ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1. Available from http://www.i6doc.com/en/.

Advanced Query Strategies for Active Learning with Extreme Learning Machine

Anton Akusok[1], Emil Eirola[1], Yoan Miche[2], Andrey Gritsenko[3] and Amaury Lendasse[3,4]

1- Arcada University of Applied Sciences - Department of Business Management and Analytics Jan-Magnus Janssonin aukio 1 - Finland

2- Nokia Solutions and Networks Group - Espoo - Finland

3- The University of Iowa - Department on Mechanical and Industrial Engineering Seamans Center for the Engineering Arts and Sciences - USA

4- The University of Iowa - The Iowa Informatics Initiative Seamans Center for the Engineering Arts and Sciences - USA

Abstract.
This work proposes three new query strategies for active learning. They are built on modern developments in Extreme Learning Machine (ELM): class-weighted ELM, prediction intervals with ELM, and mislabeled samples detection with ELM. Both ELM and active learning are important methods of applied machine learning. Combined, they offer a fast and precise tool for practical data acquisition in classification tasks where raw data is cheap but labels are expensive to get. Some proposed methods rival the state-of-the-art in performance and speed, based on testing with three real world datasets.

1 Introduction

An Extreme Learning Machine [1] (ELM) is a fast [2] and flexible machine learning method with a non-iterative solution that provides good scalability as well as theoretical performance guarantees [3]. Regularized versions of ELM avoid over-fitting problems [4] and are widely used as tools to address practical challenges [5, 6].

The ELM ecosystem is highly focused on solving real-world problems, not only on theoretical developments. Data acquisition is an integral part of the solution to such problems, in the form of getting additional data, labeling more samples, or revising previous labels. This is especially true for applied research [7], where a machine learning expert shares his own work time between research, implementation and data acquisition. Extreme Learning Machine save time on parameter tuning in research as the method is mostly parameter-free, and save efforts for code optimization as it is efficient to implement.

This work considers ELM in the last part of the whole problem solution – data acquisition and labeling. Data related problems are covered by a broad field of Active Learning[1] [8]. Existing active learning literature with ELM appears only recently [9], and mostly considers simple methods like committee of

[1] `http://active-learning.net`

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

models [10]. The goal of this paper is to compare the state-of-the-art pool-based active learning methods [11] with various ELM-based sampling algorithms derived from weighted ELM [12], mislabeled samples detection [13], and prediction intervals for ELM [14].

2 Active Learning Query Strategies

Active learning is querying (expensive) labels for selected (cheap) unlabeled samples from a large pool or stream of such, in a way that allows a classification method to achieve the target accuracy with much less labeled training data. A structured overview of active learning methods is available in a recent survey [8].

The reference query strategies include three state-of-the-art methods as Querying Informative and Representative Examples (QUIRE) [15], SVM with Hint information (HintSVM) [16] and Active Learning By Learning (ALBL) [11]. Two baseline reference methods are random sampling, and uncertainty sampling that queries samples with least confident predicted labels from a committee of models.

3 Methodology

An Extreme Learning Machine has an easy and compact mathematical formulation. The solution of ELM minimizing $L2$-norm of error is formulated as:

$$\boldsymbol{\beta} = \boldsymbol{H}^{\dagger}\boldsymbol{Y} \quad \text{where} \quad \boldsymbol{H} = f(\boldsymbol{X}\boldsymbol{W}) \tag{1}$$

Here $\boldsymbol{X}$ is an input data matrix including a dummy column of $+1$ for bias, $\boldsymbol{W}$ is a random projection matrix, $\boldsymbol{Y}$ is a matrix of training outputs, $(\cdot)^{\dagger}$ denotes matrix pseudo-inverse, and $f(\cdot)$ is a non-linear activation function applied element-wise. Number of hidden neurons in ELM is given by the number of columns in $\boldsymbol{W}$.

3.1 Committee of Class-Weighted ELMs

This query strategy (denoted as *W-ELM*) uses uncertainty sampling framework, querying label for the sample that has the highest entropy of outputs from a committee of models. The committee is build from ELM models that are fast to re-train as more data becomes available. Additionally, each ELM model utilizes different class weights [12] that normally counter class imbalance in training set. Here the weights allow for a more diverse set of predicted labels by a model committee even with very little training data available. This increasing variation in prediction entropy between unlabeled samples and helps to select the most uncertain one.

