Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Correspondence:

Uncertainty-aware complementary label queries for active learning*

Shengyuan LIU[1], Ke CHEN[‡2], Tianlei HU[1], Yunqing MAO[3]

[1]*Key Lab of Intelligent Computing Based Big Data of Zhejiang Province, Zhejiang University, Hangzhou 310027, China*

[2]*State Key Laboratory of Blockchain and Data Security, Zhejiang University, Hangzhou 310027, China*

[3]*City Cloud Technology (China) Co., Ltd., Hangzhou 310000, China*

E-mail: liushengyuan@zju.edu.cn; chenk@cs.zju.edu.cn; htl@zju.edu.cn; myq@citycloud.com.cn

Received Nov. 23, 2022; Revision accepted Apr. 20, 2023; Crosschecked Sept. 20, 2023

https://doi.org/10.1631/FITEE.2200589

Many active learning methods assume that a learner can simply ask for the full annotations of some training data from annotators. These methods mainly try to cut the annotation costs by minimizing the number of annotation actions. Unfortunately, annotating instances exactly in many real-world classification tasks is still expensive. To reduce the cost of a single annotation action, we try to tackle a novel active learning setting, named active learning with complementary labels (ALCL). ALCL learners ask only yes/no questions in some classes. After receiving answers from annotators, ALCL learners obtain a few supervised instances and more training instances with complementary labels, which specify only one of the classes to which the pattern does not belong. There are two challenging issues in ALCL: one is how to sample instances to be queried, and the other is how to learn from these complementary labels and ordinary accurate labels. For the first issue, we propose an uncertainty-based sampling strategy under this novel setup. For the second issue, we upgrade a previous ALCL method to fit our sampling strategy. Experimental results on various datasets demonstrate the superiority of our approaches.

1 Introduction

Traditional multi-class classification methods typically necessitate fully-annotated data, a process which can be time-consuming and costly. To mitigate the extensive resource allocation for annotation actions, researchers have explored various ways to learn from weakly supervised annotations (Zhou, 2018): active learning (AL) (Sinha et al., 2019; Yoo and Kweon, 2019), semi-supervised learning (Zhang T and Zhou, 2018), partial label learning (Wang et al., 2019), and others. Among these, AL assumes that different samples in the same dataset have different values for the update of the current model, and tries to select the samples with the highest value from the training set. In this paper, we explore AL and another weakly supervised learning framework named complementary-label learning (Ishida et al., 2017, 2019), as shown in Fig. 1.

Many classic AL methods assume that the ground-truth labels of unlabeled samples can be obtained from an oracle (Ren et al., 2021). The primary objective of these methods is to indirectly cut the annotation costs by minimizing the number of required label queries (Settles, 2011). However, in

‡ Corresponding author

* Project supported by the Key Research and Development Program of Zhejiang Province, China (No. 2021C01009) and the Fundamental Research Funds for the Central Universities, China

ORCID: Shengyuan LIU, https://orcid.org/0000-0002-2880-695X; Ke CHEN, https://orcid.org/0000-0002-3062-0900

© Zhejiang University Press 2023

Fig. 1 Example of an image with full annotation [a] and complementary label [b]. In the complementary-label setting, we know only that this image does not contain a wolf

several practical scenarios, acquiring full annotations for patterns from an oracle remains costly, such as for multi-class classification tasks. Consequently, it raises the question: is it feasible to directly decrease the total annotation costs by reducing the expense associated with a single annotation action?

To tackle this problem, a few active learning approaches have provided some practicable ways to reduce the annotation costs. For example, ALPF (Hu et al., 2019) learners engage in querying semantic information and acquire partial feedback from annotators. Nonetheless, ALPF needs additional prior knowledge to generate semantic questions, and post-pruning exists in the sampling process. WEAKAL (Gonsior et al., 2020) uses typical AL techniques, and uses the outputs of a network as additional weak labels. Nevertheless, when the instances are difficult to annotate, this algorithm is error-prone due to the low quality of weak labels. Zhang CC and Chaudhuri (2015) used both strong and weak labelers to save the limited budget. DuoLab (Younesian et al., 2020) uses a similar method in online active learning. However, these two algorithms still need strong labelers. Inspired by ALPF, we want to obtain weak annotation only from weak labelers without extra prior knowledge. Thus, we try to combine AL with complementary-label learning.

