Application on Case Analysis of Active Learning in Reducing Annotation Cost

Kangrui Ren
School of Computer Science
McGill University
Montreal, Quebec, Canada
kangrui.ren@mail.mcgill.ca

ABSTRACT

With the wide spread and development of CNNs, Active Learning has shown great charm in processing large data sets, reducing the number of training data and labor costs. However, because most of the current query strategies are designed manually, the ability of cross domain generalization of selection strategies is still insufficient. While reducing annotation cost with Active Learning, some potential defects still need to be improved. This paper focuses on a special medical imaging case study, in which the active learning algorithm proposed by the author of the original literature can effectively reduce the labeling cost by at least half. At the same time, this paper also discusses the problems of active learning, the orientation of development and solutions for certain latent danger.

1 INTRODUCTION TO ACTIVE LEARNING TOPICS

We now review the basic topics of Active Learning, which includes the meaning of Active Learning, Scenarios of Active Learning, and Query Strategy (including Heuristics for classification problems).

1.1 Active Learning Definition

Machine Learning can be divided into "supervised learning" and "unsupervised learning" according to whether the sample data is labeled. In addition, the algorithms that use unlabeled samples and labeled samples for machine learning can be further summarized into three categories: semi-supervised learning, transductive learning and active learning. Burr Settles[14] defines that Active Learning (sometimes called “query learning” or “optimal experimental design” in the statistics literature) is a sub-field of machine learning and, more generally, artificial intelligence.

Literature[8] briefly introduces the similarities and differences between active learning and semi-supervised learning: “Semi-supervised learning and active learning both select and mark several valuable data from unlabeled data set and add them to the labeled data set to improve the accuracy of the classifier and reduce experts’ workload. Nevertheless, they vary in learning method: semi-supervised learning generally does not require human participation. It uses a benchmark classifier with certain classification accuracy to automatically label unlabeled data; while active learning must label selected valuable data manually. Although Semi-supervised learning reduces annotation cost with automatic or semi-automatic labeling instead of manual labeling, it does not guarantee the label accuracy since the labeling result depends on the classification accuracy of the benchmark classifier trained with previous labeled examples. In contrast, manual labeling will not introduce wrong class labels.

"Learning module" and "Query Strategy" are two basic and important modules of active learning algorithms. The basic working procedure of Active Learning is this. Active learning uses "Query strategy" to actively select some (1 or N) samples from unlabeled sample sets for experts Labeling; then add the labeled samples to the training data for further training; when the "learning module" meets the termination conditions, the program ends, otherwise the above steps will repeat to obtain more labeled samples for training.

In addition, the active learning algorithm has a key assumption: “The key hypothesis is that if the learning algorithm is allowed to choose the data from which it learns—to be "curious," if you will—it will perform better with less training".

1.2 Active Learning Flow Diagram

As shown in the figure below is a common active learning flow diagram, which represents a complete iterative process. The model can be expressed as A = (C, L, S, Q, U). Where C is the classifier (1 or more), L is the labeled sample set, S is the expert who can label the sample, Q is the currently used query strategy, and U is the unlabeled sample set.

The flow diagram can be interpreted as the following steps.

(1) Select the appropriate classifier (network model) as current model, and partition data into training sample (labeled sample for training model), validation sample (labeled sample to verify the performance of the current model), and active sample (unlabeled data set corresponding to unlabeled pool).
(2) Initialization: random initialize or initialize through transfer learning; if there are annotation samples in target domain, the model is trained through these annotation samples.
(3) Use the current model to detect samples in active sample set, and get the prediction results of each sample. At this time, researchers can select uncertainty strategy to measure the labeling value of samples. The closer the prediction result is to 0.5, the higher uncertainty the sample has, that is, the more worthy the sample is to be labeled.
(4) Experts label the selected samples and add them into the training sample set.
(5) Use all currently labeled samples in training sample set for current model to perform fine-tuning and updating.
(6) Use current model to verify the validation sample set. If the performance of the current model gets the target or can no longer label new samples (no experts or no money), the iteration process will terminate. Otherwise, cycle through steps (3) - (6).

