Meta-Learning With Landmarking: A Survey

Ashvini Balte
Post Graduate Student
MIT, Poud Road
Kothrud, Pune

Nitin Pise
Associate Professor
MIT, Poud Road
Kothrud, Pune

Parag Kulkarni
Adjunct Professor
College of Engineering
Shivajinagar, Pune

ABSTRACT
Everyday large amount of data is generated. Data mining is a technique that extracts significant information from raw data using different classifiers. Meta-learning is machine learning technique that supports data mining for choosing appropriate classifier. Instead of whole data set, meta-learning extracts features from meta-data. So, it is important to identify the best suitable classifier based on meta-features. Landmarking is uncommon meta-feature extraction technique present in meta-learning. Instead of statistical measurement landmarking tries to determine position of training data in the areas of problem learning by directly measuring the performance of some simple and significant learning algorithms themselves. This paper is a survey of presented approach in landmarking area. A few research challenges are mentioned at the end of this paper.

General Terms
Data mining, Machine learning, Algorithm

Keywords
Meta-learning, Landmarking, Algorithm selection, Classifier, Accuracy

1. INTRODUCTION
Number of algorithm is presented for data mining [1]. Algorithm selection is time consuming task because it involves experimentation with different algorithms and analyzing performance of those algorithms. When some classifier performs well on some data set then we cannot predict that same algorithm is suitable for all data set according to 'No Free Lunch' theorem. So, no universal recommendation of single classifier is possible. Several techniques are presented to select appropriate classifier. Studies show that meta- learning is reliable method that accelerates algorithm selection process by using learning algorithms.

Meta-learning is sub-field of machine learning [1, 2, 3]. The main idea is learning system receives input from meta-data from machine learning experiments and obtains knowledge that is used to predict suitable classifier. In general two phases of meta-learning are present; that is meta-features are extracted from the given datasets and it is used as model building, known as the knowledge acquisition mode or model generation step [1,4,5] and further this knowledge is applied on the meta-feature of the new dataset are known as Advisory mode or testing generated model[1,4,5]. Acquired knowledge also trains meta-learning model. Different papers give the different names for it. For non-expert user meta-learning system can reduces the effort and the development time by choosing appropriate algorithm. [6,7] Different features are presented in meta-learning. These features are divided into several categories [1,11],

- *Simple/general meta-features* [1,8] are directly and easily accessible from the given dataset. They include number of attributes, number of classes and amount of observation, amount of output values and dataset dimensionality.
- *Statistical meta-features* [1,8] Data characterization has done with statistical approach such as liner correlation coefficient, skewness, kurtosis, standard deviation, coefficient of variation, covariance and 1-D variance fraction coefficient.
- *Information theoretic meta-features* [1,8] uses class label and entropy measures of attributes such as normalize class entropy, normalize attribute entropy, joint entropy of class and attribute, mutual information of class and attribute, equivalent number of attributes and signal to noise ratio.
- *Model based meta-features* [1,11] are trained decision tree without pruning. Different properties of this tree are used as feature values such as number of leaves, number of nodes, nodes per attribute, nodes per sample and leaf correlation.
- *PCM meta-features* [16] perform principle component analysis and compute various statistics of principal component. It includes pca skewness, first pc and pca kurtosis first pc.
- *Landmarking meta-features* [1,3,4] are computed by simple and significant machine learning algorithm to characterize properties of dataset. They include one nearest learner, decision node, naïve bayes, linear discriminant, worst node, average node, random node.

Major focus of research has been identifying what kind of meta-feature is suitable for dataset characterization. The central idea is high quality data characteristics provide some information to differentiate the performance of a set of given learning strategies [18]. There are two meta-learning approaches: classification and regression. Classification separates dataset into responsible categories. Regression predicts type of problem. Besides this ranking is third approach. Instead of predicting single classifier it gives the rank of classifier [1,11]. Main aim of this paper is to presents survey of landmarking in meta-learning. Landmarking exploits the performance of very simple algorithms from different classes of learners and uses the accuracy as optimization criteria. Another approach of landmarking consists of using the performance of classifier on small samples of dataset to characterize the problem.

The remainder of this paper is organized as follow. Section 3 describes the Landmarking. Section 4 describe the different approach present using meta-learning by Landmarking for algorithm selection. Section 5 mentions current research challenges and section 6 concludes this paper.

