Active Learning with Hinted Support Vector Machine

Chun-Liang Li Chun-Sung Ferng Hsuan-Tien Lin

National Taiwan University

2012/11/6

Pool-Based Binary Classification Active Learning

Given

- The labeled pool $\mathcal{D}_l = \{(\textit{feature } \mathbf{x}_i, \textit{label } y_i)\}_{i=1}^{N}$, $y_i \in \{+1, -1\}$
- The unlabeled pool $\mathcal{D}_u = \{\tilde{\mathbf{x}}_j\}_{j=1}^{M}$

A pool-based active learning algorithm **iteratively**

- use **querying algorithm** $\mathcal{Q}$ to query $\tilde{\mathbf{x}}_s \in \mathcal{D}_u$
- update $\mathcal{D}_l$ and $\mathcal{D}_u$
- learn a **decision function** $f^{(r)}$ by **learning algorithm** $\mathcal{L}$

and improve the performance of $f^{(r)}$ w.r.t #queries

Goal

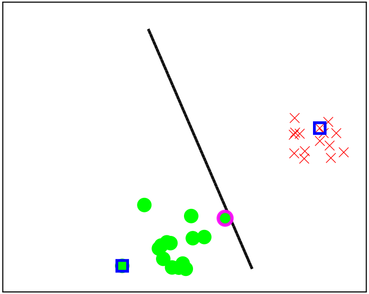
use few queries to improve performance of decision function

Uncertainty Sampling (A Popular Paradigm)

In each iteration, query the least certain one

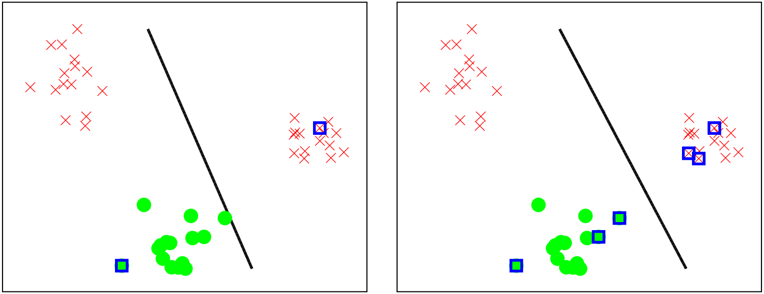
Tong and Koller (2000)

- learn a SVM hyperplane for choosing the instance closest to the boundary
- use the same hyperplane for querying and learning



- blue framed:labeled instances
- magenta circled: to be queried

Potential Drawback



(a) Initial Stage (b) After #iterations

be **overly confident** to unknown area

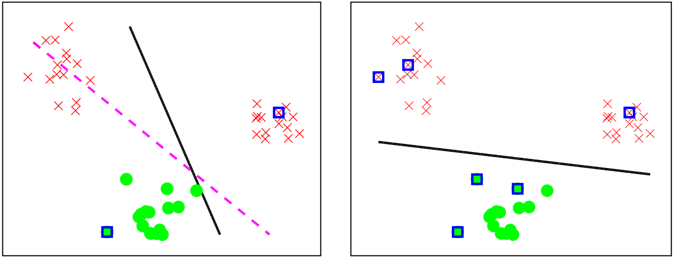
Representative Sampling

- clustering-based algorithms (Donmez et al., 2007)
- label estimation in semi-supervised learning (Huang et al., 2010)

Hinted Sampling

Intuition

Use some unlabeled instances $\mathcal{D}_h \subseteq \mathcal{D}_u$ as hints (Abu-Mostafa, 1995) to make querying boundary be aware of (pass through) unknown areas



(c) Initial Stage (d) After #iterations

querying boundary is different from the decision boundary (black)

Active Learning with Hinted SVM (ALHS)

- Separate $\mathcal{D}_l \rightarrow$ classification problem
- Pass through $\mathcal{D}_h \rightarrow$ regression problem