3.2 Prediction Interval Size

This recent approach (denoted as *ELM PI*) for estimating individual prediction interval sizes for test data samples [14] is directly applicable to active learning. An algorithm may query samples with the largest prediction interval, that

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

correspond to most uncertain data space areas. This is a pure "exploration" strategy [8], but a rather simple one, and it also benefits from a fast training speed of ELM.

3.3 Mislabel Detection Score

An ELM-based algorithm (denoted as *MD-ELM*) that can automatically detect possible mislabeled samples in a dataset by randomly changing a few labels and re-estimating error, increasing scores of samples such that changing their label results in a lower error. ELM provides a precise non-linear classification error estimator, as well as a fast way of computing Leave-One-Out error [13].

Samples with the highest mislabel score can be candidates to query. Each single unlabeled sample is added to the training dataset, and gets a random class label on every iteration of an algorithm. A high score associated with a particular label means that adding that sample with this label to the dataset will improve the global accuracy.

4 Experiments

Performance of the proposed methods is evaluated on three real-world binary classification datasets from LibSVM repository[2]: Pima Indians *Diabetes* (768 samples with 8 features), *Australian* credit approval (690 samples with 14 features) and *Heart* disease (270 samples with 13 features). All features are scaled to be within $[-1, 1]$ range.

4.1 Experimental Setup

The datasets are randomly split into 2/3 training and 1/3 test parts, then all the methods are trained on the same training set and evaluated on the same test set; this repeated 100 times and average results reported. Training set starts with only 10 labeled samples, additional labels are provided to the queried samples from the remaining unlabeled training pool. The accuracy of all query strategies is evaluated by the same linear SVM model [17] for comparable results. Only the order of adding labeled training samples differ between the methods.

All ELM models use 8 neurons as more will lead to extreme over-fitting with the initial 10 labeled samples without aggressive L2-regularization. Activation function is the hyperbolic tangent. Committee of class-weighted ELMs has 7 models with the following class weights: $\{5,1\}$, $\{3,1\}$, $\{2,1\}$, $\{1,1\}$, $\{1,2\}$, $\{1,3\}$, $\{1,5\}$. MD-ELM evaluates every unlabeled training sample on 6 different ELM models with 10 iterations each. This is the slowest method as it has to get sufficient statistics for every unlabeled sample in the pool. All results are presented on Figure 1.