2. LANDMARKING

The general meaning of landmark is boundary or identity of a particular locality. Likewise In meta-learning, landmarking identifies the location of task in their area [9]. Each learner has the category of task on which it has measurable performance under some circumstances. This is known as an area of expertise [9,10]. Learner on task discovers the information about the nature of the task. In short task can be described by the collection of areas of expertise that belongs to landmarker or landmarker learner [9,10]. A learning mechanism is one whose performance is used to describe task. Landmarking uses those learners to locate task in *expertise space* which is the space of all areas of expertise. Landmarking thus view meta-learning as intended to find location of tasks in expertise space. For that purpose landmarking uses Expertise map [9] as its main source of information. Basic idea is to find expertise space of task in area of expertise; then it indirectly uncovers the information about nature of task [9,10,11]. It gives the sign spot of location of problem in space of learner expertise. Landmark run parts of task instead of running full-blown mechanism as compared to cross validation of model selection. For suitable landmark two main criteria are put forward [14, 19]

- Efficiency [14,19]: Landmarker should be inexpensive. If costly computation is required to obtain landmark, then it is better to use brute force approach i.e., directly testing of candidate algorithm on given dataset.
- Bias Diversity [14]: The good landmarker should consist of landmark with different prejudice and both have similar performance measure on all data sets then it would be sufficient to choose any one.

Based on these criteria, typical choice of landmarkers are simple, significant algorithms with a high prejudice.

2.1 Landmarking Representation

Landmarking representation is important because instead of calculating absolute performance, relation in landmark captures landmark's performance relative to one another [11, 14]. Landmarking representation gives the idea of what type of relations are present in landmarking. Those relations are described below [14]:

- Absolute (LM): This is conventional strategy in landmarking, where estimated accuracy of landmark algorithms is directly used.
- Ranks (RK): It is possible that one attribute corresponds to more than one landmark. Such attributes does not encode as the accuracy estimate but it is used as performance rank among competitors.
- Order (OR): Each possible rank contains at least one attribute and its value is the landmark that obtains that rank. For a tie (i.e. more than one algorithm achieves the same rank) they are broken randomly by assigning one algorithm to exactly one of the ordered attribute.
- Pair wise (RL): It enables learner for pair-wise comparison between accuracy of the landmarkers at the meta-level. For each pair of landmarkers such relation returns +1 when first value is bigger, -1 if second value is bigger and for missing value? Otherwise.
- Ratios (RT): This is the generalization of previous one. For all pairs of landmark, the pair wise accuracy ration is encoded. Here difference relative to order of magnitude of the error is calculated.

Obviously, these different types of encoding relation knowledge can be used in combination with each other.

2.2 Landmarker

Here different landmarkers are describing, which is used in meta-learning process [11,13,14] :

- Average Node Learner: Calculates the average accuracy of single node decision tree where each node relates to one value.
- Best Node Learner or Decision Node Learner: Based on information gain ratio, it shows how informative is an attribute with respect to classification task using its entropy. It chooses attribute which have highest information gain.
- Worst Node Learner: In this the information gain criteria uses the attribute which represents lowest selected value.
- One Nearest Node Learner: This landmark learner classifies how near the test point that belongs to same class are.
- Elite 1-Nearest Neighbor Learner: This subset is comprised of the most informative attributes, if information gain ratio among attribute is smaller than 0.1. Otherwise the elite subset is singleton and learner acts like a decision node learner.
- Randomly Choose Node Learner: This result is based on randomly choose attribute. This node is used to split the training set and classifies given test examples.
- Naïve Bayes Learner: Training set uses bayes theorem to classify test cases. Bayes theorem:

  $p(x/y)= p(y/x)*p(x)/p(y)$

  Where $p(X)$ is prior probability and $p(X/Y)$ is posterior probability.

The landmark learner measures conditionally the independence of attributes for the given class. This landmarker is used to estimate accuracies of landmarkers.

3. CURRENT ONGOING WORK ON LANDMARKING

The research work in the area of landmarking is presented in pattern recognition by either improving the accuracies or by choosing the appropriate landmark. The second approach is relating to landmark in different areas. In this section, first explained the work has been done in landmarking.
The landmarking approach investigates simple and significant computable classifier. These classifiers are applied on the dataset and the resulting performance values are used as meta-features of the dataset.

The landmarking performance is compared with the information that is theoretic based and directly predicts the best classifier [9]. The result shows how landmark performance is better than the information theory for predicting accuracy of classifier in expertise space. Artificial dataset is used [10] for training or creating decision tree and UCI repository dataset is used for testing. The results are compared with information theoretic approach and combination of both. Using this technique authors predict the best classifier out of pair of classifiers as above landmarking performance is measured [13] for locating tasks. Landmarking also discovers the neighborhood and complete map of expertise space.

