Active Learning for Large-scale Object Classification: From Exploration to Exploitation

Ho-Gyeong Kim
KAIST
Daejeon, Republic of Korea
hogyeong@kaist.ac.kr

Jihyeon Roh
KAIST
Daejeon, Republic of Korea
rohleejh@kaist.ac.kr

Hwaran Lee
KAIST
Daejeon, Republic of Korea
hwaran.lee@kaist.ac.kr

Geonmin Kim
KAIST
Daejeon, Republic of Korea
gmkim90@kaist.ac.kr

Soo-Young Lee
KAIST
Daejeon, Republic of Korea
sy-lee@kaist.ac.kr

ABSTRACT

Information and communication technologies supply data every day at incredibly increasing rate, however, almost all of the accumulated data are unlabeled and obtaining their labels is expensive and time-consuming. Among the raw data, selecting and labeling some samples expected to be more informative than others can enhance machines without high cost. This process is called selective sampling, essential part of active learning. So far, most researches have concentrated on classical uncertainty measures to acquire informative data, which is related to 'exploitation' process of learning. However, when the initial labeled dataset is too small or biased, the early stage model can be unreliable and its decision boundary would be over-fitted to the initial data. Moreover, the obtained data by the exploitation strategy may exacerbate the model further. We introduced 'exploration' strategy as well as 'exploitation' strategy. In this paper, we employ Self-Organizing Maps (SOM), one of neural networks to estimate and explore data distribution. For exploitation, margin sampling is applied to the classifier, neural network with soft-max output layer. The effectiveness proposed methods are demonstrated on ILSVRC-2011 image classification task based on features extracted from well-trained Convolutional Neural Networks (CNN). Active learning with exploration strategy shows its potential by stabilizing the early stage model and reducing the classification error rate, and finally making it to be high-quality models.

Author Keywords

Active learning; exploration-exploitation; self-organizing maps; neural network; large-scale image classification

HAI 2015, October 21 - 24, 2015, Daegu, Republic of Korea.

ACM 978-1-4503-3527-0/15/10…$15.00.
http://dx.doi.org/10.1145/2814940.2814989

ACM Classification Keywords

H.5.1. Information interfaces and presentation (e.g., HCI): Multimedia Information Systems

INTRODUCTION

For intelligent system to be trained with existing big data, most learning algorithm emphasizes to obtain large number of labelled data. Even though the communication and computing technologies provide enormous data, they are unlabeled unfortunately, and it is expensive and time-consuming to acquire all labels of them. In this circumstance, labeling samples drawn from unlabeled data for the supervised training is significant to construct a high performance. One possible solution for this problem is to choose samples which are expected to be more helpful than others for the training with a less number of training samples. This method is called selective sampling, firstly proposed by Cohn [1], which is an essential component of active learning algorithm that controls the selection of the next training samples.

In active learning, a learner begin with initial samples in the labelled training set. As learning proceeds, it requests one or more other labelled examples. Then, the current model is learned by using new labelled dataset and selects next samples from unlabeled data pool. The new labelled samples are simply added to the labelled set, and the learner proceeds in a supervised way. The sampling method can be varied depending on what aspects of active learning focus on. For example, if active learning focuses on reducing uncertainty of classification boundary (i.e. seeking exploitation), sampling from data near margin of classifier is widely used. For example, Schohn and Cohn [2] sampled data with small margin from the classification boundary of SVM. The intuition behind this approach, called margin sampling, is that small margin data are more likely to be misclassified compared to large margin data. Also, Xu [3] proposed developed margin sampling method which utilizes K-means clustering to obtain representative data near margin. In this paper, classifier is also based on Neural Network with soft-max output layer. We employ sampling criteria such as probability difference between 1st and 2nd best class [8].

For active learning to be successful, it is important to starts from initial model with not too low accuracy. However, active learning with only margin sampling cannot consider this criteria. The initial model can be unreliable when the size of dataset is too small, because those labelled data may be biased and make decision boundary to be over-fitted to only initial training data. To overcome this limitation, not only seeking exploitation but also exploration is important. For exploration, the learner tries to find data distribution by obtaining samples which improve data distribution estimated from current samples. Recent works in active learning ([4]-[7]) consider both 'exploitation' and 'exploration' to compromise limitations of each approach and control the balance between the two strategies. For exploration, random sampling is widely used to obtain samples regardless of classification boundary. Other than random sampling, there were methods to seek finding feature distribution. Therefore, this paper employs Self-Organizing Maps (SOM), one of neural network which is trained by competitive/cooperative learning, to estimate data distribution without prior probability distribution.