In summary, we tackle a novel ALCL (Liu et al., 2023), as shown in Fig. 2. There are two challenging issues in ALCL: the first is how to learn from complementary and ordinary labels, and the second is how to sample instances to be queried (Liu et al., 2023). To solve the first issue, we upgrade the method named weight redistribution based on the balance of category contribution from Liu et al. (2023), which treated the candidate and complementary labels differently. This approach incorporates instance-wise reweighting into the loss function to concern candidate labels, thereby emphasizing instances with a reduced number of candidate labels. For the second issue, we design a new sampling strategy named uncertainty for sampling and deep learning (USD) based on the strength of ALCL and uncertainty in deep learning. Comprehensive experimental results illustrate that this revised weight redistribution based on the balance of category contribution (WEBB) method outperforms state-of-the-art complementary-label learning algorithms. Furthermore, our USD sampling strategy has indeed been improved upon existing sampling strategies in ALCL.

2 Preliminaries

Our work expands upon the AL framework (Settles, 2009), complementary-label learning (Ishida et al., 2017), and ALCL (Liu et al., 2023), and our proposed USD sampling strategy is based on the uncertainty in deep learning. In this section, we briefly review the uncertainty in deep learning.

There are two main kinds of uncertainty for deep neural networks: epistemic uncertainty and aleatoric uncertainty. Epistemic uncertainty accounts for uncertainty in the model parameters and can be improved with enough data, so it is often referred to as model uncertainty. Aleatoric uncertainty can be further categorized as homoscedastic uncertainty and heteroscedastic uncertainty. Homoscedastic uncertainty remains constant for different inputs, whereas heteroscedastic uncertainty depends on the input samples of the model. For example, some inputs could be more challenging to recognize in image recognition. A model usually has high heteroscedastic uncertainty on inputs with low confidence.

Recently, many researchers have investigated how to estimate and use uncertainty in deep learning (Blundell et al., 2015; Gal and Ghahramani, 2016). With these techniques, uncertainty-based methods are introduced into multi-view learning (Geng et al., 2021), few-shot classification (Zhang ZZ et al., 2020), and multi-task learning (Cipolla et al., 2018). Geng et al. (2021) assumed that inputs are sampled from different Gaussian distributions, and that the variance in Gaussian distributions reflects the uncertainty of inputs. Zhang ZZ et al. (2020) characterized the uncertainty by the variance between inputs and

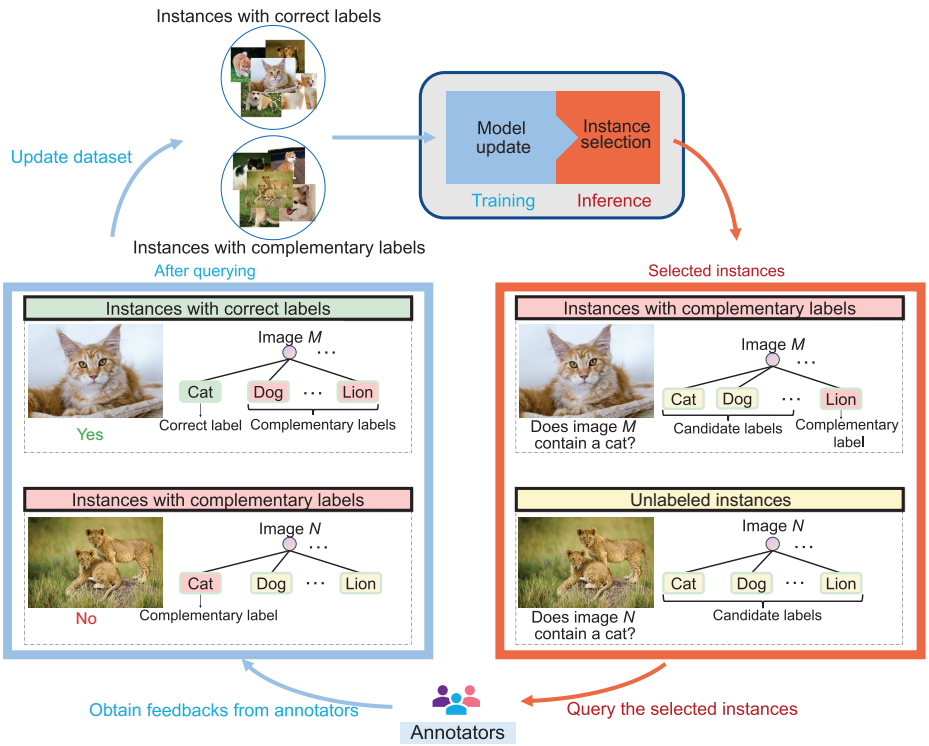

Fig. 2 Workflow for an ALCL learner (ALCL: active learning with complementary labels)