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

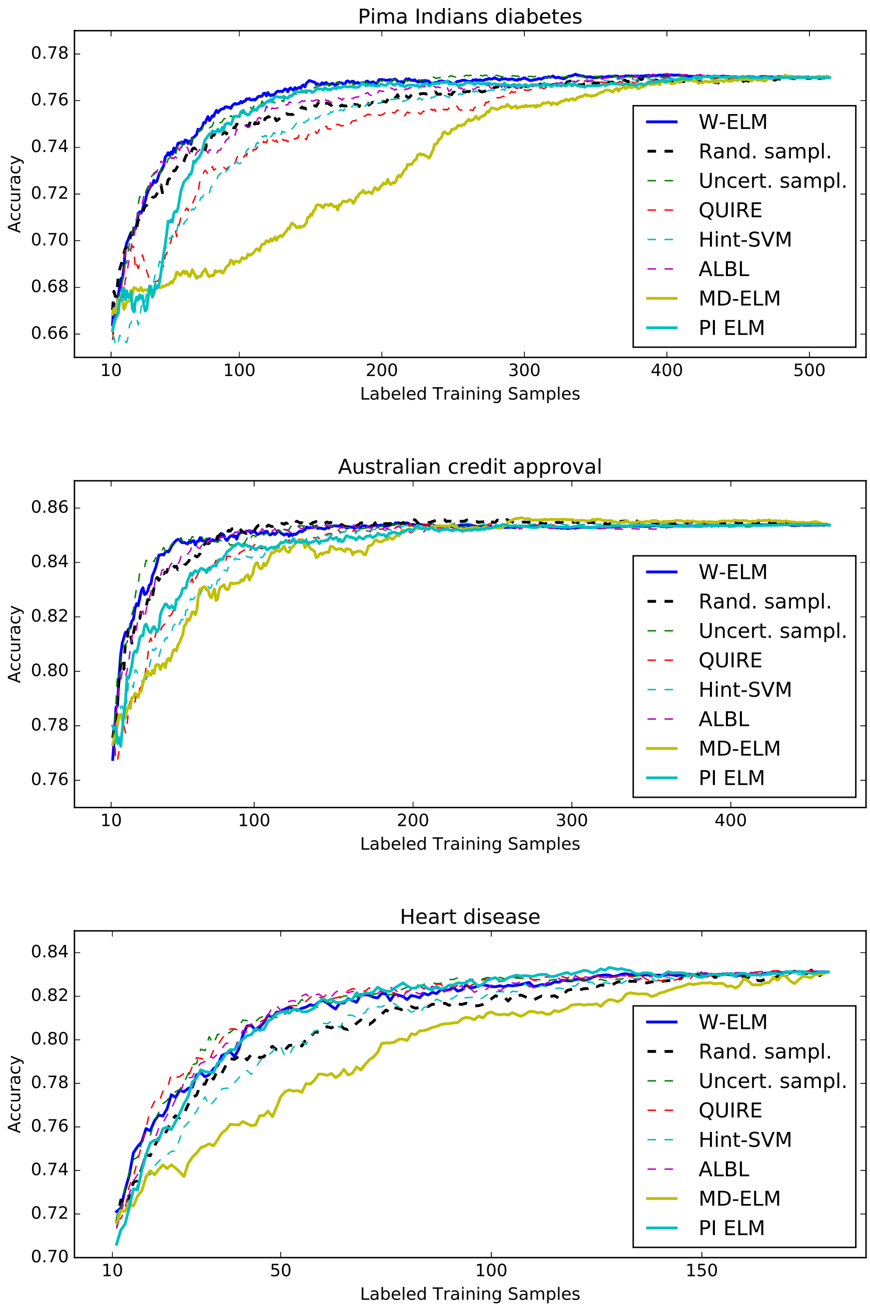

Fig. 1: Average accuracy over 100 runs for Pima Indians diabetes (*top*), Australian credit approval (*middle*) and Heart disease (*bottom*) datasets using active learning with different query strategies.

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1. Available from http://www.i6doc.com/en/.

4.2 Results

In the Pima Indians diabetes dataset (Figure 1, *top*) a near-optimal result is achieved using only 150 out of 512 training samples. Uncertainty based sampling methods perform the best, disregard the underlying model (W-ELM or linear SVM). They are followed by the Prediction Interval-based querying with ELM, which can be considered as another view of an uncertainty-based method. Other methods except MD-ELM perform similar to random sampling.

In the Australian credit approval dataset (Figure 1, *middle*) uncertainty based querying again shows good results, reaching a good accuracy with only 50 out of 460 training samples. Surprisingly, MD-ELM method reaches the best accuracy at around 60%-70% of the training data, even surpassing the accuracy with the whole dataset. This may be explained by some incorrectly labeled data, that stays filtered out by MD-ELM.

The Heart disease dataset (Figure 1, *bottom*) has less data samples than the other two, and the performance keeps increasing constantly as more labeled data becomes available. Here W-ELM performs similarly to other methods like the state-of-the-art QUIRE, and only MD-ELM under-performs at the whole data range.

Considering runtime, W-ELM is the fastest method (excluding random sampling that only generates a random number). Other fast methods are uncertainty-based sampling, Hint SVM and prediction intervals ELM, which are 2-3 times slower. QUIRE and ALBL are more than a hundred times slower than W-ELM, and MD-ELM is the slowest at >300 times the runtime of W-ELM.

5 Conclusions

The paper proposes three novel query strategies for active learning based on recent developments in Extreme Learning Machine, comparing them on three real world datasets. The proposed committee of differently class-weighted ELM performs at the state-of-the-art level in accuracy and in a fast runtime. The second strategy of prediction interval-based sampling is slightly inferior. The last method of mislabeled detection-based sampling may have a specific usage of avoiding incorrectly labeled samples in the dataset, but it suffers from a slow runtime and below-random performance with small amount of labeled data.

