Learn to Active Learn: Dynamic Active Learning with Attention-based Strategies Selection

Seungwhan Moon [1] **Jaime Carbonell** [1]

Abstract

We propose a new dynamic proactive learning framework where the near-optimal instance selection strategy is learned from past annotation history and updated adaptively to the time-varying progress of learner. We motivate our approach from the observation that no single static strategy (*e.g.* uncertainty, density-based, annotator expertise-based, etc.) can choose the optimal samples at every iteration, and that different selection criteria lead to better improvement at different phases of proactive learning. We thus formulate the problem of finding optimal strategy as dynamic attention learning of weights over an ensemble of multiple selection strategies, where attention weights are optimized using a structural SVM framework. Our empirical results on several datasets show that the proposed approach outperforms other baseline ensemble methods as well as strong state-of-the-art active learning strategies that adhere to a single policy. The framework we propose is flexible and thus can work with any other active learning selection criteria, extending the scope beyond the baselines and the simple exploration-exploitation trade-off.

1. Introduction

Active learning has been widely studied and applied in a number of domains to reduce annotation cost and sample complexity in many machine learning or data mining applications (McCallum & Nigam, 2001; Roy & Mccallum, 2001; VanHoudnos et al., 2017; Moon et al., 2014). To practice active learning in real-world scenario, *proactive learning* has been recently proposed, which accommodates for more practical constraints such as different availability or expertise of annotators, extending the scope of the problem to wider application (Donmez & Carbonell, 2010; Yan et al., 2014; Moon & Carbonell, 2014). Most of the recent work on active learning does not address the time-varying progress of learner knowledge state adaptively, and instead adheres to a static policy invariant with respect to new observations, resulting in non-optimal instance selection.

Figure 1 illustrates the motivation of this work and the efficacy of the proposed approach. It is clearly observed in this illustration that (1) no single strategy can directly predict the best actual future improvement all the time, that (2) ground-truth *trends* of the preferred mode of strategies change over time as more samples are annotated, and that (3) sometimes there is no single selection strategy that can single-handily predict future improvement, whereas there exists an ensemble of strategies that predicts better. In reality there are numerous other factors that affect a true underlying optimal strategy than what it is illustrated in this figure, such as the time-varying improvement of labeler accuracy, many of which are subject to change as new observations are made. We therefore aim to design an adaptive proactive learning framework that can address beyond determining when to explore or to exploit, and aim at finding exactly what and how much to explore or exploit, with a careful balance that can approximate the ground-truth optimal strategy. Note that this balance pattern is highly dependent on underlying distribution of a dataset as well as stream of labels obtained from annotators, and thus optimal selection strategies cannot be known *a priori*. As such, we propose to learn a new strategy as an attention-based aggregation of existing strategies, where optimal combination is assumed to be conditional on the labels obtained so far. Additionally, we propose to update this strategy adaptively to a new trend as the true optimal strategy varies with progress of learner and new observations.

In this paper, we formulate the problem of finding optimal proactive learning strategy as dynamic attention learning of weights on ensemble of multiple selection strategies. In essence, we assume that utility of an unlabeled sample can be composed of weighted combination of values measured by multiple strategies, where the optimal attention over strategies is dynamically predicted. The active learning process then reduces to a utility maximization task where utility values of samples are computed based on the learned

[1]Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, U.S.A.. Correspondence to: Seungwhan Moon <me@shanemoon.com>.

preprint (2018)

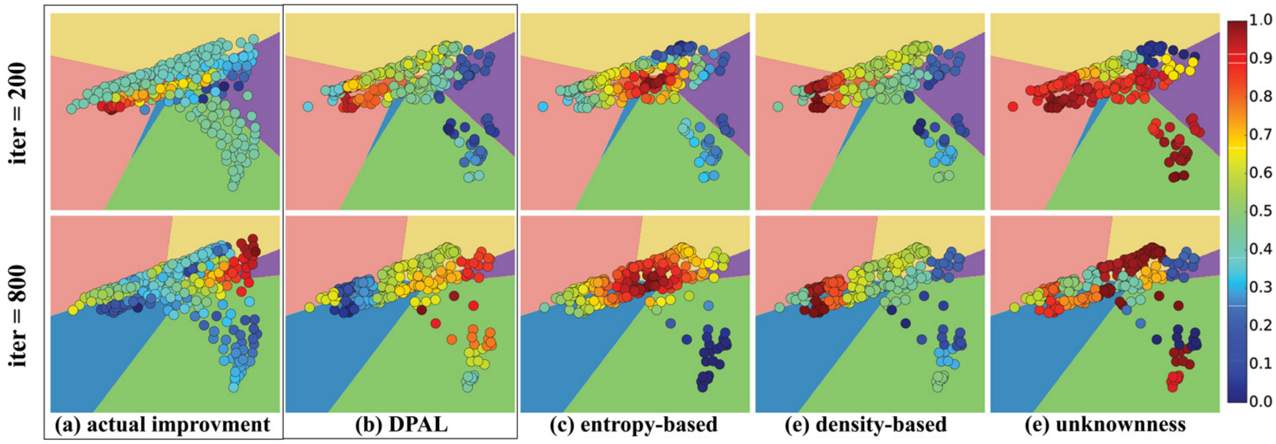


Figure 1. Motivation for dynamic proactive learning, illustrated with UCI Landsat Satellite Dataset (Lichman, 2013) projected on 2-dimensional space via PCA. Each row represents different stages (iteration=200 and 800) of active learning. Drawn in the background are the 5 class-dividing hyperplanes learned thus far, and each dot represents an unlabeled sample. (a) shows actual future improvement of the learner in cross-validation performance when a ground-truth label for a corresponding sample is obtained. (c)-(e) show utility values measured by various active learning strategies. While none of the single strategies serves as a consistently reliable indicator, (b) the dynamic proactive learning framework (DPAL) predicts utility values that match closely to actual future improvement.

optimal strategy.