1.3 Scenarios of Active Learning

We will briefly introduce three Active learning scenarios according to [11].

1.3.1 *Membership Query Synthesis.* :

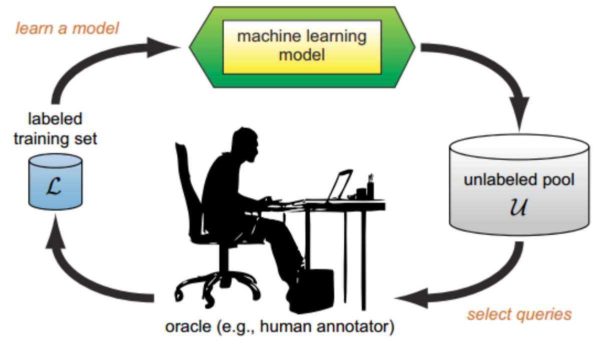

Figure 1: Active Learning Diagram

It means that the model can generate new samples, that is, the model can control the generation of samples.

The first Active Learning scenario "Membership Queries" [1]: The learner can generate labels for unlabeled data sets in the input space, including queries that the learner generates de novo, instead of these samples from the same potential distribution.

Effective query synthesis is usually easier to handle and efficient in solving specific domain problems [2].

Query synthesis can be applied to many problems in many fields. However, if the expert link in active learning is artificial, it is often inappropriate to label random samples. As [12] shows, many query images generated by the learner do not contain recognizable symbols, and the forged mixed characters have no semantic meaning. Similarly, when membership queries are applied to NLP tasks, it may produce some disordered streams of text or speech.

1.3.2 *Stream-Based Selective Sampling.* :

In order to solve the aforementioned problems, stream based and pool based scenarios are proposed.

Stream-Based Selective Sampling means that we select one unmarked instance at a time and let the model decide whether to query the tag of the instance or based on its content. To determine the amount of information for an instance, It should use query strategy (which will be explained in following contents).

There is a crucial assumption for Stream-Based Selective Sampling that unlabeled samples can be obtained with inexpensive costs. Then researchers can collect unlabeled data from real distribution and learners can decide whether to select these unlabeled samples for experts labeling. It should be noted that if the input sample belongs to uniform distribution, selective sampling may be as effective as membership query learning.

In addition, because there is a preset annotation condition for Stream-Based Selective Sampling and the condition often needs to be adjusted according to different tasks [15], thus it is difficult to use this scenario as a general strategy.

1.3.3 *Pool-Based Sampling.* :

Pool-Based Sampling means the instance is extracted from the large unmarked data pool, which stores all data, according to information measurement methods. This measure applies to all instances in the pool (or to a subset if the pool is very large) and then selects the most informative instance.

Compared with Stream-Based Selective Sampling, Pool-Based Sampling can select the samples with the highest contribution to classification in the current sample pool , which reduces not only the query cost but also the annotation cost. This makes the Pool-Based Sampling strategy widely used.

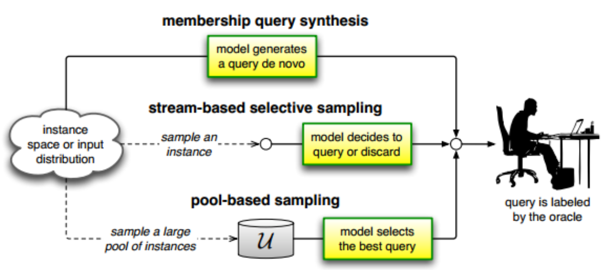

Figure 2: Active Learning Scenarios

1.4 Query Strategy

Query strategy is one of the cores of active learning. We will mainly discuss some typical Query Strategies and Heuristics for classification problems in this section.

