Active learning for time instant classification

Nauman Ahad [*1] **Namrata Nadagouda** [*1] **Eva L. Dyer** [1 2] **Mark A. Davenport** [1]

Abstract

Active learning is a common strategy for reducing the dependency of model training on large labeled datasets by selecting only the most useful data for labeling. In this work, we consider the problem of actively selecting labels for time instant classification using neural network classifiers. We propose a novel method that selects samples based on a combination of factors that includes uncertainty, diversity, and data density.

1. Introduction

Time series classification has applications in diverse fields such as healthcare, climate science, neuroscience, and financial analysis. Learning classifiers can often require large amounts of labeled data that can be difficult and expensive to obtain. To reduce this label dependency, we can use *active learning* (AL) to acquire labels that are most informative for learning classifiers (Settles, 2009). In this work, we consider the problem of active time series instant classification which involves assigning data, at individual time instants or time stamps, to different classes. While much of prior work considers actively classifying entire time sequences into a single class, there isn't much work in literature that deals with actively classifying time instants. Additionally, there is limited work that explores AL for deep networks, as prior works have studied AL for time series data using nearest neighbour classifiers (He et al., 2015; Peng et al., 2017).

Majority of the works that have explored using deep networks for active classification focus on non-correlated data such as images and text. These works are not directly applicable to time instant classification since they do not consider the inter sample correlations occurring within a time sequence owing to its periodicity. Standard deep AL methods are either uncertainty or diversity based which fail to perform well. For e.g., uncertainty based methods prioritize noisy or class transition time instants, leading to subpar performance. Our proposed method addresses these problems by designing a method that incorporates uncertainty, diversity, and data density to improve active time instant classification for deep networks.

2. Method

Problem setup: Suppose that we wish to classify each time instant $\boldsymbol{x}_i \in \mathbb{R}^d$ within a time series $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T\}$ as corresponding to a class y_i. Our goal is to devise an active acquisition strategy for labels y that allows time series instant classifiers to be trained in a label efficient manner. The proposed AL method is designed for neural network classifiers f that can be decomposed into representation learning network f_θ, which consists of stacked 1D CNNs, and a classifier network f_ψ, which consists of a linear layer followed by a softmax operator. The network f_θ transforms each time instant $\boldsymbol{x}_i \in \mathbb{R}^d$ to a latent instant representation $\boldsymbol{z}_i \in \mathbb{R}^k$, while the classifier network transforms these latent representations $\boldsymbol{z}_i$ to class prediction probability vectors $\widehat{\boldsymbol{y}}_i \in \mathbb{R}^c$, where c is the number of different classes.

Our proposed method: We first apply an uncertainty quantification function U on all softmax probability vectors $\widehat{\boldsymbol{y}}_i, \forall i \in T$. Representations $\boldsymbol{z}_i$ corresponding to the top K percentile of uncertainty scores $U(\widehat{\boldsymbol{y}}_i)$ are clustered into M different clusters through the K-means clustering algorithm. Out of these M clusters, the top B dense clusters, where B is the sampling budget in an AL round, are selected and the most representative sample for each of these clusters are chosen. Labels for the chosen samples are obtained and the classifier is then retrained on all the labeled samples. This procedure is repeated until the total labeling budget is met.

Intuition for our method: Time series instants, and their corresponding instant representations, are likely to be periodic in time. Uncertain periodic time instants are likely to yield similar representations, and hence a K-means clustering algorithm is likely to cluster periodic uncertain instants together. Clusters of higher density are likely to be associated with more frequently encountered uncertain time instants, which need to be prioritized for label acquisition. By selecting points from such clusters, we avoid noisy or class transition points which are inherently uncertain.

[*]Equal contribution [1]Department of Electrical and Computer Engineering, Georgia Tech, Atlanta, Georgia, USA. [2]Coulter Department of Biomedical Engineering, Georgia Tech & Emory University, Atlanta, Georgia, USA. Correspondence to: Nauman Ahad <nauman.ahad@gatech.edu>, Namrata Nadagouda <namrata.nadagouda@gatech.edu>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

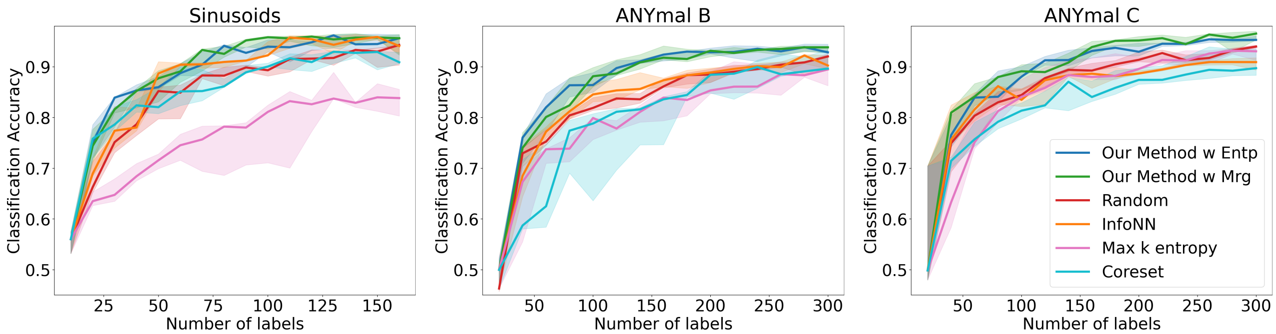

Figure 1. Our method shown in blue line (using entropy as an uncertainty function) and green line (using margin confidence as an uncertainty function) outperforms all the baseline methods. All experiments are repeated 5 times. The median score is shown in solid lines while the gap between 75[th] and 25[th] percentile scores is filled with shaded colors.