Our contributions are three-fold: (1) We propose a dynamic proactive learning (DPAL) framework that optimizes between multiple strategies based on the previous annotation history, extending the scope of adaptive active learning beyond the simple exploration-exploitation trade-off. In addition, we consider diverse dimensions of data utility in relation to the proactive learning scenario with multiple imperfect annotators, thus accounting for time-varying estimate of annotators expertise; (2) We formulate our approach using a structural SVM framework, which thus can accommodate any active learning criterion or loss function for measuring importance of different strategies; (3) We dynamically update our strategy given growing observations, thus being robust to time-varying trend of ground-truth optimal strategy.

2. Problem Formulation

We define a pool-based multi-class proactive learning scenario as follows. We have $X = \{x_1, \cdots, x_N\}$ and corresponding ground-truth multi-class labels $Z = \{z_1, \cdots, z_N\}$, each of which is identified with a category $c \in \mathcal{C}$. We are given a pool of annotators $K = \{k^{[1]}, \cdots, k^{[M]}\}$, who have expertise in different areas of the input space with varying degrees. We denote $y_n^{[m]}$ as a label of x_n annotated by $k^{[m]} \in K$. For simplicity we do not allow duplicate label assignment of a sample by multiple annotators, thus $y_n = y_n^{[m]}$ and $k_n = k^{[m]}$ if x_n has been annotated by some annotator $k^{[m]}$, and null otherwise. In a semi-supervised setting, we assume a small subset of labeled set $L = \{(x_n, y_n, k_n) \mid n \in I_L\}$ are known to the learner, where $I_L \subset \{1, \cdots, N\}$, and $|I_L| \ll N$. The unlabeled learning pool then can be defined as $UL = \{(x_n, y_n, k_n) \mid n \in I_{UL}\}$ for $I_{UL} = \{1, \cdots, N\} \setminus I_L$. The proactive learning task is then to choose a subset of UL and a sequence of $k \in \mathcal{K}$ for each instance that will best improve the learner performance, under a fixed budget constraint $B > 0$.

We employ a conventional greedy utility maximization approach (Araya-López et al., 2011; Moon & Carbonell, 2014), where expected utility of a query is measured from a real-valued function $U(x, k, L) : \mathcal{X} \times \mathcal{K} \times \mathcal{L} \rightarrow \mathbb{R}^+$, for a sample $x \in X$ annotated by an annotator $k \in K$, given a set of data labeled so far $L \in \mathcal{L}$. The objective then reduces to finding a pair (x, k) with the highest utility at each iteration, which when annotated, gives the best expected improvement to the learner, *e.g.* $(x^*, k^*) = \text{argmax}_{(x,k) \in X \times K} U(x, k, L)$.

Note that the performance of an active learner thus naturally depends on how well we estimate the true utility function of a problem. Next, we describe how we define the utility function U and learn the most optimal strategy.

3. Dynamic Proactive Learning Framework

We design our DPAL system based upon the structural SVM (*e.g.* (Joachims et al., 2009; Yu & Joachims, 2009)). We set the target utility function as our discriminant function, and assume that the discriminant function is linear in the feature vector $\Psi(x, k, L)$, which describes the utility features mea-

Algorithm 1 Weight Update for DPAL

Input: current $\mathbf{w}$, labeled data L, parameters Q (number of clusters), G (maximum number of comparing pairs)
Output: Updated $\mathbf{w}$
def `updateDPAL`($\mathbf{w}$, L)
  Choose $L' \subset L$ by sequential order
  Let $\{L_q\}_{q=1,\cdots,Q}$ be Q clusters within $L \backslash L'$
  Generate $\min(G, \binom{Q}{2})$ pairs from $\{(L_i, L_j)\}_{i \neq j}$
  $\mathbf{w} := \mathrm{argmin}_{\mathbf{w}} \ \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{N}\sum_{n=1}^{N} \xi_n \ s.t.$ Eq. 8
  return w
end def

sured by multiple strategies (as defined in Section Feature Space) given a sample x, annotator k, and labeled set L. Therefore, the objective of the problem at every proactive learning iteration can be formulated as:

$$(x^*, k^*) = \underset{(x,k)\in X\times K}{\mathrm{argmax}}\ U(x,k,L) = \underset{(x,k)\in X\times K}{\mathrm{argmax}}\ \mathbf{w}\cdot\Psi(x,k,L)$$

Note that $\mathbf{w}$ determines the weighted importance of each feature (strategy) to the learner, and that Φ is a function of L, which is subject to change at every iteration. We therefore learn the best weight distribution $\mathbf{w}^*$ through training and *repeat* this process to dynamically adjust weights as the labeled set L expands. Algorithm 1 explains in detail how we learn and update the optimal $\mathbf{w}^*$ by leveraging the past annotation history. At each active learning iteration we learn a new classifier f using the labeled set L with a support vector machine with RBF kernels (Cristianini & Shawe-Taylor, 2000).