Two types of landmarking techniques [14] evaluate the relative and sub-sampling landmarkers. In relative landmark estimate five different representations of landmarking which are absolute, pairwise, ranks, order and ratio. Comparisons between these landmarkers are investigated. Besides these sub-sampling landmarkers are use the performance of the target classifier on a sub-sampled dataset as meta-feature.

All meta-features are considered [11] and suitably predict accuracies of classifier based on regression learner. Evolution of meta-feature for meta-learning using regression was performed based on 54 datasets. Five different categories of meta-feature namely simple, statistical, information theoretic, model based and landmarking were used and comparatively evaluated. For selecting meta-feature the performance of meta-learner root mean squared error and the person product-moment correlation coefficient is evaluated. A software wizard [12] for classifier recommendation using regression and landmarking feature was developed and integrated into RapidMiner. This meta-learning tool shows improvements in development time and achieves a classification accuracy compared to standard pattern reorganization approach.

New approach DecT [8] for dataset characterization in meta-learning is proposed. Data sets are processed using standard tree induction algorithm. Further it captures information regarding data characterization. The DecT is compared with DCT and landmarking both, but results are not better than DecT. Framework for automatic classifier selection [18] is presented. The system continuously improves this knowledge by storing the result of tests run on the datasets. It consist the combination of landmarking with data characterization. It increases the accuracy of the prediction.

Instead of ad-hoc landmark new version of landmark i.e. hill climbing landmark is proposed in [19]. The performance of traditional landmark is associated with single algorithm at low computational complexity but hill climbing landmark is on correlation of algorithm and efficient. The results gives improvement in efficiency gained rank order correlation and algorithm pair ordering.

Rather than single instance of classifier, performance is measured on multi instance classifier [20] in which it uses two cost effective approaches. The first approach is single instance classifier. It is inexpensive approach and second is multi instance classifier which is expensive approach. The result gives that landmark has best predictive performance for domain of data set and multi-instance learner are compared. Following table describes in detail the current ongoing work.

Table 1. Literature Survey

Title	Year	Advantage/ Methods	Limitation
Meta learning a Survey Trends and Technologies[4]	2013	-Overview of research direction of meta-learning	-Experimental result is not checked
A Survey and Current Research Challenges in Meta-learning Approaches Based on-Dataset Characteristics [5]	2012	-Various approach of meta-learning -ARR for Ranking	-Few data characteristics are used
On Learning Algorithm Selection for Classification[7]	2004	-Rule based classifier selection	-Limited to C5.0, NN, SVM -It suggest to consider more classifier such as IBK, PART, KD, NB, OneR
Improved Dataset Characterisation for Meta-learning[8]	2002	-DCT is used for data characterization. -Rank the learning algorithm based on accuracy and time	-Use Only C5 tree algorithm -K-NN in Zoom ranking is also useful
Meta-learning by Landmarking Various Learning Algorithm[10]	2000	-Root Mean Square Error -Compare landmark performance with information theoretic	-Real data set are for testing not for training
Automatic Classifier Selection for Non-expert [11]	2012	-Meta-learning using regression. -Root Mean Squared Error	-A level of expertise required for building pattern recognition system
Discovering Task Neighbourhoods Through Landmark Learning Performance[13]	2000	-Performance measure with the expertise space	-Limited to expertise space
An Evaluation of Landmarking Variant[14]	2001	-Pair wise comparison -Simple Rule Predict accurately in Boost decision tree	-Most results are negative

Using Meta-Learning to Initialize Bayesian Optimization of Hyper parameter[16]	2014	-Sequential model based Optimization (SMBO) approach is used with Meta-learning based approach (MI-SMBO). -MI-SMAC	-Meta-learning is not integrated with SMBO. -Evaluated MI-SMAC on small configure space
A Hill Climbing Landmarker Generation Algorithm Based on Efficiency and Correlativity criteria[19]	2005	-Hill Climbing landmark based on the forward selection method for variable selection in regression -Require Polynomial training time complexity	-Hill climbing landmarker generation algorithm not fully matched with predictive performance for all-subsets version
Auto-WEKA Combined Selection and Hyperparameter Optimization of Classification Algorithm[23]	2013	- Auto WEKA based on bayesion optimization method	- Need large improvements in cross-validation performance
MetaStream: A Meta-learning Based Method for Periodic Algorithm Selection in Time Changing Data[24]	2014	-Metastream method used for time changing data	- Need to find out meta-features related to time series analysis

4. TOOLS FOR LANDMARKING

There are mainly two popular tools used for landmarking task description. These tools are open source and developed in java programming language. It minimizes effort of user to use machine learning algorithm for data mining task.