1.4.1 *Query Strategies.*

- Random Sampling (RS) : Select Sample Randomly
- Uncertainty Sampling (US) :
  Selecting the most uncertain samples in the current model (e.g. In classification problem, p = 0.5 means that the sample is ambiguous and highly uncertain), labeling such samples is the most helpful approach to improve the current model. Thus it is also one of the most commonly used strategies in Active Learning. However, US strategy only considers the information of a single sample rather than the overall distribution of the sample space, so it will find outlier or some redundant samples. For example, literature [17] combines US with diversity to ensure the diversity of the selected samples(See Section 2 for details).
- Kapoor Algorithm:
  An algorithm [9] that balances exploration and exploitation by incorporating mean and variance estimation of the GP classifier.
- ALBE Strategy:
  A recent example [6] of meta-AL that adaptively uses a combination of strategies, including [7].

1.4.2 *Heuristics for classification problems.* :

In this section, we will mainly explore some common heuristics (query strategies also) , which can be divided into three categories: QBC Based, Margin Sampling Based, and Posterior Probability Based.

- Query-By-Committee Based Algorithms :
  In the QBC algorithm, labeled samples are used to train multiple hypothesis models with different parameters and to predict unlabeled samples. Therefore, the QBC algorithm needs to train a certain number of classifiers, and its computational complexity is quite large in practical applications. In order to restrict the calculation, EQB method is used to simplify the calculation. In the case of high-dimensional data, AMD algorithm can divide the feature space into subspace, which is the deformation of EQB algorithm. Different classification methods classify the same samples in different regions, avoiding the problem of dimension disaster in the calculation process. The advantages of the algorithm is the classifier can use a variety of classification models and combination patterns such as: neural network, Bayesian law and so on.
- Margin Sampling Based Algorithms :
  The Margin Sampling mainly focuses on the case of support vector machine. According to the classification model, the distance between the sample and the classification interface is calculated to select the sample. In MS algorithm, only the sample nearest to the classification interface is

selected to join the training set, which is the simplest Margin Sampling method. The difference between MCLU algorithm and MS is that the distance difference between the two most likely samples farthest from the classification interface is selected as the evaluation criterion for MCLU. In the mixed category region, MCLU can select the sample with the most uncertainty, but MS has poor effect. In some cases, MS and MCLU will select redundant samples, introduce diversity criteria, eliminate similar samples, and reduce the number of iterations. The common diversity criterion is the similarity between samples, that is, the higher the similarity between samples, the more consistent the data characteristics reflected by the samples, then the samples need to be removed, otherwise, the lower the similarity. The similarity coefficient can be used to characterize the similarity of sample points.

- Posterior Probability Based Algorithms :
The calculation of this method is the fastest. The disadvantage of KL method is that in the iterative optimization process, it can only select one sample, which increases the number of iterations. In addition, if the classification model can not provide accurate probability evaluation value, it depends on the subsequent optimization evaluation value. In BT algorithm, the idea is similar to EQB. In multi classifier, the difference between the two maximum probabilities of samples is selected as the criterion. When the two maximum probabilities are close enough, the classification accuracy of the classifier is the lowest.

2 APPLICATION OF ACTIVE LEARNING IN REAL INDUSTRIAL DATA

This section mainly refers to the literature[17] [16]. The algorithms proposed in these literature aim to solve an important problem in the application of deep learning: How to train a model with as few labeled data sets as possible? And the performance of this model can reach the performance of a model trained by a large number of labeled data sets according to the common method (randomly selected training data).

In this section, we will mainly analyze the application of active learning in medical image. We will introduce the relevant background, significant observations, the newly proposed algorithm with related advantages, etc.

2.1 Background

Convolutional neural network (CNN) has brought great changes in computer vision, which cannot be separated from the generation of large-scale annotated data-sets (such as ImageNet and Places). The success of CNN has also aroused strong interests in biomedical image field, and this trend is widely spread. However, there is no large-scale and annotated data set in the field of biomedical images. It is precisely because of the lack of large-scale annotated data sets that this trend is in a blocked state, and current researches can not achieve the desired results.

The labeling of biomedical image samples is not only boring and time-consuming, but also requires a lot of manpower and material resources. What's more, the labeling of data sets in this field is not competent for ordinary people, but experts with certain professional knowledge and skills.