3.1. Feature Space

The feature vector Ψ defines various factors or strategies that account for utility of a query, and thus the exact design is subject to user's choice. In this work, we decompose the feature vector into two main components:

$$\mathbf{w}\cdot\Psi(x,k,L) = \boldsymbol{\alpha}\cdot\Phi(x,k,L) + \boldsymbol{\beta}\cdot\Pi(x,k,L) \tag{1}$$

where the first component $\Phi(x,k,)$ includes a set of features that describe the inherent value of sample x in relation to the currently available labeled set L (*e.g.* uncertainty (Lewis & Catlett, 1994), density (Nguyen & Smeulders, 2004; Zhu et al., 2010), etc.), whereas $\Pi(x,k,L)$ consists of features that relate to an annotator k (*e.g.* probability of obtaining a correct answer from a labeler (Yan et al., 2014), etc.).

In this work, we choose to decompose the first component $\Phi(x,L) = \Phi(x,k,L) \in \mathbb{R}^4$ into four elements, each of which measures uncertainty, density, unknownness, and conflictivity of a sample x, respectively:

$$\Phi_1(x,L) = -\sum_{c\in\mathcal{C}} P(y=c|x;L)\cdot\log P(y=c|x;L) \tag{2}$$

$$\Phi_2(x,L) = \rho(x|X) \tag{3}$$

$$\Phi_3(x,L) = -\rho(x|\{x_n\,|\forall n \in I_L\}) \tag{4}$$

$$\Phi_4(x,L) = -\sum_{c\in\mathcal{C}} P(y=c|q;L)\cdot\log P(y=c|q;L) \tag{5}$$

where $\Phi_1(x,L)$ is the entropy of class posterior distribution of x given a labeled set L which measures uncertainty (Settles & Craven, 2008), $\Phi_2(x,L)$ is the inherent density of samples around x in its distribution, $\Phi_3(x,L)$ is the unknownness of a sample x given the labeled data L that penalizes local regions that have already been explored, which we estimate as inverse of observed density of labeled samples (independent of their labels), and $\Phi_4(x,L)$ is the conflictivity of labels within its local cluster $q \in Q$ (Moon & Carbonell, 2014). We estimate density for Eq.3 and Eq.4 with the non-parametric Gaussian kernel method.

The intuition is that at the beginning of active learning phase $\mathbf{w}$ will put a higher importance to $\Phi_2(x,L)$ and $\Phi_3(x,L)$ to encourage exploration, and gradually put more emphasis on $\Phi_1(x,L)$ to reduce the global entropy. $\Phi_4(x,L)$ reduces the local entropy at conflicting regions (clusters), thus fine-tuning the decision boundaries towards the end of the active learning phase.

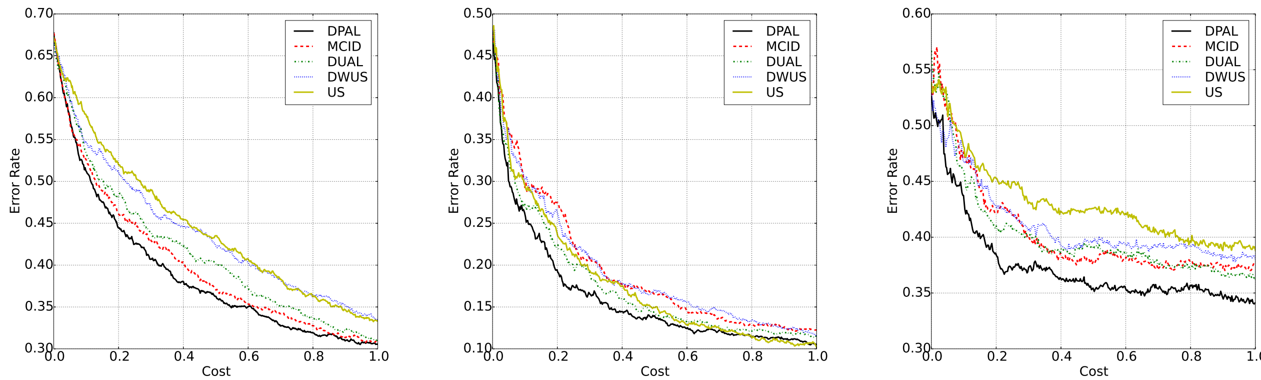
3.2. Learning

We choose to employ a single-valued metric for the second component $\Pi(x,k,L)$, which measures the expected probability of getting a correct answer given an annotator, thus $\Pi(x,k,L) \in \mathbb{R}^1$. Specifically, we define:

$$\begin{aligned}\Pi_1(x,k,L) &= P(ans|x,k) \\ &= \frac{1}{Z_\eta}\sum_{c\in\mathcal{C}} P(y=c|x;L^{[k]})\cdot\log P(y=c|x;L^{[k]})\end{aligned} \tag{6}$$

