Active learning (machine learning)

This article is about a machine learning method. For active learning in the context of education, see active learning.

Active learning is a special case of semi-supervised machine learning in which a learning algorithm is able to interactively query the user (or some other information source) to obtain the desired outputs at new data points. In statistics literature it is sometimes also called optimal experimental design.[1][2]

There are situations in which unlabeled data is abundant but manually labeling is expensive. In such a scenario, learning algorithms can actively query the user/teacher for labels. This type of iterative supervised learning is called active learning. Since the learner chooses the examples, the number of examples to learn a concept can often be much lower than the number required in normal supervised learning. With this approach, there is a risk that the algorithm be overwhelmed by uninformative examples. Recent developments are dedicated to hybrid active learning[3] and active learning in a single-pass (online) context,[4] combining concepts from the field of Machine Learning (e.g., conflict and ignorance) with adaptive, incremental learning policies in the field of Online machine learning.

1 Definitions

Let T be the total set of all data under consideration. For example, in a protein engineering problem, T would include all proteins that are known to have a certain interesting activity and all additional proteins that one might want to test for that activity.

During each iteration, i, T is broken up into three subsets

1. $\mathbf{T}_{K,i}$: Data points where the label is **known**.

2. $\mathbf{T}_{U,i}$: Data points where the label is **unknown**.

3. $\mathbf{T}_{C,i}$: A subset of $T_{U,i}$ that is **chosen** to be labeled.

Most of the current research in active learning involves the best method to choose the data points for $T_{C,i}$.

2 Query strategies

Algorithms for determining which data points should be labeled can be organized into a number of different categories:[1]

- Uncertainty sampling: label those points for which the current model is least certain as to what the correct output should be

- Query by committee: a variety of models are trained on the current labeled data, and vote on the output for unlabeled data; label those points for which the "committee" disagrees the most

- Expected model change: label those points that would most change the current model

- Expected error reduction: label those points that would most reduce the model's generalization error

- Variance reduction: label those points that would minimize output variance, which is one of the components of error

- Balance exploration and exploitation: the choice of examples to label is seen as a dilemma between the exploration and the exploitation over the data space representation. This strategy manages this compromise by modelling the active learning problem as a contextual bandit problem. For example, Bounef-fouf et at.[5] propose a sequential algorithm named Active Thompson Sampling (ATS), which, in each round, assigns a sampling distribution on the pool, samples one point from this distribution, and queries the oracle for this sample point label.

- Exponentiated Gradient Exploration for Active Learning:[6] In this paper, the author proposes a sequential algorithm named exponentiated gradient (EG)-active that can improve any active learning algorithm by an optimal random exploration.

A wide variety of algorithms have been studied that fall into these categories.[1][2]

3 Minimum Marginal Hyperplane

Some active learning algorithms are built upon Support vector machines (SVMs) and exploit the structure of the

SVM to determine which data points to label. Such methods usually calculate the margin, W, of each unlabeled datum in $T_{U,i}$ and treat W as an n-dimensional distance from that datum to the separating hyperplane.

Minimum Marginal Hyperplane methods assume that the data with the smallest W are those that the SVM is most uncertain about and therefore should be placed in $T_{C,i}$ to be labeled. Other similar methods, such as Maximum Marginal Hyperplane, choose data with the largest W. Tradeoff methods choose a mix of the smallest and largest W s.

4   See also

- Proactive learning

5   Notes

[1] Settles, Burr (2010), "Active Learning Literature Survey" (PDF), *Computer Sciences Technical Report 1648. University of Wisconsin–Madison*, retrieved 2014-11-18.

[2] Olsson, Fredrik. "A literature survey of active machine learning in the context of natural language processing".

[3] E. Lughofer (2012), Hybrid Active Learning (HAL) for Reducing the Annotation Efforts of Operators in Classification Systems. Pattern Recognition, vol. 45 (2), pp. 884-896, 2012.

[4] E. Lughofer (2012), Single-Pass Active Learning with Conflict and Ignorance. Evolving Systems, vol. 3 (4), pp. 251-271, 2012.

[5] Bouneffouf et .al (2014), Contextual Bandit for Active Learning: Active Thompson Sampling. Neural Information Processing - 21st International Conference, ICONIP 2014

[6] Bouneffouf et .al (2016), Exponentiated Gradient Exploration for Active Learning. Computers, vol. 5 (1), 2016, pp. 1-12

6   Other references

- N. Rubens, D. Kaplan, M. Sugiyama. Recommender Systems Handbook: Active Learning in Recommender Systems (eds. F. Ricci, P.B. Kantor, L. Rokach,B. Shapira). Springer, 2011 , .

- Active Learning Tutorial, S. Dasgupta and J. Langford.

7 Text and image sources, contributors, and licenses

7.1 Text

- **Active learning (machine learning)** *Source:* https://en.wikipedia.org/wiki/Active_learning_(machine_learning)?oldid=699254607 *Contributors:* Bearcat, WorldsApart, Johndburger, SmackBot, R'n'B, Llorenzi, Alcmaeonid, Metaxal, Pcasari, SchreiberBike, Galzigler, Jncraton, Yobot, AnomieBOT, Ambertk, Dithridge, X7q, Ranumao, John of Reading, Tdietterich, Snotbot, BG19bot, SandraEbert, Rufini and Anonymous: 8

7.2 Images

7.3 Content license