1. WEKA [21]: WEKA (Waikato Environment for Knowledge Analysis) developed at the University of Waikato; New Zealand is open source software under General Public License (GNU). WEKA is popular in machine learning. It is collection of machine learning algorithms for data mining. It is written in JAVA and it becomes the platform independent. Tools are present in WEKA for classification, regression, data-pre-processing and clustering, visualization and association rules. It is well suitable for developing well suited schemes for machine leaning. Classifiers present in WEKA are bayes, functions, lazy, meta, misc, rules and trees. Package manager allows installing different packages for plug-ins, classifier and clustering are available. The RapidMiner is also fully integrated with WEKA.

2. RapidMiner [22]: the old formal name is YEAL (Yet Another Learning Environment). It is open source machine learning tool written in java. Rapid miner is distributed under AGPL open source license. The RapidMiner integrated environment is for data mining, text mining, machine learning, business analytics and predictive analytics. The plug-in of RapidMiner Rapid-I and GambH are popular in machine learning. In general RapidMiner is client server module; it provides software as service in cloud environment. RapidMiner provides a GUI to design and execute analytical workflows known as process and they consist of multiple “Operators”. Each operator is performing a single task within the process and the output of each operator forms the input of the next one. The new operator Landmarking [12] is developed in RapidMiner.

5. DISCUSSION

Landmarking is interesting approach in meta-learning. Though much work is done in this area there are also many open challenges for further research. Much of the problem use artificially generated data sets [9,10], there are many more real data sets are present then it is necessary to test landmark on it. In previous work, few meta-learners are tested for classifiers. This can be improved by the selecting more meta-learners. The learning space of landmark indirectly discovers the task information by characterizing the problem. So often it concentrates on particular area of expertise. So, this gives the advantage on instead of processing the whole data set it slowly concentrates on cartography area. Landmarking is method which directly calculates this area on expertise space. This improves the accuracy of classifier and avoids the unusable pruning of dataset in learning space. This requires less computation time and accuracy remains as it is.

Furthermore, it necessary to generate more artificial datasets which gives more problem space, it then becomes challenging for the landmarking. Multi-instance learner landmarks are more expensive. This can be reduced by extracting the appropriate features which give accuracy. Most papers present the ad-hoc approach. This can be improved by improving the landmarkers in learning space.

6. CURRENT RESEARCH CHALLENGES

1. In most of the papers the landmark meta-features are compared with information theoretic for predicting performance of suitable classifier. In future work it is possible it to compare with the model based or simple or statistical or combination of all this meta-features.

2. Most papers give the pairwise comparisons which are most suitable for their problem definition. It may be possible that other landmarker (e.g. absolute, ranks, order, pairwise, ratios) are more suitable in different problems. Those cases are needed to be identified.

3. In RapidMiner new operator landmarking are developed, likewise it is possible that new operator in WEKA can be developable.

4. Most cases meta-learning algorithms Ripper, C5.0, C4.0 and Ltree are used. Landmarking performance can be measured using more than these algorithms.

5. In most of the papers, there are no automatic features that are selected. Only limited features are selected. It is possible to implement the automatic feature selection based on classifier accuracies, for generalizing recommendation system.

7. CONCLUSION

The intention of this paper was to present review of current work related to Landmarking and finding the future scope of the landmarking in Meta-Learning. Landmarking is data characterization technique present in Meta-Learning. Most research proves that data characterization plays an important role during selection of suitable algorithm. This paper is review of different techniques which are used for extracting landmarking meta-features. As result of our survey, the comparative analysis of present landmarking techniques is given in this paper. Current research challenges are listed in this paper. This analysis is useful for future research related to landmarking.

8. REFERENCES

[1] Vilalta, R, Giraud-carrier, C, Brazdil, P, Soares, C. 2004. Using meta-learning to support data mining, International Journal of Computer Science and Applications, 1(1), 31–45.

[2] Smith-Miles, K. A. 2008. Cross-disciplinary perspectives on meta-learning for algorithm selection", ACM Computing Surveys, 41(1), 1–25.

[3] Vilalta, R, Drissi, Y. 2002. A perspective view and survey of meta-learning, Journal of Artificial Intelligence Review, 18(2):77–95.

[4] Lemke, C, Budka, M, Gabrys, B. 2013. Metalearning: a survey of trends and technologies, Artificial Intelligence Review, Springer Netherlands1-14.

[5] Bhatt, N, Thakkar, A, Ganatra, A. 2012. A survey and current research challenges in meta-learning Approaches based on dataset characterisstic, International Journal of soft computing and Engineering, Volume-2, Issue-1.