where $L^{[k]} \subset L$ is a set of labels annotated by an annotator k, Z_η is an entropy normalization constant, and $P(y|x;L^{[k]})$ is the class-posterior probability of an annotator model trained on the annotation observations by k. Note that if an annotator model is confident of its label, its class-posterior entropy will be low, and vice versa. The probability $P(ans|x,k)$ of getting a correct answer for x given k is thus estimated as negative entropy of class-posterior probability of its annotator model, assuming annotator confidence correlates with its accuracy. Note that this approach allows for estimation of $P(ans|x,k)$ without ground-truth samples to compare against, and that the accuracy of this estimation generally improves over time as we observe more labels from each annotator.

In order to learn the optimal vector $\mathbf{w}$ in Eq.(1) that represents the true current weight preference, we leverage



(a) 20 Newsgroups (b) MNIST (c) Covertype

Figure 2. Error rates at normalized cost of queried instances annotated by noiseless labelers, on (a) 20 newsgroups, (b) MNIST, (c) Covertype datasets.

the most recently annotated data from L. Specifically, we choose a subset $L' \subset L$ that contains the oldest samples by their sequential order of annotation, and re-evaluate how informative the most recent annotations $(x, y, k) \in L \backslash L'$ were, given a model trained on L'. We typically choose L' such that $|L'|/|L| = 0.8$. The learning objective of the structural SVM (Yu & Joachims, 2009; Kim et al., 2015) is then:

$$\min_{\mathbf{w},\xi\geq 0} \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{N}\sum_{n=1}^{N}\xi_n \tag{7}$$

$$\begin{aligned} \text{s.t. } \mathbf{w}\cdot(\Psi(x_i,k_i,L') - \Psi(x_j,k_j,L')) \quad (8) \\ \geq \Delta(x_i,k_i,L'),(x_j,k_j,L')) - \xi_i, \\ \forall_{i\neq j}(x_i,k_i),(x_j,k_j) \in L\backslash L' \end{aligned}$$

where ξ_n is a slack variable and C is a regularization parameter, (x_i, k_i) and (x_j, k_j) are two samples being compared drawn from the set $L\backslash L'$. The loss function $\Delta((x_i, k_i, L'), (x_j, k_j, L'))$ is defined such that it penalizes a less optimal pair of the two given L' that leads to a less significant performance improvement (once it is added to L'). We propose to use the difference in total entropy decrease as a loss function:

$$\begin{aligned} &\Delta((x_i,k_i,L'),(x_j,k_j,L')) \quad (9) \\ &= -\frac{1}{Z_\eta}\sum_{n\in I_{UL}}\Big(\eta\big(x_n|L'\cup(x_i,y_i)\big) - \eta\big(x_n|L'\cup(x_j,y_j)\big)\Big) \end{aligned}$$

where η is the entropy of class posterior probability of a sample given a labeled set, and Z_η is a normalization constant. In essence, Eq.(9) measures the observed relative increase in total entropy (uncertainty), thus penalizing a less optimal choice of a sample and an annotator.

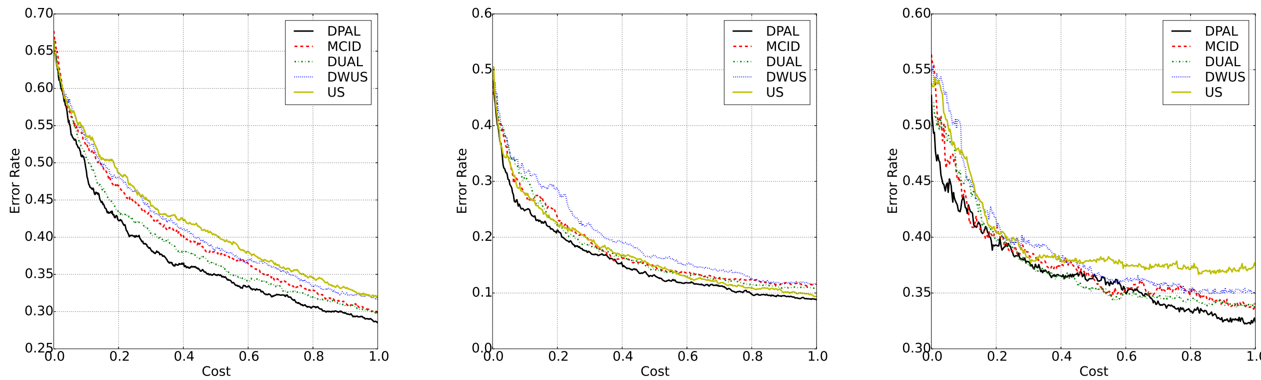
In Eq.(8), $(x_i, k_i), (x_j, k_j) \in L\backslash L'$ can be any possible combination of two instance and annotator pairs, and thus the size of comparing pairs of (x, k) is exponential. To cope with this issue, we limit the generation of negative $(x, k) \in L\backslash L'$ to a fixed number G according to the allowed computational resource. Another challenge to the proposed loss function is that performance improvement signal is often not strong enough when only a single data point is added to a training set. Therefore, instead of comparing a pair of single samples per loss computation, we take a batch approach by first clustering samples in $L\backslash L'$ by their proximity in utility space Ψ, and then treating an average of each cluster as an input to the optimization problem in Eq.(7). This batch update approach not only boosts improvement signal for more robust loss function computation, but also further reduces possible combination pairs for faster optimization. We use alternating optimization that has been widely used for solving structural SVM problems (*e.g.* (Lan et al., 2012; Yu & Joachims, 2009)) at first with random initialization of $\mathbf{w}$, and update the optimal $\mathbf{w}$ periodically (every 10% budget consumption). For each periodic update for $\mathbf{w}_t$ at step t we use the previous weight vector $\mathbf{w}_{t-1}$ as an initialization point, and add a smoothing regularization term $\lambda||\mathbf{w}_t - \mathbf{w}_{t-1}||_2$ to Eq.7.

