Active Multitask Learning with Committees

Jingxi Xu[1], Da Tang[1] and Tony Jebara[12]
Columbia University[1], Netflix Inc.[2]

Motivation

1. The cost of annotating training data has traditionally been a bottleneck for supervised learning approaches.
2. The problem is further exacerbated when supervised learning is applied to a number of correlated tasks simultaneously since the amount of labels required scales with the number of tasks.

Contribution

- Propose an active multitask learning algorithm that achieves knowledge transfer between tasks.
- Our approach reduces the number of queries needed during training while maintaining high accuracy on test data.

Proposed Method

Active Multitask Learning with Committees (AMLC)

- Forms a *committee* for each task that jointly makes decisions.
- Learns committee weights on the fly.
- Directly shares data across similar tasks.

function AMLC (b, C, T)
  Initialize $w_m^{(0)} = \mathbf{0}_D, \forall m \in [K], \boldsymbol{\tau}^{(0)} = \frac{1}{K}\mathbf{1}_{K\times K}$
  for $t = 1, 2, .., T$ **do**
    Receive $(x^{(t)}, k)$
    Compute $p_{km}^{(t)} = \langle x^{(t)}, w_m^{(t-1)} \rangle$ for $m \in [K]$
    $p = \sum_{m\in[K]} p_{km}^{(t)} \tau_{km}^{(t-1)}$
    Predict $\hat{y}^{(t)} = \text{sign}(p)$
    Draw $P^{(t)} \sim$ Bernoulli $\left(\frac{b}{b+|p|}\right)$.
    if $P^{(t)} = 1$ **then**
      Query true label $y^{(t)}$ and set $M^{(t)} = \mathbb{1}\left[y^{(t)} \neq \hat{y}^{(t)}\right]$
      Update $w_k^{(t)} = w_k^{(t-1)} + P^{(t)} M^{(t)} y^{(t)} x^{(t)}$
      Update $\boldsymbol{\tau}$:

$$\tau_{km}^{(t)} = \frac{\tau_{km}^{(t-1)} e^{-C \cdot \frac{l_{km}^{(t)}}{\lambda}}}{\sum_{m'\in[K]} \tau_{km'}^{(t-1)} e^{-C\cdot \frac{l_{km'}^{(t)}}{\lambda}}}, m \in [K]$$

      for $\forall m \in [K]$, and $m \neq k$ **do**
        Set $S_m^{(t)} = \mathbb{1}\left[\text{sign}\left(p_{km}^{(t)}\right) \neq \hat{y}^{(t)} \wedge \tau_{km}^{(t)} \geq \tau_{kk}^{(t)}\right]$
        Update $w_m^{(t)} = w_m^{(t-1)} + S_m^{(t)} y^{(t)} x^{(t)}$
      end for
    end if
  end for
  return $\boldsymbol{\tau}^{(t)} \boldsymbol{w}^{(t)}$
end function

Comparison with PEER

The most related work to ours is ***active learning from peers (PEER)*** (Murugesan & Carbonell, 2017). There are two main differences.

1. AMLC does not treat the task itself and its peer tasks separately. This helps the learner to recover from **blind confidence** and encourages information-poor tasks to learn more from information-rich tasks.
2. AMLC enables the sharing of training data across similar tasks directly. When tasks are identical, it converges to a relationship matrix with equal elements and effectively all tasks are trained on all training examples.

Experiments

- Compare the accuracy and number of queries after traversing the whole training set (Table 1).
- Compare the accuracy given a limited number of queries (Figure 1).

Models	Landmine Detection		Spam Detection		Sentiment	
	Accuracy	*#Queries*	*Accuracy*	*#Queries*	*Accuracy*	*#Queries*
Random	0.8914 ±0.0126	2323.5 ±11.5	0.7940 ±0.0131	751.8 ±14.3	0.6068 ±0.0065	1092.0 ±16.4
Independent	0.9070 ±0.0079	2770.3 ±25.8	0.8232 ±0.0159	1188.6 ±6.6	0.6404 ±0.0050	1987.5 ±7.1
PEER	0.9362 ±0.0025	1206.0 ±23.2	0.8334 ±0.0134	1085.7 ±13.9	0.6425 ±0.0067	1979.7 ±10.3
PEER+Share	0.9231 ±0.0112	1885.8 ±71.3	**0.8766 ±0.0135**	935.3 ±18.3	**0.7645 ±0.0077**	1754.5 ±9.2
AMLC	**0.9367 ±0.0020**	**189.1 ±12.0**	**0.8706 ±0.0102**	**321.3 ±15.6**	0.7358 ±0.0093	**633.9 ±11.4**

Table 1: Accuracy on test set and total number of queries during training over 10 random shuffles of the training examples. The 95% confidence level is provided after the average accuracy. The best performance is highlighted in bold.

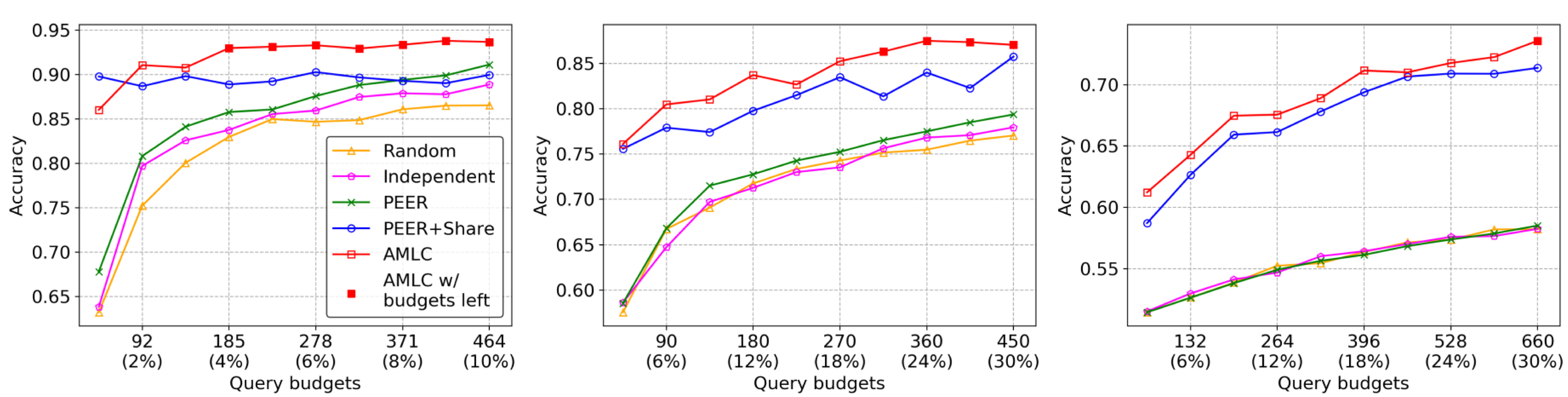

Figure 1: Accuracy on test set w.r.t. query budget. Left, middle and right are Landmine Detection, Spam Detection and Sentiment respectively. The filled markers indicate that there are still queries left in the budget.

Future Research

- Theoretical analysis of the error bound.
- Handle unbalanced task data.

Contact Information

Web https://jxu.ai
Email jingxi.xu@columbia.edu

References

[1] Keerthiram Murugesan and Jaime Carbonell. Active learning from peers. In *Advances in Neural Information Processing Systems*, pages 7008–7017, 2017.