[6] Brazdil, P, Soares, C, da Costa J.P. 2003. Ranking learning algorithms: Using IBL and meta-learning on accuracy and time results, Machine Learning 50(3), 251-277.

[7] Ali, S, Smith, K. A. 2006. On learning algorithm selection for classification, Applied Soft Computing 6, 119-138.

[8] Peng, Y, Flach, P, Soares, C, Brazdil, P.2002. Improved dataset characterization for meta-learning, S. Lange, K. Satoh, C. Smith (eds.) Discovery Science, Lecture Notes in Computer Science, vol. 2534, pp. 193-208. Springer Berlin / Heidelberg.

[9] Bensusan, H. Giraud-Carrier, C. 2000. Casa batlo is in passeig de gracia or how landmark performances can describe tasks, Proceedings of the ECML-00 Workshop on Meta-Learning: Building Automatic Advice Strategies for Model Selection and Method Com-bination, pp. 29-46.

[10] Pfahringer, B, Bensusan, H. Giraud-Carrier, C. 2000. Meta-learning by landmarking various learning algorithms, In Proceedings of the Seventeenth International Conference on Machine Learning, pp. 743-750. Morgan Kaufmann.

[11] Reif, M, Shafait, F, Goldstein, M, Breuel, T, Dengel, A. 2012. Automatic classifier selection for non-experts. Pattern Analysis and Applications pp. 1–14

[12] Abdelmessih, S. D, Shafait, F, Reif, M, Goldstein M. 2010. Landmarking for meta-learning using RapidMiner, RapidMiner Community Meeting and Conference.

[13] Bensusan, H, Giraud-Carrier, C. 2000. Discovering task neighbourhoods through land-mark learning performances, PKDD 00: Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery, pp. 325-330. Springer-Verlag, London, UK.

[14] Furnkranz, J, Petrak, J. 2001 An evaluation of landmarking variants, In: C. Giraud-Carrier, N. Lavrac, S. Moyle, B. Kavsek (eds.) Proceedings of the ECML/PKDD Workshop on Integrating Aspects of Data Mining, Decision Support and Meta-Learning (IDDM-2001), pp. 57-68. Freiburg, Germany.

[15] Bensusan, H, Kalousis, A. 2001. Estimating the predictive accuracy of a classier, in: L. De Raedt, P. Flach (eds.) Machine Learning: ECML 2001, Lecture Notes in Computer Science, vol. 2167, pp. 25-36. Springer Berlin / Heidelberg,

[16] Feurer, M, Springenberg, J. T, Hutter F. 2014. Using Meta-Learning to Initialize Bayesian Optimization of Hyperparameters, In: Proceedings 21st European Conference on Artificial intelligence 2014, workshop on Meta-Learning and Algorithm Selection pp. 3-10.

[17] Asuncion, A, Newman, D. 2007.D: UCI machine earninRepository http://www.ics.uci.edu/~mlearn/MLRepository.html University of California, Irvine, School of Information and Computer Sciences.

[18] Cacoveanu, S, Vidrighin C, Potolea R. 2009, Evolution Meta-Learning Framework For automatic Classifier Selection, IEEE 5th International Conference on Intelligent Computer Communation and Peocessing (ICCP).

[19] Ler, D, Koprinska, I, and Chawla, S. (2005). A hillclimbing landmarker generation algorithm based on efficiency and correlativity criteria. Proceedings of the 18th International FLAIRS Conference, Machine Learning Track.

[20] Vanwinckelen, G, Blockeel, H. 2014. A meta-learning system for multi-instance classification, in:Proceedings of the ECML-14 Workshop on Learning over Multiple Contexts, 1-14.

[21] Hall, M, Frank, E, Holmes, G,. Pfahringer, B, Reutemann P, Witten I.H.2009. The WEKA Data Mining Software: An Update, SIGKDD Explorations, Volume 11, Issue 1.

[22] Mierswa, I, Wurst, M, Klinkenberg, R, Scholz, M, Euler, T, August 2006. Yale: Rapid prototyping for complex data mining tasks, In Lyle Ungar, Mark Craven, Dimitrios Gunopulos, and Tina Eliassi-Rad, editors, KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 935–940, New York, NY, USA.

[23] Thornton, C, Hutter, F, Hoos, H, 2013. Auto-WEKA Combined Selection and Hyperparameter Optimization of Classification Algorithm, KDD 2013 Chicago, Illinois, USA. ACM.

[24] Rossi, A, Carvalho, A, Soares, C, Souza, B, 2014. MetaStream: A Meta-learning Based Method for Periodic Algorithm Selection in Time Changing Data, Neurocomputing, 127: 52-64.