2.2 Introduction

Annotate data is a complicated and difficult work, especially in the field of bio-medicine: 1) numerous doctors with relevant professional knowledge; 2) high annotation cost; 3) long training cycle. Then, if we can train a good classifier with a small amount of labeled data, it will be a very meaningful work.

In order to significantly reduce the annotation cost, the author proposes a novel method called AIFT(active, incremental fine-tuning). AIFT naturally integrates active learning and transfer learning into one structure. First, AIFT uses a pre-trained CNN (from ImageNet) to find the samples that are most helpful to improve the current CNN performance. Then experts label the selected samples and add them to the fine-tuning training set. After fine-tuned, the performance of CNN will gradually improved. This significant improvement mainly comes from the contribution of "active" and "incremental" in AIFT algorithm.

It is worth noting that AIFT will conduct fine-tuning in each iteration, which is the innovation point of this paper (fine-tuning is performed on the current CNN Rather than on the basis of the original pre-trained CNN. This provide the model a little memory of previous models. At the same time, however, the parameters of fine-tuning are hard to control, which is the further work of this paper.)

2.3 Insightful Observations

The remarkable effect of AIFT is due to some following simple but powerful observations.

(1) In order to improve the performance of CNN in biomedical image processing, we usually think of data augmentation. Each candidate sample is divided into multiple patches and these patches are generated automatically.

(2) Patches from the same candidate should have the same tag. Naturally, these patches (before they were added to the training data-set) should have the same tags as predicted by current CNN.

(3) The entropy of each patch from the same candidate and the diversity between two patches are taken as two meaningful indicators to measure the beneficial degree of the candidate to improve the current CNN performance (In addition to entropy and diversity, there are many indicators available, but this is not the focus of this paper.)

(4) When data augmentation generates some "hard" samples which are more helpful to improve the current CNN, it will inevitably bring some noise labels. Therefore, in order to enhance the robustness of AIFT, only part of the patches are selected to participate in the calculation of entropy and diversity according to the current prediction results of CNN.

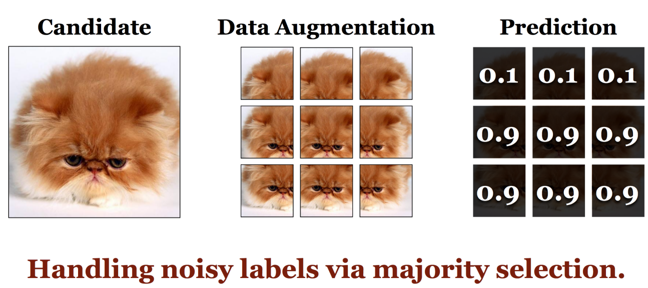

Figure 3: Noise Label Issue

It can be seen that in Figure 3, the consistency of prediction is very low for a candidate, which means that the value of diversity is very large, that is, it should be considered as a hard sample. However, a close look at the nine patches shows that even a good classifier can hardly distinguish 1, 2, 3 in the middle graph as a cat. But for 4-9 patches, the network makes a good classification. Therefore, we call patches similar to 1-3 as noise labels generated by random data augmentation.

The solution here is to give a majority voting for the predictions of all patches. As long as the general label of network prediction is unified, the prediction is trustworthy and unified.

2.4 AIFT Algorithm

Algorithm 1: Active incremental fine-tuning method.

