



Contribution ID: 122

Type: **not specified**

Extending Inverse Distance-based Exploration for Active Learning to Streaming: Online Active Learning for Regression

In statistical and machine learning, efficient data acquisition is pivotal to model performance, particularly when labeled data are costly or time-intensive to obtain. This motivates active learning, in which the learning algorithm selectively queries maximally informative data points to accelerate training and improve predictive efficiency. While many active learning strategies consider query synthesis or pool-based sampling, we address the problem of online active learning in regression scenarios. In the stream-based scenario, unlabeled instances arrive continuously and must be either queried (labeled) or skipped on the fly. We introduce a new algorithm that adaptively selects which instances to label from a data stream by thresholding a suitably designed acquisition function. The method transfers and extends the Inverse Distance-based Exploration for Active Learning (IDEAL) principle, originally developed for pool-based settings, to the streaming context. This transfer preserves IDEAL's balance between exploitation of the current model structure and exploration to promote diversity in the feature space. We benchmark our method against state-of-the-art active learning strategies and against a passive baseline that labels incoming stream instances at random with a fixed probability, providing a clear reference for gains attributable to targeted querying. Performance is assessed through controlled numerical experiments on illustrative synthetic regression problems. We further demonstrate practical utility on a real chemometric data set.

Special/ Invited session

Classification

Both methodology and application

Keywords

online active learning, stream-based selective sampling, regression

Primary author: SPANGL, Bernhard (BOKU University, Vienna)

Co-authors: OTTO, Wilhelm (University of Applied Sciences Wiener Neustadt); XU, Xiaojian (Brock University, St. Catharines)

Presenter: SPANGL, Bernhard (BOKU University, Vienna)

Track Classification: AI: Machine Learning and Predictive Analytics