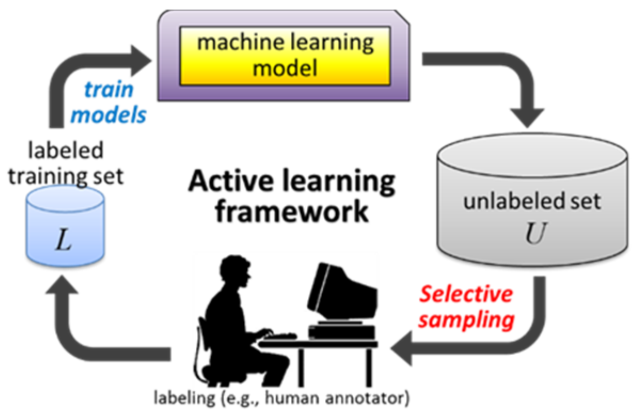

Figure 1. General framework of active learning

Furthermore, it is important to balance between exploration and exploitation. This paper suggests a method to control the ratio between exploration and exploitation as active learning epoch proceeds.

ACTIVE LEARNING ALGORITHM

In this paper, we propose sample selection criteria for both exploration and exploitation. At each sampling iteration, which we call active learning iteration, in the paper, fixed amount of samples with higher uncertainty based on selection criteria are labeled first. Then classifier will be trained with whole training samples including newly selected ones. Proposed Active learning algorithm is summarized in Figure 2.

First, we assume that we only have small amounts of labeled set L_0, and numerous unlabeled dataset U_0 is available other than L_0. The proposed active learning system select samples to make set S_k and obtain label from the external source iteratively. Then, classifier will be trained with updated dataset $L_k = \{L_{k-1} \cup S_k\}$ and unlabeled set is remained as $U_{k+1} = U_k - S_k$.

The set of selected samples, S_k is comprised of S_k^r and S_k^i, which comes from selection criteria of exploration and exploitation.

Active learning from exploration to exploitation:

Initialize: Randomly select an initial labelled set L_0,
Train SOM from a subset of $\{L_0 \cup U_0\}$.
While $U_k \neq \emptyset$:
$S_k = \emptyset$.
Train a classifier C_k using current labelled set L_k.
1) Exploration: select $\mathbf{x} = \text{argmin}_{\mathbf{x} \in U_k} D(\mathbf{x})$ in SOM and add to S_k^r.
Repeat selection until we collect n_r samples
2) Exploitation: select $\mathbf{x} = \text{argmin}_{\mathbf{x} \in U_k} U(\mathbf{x})$ in Softmax and add to S_k^i.
Repeat selection until we collect n_i samples
Update an unlabeled set as $U_{k+1} = U_k - S_k$ and a labelled set
$L_{k+1} = \{L_k \cup S_k\}$, where $S_k = \{S_k^r \cup S_k^i\}$, $n_r + n_i = n$
$k = k + 1$.
End While

Figure 2. Pseudo-code of algorithm

SELECTIVE SAMPLING

i) Exploration

For all unlabeled data, a winner neuron from a learned SOM (Self-organizing map) map using following equation.

$$\mathbf{w}_{j^*}^{u} = \text{argmin}_{j=1\ldots N} \|\mathbf{w}_j - \mathbf{x}\| \tag{1}$$

After all unlabeled data are assigned to winner neurons, each has candidate set which includes the closest or most similar data point to the neuron.

The number of selected data per a grid can be described as follows:

$$n_{explore}^{j} = \frac{n^j}{n_U} \times n_{explore} \tag{2}$$

where $n_{explore}$ is the total number of selected data by exploration strategy, U indicates a pool of unlabeled data, and j is the index of neuron in SOM. A neuron on which data are located denser samples more data than on which sparse data.

ii) Exploitation (Margin sampling)

Exploitation strategy aims to refine classification boundary by selecting samples located near boundary. Margin sampling is based on the idea that most incorrect classification comes from confusing between top-1 and top-2 prediction. In this measures, samples with less confidence value is selected first since they are more uncertain to identify label from the current classifier.

$$M_{margin}(\mathbf{x}) = P_{\mathbf{W}}(\hat{y}_1|\mathbf{x}) - P_{\mathbf{W}}(\hat{y}_2|\mathbf{x}) \tag{3}$$

where $\hat{y}_i$: top i-th prediction class output.

