (2008)
10. Ravi, S., Larochelle, H.: Optimization as a model for few-shot learning. In: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, 24–26 April 2017, Conference Track Proceedings. OpenReview.net (2017). https://openreview.net/forum?id=rJY0-Kcll
11. Rivolli, A., Garcia, L.P., Soares, C., Vanschoren, J., de Carvalho, A.C.: Towards reproducible empirical research in meta-learning. arXiv preprint arXiv:1808.10406 (2018)
12. Vanschoren, J.: Meta-learning. In: Hutter, F., Kotthoff, L., Vanschoren, J. (eds.) Automated Machine Learning. TSSCML, pp. 35–61. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-05318-5_2. http://automl.org/book
13. Vilalta, R., Drissi, Y.: A perspective view and survey of meta-learning. Artif. Intell. Rev. **18**(2), 77–95 (2002)