prototypes. Cipolla et al. (2018) used Boltzmann distribution to build the relationship between inputs and prediction, and the conditional probability to handle different tasks. Arnab et al. (2020) introduced uncertainty to weigh the samples with weak annotations in action recognition. Inspired by these studies, our method attempts to use the uncertainty in deep learning to guide the queries of active learning. In other words, the uncertainty in our approach has a dual purpose, deciding which sample to query and guiding how to learn from samples.

3 The proposed framework

In this section, we first introduce two reasonable sampling strategies as baselines. Then, we introduce our USD strategy. Finally, we upgrade the previous ALCL method WEBB.

3.1 Baselines

In the ALCL problem, the number of complementary labels should be considered in sampling strategies. We use a weighting mechanism in Eq. (1) that considers this particularity:

$$w_i = 1 + \alpha \sum_{m \in \bar{Y}_i} p(m|x_i), \tag{1}$$

where $p(m|x_i)$ represents the conventional confidence level associated with instance x_i for label m, and $\bar{Y}_i$ is the complementary label set. The hyperparameter α signifies a preference factor, and when $\alpha > 0$, it indicates a proclivity towards selecting instances exhibiting a smaller number of complementary labels.

In our previous study (Liu et al., 2023), we have applied this weighting scheme on two conventional uncertainty-based sampling strategies: least-confidence (LC) (Culotta and McCallum, 2005) and margin sampling (MS) (Scheffer et al., 2001). From these, two baselines have been established: weighted least-confidence (WLC) and weighted margin sampling (WMS) (Liu et al., 2023).

3.2 Uncertainty for sampling and deep learning

In AL, many uncertainty-based methods use various types of data uncertainty to decide which

sample to query. Although they indeed obtain valuable samples to query, they usually ignore the possible usage of uncertainty in deep learning. Inspired by these, we use one kind of uncertainty to decide which sample to query and guide how to learn from samples.

Our primary purpose is, for the network, to learn from samples with low uncertainty or generate high uncertainty and avoid being penalized heavily for noisy samples where the information is insufficient. The samples with high uncertainty will be queried later. With the original loss, we define our uncertainty loss as

$$\mathcal{L}_{\mathrm{u}}(x,y)=\frac{1}{\sigma^2(x)}\mathcal{L}(x,y)+\log\sigma^2(x), \tag{2}$$

where $\mathcal{L}(x,y)$ is the original loss function and $\sigma^2(x)$ is the uncertainty of x. When $\mathcal{L}(x,y)$ is cross entropy, this corresponds to assuming a Boltzmann distribution on the output of the network with a temperature of σ^2, and approximately minimizing its log-likelihood (Cipolla et al., 2018). For σ, we use an uncertainty prediction module to additionally predict the uncertainty $\sigma^2(x)$ for samples. The number of complementary labels will influence the uncertainty.

In fact, we use the uncertainty prediction module to directly predict $v:=\log\sigma^2$ and obtain $\frac{1}{\sigma^2}=\exp(-v)$ in the following experiments. Because if we predict σ or σ^2, dividing by 0 may happen when computing $\frac{1}{\sigma^2}$.

3.3 WEBB with USD

From Section 3.2, we know that different inputs have different uncertainties in USD, so we cannot obtain loss $\mathcal{L}_{\mathrm{u}}$ from candidate label classification loss in WEBB (Liu et al., 2023). When we use USD as the sampling strategy, candidate label classification loss will be written as

$$\mathcal{L}_{1\mathrm{u}}=\frac{1}{K}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{w_{ik}}{\sigma^2(x_i)\sum\limits_{j=1}^{n}w_{jk}}\left(-\log\sum_{m\notin\bar{Y}_i}p(m|x_i)\right)+\sum_{i=1}^{n}\log\sigma^2(x_i), \tag{3}$$

where w_{ik} is the weight of instance i with lable k, n is the size of the dataset, and K is the number of labels.

