Types of Learning

• Supervised
• Unsupervised
• Semi-supervised
• Reinforcement
Why AL?

world $\rightarrow$ passive learner $\rightarrow$ model
Why AL?

world $\rightarrow$ active learner model
Why AL?
Applications

Information Extraction, Classification and Filtering

- Bioinformatics
- Speech Recognition
- Web
- Image retrieval
  and many others
Active researchers in AL

Burr Settles
“Active Learning Literature Survey”
http://www.cs.cmu.edu/~bsettles/

Sanjoy Dasgupta
http://cseweb.ucsd.edu/~dasgupta/
Approaches to AL
Query Strategies

- Uncertainty Sampling
- Query-by-Committee
- Expected Model Change
- Expected Error Reduction
- Variance Reduction
- Density-Weighted Methods
Query Strategies: uncertainty sampling

• Least Confident Prediction

\[ x_{LC}^* = \arg\max_x 1 - P_{\theta}(\hat{y}|x) \]

\[ \hat{y} = \arg\max_y P_{\theta}(y|x) \]
Query Strategies:
uncertainty sampling

• Entropy

\[ x^*_H = \arg\max_x - \sum_i P_\theta(y_i|x) \log P_\theta(y_i|x) \]
Query by committee
Query by committee

• Construct a committee of models
• Define disagreement measure
Query by committee

• Construct a committee of models
  – Sampling from posterior distribution of models given labeled data
  – Partitioning feature space
  – Diversity

• Define disagreement measure
Query by committee: disagreement measures

- Vote entropy

$$x_{VE}^* = \arg\max_x - \sum_i \frac{V(y_i)}{C} \log \frac{V(y_i)}{C}$$
Query by committee: disagreement measures

- K-L divergence

\[
x_{KL}^* = \arg\max_x \frac{1}{C} \sum_{c=1}^{C} D(P_{\theta(c)} \parallel P_c),
\]

\[
D(P_{\theta(c)} \parallel P_c) = \sum_i P_{\theta(c)}(y_i|x) \log \frac{P_{\theta(c)}(y_i|x)}{P_c(y_i|x)}
\]
Analysis of AL

Does Active learning work?
Analysis of AL

Empirical Analysis

Caveats:
1. A training set with an active learner inherently tied to a model that was used to generate it
2. Learning correlated with the proficiency of the learner
3. 91% researcher who used active learning had their expectations fully or partially met [Tomanek and Olson 2009]

Theoretical Analysis
1. Number of queries required to learn a model
2. This number less than passive supervised learning?
3. Active learning improves on passive learning if noise rate is bounded
Practical Considerations in AL

- Batch Mode AL
- Noisy Oracles
- Variable Labeling Costs
- Alternative Query Types
- Multi-Task Active Learning
- Unknown / Changing Model Class
- Stopping Criteria
Practical Considerations in AL

Batch Mode AL

- Serial query is slow or expensive
- Batch mode AL allows query instances in groups
- How to properly assemble optimal query set $Q$
- Instances in the batch are both diverse and informative
- Better than random batch sampling, simple “Q-best” batch
Practical Considerations in AL

Noisy Oracles

- Quality of labeled data is high
- Labels can come from empirical / human experts
- Crowd sourcing
- When to repeat label to de-noise an existing training instance
- Query only the more reliable annotators in subsequent iteration
Practical Considerations in AL

Noisy Oracles

Open Research Questions

- Noisy Oracle whose quality varies over time
- How payment influence annotation quality
- Inherently noisy instance, repeated annotation doesn’t help
- Crowdsourcing annotators are not readily available, how can learning progress?
Practical Considerations in AL

Variable Labeling Costs

- Labeling cost and quality can vary from instance to instance
- Reducing number of labeled instances doesn’t guarantee reduction of overall labeling cost
- Take into account both labeling cost + misclassification cost
- Real world Human annotation cost on four data-sets [Settles et al 2008a]
  - Annotation cost can vary considerably
  - Active learning ignoring cost may not perform better than passive learning (random sampling)
  - Annotation cost may vary on person doing it
  - Annotation cost also may depend on jitter and pause

Alternate query types:
- Labeling features than instances
Practical Considerations in AL

Alternative Query Types

- Multiple Instance (MI) active learning
- Instances can be grouped into bags
- Bags can be labeled (positive / negative)

Figure 1. Context based Image retrieval (CBIR) 2. Text classification[1]
Practical Considerations in AL

Multi-Task Active Learning

- Same data instance can be labeled in multiple ways for different subtasks
- Single query will be labeled for multiple tasks
- Two-task active learning scenario
  - alternating selection
  - rank combination
- Images can be labeled for several binary classification tasks in parallel
e.g. image = {beach, sunset, mountain, field}
Practical Considerations in AL

Unknown / Changing Model Class

- Training set built via active learning comes from biased distribution which is tied to a class
- Reuse this training data with different model type or unknown class
- Decision Tree performs well on Naïve Bayes trained data using uncertainty sampling
- If best model / features are known in advance, or unlikely to change in future – AL can be used safely
- Use of heterogeneous ensembles in selecting queries
Practical Considerations in AL

Stopping Criteria

- Cost of acquiring new training data is more than cost of errors made by current model
- Accuracy of learning reached plateau
- Active learning may decide to stop asking question to conserve resource
- Real stopping criteria based on practical applications based on economic, external factors
References