4. Empirical Evaluation

Below we demonstrate the efficacy of the proposed DPAL framework on several datasets ((Lecun & Cortes; Lichman, 2013): Table 1) on a proactive learning task against several baseline methods. The results are averaged over 10-fold runs for every experiment.

4.1. Task

We evaluate the performance of each baseline in a proactive learning scenario (Yan et al., 2014; Moon & Carbonell, 2014), given a dataset and a pool of simulated annotators each with different expertise, as well as in a traditional



(a) 20 Newsgroups (b) MNIST (c) Covertype

Figure 3. Error rates at normalized cost of queried instances annotated with a pool of noised labelers (labeler noise ratio = 0.3 for non-expertise classes, and 0 for expertise classes.), on (a) 20 newsgroups, (b) MNIST, (c) Covertype datasets.

Table 1. Overview of datasets.

Dataset	# Experts	# Classes	Size
20 Newsgroups	5	20	18846
MNIST	5	10	70000
Covertype	5	10	58000

Table 2. Normalized proactive learning costs at error rate convergence for each dataset with varying DPAL $|L'|/|L|$ ratios. Bold denotes the best performance for each test, and * denotes the statistically significant improvement ($p < 0.05$).

	$\lvert L'\rvert/\lvert L\rvert$			
Dataset	0.6	0.7	0.8	0.9
20 Newsgroups	1.13	1.07	**1.00***	1.16
MNIST	1.08	**0.97**	1.00	0.99
Covertype	1.04	1.10	**1.00**	1.21

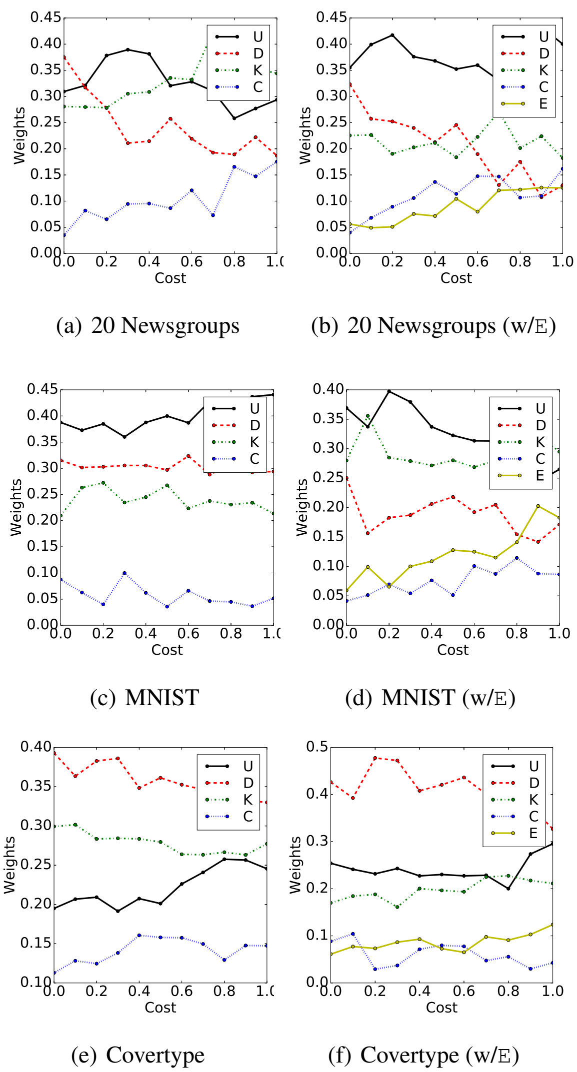
active learning scenario given a single noiseless annotator. We start with a small percentage of labeled samples (0.5%), and at each iteration each learning algorithm chooses a pair of a sample and an annotator to expand its labeled samples pool. The goal is to reach a desired accuracy exhausting as little budget as possible. We generate multiple class-sensitive simulated experts for each dataset (as in Table 1) with varying noise levels, by building classifier models each of which is trained on a set of samples that has ground-truth labels for its expertise area and random labels (at a given noise ratio) for non-expertise areas.