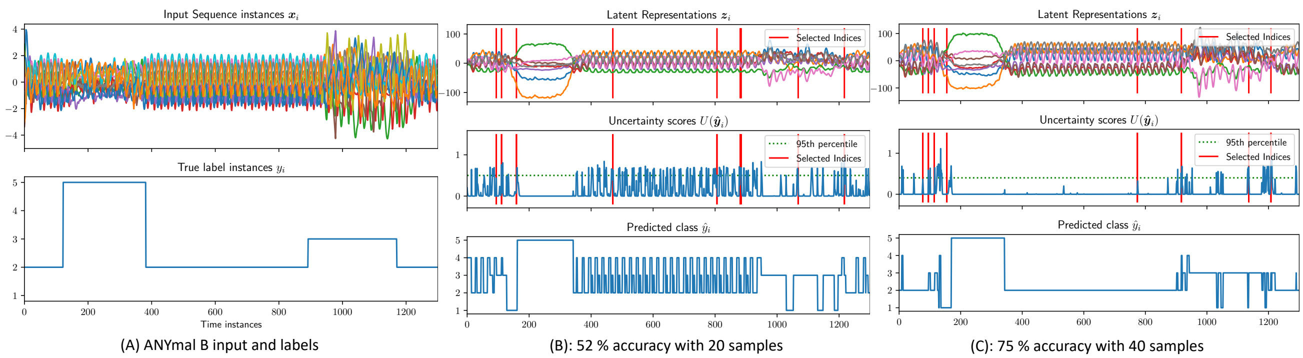

Figure 2. Our method on ANYmal B. (A) shows the input sequences instants $\boldsymbol{x}_i$ and their corresponding labels y_i. The red vertical lines show sample instants chosen for labeling at the end of a training round. It can be seen in (B) that our method prioritizes selection of uncertain periodic instants which are frequently encountered (class 2) for labelling, leading to correct prediction of these instants in (C).

3. Experiments

Synthetic data: We simulate sequences of sinusoidal waves $y = \sin(2\pi f x) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$ that switch randomly between 5 different frequency (f) values of $12, 15, 20, 24, 30$ which the classifier predicts.

Robot activity datasets: We use the Isaac gym environment (Liang et al., 2018) to generate sequences of two different types of quadrupedal robots, ANYmal B and ANYmal C, walking through a predefined map which consists of 5 different terrains (Azabou et al., 2022). These terrain classes to be predicted include walking down a staircase, walking on a flat surface, walking up a staircase, walking down a slope, and walking up a slope.

Baselines and results: We experiment with two variations of our method, using entropy and margin sampling for uncertainty quantification. The baseline methods we use are random (where samples are selected uniformly at random), max-k entropy (an informativeness based method), coreset selection (Sener & Savarese, 2018) (a diversity based method), and InfoNN (Nadagouda et al., 2023) (a method proposed for active image classification that considers both informativeness and diversity). On an average, our method with margin sampling seems to be the best performing in all the experiments. Since our method outperforms methods that use either uncertainty (entropy) or diversity (coreset) and also a combination of the two factors (InfoNN), we believe that incorporating data density indeed is very beneficial in selecting informative samples.

4. Discussion and future directions

In our method, we employ clustering of representations to find densely distributed sample regions. It would be interesting to think of more principled methods that capture the correlation in the time series data to estimate these dense regions. Possible directions towards this include finding distributional modes of smaller windowed segments, or the short term Fourier transforms for these windows, through the kernel mean-shift algorithm. Also, the methods we currently use for uncertainty quantification sometimes fail to accurately measure a sample's informativeness. For example, some of the incorrectly classified samples have very low entropy values, thus avoiding their selection. This necessitates the usage of other methods, potentially specific to time series data (Peng et al., 2017) to quantify the uncertainty.

References

Azabou, M., Mendelson, M., Sorokin, M., Thakoor, S., Ahad, N., Urzay, C., and Dyer, E. L. Learning behavior representations through multi-timescale bootstrapping. 2022.

He, G., Duan, Y., Li, Y., Qian, T., He, J., and Jia, X. Active learning for multivariate time series classification with positive unlabeled data. In *2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 178–185. IEEE, 2015.

Liang, J., Makoviychuk, V., Handa, A., Chentanez, N., Macklin, M., and Fox, D. Gpu-accelerated robotic simulation for distributed reinforcement learning. In *Conference on Robot Learning*, pp. 270–282. PMLR, 2018.

Nadagouda, N., Xu, A., and Davenport, M. A. Active metric learning and classification using similarity queries. In *Uncertainty in Artificial Intelligence (UAI)*, 2023.

Peng, F., Luo, Q., and Ni, L. M. Acts: an active learning method for time series classification. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pp. 175–178. IEEE, 2017.

Sener, O. and Savarese, S. Active learning for convolutional neural networks: A core-set approach. In *International Conference on Learning Representations*, 2018.

Settles, B. Active learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2009.