Input:
$\mathcal{U} = \{\mathcal{C}_i\}, i \in [1, n]$ $\{\mathcal{U}$ contains n candidates$\}$
$\mathcal{C}_i = \{x_i^j\}, j \in [1, m]$ $\{\mathcal{C}_i$ has m patches$\}$
$\mathcal{M}_0$: pre-trained CNN
b: batch size
α: patch selection ratio
Output:
$\mathcal{L}$: labeled candidates
$\mathcal{M}_t$: fine-tuned CNN model at Iteration t
Functions:
$p \leftarrow P(\mathcal{C}, \mathcal{M})$ $\{$outputs of $\mathcal{M}$ given $\forall x \in \mathcal{C}\}$
$\mathcal{M}_t \leftarrow F(\mathcal{L}, \mathcal{M}_{t-1})$ $\{$fine-tune $\mathcal{M}_{t-1}$ with $\mathcal{L}\}$
$a \leftarrow mean(p_i)$ $\{a = \frac{1}{m}\sum_{j=1}^{m} p_i^j\}$
Initialize:
$\mathcal{L} \leftarrow \varnothing$, t$\leftarrow 1$
repeat
  for *each* $\mathcal{C}_i \in \mathcal{U}$ **do**
    $p_i \leftarrow P(\mathcal{C}_i, \mathcal{M}_{t-1})$
    if $mean(p_i) > 0.5$ **then**
      $\mathcal{S}_i' \leftarrow$ top α percent of the patches of $\mathcal{C}_i$
    else
      $\mathcal{S}_i' \leftarrow$ bottom α percent of the patches of $\mathcal{C}_i$
    end
    Build matrix R_i using Eq. 3 for $\mathcal{S}_i'$
  end
  Sort $\mathcal{U}$ according to the numerical sum of R_i
  Query labels for top b candidates, yielding $\mathcal{Q}$
  $\mathcal{L} \leftarrow \mathcal{L} \bigcup \mathcal{Q}$; $\mathcal{U} \leftarrow \mathcal{U} \setminus \mathcal{Q}$
  $\mathcal{M}_t \leftarrow F(\mathcal{L}, \mathcal{M}_{t-1})$; $t \leftarrow t + 1$
until *classification performance is satisfactory*;

Figure 4: AIFT Algorithm

Several researchers [4] [3] have proved the practicability of fine-tuning CNN in biomedical image analysis [5], but they only use fine-tuning once during training process, which means they simply apply fine-tuning on

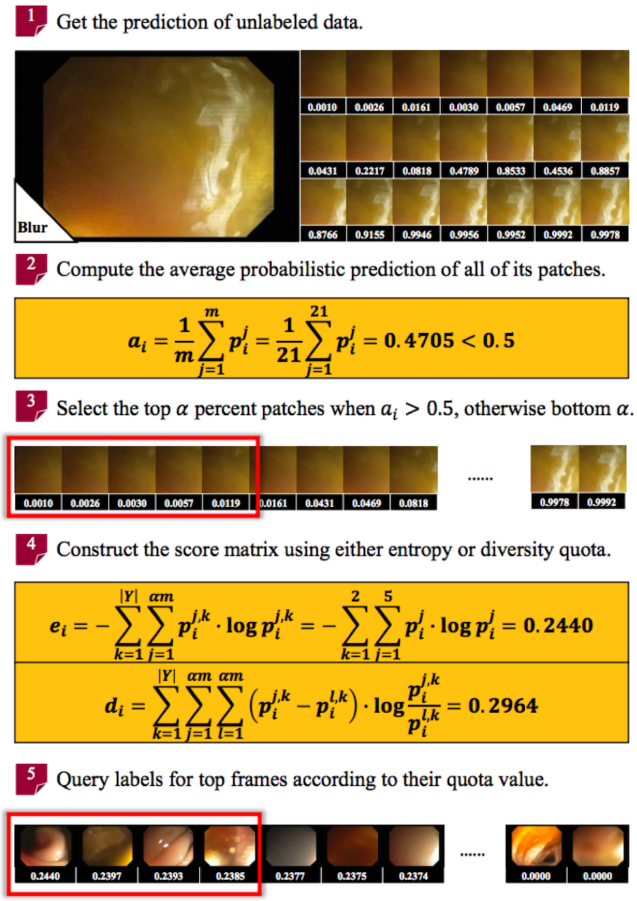

Figure 5: Entropy and Diversity Calculations

a pre-trained CNN without introducing active selection process.

The AIFT integrates Active Learning into fine-tuning CNNS .By means of continuous fashion [13], it makes the application of CNNs in biomedical images more impressive and significantly reduces the cost of labeling data.

