Investigating Active Learning for Concept Prerequisite Learning

Chen Liang[1], Jianbo Ye[1], Shuting Wang[1], Bart Pursel[2], C. Lee Giles[1]

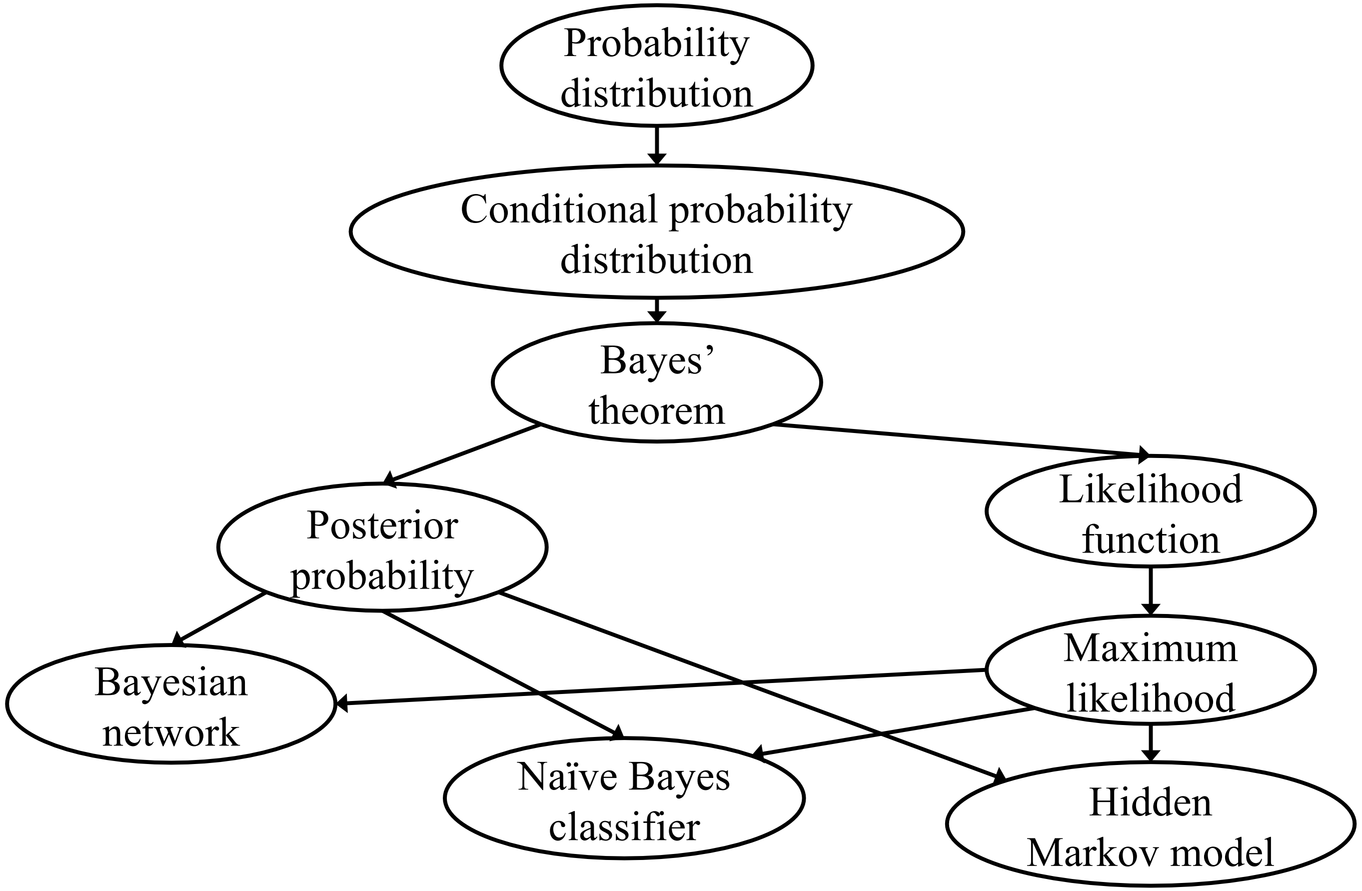
[1]Information Sciences and Technology

[2]Teaching and Learning with Technology

Pennsylvania State University

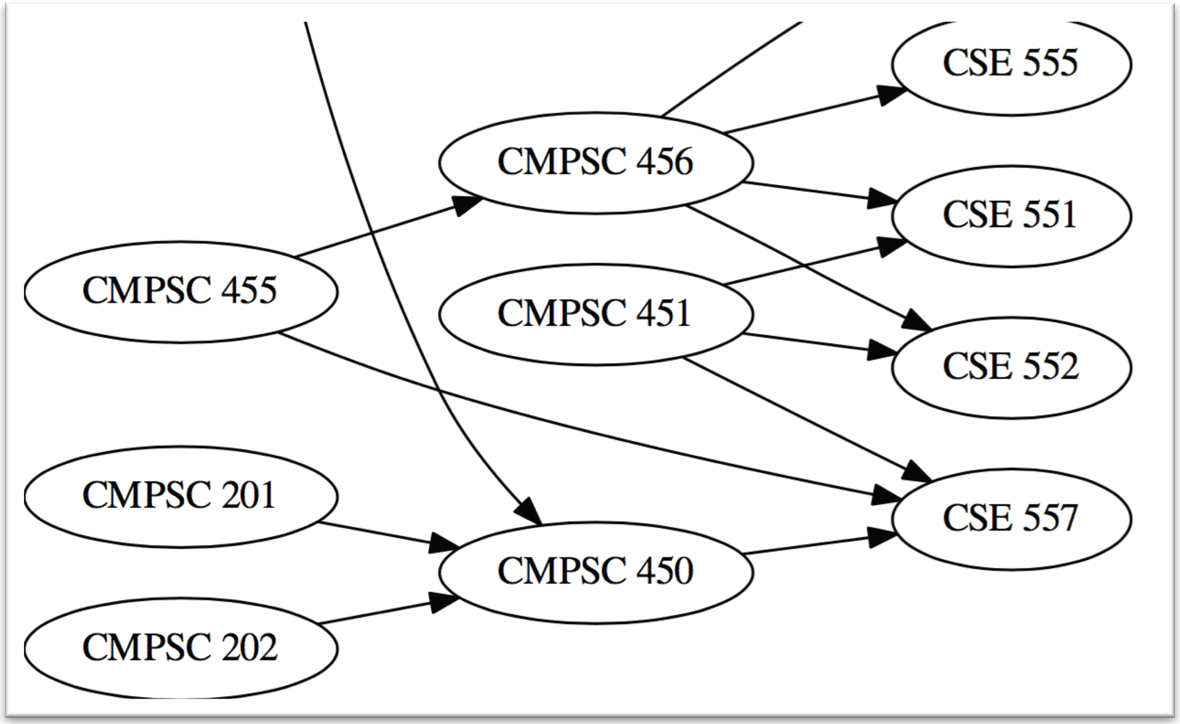
Concept Prerequisite Relations

- A fundamental relation among concepts in knowledge structures



Motivation

- Prerequisites are important for many educational applications
  - Intelligent tutoring system; curriculum planning; transferring credits, …
- Time-consuming process



Manually Created

Motivation

- Prerequisites are important for many educational applications
  - Intelligent tutoring system; curriculum planning; transferring credits, …
- Time-consuming process

What if we want to organize concepts from thousands of MOOCs?

Concept Prerequisite Learning

Goal:

Predict whether a concept *A* is a prerequisite of a concept *B* given the pair (*A*,*B*). [Talukdar, BEA’12; Liang, EMNLP’15]

Binary classification problem

Related Work

- Earlier work in educational data mining [Vuong, EDM’11; Scheines, EDM’14; Chen, EDM’15]
  - Analyzes student assessment data which records the performance of students on different items (e.g. units, sections, etc.)
  - Requires that the association between test items and handcrafted knowledge components is set beforehand.
  - Often subject to a specific assessment pool, not applicable for processing a large concept set.

Related Work (cont.)

- Classification framework

Lack of large-scale prerequisite labels

  - Wikipedia-based features [Talukdar, BEA'12; Liang, EMNLP'15; Wang, CIKM'16]
  - Information-theoretic metrics [Gordon, ACL'16]
  - MOOC-related features [Pan, ACL'17]
- Learning from course prerequisites [Yang, WSDM'15; Liang, EAAI'17]

Active learning has not been applied to the concept prerequisite learning problem

Our Contribution

- The first to apply active learning to concept prerequisite learning
- Investigate three families of query selection strategies
- Propose a novel set of features for representing concept pairs

Pool-based Active Learning

Algorithm 1 Pseudocode for pool-based active learning.