Further research will be focused on incremental active learning with ELM. Recent advances include a fast batch inclusion and even removal of training samples in the existing model. That will allow for an automatic data processing and re-evaluation of samples mislabeled by the automatic process. Also speedup considerations will be explored in a computationally heavy applications like an image processing with the sliding window.

[2] `https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html`

References

[1] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: Theory and applications. *Neural Networks Selected Papers from the 7th Brazilian Symposium on Neural Networks (SBRN '04)7th Brazilian Symposium on Neural Networks*, 70(1–3):489–501, December 2006.

[2] Anton Akusok, Kaj-Mikael Björk, Yoan Miche, and Amaury Lendasse. High-Performance Extreme Learning Machines: A Complete Toolbox for Big Data Applications. *IEEE Access*, 3:1011–1025, July 2015.

[3] Guang-Bin Huang, Lei Chen, and Chee-Kheong Siew. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Transactions on Neural Networks*, 17(4):879–892, July 2006.

[4] Federico Montesino Pouzols and Amaury Lendasse. Evolving fuzzy optimally pruned extreme learning machine for regression problems. *Evolving Systems*, 1:43–58, 2010.

[5] Paul Merlin, Antti Sorjamaa, Bertrand Maillet, and Amaury Lendasse. X-SOM and L-SOM: A double classification approach for missing value imputation. *Advances in Computational Intelligence and Learning, ESANN 2009*, 73(7–9):1103–1108, March 2010.

[6] Rui Nian, Bo He, Bing Zheng, Mark van Heeswijk, Qi Yu, Yoan Miche, and Amaury Lendasse. Extreme learning machine towards dynamic model hypothesis in fish ethology research. *Neurocomputing*, 128:273–284, March 2014.

[7] A. Lendasse, M. Cottrell, V. Wertz, and M. Verleysen. Prediction of electric load using Kohonen maps - Application to the Polish electricity consumption. In *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*, volume 5, pages 3684–3689 vol.5, 8-10 May 2002.

[8] Burr Settles. Active Learning Literature Survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.

[9] Yong Zhang and Meng Joo Er. Sequential active learning using meta-cognitive extreme learning machine. *Neurocomputing*, 173, Part 3:835–844, January 2016.

[10] Hualong Yu, Changyin Sun, Wankou Yang, Xibei Yang, and Xin Zuo. AL-ELM: One uncertainty-based active learning algorithm using extreme learning machine. *Neurocomputing*, 166:140–150, October 2015.

[11] Wei-Ning Hsu and Hsuan-Tien Lin. Active Learning by Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 2659–2665, January 2015.

[12] Anton Akusok, Yoan Miche, Jozsef Hegedus, Rui Nian, and Amaury Lendasse. A Two-Stage Methodology Using K-NN and False-Positive Minimizing ELM for Nominal Data Classification. *Cognitive Computation*, 6(3):432–445, March 2014.

[13] Anton Akusok, David Veganzones, Yoan Miche, Eric Séverin, and Amaury Lendasse. Finding Originally Mislabels with MD-ELM. In *Proceedings of ESANN2014*, pages 689–694. d-side publi., Bruges, Belgium, 23-25 April 2014.

[14] Anton Akusok, Yoan Miche, Kaj-Mikael Björk, and Amaury Lendasse. Per-sample Prediction Intervals for Extreme Learning Machines. In *Proceedings of ELM-2016*. Singapore, Accepted 2016.

[15] Sheng-jun Huang, Rong Jin, and Zhi-hua Zhou. Active Learning by Querying Informative and Representative Examples. In J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 892–900. Curran Associates, Inc., 2010.

[16] Chun-Liang Li, Chun-Sung Ferng, and Hsuan-Tien Lin. Active Learning Using Hint Information. *Neural Computation*, 27(8):1738–1765, June 2015.

[17] Amaury Lendasse, Yongnan Ji, Nima Reyhani, and Michel Verleysen. LS-SVM Hyperparameter Selection with a Nonparametric Noise Estimator. In Włodzisław Duch, Janusz Kacprzyk, Erkki Oja, and Sławomir Zadrożny, editors, *Artificial Neural Networks: Formal Models and Their Applications*, pages 625–630. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.