Figure 4 and 5 show the specific process of AIFT algorithm and the calculation process of diversity mentioned in Section 2.3 respectively.

Through the analysis of this algorithm, we could discover that compared with the traditional Active Learning, the AIFT has the following advantages:

(1) The initial data-set of AIFT is a completely unlabeled sample set.
(2) Improve the learner by continuous fine-tuning instead of repeating training.
(3) Each candidate corresponds to multiple patches, and the expected consistency of these patches is used to select the samples that are more worthy of priority labeling.
(4) Automatically remove the noisy Tags: for each candidate, only a part of the patches participate in the selection process.
(5) For each candidate, only a small number of patches' enrollment and diversity are calculated, which saves a lot of calculation time.

More importantly, the AIFT has great potential in the field of biomedical images, and has an important impact on the application of computer-aided diagnosis system. Because the current regulations require the computer-aided diagnosis system to be deployed in a closed environment, in which all CAD results will be reviewed and errors if any will be corrected by radiologists. As a result, all false positive results should be discarded and all false negative results should be supplied. Real time online feedback can make the computer-aided diagnosis system learning themselves, and the introduction of continuous fine-tuning proposed improves this possibility.

3 EXPERIMENT VERIFICATION

This section conducted a verification experiment for testing the ability to reduce annotation cost of active learning. We use several data-sets related to classification tasks for this experiment: including MNIST, Cifar-10, Dog-Cat from kaggle, etc. The detailed algorithm is shown in 3.1.

3.1 Experiment Process

Here we show a example of Cifar-10.

1) Separate data: cardinality of active samples set = 50000, cardinality of validation samples set = 10000, cardinality of training samples set = 0.

2) Initialization of the AlexNet model: randomly assign weight to get the initial model with cardinality of training samples set = 0.

3) Predict (50000 - cardinality of training samples set) data from active samples set and get ten probability values corresponding to 10 categories.

4) Focus on the maximum probability value of prediction of each sample: Pmax. We preliminarily believe that if Pmax is greater than 0.5, it means that the current model has a certain classification result for the

sample (the correctness of the classification result is not considered); on the contrary, the prediction of the current model for the sample is ambiguous and will be marked as hard sample.

5) Sort samples that P(real label) < Threshold) (in this case, the 10 classification tasks take threshold = 0.5), and take first n samples to training samples set.

6) Fine-tuning current model with training samples set and get a new model.

7) Repeat steps (3) to (6) until the cardinality of active samples set is 0 or the current model achieves ideal effect.

According to the above algorithm, experiments were carried out on MNIST, Cifar-10 and Dog-Cat data sets. The experimental results show that the introduction of active learning can not only reduce the cost of sample labeling, but also improve the accuracy of classification.

For example, in MNIST experiment (training samples number = 50000, validation samples number = 10000) 1) the model was directly trained 120 times with all training data with accuracy = 98.46%; 2) using active learning and uncertainty strategy (US), only 2600 labeled data Val were needed to reach accuracy = 99.04%.The remaining 47400 samples are put into the trained model for prediction, and the effect of 99.59% is obtained.

It can be seen that for MNIST data sets, only US strategy can achieve significant results. Why can active learning improve the accuracy of classification model? Reference [15] mentioned an explanation that "labeled training data may contain some low-quality samples (noise points), but it will reduce the robustness of the model (model transition fitting noise points). Therefore, the number of labeled training samples should be increased under the condition of ensuring the quality". How to efficiently select unlabeled samples with high contribution to classification and add them to the existing training set is a key problem to be solved in active learning. The experimental results of the above three data-sets are as follows:

3.2 Explanation

1) After the introduction of active learning, the annotation cost can be reduced by more than half on MNIST, Cifar-10 and Dog-Cat data-sets. Although the data set used in the above experiments is relatively simple, it can also prove the effect of active learning.