Then we can calculate the final training loss with complementary-label classification loss.

4 Experiments

4.1 Experimental setup

1. Datasets and optimization. We use four widely-used benchmark datasets: MNIST (LeCun et al., 1998), Fashion-MNIST (Xiao et al., 2017), Kuzushiji-MNIST (Clanuwat et al., 2018), and CIFAR-10 (Krizhevsky and Hinton, 2009).

In subsequent experiments, the allocated query limit for each dataset is 20 000. For datasets MNIST, Fashion-MNIST, and Kuzushiji-MNIST, a multilayer perceptron (MLP) model ($d-500-k$) is trained as the primary framework, while ResNet (18 layers) (He et al., 2016) is used in CIFAR-10.

The optimization process is conducted with Adam (Kingma and Ba, 2015), applying a weight decay of 1×10^{-4} on weight parameters. Learning rates for each dataset are evaluated from $\{1\times10^{-5},5\times10^{-5},1\times10^{-4},5\times10^{-4},1\times10^{-3}\}$, and reduced by half every 50 epochs. The suggested learning rates are selected by the performance on a validation subset extracted from the training dataset.

2. Baselines and the proposed method. The efficacy of the updated WEBB is evaluated in contrast with two complementary-label learning algorithms, LOG and EXP (Feng et al., 2020). Each of them is parameterized using experimental results acquired from a validation dataset.

For the sampling strategies, we assess the performance of our approach USD in contrast to random sampling, LC (Culotta and McCallum, 2005), MS (Scheffer et al., 2001), WLC, WMS, sequence entropy (SE) (Settles and Craven, 2008), and the learning loss strategy (Yoo and Kweon, 2019). In the SE approach, the candidate labels are taken into consideration collectively for entropy calculation. The learning loss strategy introduces an additional auxiliary module, named the loss prediction module, with the purpose of predicting the loss associated with the target model. This auxiliary module is employed to identify challenging instances for the target model. In the following experiments, the variable α is set to

0.1 for both WLC and WMS. The hidden dimension of the loss prediction module in learning loss is kept at 128 as recommended.

4.2 Comparison among complementary-label learning approaches

We record the comprehensive mean and standard deviation of the test accuracy after every training epoch. In Fig. 3, the mean and standard deviation of test accuracy for LOG, EXP, and our revised WEBB are depicted during the last 250 training epochs.

Fig. 3 shows that the performance of EXP is inconsistent due to its nearly equal treatment of instances with varying numbers of complementary labels as mentioned previously. Equipped with each sampling strategy, the refined WEBB and LOG work well in ALCL, and the refined WEBB outperforms all other methods on various datasets.

4.3 Comparison among active sampling strategies

Fig. 4 presents the mean and standard deviation of accuracy enhancement compared to the random sampling baseline (based on five trials) over the number of queries on CIFAR-10. The random sampling baseline means that the model is trained to converge with data from previous queries, and the sampling of data is random. The improvements of the other methods over the random sampling baseline reflect the quality of sampling under the same number of queries.

Fig. 4 shows that USD outperforms the other sampling strategies substantially, especially in the early stages, because the uncertainty in USD does not significantly depend on the accuracy of the current network. Specifically, USD uses an uncertainty prediction module to guide the sampling of instances. The training of the uncertainty prediction module is easier than the training of the original network, so we can obtain more valuable instances in the early stages with USD.

5 Conclusions

In this paper, we tackle the problem of ALCL (Liu et al., 2023). The objective of ALCL

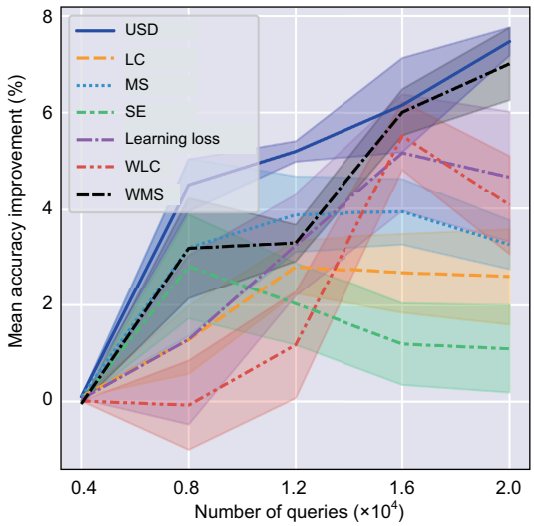

Fig. 4 Mean accuracy enhancements with standard deviation (denoted by shading) of active learning methods compared to the random sampling baseline, across the number of queries on CIFAR-10. Dark colors represent the mean accuracy of the five trials, while light colors indicate the standard deviation. The loss function used in these trials is from WEBB