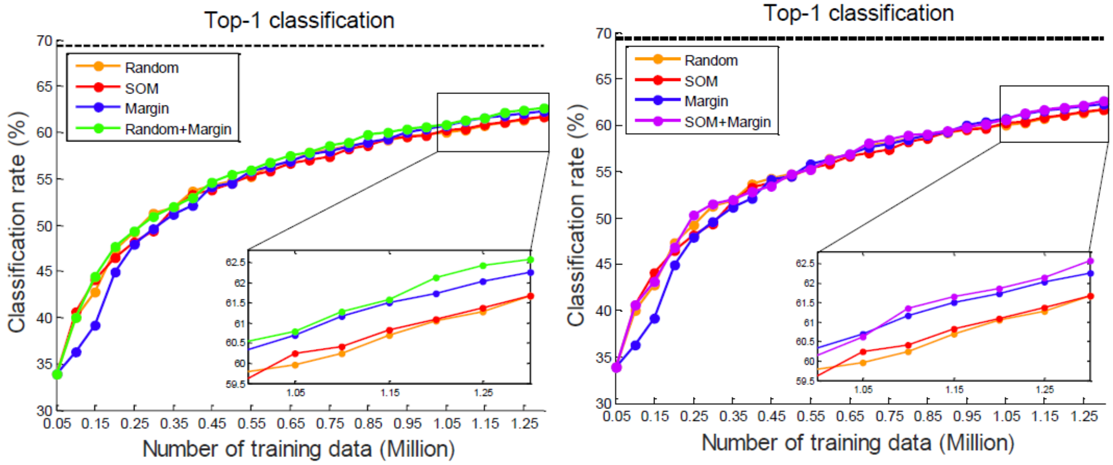

Figure 3. Top-1 classification rate with active learning epoch.

Figure 4. Example of selected images for (left) red wine, (right) orange. Margin (a) uncertain, (b) certain, (c) SOM sampling

EXPERIMENT

i) Dataset

For verifying the proposed active learning system on big data, we use ILSVRC 2011 dataset for the experiments. ILSVRC is an annual competition for classifying large number of images with 1000 classes, which comes from subsets of ImageNet. A lot of researcher tries to compete for classification results on this dataset. Especially, it has 1.3million images for training.

ii) Results

The effect of each sampling method is examining by experiment, starting with 50 images per class and augmenting 0.025 million images, which is shown in Figure 3. With this initial condition, several sampling methods which combines both exploration and exploitation are compared. Combination of exploration-exploitation along sampling epoch, outperforms other sampling methods.

Figure 4 shows example images of class ‘red wine’ and ‘orange’ with having certain and uncertain for margin sampling methods. As we see, images with high confidence (Figure 4(b)) looks easily identified. But images with low confidence (Figure 4(a)) looks difficult to classify even for human. As we see in Figure 4(c), SOM sampling selects images with representing red wine and orange classes.

CONCLUSION

In this article, we have shown the selective sampling methods for both exploration and exploitation. For exploration strategy, we have adopted a SOM to learn representative features exists in input data, and compares with random sampling which is widely used exploration strategy. For exploitation strategy, we have used neural network classifier to fine-tune the classification boundaries via selective sampling. Moreover, we apply the combination of two strategies as sampling proceeds. The effectiveness of proposed methods are demonstrated on ILSVRC-2011 image classification task based on features extracted from CNNs. Active learning with exploration strategy shows its potential by stabilizing early stage models, reducing error rates, and finally obtaining high-quality models.

ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012-0008681).

REFERENCES

1. COHN, David; ATLAS, Les; LADNER, Richard. Improving generalization with active learning. Machine Learning, 1994, 15.2: 201-221.

2. SCHOHN, Greg; COHN, David. Less is more: Active learning with support vector machines. In: ICML. 2000. p. 839-846.
3. Xu, Zhao, et al. "Representative sampling for text classification using support vector machines." ECIR. Vol. 3. 2003.
4. FREUND, Yoav, et al. Selective sampling using the query by committee algorithm. Machine learning, 1997, 28.2-3: 133-168.
5. IYENGAR, Vijay S.; APTE, Chidanand; ZHANG, Tong. Active learning using adaptive resampling. In: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2000. p. 91-98.
6. LINDENBAUM, Michael; MARKOVITCH, Shaul; RUSAKOV, Dmitry. Selective sampling for nearest neighbor classifiers. Machine learning, 2004, 54.2: 125-152.
7. ROY, Nicholas; MCCALLUM, Andrew. Toward optimal active learning through monte carlo estimation of error reduction. ICML, Williamstown, 2001.
8. T. Scheffer, C. Decomain, and S. Wrobel. Active hidden Markov models for information extraction. In Proceedings of the International Conference on Advances in Intelligent Data Analysis (CAIDA), pages 309–318. Springer-Verlag, 2001.