Method	DataSet	TrainSetSize	ValidSetSize	ValAcc	Acc_SActiveSample
Normal Train	Dog-Cat	30000	10000	93.84%	\
Active Learning	Dog-Cat	7000	10000	94.06%	96.97%
Normal Train	Cifar10	50000	10000	90.55%	\
Random Select	Cifar10	45000	10000	91.02%	\
Active Learning	Cifar10	18000	10000	91.27%	98.66%
Normal Train	MNIST	50000	10000	98.46%	\
Random Select	MNIST	50000	10000	98.44%	\
Active Learning	MNIST	4500	10000	98.89%	99.59%

Figure 6: Experiment Result

2) Active select achieves higher performance than random select with fewer labeled samples.

3) There is another link in the active learning experiment that is using the training model to predict the unselected samples. The AccSActiveSample expresses high accuracy (not listed in the table. From my point of view: AccSActiveSample represents the unselected samples in the active selection process, that is, the current model is enough to distinguish the categories of these samples, so it is no longer necessary to use these samples to fine-tune the model.

4 FURTHER THINKING

4.1 Performance Bottleneck

If some scenarios are encountered in the business requirements and the data needs to be annotated manually, we don't know how much labeled data are needed to get the desired effect, so we hope to get as many annotation samples as possible. But in fact, as shown in the figure below, the performance of the model does not increase infinitely with the increase of the amount of annotation data. The performance of the model will have corresponding bottlenecks, and we focus on how to use as little annotation data as possible to achieve this bottleneck.

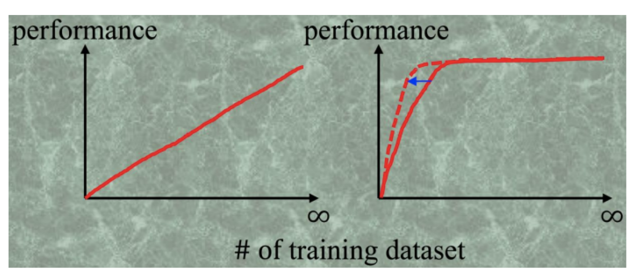

Figure 7: Performance Bottleneck

4.2 How to decrease the critical value for reaching the peak performance?

As explained in section 4.1, the business community can first select a certain amount of data for annotation, and trains the model to see how the effect is; if the performance fails to achieve the expected effect, then increase the annotation samples until the model achieves the expected effect. In fact, it is pretty similar to the above-mentioned active learning process. However, they are essentially different.

1) Generally speaking, the criterion for selecting samples is random selection, and it is more ideal to select some samples that people think are more complicated.

2) Active learning can select the samples which are considered to be the most difficult to distinguish by the current model through some selection strategies and label them to experts in related fields.

To sum up, if some actual business needs to manually label data, it should consider the following conditions to decide whether to use active learning:

1) It is convenient to obtain unlabeled source data, and there are annotation experts in related fields.

2) For some complex tasks, we need to carefully consider whether there are relevant algorithms or models to solve them, otherwise, no matter how much annotation data are spent, the expected effect will not be achieved.

3) For different task types (e.g.classification, detection, NLP, etc.), what kind of indicators should be used to measure samples (hard sample or easy sample).

4) Last but not least: After the algorithm design is completed, we should also consider how to design a complete system. For example, you need to provide an interface for experts to pass the annotated data to the input end of the model.

5 CONCLUSION

1) Active learning has been studied as early as the 1990s, but it seems that it has failed to be popular for a time. It may be due to the lack of some crucial conditions (e.g. data and computation power). With the advent of the era of big data and the rapid development of computation power, deep learning has made great achievements in academic and industrial circles. Various achievements of deep learning also bring possibilities to many fields. In recent years, active learning has begun to stir up in academic circles. Combined with deep learning, active learning has broken through the bottleneck to a certain extent, and achieved good results in some fields (mainly used to reduce the annotation cost).

2) The key of active learning is "query strategy". At present, there are mainly some manual design strategies. A query strategy can only be applied to some specific fields, which is similar to the manual design features of machine learning. Some similar routine: in view of the limitations of manual design features, deep learning combines feature selection with classifier. It no longer need to input manually designed features for classifier and has made a qualitative leap; similarly, the "query strategy" obtained from [10] can be applied to many different fields at the same time, which overcomes the deficiency of cross domain generalization ability of manually designed query strategy.