4.2. Baselines

We compare our DPAL method against the following methods from literature: US (uncertainty sampling; (Settles & Craven, 2008) for the single annotator scenario, and (Yan et al., 2014) for the multiple noised annotators scenario), DWUS (density weighted uncertainty sampling; (Zhu et al., 2010)), DUAL (exploration-exploitation switch; (Donmez et al., 2007)), and MCID (multi-class information density; (Moon & Carbonell, 2014)). The difference between DWUS and DUAL is that while DWUS measures utility of a sample as multiplicative composition of its entropy and density, DUAL alternates between uncertainty-based and density-based sampling strategies around an optimal switching point. MCID is one of the state-of-the-art sampling strategies that combines DWUS with unknownness and conflictivity as a multiplicative ensemble to better handle multi-classification active learning problems. Note that most of the referenced papers for the baselines above do not address optimal selection of an annotator and a sample given estimated annotator accuracies. Therefore, we apply and use as a baseline the approach proposed by (Yan et al., 2014), which is to first select a sample based on its distribution-dependent strategy and then to delegate an annotator with the highest estimated probability of giving the correct answer for the chosen sample. Unlike the proactive annotator selection scheme proposed by (Yan et al., 2014), our DPAL framework directly integrates estimated annotator accuracy with base utility of a sample in a jointly optimal way. To separate out the effect of DPAL's joint selection of annotator-sample pairs, we run our experiments with varying noise levels of annotator expertise, including a noise-free case (a perfect oracle annotator).

4.3. Results

Main results: Figure 2 shows the active learning curve at varying amount of annotation cost on different datasets, when labels were annotated by a single noiseless oracle. This is a conventional active learning scenario, and thus we do not consider optimal selection of annotators in this experiment. For most of the datasets, it can be seen that our DPAL method reaches the accuracy at near convergence faster than the baselines, significantly saving the budget it requires



(a) 20 Newsgroups (b) 20 Newsgroups (w/E)

(c) MNIST (d) MNIST (w/E)

(e) Covertype (f) Covertype (w/E)

Figure 4. Normalized DPAL weight transitions for (a), (c), (e): a single noiseless labeler scenario and (b), (d), (f): multiple noised labelers scenario.

to reach the same accuracy. The baselines which heuristically aims to leverage exploration-exploitation balance (`MCID`,`DUAL`,`DWUS`) tend to outperform the `US` baseline which solely aims to exploit the uncertain samples, however the performance boost is not as strong as with `DPAL`. Among the baselines aside from `DPAL`, there is no baseline that consistently wins across the datasets evaluated in this experiment. Note for example that `MCID` does not perform well on some of the datasets in this case, because the noiseless labels tend to neutralize the efficacy of the conflictivity term in `MCID`, which favors to dissolve region around locally heterogeneous labels. `DPAL` learns to suppress non-optimal strategies and balance preferable strategies, leading to better performance overall.

Figure 3, on the other hand, assumes a proactive learning scenario, where annotators are simulated with a class-sensitive noise ratio according to their expertise. A proactive learner thus tries to assign the most knowledgeable annotator for a chosen sample, although the accuracy of expertise estimation is not perfect at the beginning, and tends to improve over time. Because annotated samples include noised labels due to non-optimal expert assignment, it can be seen that baseline performance is different from Figure 2. This result indicates that an optimal strategy for a proactive learning problem is highly dependent on the labelling accuracy of annotators as well as the inherent distribution of each dataset. `DPAL` learns its optimal strategy from the past annotation history, and thus in general outperforms other baselines on most of the datasets more consistently. This result is consistent with the result in Figure 2, which indicates that `DPAL` is flexible to incorporate any number of sampling strategies and optimize for the best ensemble weights given the pool of multiple selection strategies. Note that in practice one can choose to halt proactive active learning process at any desired accuracy, as `DPAL` tends to be more effective towards the beginning of annotation compared to other baselines.

Dynamic weight transitions: Figure 4 shows the dynamic weight transitions for multiple selection strategies (a,c,e: a single noiseless labeler scenario (same setting as Figure 2), b,d,f: multiple domain experts scenario (same setting as Figure 3)), normalized to sum to 1 (`U`: uncertainty, `D`: density, `K`: unknownness, `C`: conflictivity, `E`: estimated annotator expertise). Note that the different components of the weight vector outweigh others at different active learning stages, confirming the observation that there is no single optimal active learning strategy consistently dominant (Figure 1). Note also that the optimal weights are different across the datasets, which shows the need for dynamic weight adjustment learned with DPAL rather than with a heuristic approach. It can be seen that the strategies that encourage exploration (density, unknownness) tend to be given higher weights at earlier stages, while the strategies that encourage exploitation (uncertainty, conflictivity) tend to be more dominant towards the end of the active learning process, which intuitively is a desirable strategy. The weight for estimated expertise of annotators is suppressed at the beginning when expertise estimation is unreliable due to the small number of labeled samples by each annotator to examine with. Later in the active learning process the expertise estimation weight is assigned higher weight, which leads to a more optimal selection of annotators and samples. These weight adjustment behaviors can explain the efficacy of the `DPAL` approach as demonstrated in Figures 2 and 3.