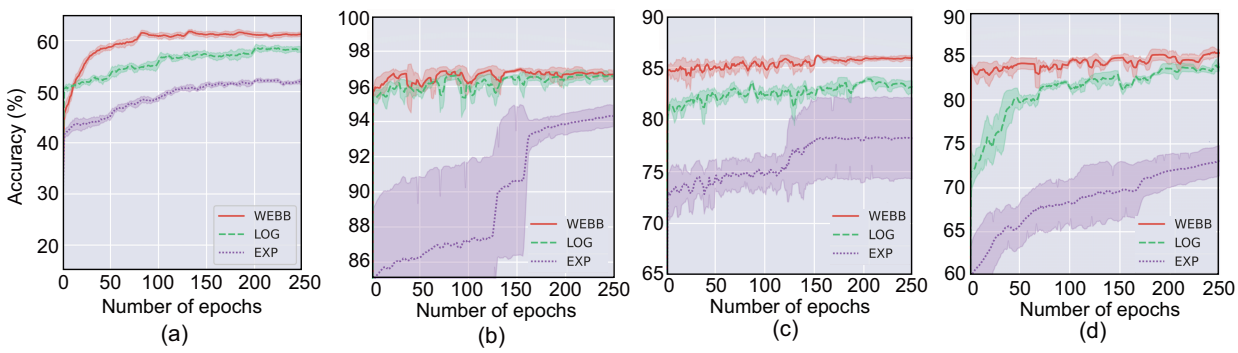

Fig. 3 Experimental results across diverse datasets and models with varied sampling strategies: (a) CIFAR-10; (b) MNIST; (c) Fashion-MNIST; (d) Kuzushiji-MNIST. Dark colors represent the average accuracy derived from the five trials, while light colors denote the corresponding standard deviation

is to directly reduce the cost of annotation actions in AL, while providing a feasible approach for obtaining complementary labels. To solve ALCL, we design a sampling strategy USD, which uses the uncertainty in deep learning to guide the queries of active learning in this novel setup. Moreover, we upgrade the WEBB method to suit this sampling strategy. Comprehensive experimental results validate the performance of our proposed approaches. In the future, we plan to investigate the applicability of our approaches to large-scale datasets and account for noise in the feedback of annotators.

Contributors

Shengyuan LIU designed the research, processed the data, and drafted the paper. Yunqing MAO helped organize the paper. Ke CHEN and Tianlei HU revised and finalized the paper.

Compliance with ethics guidelines

Shengyuan LIU, Ke CHEN, Tianlei HU, and Yunqing MAO declare that they have no conflict of interest.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Arnab A, Sun C, Nagrani A, et al., 2020. Uncertainty-aware weakly supervised action detection from untrimmed videos. Proc 16th European Conf on Computer Vision, p.751-768. https://doi.org/10.1007/978-3-030-58607-2_44

Blundell C, Cornebise J, Kavukcuoglu K, et al., 2015. Weight uncertainty in neural network. Proc 32nd Int Conf on Machine Learning, p.1613-1622.

Cipolla R, Gal Y, Kendall A, 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.7482-7491. https://doi.org/10.1109/CVPR.2018.00781

Clanuwat T, Bober-Irizar M, Kitamoto A, et al., 2018. Deep learning for classical Japanese literature. https://arxiv.org/abs/1812.01718

Culotta A, McCallum A, 2005. Reducing labeling effort for structured prediction tasks. Proc 20th National Conf on Artificial Intelligence, p.746-751.

Feng L, Kaneko T, Han B, et al., 2020. Learning with multiple complementary labels. Proc 37th Int Conf on Machine Learning, p.3072-3081.

Gal Y, Ghahramani Z, 2016. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. Proc 33rd Int Conf on Machine Learning, p.1050-1059.