3) In this paper, through a specific case of medical imaging, we explore the significant contribution of active learning in reducing the annotation cost. At the same time, it also reflects the existing problems of active learning, such as the redundancy of annotations and the poor generalization of query strategies. In the case, we also show some problems that need to be concerned in the future, such as the selection of specific evaluation indicators.

4) The experiment in this paper is only limited to using some simple "Query strategy". We are hoping to come up with more powerful strategies in the future.

REFERENCES

[1] Dana Angluin. 1988. Queries and concept learning. *Machine learning* 2, 4 (1988), 319–342.

[2] Dana Angluin. 2001. Queries revisited. In *International Conference on Algorithmic Learning Theory*. Springer, 12–31.

[3] Gustavo Carneiro, Jacinto Nascimento, and Andrew P Bradley. 2015. Unregistered multiview mammogram analysis with pre-trained deep learning models. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 652–660.

[4] Hao Chen, Qi Dou, Dong Ni, Jie-Zhi Cheng, Jing Qin, Shengli Li, and Pheng-Ann Heng. 2015. Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 507–514.

[5] Mingchen Gao, Ulas Bagci, Le Lu, Aaron Wu, Mario Buty, Hoo-Chang Shin, Holger Roth, Georgios Z Papadakis, Adrien Depeursinge, Ronald M Summers, et al. 2018. Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 6, 1 (2018), 1–6.

[6] Wei-Ning Hsu and Hsuan-Tien Lin. 2015. Active learning by learning. In *Twenty-Ninth AAAI conference on artificial intelligence*. Citeseer.

[7] Sheng-Jun Huang, Rong Jin, and Zhi-Hua Zhou. 2010. Active learning by querying informative and representative examples. In *Advances in neural information processing systems*. 892–900.

[8] Liu Kang, Qian Xu, and Wang Ziqiang. 2012. Survey on Active Learning Algorithms[J]. *Computer Engineering and Applications* 48, 34 (2012), 1–4.

[9] Ashish Kapoor, Kristen Grauman, Raquel Urtasun, and Trevor Darrell. 2007. Active learning with gaussian processes for object categorization. In *2007 IEEE 11th International Conference on Computer Vision*. IEEE, 1–8.

[10] Ksenia Konyushkova, Raphael Sznitman, and Pascal Fua. 2017. Learning active learning from data. *Advances in Neural Information Processing Systems* 30 (2017), 4225–4235.

[11] Anita Krishnakumar. 2007. Active learning literature survey. *Tech. rep., Technical reports, University of California, Santa Cruz.* 42 (2007).

[12] K. Lang and E. Baum. 1992. Query learning can work poorly when a human oracle is used.

[13] Jan Margeta, Antonio Criminisi, R Cabrera Lozoya, Daniel C Lee, and Nicholas Ayache. 2017. Fine-tuned convolutional neural nets for cardiac MRI acquisition plane recognition. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 5, 5 (2017), 339–349.

[14] Burr Settles. 2009. *Active Learning Literature Survey*. Computer Sciences Technical Report 1648. University of Wisconsin–Madison. http://axon.cs.byu.edu/~martinez/classes/778/Papers/settles.activelearning.pdf

[15] Yang Wenzhu. 2017. Research on Progress of Active Learning Algorithm. (2017).

[16] Zongwei Zhou, Jae Shin, Ruibin Feng, R Todd Hurst, Christopher B Kendall, and Jianming Liang. 2019. Integrating active learning and transfer learning for carotid intima-media thickness video interpretation. *Journal of digital imaging* 32, 2 (2019), 290–299.

[17] Zongwei Zhou, Jae Shin, Lei Zhang, Suryakanth Gurudu, Michael Gotway, and Jianming Liang. 2017. Fine-tuning convolutional neural networks for biomedical image analysis: actively and incrementally. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7340–7351.