*ECAI 2014*
*T. Schaub et al. (Eds.)*


*doi:10.3233/978-1-61499-419-0-989*

# Adaptive Active Learning as a Multi-armed Bandit Problem

**Wojciech M. Czarnecki**[1] and **Igor T. Podolak**[2]

**Abstract.** In this paper, we present a new active learning strategy whose main focus is to have the ability to adapt to the unknown (or changing) learning scenario. We introduce the learners' ensemble based approach and model it as the multi-armed bandit problem. Presented application of simple exploration-exploitation trade-off algorithms from the UCB and EXP3 families show an improvement over using the classical strategies. Evaluation on data from UCI database compare three different selection algorithms. In our tests, presented method shows promising results.

## 1 INTRODUCTION

Classical supervised machine learning methods require big labeled datasets to construct good models. Unfortunately, in real life applications it is often the case that, while large amounts of data may be available, a large portion of them misses the true labeling. Obtaining such information commonly requires substantial amounts of time/costs (like labeling video recordings, tagging text corpora or synthesis and testing of the new kind of drug). *Active learning* [7] addresses this issue by introducing the label querying step into the model's training process to minimize the total cost of building the most accurate classifier. In short learning algorithm selects samples to be labeled in such a way, that it can learn as fast as possible.

In this paper we analyze an approach to the active learniners ensembles using multi-armed bandit approach. To the author's best knowledge it is a first approach to deal with ensemble of learners' strategies, not ensemble of models [8].

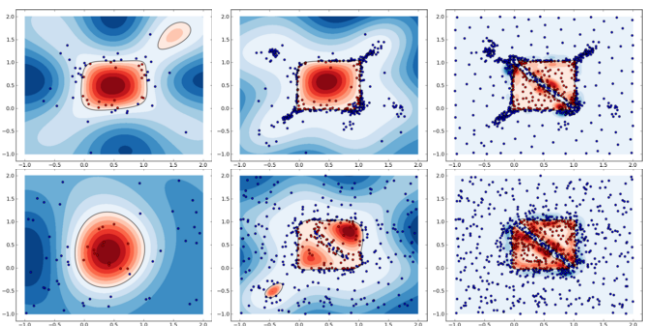

**Figure 1.** Greedy uncertainty sampler (top row) and A$^2$L strategy (bottom row) after 50,400 and 900 iterations.

[1] Faculty of Mathematics and Computer Science, Jagiellonian University, Krakow, email: wojciech.czarnecki@uj.edu.pl
[2] Faculty of Mathematics and Computer Science, Jagiellonian University, Krakow, email: igor.podolak@uj.edu.pl

## 2 PROPOSED METHOD

Mutli-armed bandit problem (MAB) has been introduced by Robbins [6] and since then has been extensively used in problems where one has to balance between the exploration of the state space and exploitation of the current knowledge. The model consists of $K$ one-armed bandit machines, on which we can "play" and get some kind of numerical reward. The aim of this game is to maximize the summarized profit from "plays" without prior knowledge about the behavior of particular machines.

Lets think about our active learners as bandit machines, on which we play, and as a result gain some new sample to label. The aim of our process is to balance the exploration-exploitation trade-off in order to maximize the usefulness of available strategies. If we assume the Markov property (each query is independent on the previous ones) of this process, we can use this model as a base of our method.

Assuming, that we have some learner utility function $v_t : AL \to \mathbb{R}$ in the iteration $t$ we choose the learner which maximizes it:

$$a^* = \arg\max_{a \in AL} v_t(a),$$

which may be seen as the analogy of the active learning strategies that, with given utility function $u_a : X \to \mathbb{R}$, select the most promising points for querying.

The main aim of MAB algorithms is to minimize the strategy regret defined as difference between the expected value of the best strategy reward and expected value of its reward. As a consequence, careful definition of the reward function is crucial for modeling any problem as MAB. In the simplest case of active learning we want to maximize the information gained from our labeled samples with the minimization of their number. Unfortunately, after obtaining some particular labeling, this is impossible to evaluate the information gained with respect to the whole training process (as it is still under way). As a result we have to select some heuristic which should lead to the development of a good learning curve.

Our main idea is to reward the strategy which leads to the biggest changes in our model's beliefs. We define reward as the difference between models prediction and the true labeling

$$r(x, y) = |m_T(x) - y|$$

where $m_T(x)$ is the model's prediction trained on the set $T$; this is the most simple (from both theoretical and computational) point of view, and has a clear interpretation as the measurement of how surprising is the particular strategy,

Fig. 1 compares our method with simple uncertainty sampling. It is easy to notice that thanks to the proposed approach selects samples in a more balanced way, trying to both exploit the labels uncertainty

**Algorithm 1** A$^2$L UCB Adaptive Active Learner

**Input:** data $U$, learners $AL$, initial training set $T$
$r[a] \leftarrow$ prior of $a, n[a] \leftarrow 1, \forall a \in AL$
$n \leftarrow |AL|$
**train model on** $T$
**repeat**
  $v[a] \leftarrow \frac{r[a]}{n[a]} + C\sqrt{\frac{2\ln(n)}{n[a]}}$
  $a^* \leftarrow \arg\max_{a \in AL} v[a]$
  **select** $x^*$ **selected by** $a^*$
  $y^* \leftarrow$ **label of** $x^*$
  $r[a^*] \leftarrow r[a^*] + r(x^*, y^*)$
  $n[a^*] \leftarrow n[a^*] + 1$
  $n \leftarrow n + 1$
  $U \leftarrow U \setminus \{x^*\}$
  $T \leftarrow T \cup \{(x^*, y^*)\}$
  **retrain model on** $T$
**until** end of resources or training data

and explore the whole input space. Compare the exploration of the space after 400th iteration, when uncertainty sampler has almost no knowledge about the space beyond the middle rectangle, while our method already explored the "outer" region and discovered that the rectangle is composed of two triangles.