HintSVM (For querying)

$$\min_{\mathbf{w},b,\xi,\tilde{\xi},\tilde{\xi}^*} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + C_l \sum_{i=1}^{|\mathcal{D}_l|} \xi_i + C_h \sum_{j=1}^{|\mathcal{D}_h|} \left(\tilde{\xi}_j + \tilde{\xi}_j^*\right)$$

$$\text{subject to} \quad \begin{aligned} y_i(\mathbf{w}^T\mathbf{x}_i + b) &\geq 1 - \xi_i \quad \text{for } (\mathbf{x}_i, y_i) \in \mathcal{D}_l, \\ \mathbf{w}^T\tilde{\mathbf{x}}_j + b &\leq \epsilon + \tilde{\xi}_j \quad \text{for } \tilde{\mathbf{x}}_j \in \mathcal{D}_h, \\ -(\mathbf{w}^T\tilde{\mathbf{x}}_j + b) &\leq \epsilon + \tilde{\xi}_j^* \quad \text{for } \tilde{\mathbf{x}}_j \in \mathcal{D}_h. \end{aligned}$$

- A convex optimization problem
- Uncertainty sampling with SVM is a **special case** of ALHS ($C_h = 0$)

Active Learning with Hinted SVM (ALHS)

ALHS

Our algorithm ALHS **iteratively**

- select $\mathcal{D}_h \subseteq \mathcal{D}_u$
- use HintSVM in **querying algorithm** $\mathcal{Q}$ to query $\tilde{\mathbf{x}}_s \in \mathcal{D}_u$
- update $\mathcal{D}_l$ and $\mathcal{D}_u$
- learn a typical SVM $f^{(r)}$ as decision function by **learning algorithm** $\mathcal{L}$

Comparison and Contribution

Comparison

	Querying Algo $\mathcal{Q}$	Learning Algo $\mathcal{L}$
Uncertainty Sampling	Typical SVM	Typical SVM
ALHS	HintSVM	Typical SVM

Contributions

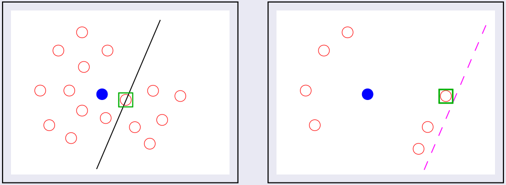
- Resolve potential drawback of uncertainty sampling
- Convex Optimization - Simpler than some representative algos
- achieve better performance

Hint Selection Strategies in ALHS

- HintSVM: $\min \frac{1}{2}\mathbf{w}^T\mathbf{w} + C_l \sum_{i=1}^{|\mathcal{D}_l|} \xi_i + C_h \sum_{j=1}^{|\mathcal{D}_h|} \left(\widetilde{\xi}_j + \widetilde{\xi}_j^*\right)$
- Balance cost parameters $C_l = \max\left(\frac{|\mathcal{D}_h|}{|\mathcal{D}_l|}, 1\right) \times C_h$

Hint Dropping

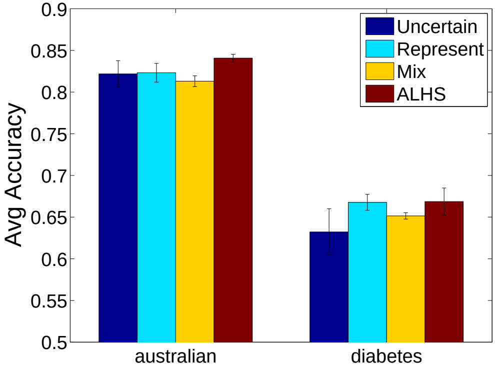
Too many hints would overwhelm HintSVM



- hints surrounding to a labeled instance are less useful
- Drop all $\mathcal{D}_h$ after T iterations (similar to Donmez et al., 2007)

Experiment I

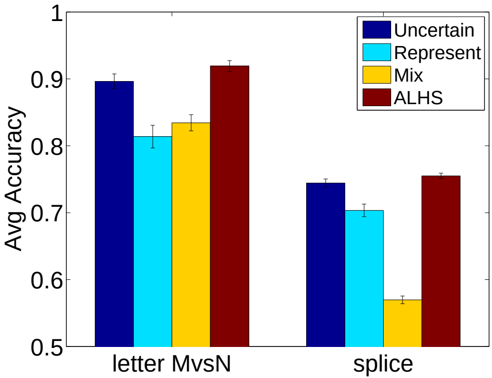
The datasets that representative sampling outperforms uncertainty sampling



ALHS can compete with or even outperforms them

Experiment II

The datasets that uncertainty sampling outperforms representative sampling



ALHS can compete with or even outperforms them

Conclusion

- general framework of active learning: Hinted Sampling
- HintSVM: convex optimization, simpler
- good experimental results
- future work: a new direction for theoretical analysis of representative sampling

Thanks, any question?