Learning to Active Learn by Gradient Variation based on Instance Importance

Sergio Flesca, Domenico Mandaglio, Francesco Scala and Andrea Tagarelli
{flesca, d.mandaglio, francesco.scala, tagarelli}@dimes.unical.it
Dept. Computer Engineering, Modeling, Electronics, and Systems Engineering (DIMES) University of Calabria, Rende (CS), Italy

Introduction

- **Active Learning** aims at selecting the data instances to be labeled by an expert, or annotation oracle, in order to train a machine learning model as quickly and effectively as possible.
- We propose a new iterative **learning-to-active-learn approach** that selects the instances to be labeled as the ones producing the maximum change to the current classifier, measured in terms of **variations in the learning gradient** of the classification model.
- Our experimental evaluation conducted on CIFAR-10 image data, and including a comparison with two baselines, has shown promising results by the proposed approach in terms of percentage increase in accuracy.

Methods

Algorithm 1: LAL-IGradV

Data: LI: set of labeled instances, UI: set of unlabeled instances, DNN: deep neural network model, R: importance score regressor, *epch*: maximum number of epochs, k: number of relevant instances to select

```
Train DNN on LI
NLI ← Select k instances from UI uniformly at random
The oracle annotates the instances in NLI
for i = 1...epch do
    Train DNN on LI ∪ NLI and compute importance score r_x, for each x ∈ NLI
    Train R on the set of pairs {⟨x, r_x⟩ | x ∈ NLI}
    LI ← LI ∪ NLI
    Apply R to UI instances to predict importance scores (r̂_x)
    topK ← Select top-k instances from UI by importance score r̂_x
    The oracle annotates the instances in topK
    NLI ← topK
```

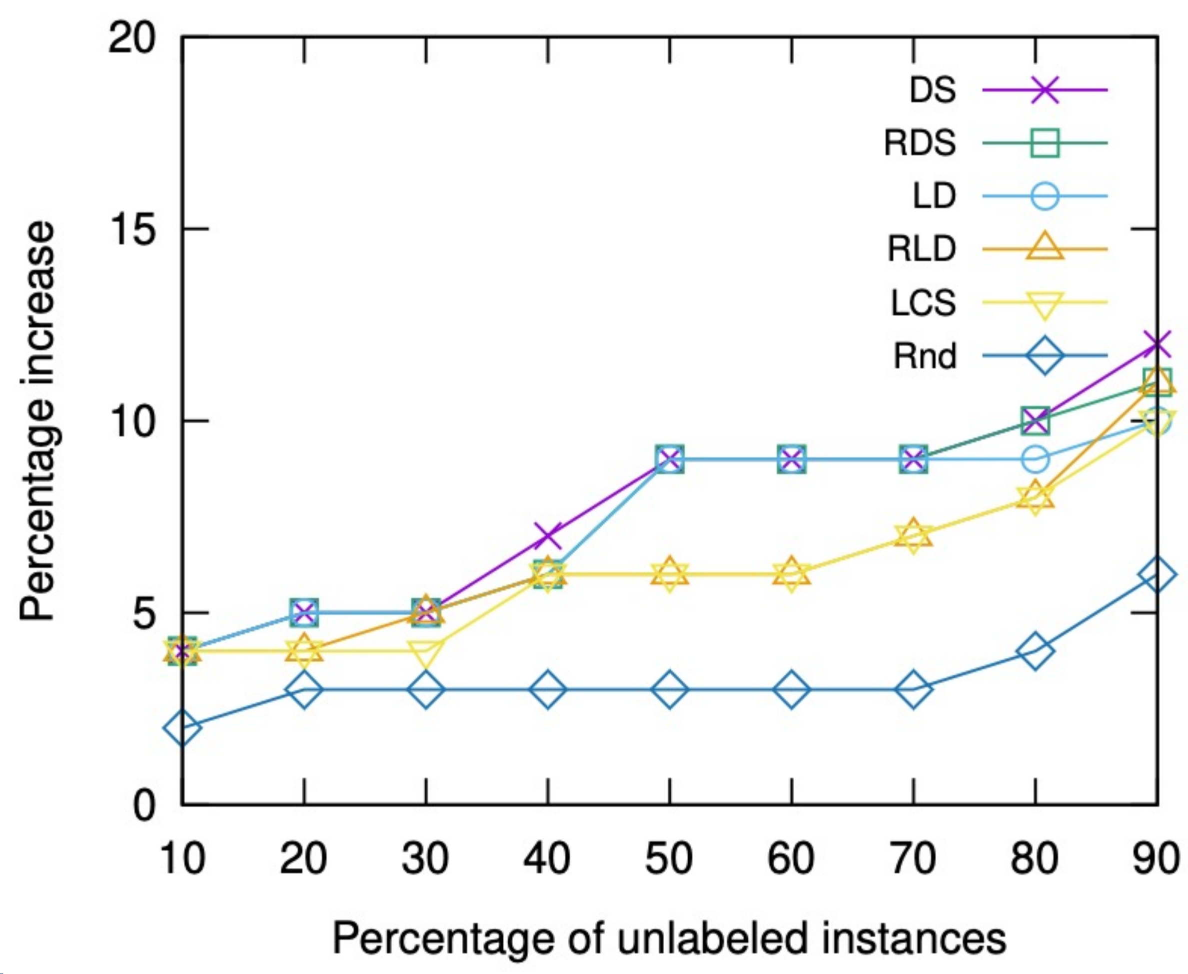
Different strategies to compute the importance score of an instance *x* at the current epoch:

- **Direct similarity** (DS): proportional to the similarity between the directions of the gradient of the neural network at the current epoch and the gradient calculated with respect to the instance *x* only.
- **Ranked direct similarity** (RDS): first applies the DS technique, then the importance scores of the new set of labeled instances (*NLI*) computed by DS are ordered and divided in three bins.
- **Leave-one-out distance** (LD): proportional to the difference between the directions of the gradient of the neural network at the current epoch and the gradient calculated when leaving out *x*.
- **Ranked leave-one-out distance** (RLD): first applies the LD technique, then the importance scores of the instances in *NLI* computed by LD are ordered and divided in three bins.

Results

Performance of our proposed methods: initial and final accuracy $A^{(0)}$ and A, percentage increase w.r.t. a random baseline (Rnd) and w.r.t. the Least Confidence Sampling (LCS) method, and active learning time (sec) averaged over the epochs, for various percentage values of unlabeled instances

	$A^{(0)}$	*DS*				*RDS*				*LD*				*RLD*			
		A	$\%_{Rnd}$	$\%_{LCS}$	time	A	$\%_{Rnd}$	$\%_{LCS}$	time	A	$\%_{Rnd}$	$\%_{LCS}$	time	A	$\%_{Rnd}$	$\%_{LCS}$	time
10%	0.793	0.831	2.32	0.43	186	0.832	2.44	0.54	191	0.831	2.28	0.39	625	0.828	1.90	0.01	769
20%	0.783	0.826	1.90	0.75	178	0.825	1.79	0.65	217	0.824	1.72	0.57	623	0.822	1.46	0.32	796
30%	0.784	0.827	1.95	0.50	170	0.828	2.06	0.61	250	0.826	1.75	0.30	620	0.822	1.46	0.32	827
40%	0.763	0.819	4.01	1.08	170	0.811	3.04	0.13	295	0.811	3.02	0.11	620	0.811	2.96	0.05	872
50%	0.733	0.801	5.97	2.84	162	0.800	5.82	2.70	352	0.799	5.80	2.67	619	0.779	3.07	0.03	1002
60%	0.728	0.801	6.32	3.21	162	0.798	5.96	2.86	423	0.795	5.57	2.48	614	0.777	3.20	0.18	1089
70%	0.708	0.778	6.49	2.50	154	0.778	6.38	2.40	513	0.773	5.82	1.86	607	0.760	4.01	0.12	1190
80%	0.640	0.705	5.39	1.82	139	0.704	5.27	1.71	613	0.700	4.62	1.08	604	0.694	3.78	0.27	1310
90%	0.570	0.644	5.89	2.22	129	0.636	4.60	0.98	732	0.632	3.95	0.35	602	0.636	4.59	0.97	1395



Percentage increase of accuracy for the various active learning methods, with varying percentage of unlabeled instances, and number of selected instances at each epoch equal to 500