Sensitivity to DPAL hyperparameters: Table 2 shows the `DPAL` performance at varying $|L'|/|L|$ ratios ($= 0.6, 0.7, 0.8, 0.9$), averaged over 10-fold cross validation runs. Each row represents the normalized proactive learning cost to reach convergence in error rate for each dataset, showing how each configuration saves learning budget compared to others. Intuitively, when $|L'|/|L|$ is lower, DPAL evaluates the contribution of each sample conditioned on L' further back in the annotation history, and thus the learned weights $\mathbf{w}$ might not be optimal for current evaluation. When $|L'|/|L|$ is higher, there is less annotation history to leverage ($L \backslash L'$) for learning the optimal strategy, thus being more prone to over-fitting. We do not observe statistically significant improvement for any particular ratio consistent across all of the datasets, and thus for all of our experiments we simply choose $|L'|/|L| = 0.8$ which yields the best average value on the three datasets.

5. Related Work

Exploration-exploitation balance in active learning is well studied and closely related to our work (Yi Zhang, 2003; Melville & Mooney, 2004; Baram et al., 2004; Osugi et al., 2005; Bondu et al., 2010; Loy et al., 2012). The main idea behind these work is that a learning model can quickly improve its performance by first *exploring* diverse areas in its data distribution, and then by "fine-tuning" its hypothesis hyper-plane via *exploitation*. For example, (Donmez et al., 2007) proposes the dual strategy for active learning (DUAL) which switches between the two alternating policies (uncertainty and density) pivoting around a pre-defined threshold value. Our approach extends their work in terms of scalability and flexibility because we avoid use of heuristics, and instead optimize strategy based on stream of observed labels. (Loy et al., 2012) obtains a similar exploration-exploitation balance by extending the conventional query-by-committee (QBC) methods (Seung et al., 1992; Freund et al., 1997; Mamitsuka, 1998) under a stream-based active learning setting. More recently, several studies have investigated active learning in a multi-armed bandit framework (Ganti & Gray, 2013; Salganicoff & Ungar, 2014), which allow for exploration in hypothesis space by calculating lower confidence bounds on the risk of pulling each hypothesis. (Moon & Carbonell, 2016; 2017) apply the idea of exploration-exploitation on joint transfer and active learning settings, in which the active strategy selection applies on the high-level decision of whether to transfer data or to acquire more data. However, most of these work do not address other diverse dimensions that can lead to better active improvement, such as time-varying expertise level of annotators or learner's estimation of annotator expertise given a growing number of observations. Our approach takes into account multiple strategies of user's choice, extending the previous work on exploration-exploitation balance by delving more deeply and precisely into the question of exactly where in data to explore, how much to exploit, and *by asking whom*.

Attention-based learning has been widely applied to various types of deep neural networks (Xu et al., 2015; Sukhbaatar et al., 2015; Yao et al., 2015; Moon et al., 2018a;b). Typically, a simple neural network (*e.g.* feedfoward network) with a softmax layer is added as a submodule to a model which is used to predict importance weights of certain parameters of the main model. Our work is inspired by this line of work, and applies the notion of attention-based learning to the strategy selection problem for active learning.

6. Conclusions

We proposed a new approach that dynamically adjusts the optimal ensemble proactive learning strategy based on the past annotation history. While conventional active learning approaches aim to optimize only for the selection of samples estimated given a static strategy, we optimize for the near-optimal selection or ensemble of multiple strategies adaptive to the time-varying progress of the active learner. In order to achieve this, we designed alternating optimization over SVM problems where an optimal weight vector combining multiple strategies can be learned from past anotation history. We demonstrated that the proposed approach outperforms or matches the performance of other baselines over several datasets by dynamically adjusting weights according to desirable behaviors at each phase.

We note that while the proposed approach learns the optimal strategy for the immediately-next sampling, the myopic approximation does not always guarantee a global optimum. Future work will explore different strategies or reward functions that favor more globally-optimal strategies in order to further improve the performance.

References

Araya-López, M., Buffet, O., Thomas, V., and Charpillet, F. Active learning of mdp models. In *Recent Advances in Reinforcement Learning*, pp. 42–53. Springer, 2011.

Baram, Y., El-Yaniv, R., and Luz, K. Online choice of active learning algorithms. *Journal of Machine Learning Res.*, 5:255–291, December 2004. ISSN 1532-4435.

Bondu, A., Lemaire, V., and Boulle, M. Exploration vs. exploitation in active learning : A bayesian approach. In *Neural Networks (IJCNN), The 2010 International Joint Conference on*, pp. 1–7, July 2010. doi: 10.1109/IJCNN.2010.5596815.

Cristianini, N. and Shawe-Taylor, J. *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods*. Cambridge University Press, New York, NY, USA, 2000. ISBN 0-521-78019-5.

Donmez, P. and Carbonell, J. G. From Active to Proactive

Learning. *Advances in Machine Learning I*, 262:97 – 120, 2010.

Donmez, P., Carbonell, J. G., and Bennett, P. Dual strategy active learning. *In Proceedings of the European Conference on Machine Learning (ECML)*, pp. 116 – 127, 2007.

Freund, Y., Seung, H. S., Shamir, E., and Tishby, N. Selective sampling using the query by committee algorithm. *Journal of Machine learning*, 28(2-3):133–168, 1997.

Ganti, R. and Gray, A. G. Building bridges: Viewing active learning from the multi-armed bandit lens. *Uncertainty in Artificial Intelligence (UAI)*, 2013.