## 3 EVALUATION

We used SVM with RBF kernel implemented using `scikit-learn` library [5] in Python. At each iteration, model fitted best parameters based on the 5-fold cross validation technique on the labeled data using grid search ($\log_{10}(C) \in [-5, 10]$, $\log_{10}(\gamma) \in [-15, 5]$). Evaluation metric used for both parameters selection and further analysis is the Mathew's Correlation Coefficient (MCC) due to its good statistical properties and ability to deal with skewed classes' distributions. We also used the class weighting technique in order to deal with unbalanced training data. All features were linearly scaled to fixed intervals to overcome model biasing. Simulation starts with 5 positive and 5 negative points from a given dataset and the pool of unlabeled samples. Active learner is supposed to choose one unlabeled sample per iteration whose label is reveled and added to the training set. We compare ensemble methods consisting two basic strategies. One selects the most uncertain sample and the complementary one, which selects the certain samples (in order to ensure exploration of the inpur space). We consider three types of A$^2$L algorithms with different MAB strategies, namely UCB [3], EXP3 [4] and ShiftBand [2] (all with default parameters) and one baseline method (which selects one strategy in odd iterations, and the second one in even ones).

We performed experiments on splice, sonar and australian datasets (from UCI [1] repository) with $N = 300$ ($N = 200$ in case of sonar dataset) iterations limit. To compare the quality of the learning process we computed the area under the curve (AUC) of each models' MCC till each iteration. In other words, for each model $m$ we analyze the estimation of the function AUC$(m)$ defined as expected value of the integral of MCC value:

$$\text{AUC}(m) := \mathbb{E}\left[\int_0^N \text{MCC}_m(t)dt\right].$$

Final results of AUC, summarized in Table 1, show that UCB strategy performs consistently better than all other tested strategies. As it was previously stated, this results' importance lies particularly in the fact, that this method uses no knowledge about the experiment's length. The only tunable parameter seems to work well even if it is set to the default value which makes it a good candidate for the active learning scenario.

**Table 1.** Summary of estimations of AUC values for each strategy.

| | A$^2$L UCB | A$^2$L EXP3 | A$^2$L ShiftBand | Baseline |
|---|---|---|---|---|
| Splice | **170.43** | 165.62 | 169.11 | 158.42 |
| Sonar | **115.24** | 114.46 | 110.72 | 111.00 |
| Australian | **217.91** | 214.44 | 210.48 | 211.14 |

## 4 DISCUSSION AND CONCLUSIONS

In this paper, we have presented a novel approach to the active learning strategy construction problem by using ensemble of active learners modeled as a MAB problem. Due to its agnostic approach to the underlying learners' strategies, this generic framework can be used with any kind of strategies.

We also showed, that even though the true nature of the process violates the basic MAB assumptions, even the most basic solutions to the problem can yield noticeably better results than the baseline methods. The most important result is a very good performance of the UCB strategy, which requires no information about the actual process being solved (as opposed to EXP3 family whose performance is heavily dependent on the knowledge of the experiment length). This is a very practical result, as in real life applications we rarely know the number of iterations beforehand.

## 5 ACKNOWLEDGMENTS

The paper was partially funded by National Science Centre Poland Found grant no. 2013/09/N/ST6/03015.f

## REFERENCES

[1] Arthur Asuncion and David Newman. Uci machine learning repository, 2007.
[2] Peter Auer, 'Using confidence bounds for exploitation-exploration tradeoffs', *Journal of Machine Learning Research*, **3**, 397–422, (2003).
[3] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer, 'Finite-time analysis of the multiarmed bandit problem', *Machine Learning*, **47**(2-3), 235–256, (2002).
[4] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire, 'The nonstochastic multiarmed bandit problem', *SIAM Journal on Computing*, **32**(1), 48–77, (2003).
[5] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al., 'Scikit-learn: Machine learning in python', *The Journal of Machine Learning Research*, **12**, 2825–2830, (2011).
[6] Herbert Robbins, 'Some aspects of the sequential design of experiments', in *Herbert Robbins Selected Papers*, eds., T.L. Lai and D. Siegmund, 169–177, Springer New York, (1985).
[7] B. Settles, *Active Learning*, volume 6, Morgan & Claypool Publishers, 2012.
[8] H. S. Seung, M. Opper, and H. Sompolinsky, 'Query by committee', in *Proceedings of the fifth annual workshop on Computational learning theory*, COLT '92, pp. 287–294. ACM, (1992).