Geng Y, Han ZB, Zhang CQ, et al., 2021. Uncertainty-aware multi-view representation learning. Proc 35th AAAI Conf on Artificial Intelligence, p.7545-7553. https://doi.org/10.1609/aaai.v35i9.16924

Gonsior J, Thiele M, Lehner W, 2020. WEAKAL: combining active learning and weak supervision. Proc 23rd Int Conf on Discovery Science, p.34-49. https://doi.org/10.1007/978-3-030-61527-7_3

He KM, Zhang XY, Ren SQ, et al., 2016. Deep residual learning for image recognition. IEEE Conf on Computer Vision and Pattern Recognition, p.770-778. https://doi.org/10.1109/CVPR.2016.90

Hu PY, Lipton ZC, Anandkumar A, et al., 2019. Active learning with partial feedback. https://arxiv.org/abs/1802.07427v2

Ishida T, Niu G, Hu WH, et al., 2017. Learning from complementary labels. Proc 31st Conf on Neural Information Processing Systems, p.5639-5649.

Ishida T, Niu G, Menon A, et al., 2019. Complementary-label learning for arbitrary losses and models. Proc 36th Int Conf on Machine Learning, p.2971-2980.

Kingma DP, Ba J, 2015. Adam: a method for stochastic optimization. Proc 3rd Int Conf on Learning Representations, p.14-17.

Krizhevsky A, Hinton G, 2009. Learning Multiple Layers of Features from Tiny Images. MS Thesis, University of Toronto, Toronto, Canada.

LeCun Y, Bottou L, Bengio Y, et al., 1998. Gradient-based learning applied to document recognition. *Proc IEEE*, 86(11):2278-2324. https://doi.org/10.1109/5.726791

Liu SY, Hu TL, Chen K, et al., 2023. Complementary label queries for efficient active learning. Proc 6th Int Conf on Image and Graphics Processing, p.1-7. https://doi.org/10.1145/3582649.3582667

Ren PZ, Xiao Y, Chang XJ, et al., 2021. A survey of deep active learning. *ACM Comput Surv*, 54(9):180. https://doi.org/10.1145/3472291

Scheffer T, Decomain C, Wrobel S, 2001. Active hidden Markov models for information extraction. Proc 4th Int Conf on Intelligent Data Analysis, p.309-318. https://doi.org/10.1007/3-540-44816-0_31

Settles B, 2009. Active Learning Literature Survey. Technical Report No. 1648, University of Wisconsin-Madison, USA.

Settles B, 2011. From theories to queries: active learning in practice. Active Learning and Experimental Design Workshop in Conjunction with AISTATS, Article 18.

Settles B, Craven M, 2008. An analysis of active learning strategies for sequence labeling tasks. Conf on Empirical Methods in Natural Language Processing, p.1070-1079.

Sinha S, Ebrahimi S, Darrell T, 2019. Variational adversarial active learning. IEEE/CVF Int Conf on Computer Vision, p.5971-5980. https://doi.org/10.1109/ICCV.2019.00607

Wang HB, Liu WW, Zhao Y, et al., 2019. Discriminative and correlative partial multi-label learning. Proc 28th Int Joint Conf on Artificial Intelligence, p.3691-3697.

Xiao H, Rasul K, Vollgraf R, 2017. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms. https://arxiv.org/abs/1708.07747

Yoo D, Kweon IS, 2019. Learning loss for active learning. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.93-102. https://doi.org/10.1109/CVPR.2019.00018

Younesian T, Epema D, Chen LY, 2020. Active learning for noisy data streams using weak and strong labelers. https://arxiv.org/abs/2010.14149

Zhang CC, Chaudhuri K, 2015. Active learning from weak and strong labelers. Proc 28th Int Conf on Neural Information Processing Systems, p.703-711.

Zhang T, Zhou ZH, 2018. Semi-supervised optimal margin distribution machines. Proc 27th Int Joint Conf on Artificial Intelligence, p.3104-3110.

Zhang ZZ, Lan CL, Zeng WJ, et al., 2020. Uncertainty-aware few-shot image classification. Proc 30th Int Joint Conf on Artificial Intelligence, p.3420-3426.

Zhou ZH, 2018. A brief introduction to weakly supervised learning. *Nat Sci Rev*, 5(1):44-53. https://doi.org/10.1093/nsr/nwx106