Joachims, T., Finley, T., and Yu, C.-N. J. Cutting-plane training of structural SVMs. *Journal of Machine Learning*, 77:27—-59, 2009.

Kim, G., Moon, S., and Sigal, L. Joint photo stream and blog post summarization and exploration. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

Lan, T., Yang, W., Wang, Y., and Mori, G. Image Retrieval with Structured Object Queries Using Latent Ranking SVM. In *European Conference on Computer Vision (ECCV)*, 2012.

Lecun, Y. and Cortes, C. The MNIST database of handwritten digits. URL http://yann.lecun.com/exdb/mnist/.

Lewis, D. and Catlett, J. Heterogeneous uncertainty sampling for supervised learning. *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 148 – 156, 1994.

Lichman, M. UCI machine learning repository, 2013. URL http://archive.ics.uci.edu/ml.

Loy, C., Hospedales, T., Xiang, T., and Gong, S. Stream-based joint exploration-exploitation active learning. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 1560–1567, June 2012. doi: 10.1109/CVPR.2012.6247847.

Mamitsuka, N. A. H. Query learning strategies using boosting and bagging. In *Machine Learning: Proceedings of the Fifteenth International Conference (ICML)*, volume 1. Morgan Kaufmann Pub, 1998.

McCallum, A. and Nigam, K. Employing EM and pool-based active learning for text classification. *International Conference on Machine Learning (ICML)*, pp. 359 – 367, 2001.

Melville, P. and Mooney, R. J. Diverse ensembles for active learning. *International Conference on Machine Learning (ICML)*, pp. 74, 2004.

Moon, S. and Carbonell, J. Proactive learning with multiple class-sensitive labelers. *International Conference on Data Science and Advanced Analytics (DSAA)*, 2014.

Moon, S. and Carbonell, J. Proactive transfer learning for heterogeneous feature and label spaces. *ECML-PKDD*, 2016.

Moon, S. and Carbonell, J. Completely heterogeneous transfer learning with attention: What and what not to transfer. *IJCAI*, 2017.

Moon, S., McCarter, C., and Kuo, Y.-H. Active learning with partially observed data. *Proceedings of International World Wide Web Conference (WWW)*, 2014.

Moon, S., Neves, L., and Carvalho, V. Multimodal named entity recognition for short social media posts. *NAACL*, 2018a.

Moon, S., Neves, L., and Carvalho, V. Zeroshot multimodal named entity disambiguation for noisy social media posts. *ACL*, 2018b.

Nguyen, H. and Smeulders, A. Active learning using pre-clustering. *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 79, 2004.

Osugi, T., Kim, D., and Scott, S. Balancing exploration and exploitation: a new algorithm for active machine learning. In *Data Mining, Fifth IEEE International Conference on*, pp. 8 pp.–, Nov 2005. doi: 10.1109/ICDM.2005.33.

Roy, N. and Mccallum, A. Toward optimal active learning through sampling estimation of error reduction. In *In Proc. 18th International Conference on Machine Learning (ICML)*, pp. 441 – 448. Morgan Kaufmann, 2001.

Salganicoff, M. and Ungar, L. H. Active exploration and learning in real-valued spaces using multi-armed bandit allocation indices. In *Proceedings of the Twelfth International Conference on Machine Learning (ICML)*, pp. 480–487, 2014.

Settles, B. and Craven, M. Training text classifiers by uncertainty sampling. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1069 – 1078, 2008.

Seung, H. S., Opper, M., and Sompolinsky, H. Query by committee. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pp. 287–294. ACM, 1992.

Sukhbaatar, S., Weston, J., Fergus, R., et al. End-to-end memory networks. In *NIPS*, pp. 2440–2448, 2015.

VanHoudnos, N., Casey, W., French, D., Lindauer, B., Kanal, E., Wright, E., Woods, B., Moon, S., and Jansen, P. "this malware looks familiar: Laymen identify malware run-time similarity with chernoff faces and stick figures". In *Proceedings of the 10th EAI International Conference on Bio-inspired Information and Communication Technologies*, 2017.

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., Zemel, R. S., and Bengio, Y. Show, attend and tell: Neural image caption generation with visual attention. *arXiv preprint arXiv:1502.03044*, 2(3):5, 2015.

Yan, Y., Rosales, R., Fung, G., Subramanian, R., and Dy, J. Learning from multiple annotators with varying expertise. *Journal of Machine Learning*, 2014.

Yao, L., Torabi, A., Cho, K., Ballas, N., Pal, C., Larochelle, H., and Courville, A. Describing videos by exploiting temporal structure. In *ICCV*, pp. 4507–4515, 2015.

Yi Zhang, Wei Xu, J. C. Exploration and exploitation in adaptive filtering based on bayesian active learning. *International Conference on Machine Learning (ICML)*, 2003.

Yu, C.-N. J. and Joachims, T. Learning Structural SVMs with Latent Variables. In *International Conference on Machine Learning (ICML)*, 2009.

Zhu, J., Wang, H., Tsou, B., and Ma, M. Active Learning With Sampling by Uncertainty and Density for Data Annotations. *IEEE Transactions on Audio, Speech, and Language Processing*, 18, 2010.