Input:
  $\mathcal{D} \leftarrow \{1, 2, ..., n\}$ % a data set of n instances
Initialize:
  $\mathcal{D}_l \leftarrow \{s_1, s_2, ..., s_k\}$ % initial labeled set with k seeds
  $\mathcal{D}_u \leftarrow \mathcal{D} \backslash \mathcal{D}_l$ % initial unlabeled set
while $\mathcal{D}_u \neq \emptyset$ **do**
    Select s^* from $\mathcal{D}_u$ % according to a query strategy
    $\mathcal{D}_u \leftarrow \mathcal{D}_u \backslash \{s^*\}$
    Query the label y_{s^*} for the selected instance s^*
    $\mathcal{D}_l \leftarrow \mathcal{D}_l \cup \{s^*\}$
end while

Query Strategies

- In general, different strategies follow a greedy framework

$$s^* = \operatorname*{argmax}_{s \in D_u} \min_{y \in \{-1,1\}} f(s; y, D_l)$$

- We investigate four commonly used query strategies
  - Uncertainty sampling [Lewis, ICML'94]
  - Query-by-committee [Seung, COLT'92]

  informativeness

  - QUIRE [Huang, TPAMI'14] — Informativeness and representativeness
  - Diversity sampling — diversity

Query Strategies (cont.)

- Uncertainty sampling

$$f(s;y,\mathcal{D}_l) = 1 - P_{\Delta(\mathcal{D}_l)}(y_s = y|\mathbf{x}_s)$$

- Query-by-committee

$$f(s;y,\mathcal{D}_l) = \sum_{k=1}^{C} \mathbf{1}[y \neq g^{(k)}(\mathbf{x}_s)]$$

- QUIRE

$$f(s;y,\mathcal{D}_l) = (L_{u,l}\mathbf{y}_l + L_{u,s}y)^T L_{u,u}^{-1}(L_{u,l}\mathbf{y}_l + L_{u,s}y) - 2yL_{s,l}\mathbf{y}_l - L_{s,s},$$

- Diversity sampling

$$f(s;y,\mathcal{D}_l) = \sum_{i\in\mathcal{D}_l} 1 - \cos(\mathbf{x}_s, \mathbf{x}_i)$$

Experiments – Dataset

- Wiki concept map dataset [Wang, CIKM'16]

Domain	**# Concepts**	**# Pairs**	**# Prerequisites**
Data Mining	120	826	292
Geometry	89	1681	524
Physics	153	1962	487
Precalculus	224	2060	699

Table 1: Dataset statistics.

Experiments – Features

- For each concept pair (A, B), we calculate
  - Graph-based features
  - Text-based features
- Use a Wikipedia dump of Oct. 2016

Graph-based Features

Feature	Description
In/Out Degree	The in/out degree of A/B.
Common Neighbors	# common neighbors of A and B.
# Links	# times A/B links to B/A.
Link Proportion	The proportion of pages that link to A/B also link to B/A.
NGD	The Normalized Google Distance between A and B [Witten, AAAI'08].
PMI	The Pointwise Mutual Information relatedness between the incoming links of A and B [Ratinov, ACL'11].
RefD	A metric to measure how differently A and B's related concepts refer to each other [Liang, EMNLP'15].
PageRank	The difference between A and B's PageRank scores. [Page'99]
HITS	The difference between A and B's hub/authority scores. [Kleinberg, JACM'99]

Text-based Features

Feature	Description
1st Sent	Whether A/B is in the first sentence of B/A.
In Title	Whether A appears in B's title.
Title Jaccard	The Jaccard similarity between A and B's titles.
Length	# words of A/B's content.
Mention	# times A/B are mentioned in the content of B/A.
NP	# noun phrases in A/B's content; # common noun phrases.
Tf-idf Sim	The cosine similarity between Tf-idf vectors for A and B's first paragraphs.
Word2Vec Sim	The cosine similarity between vectors of A and B trained by Word2Vec.
LDA Entropy	The Shannon entropy of the LDA [Blei, JMLR'03] vector of A/B.
LDA Cross Entropy	The cross entropy between the LDA vector of A/B and B/A [Gordon, ACL'16].

Classification Results

- Compared classifiers:
  - Naïve Bayes (NB)
  - Logistic Regression (LR)
  - Support Vector Machine (SVM)
  - Random Forest (RF)
- 5-fold cross validation

Classification Results (cont.)

Classifier	Metric	Data Mining	Geometry	Physics	Precalulus
NB	P	71.5	84.4	54.3	85.7
	R	28.5	44.3	71.9	66.9
	$F1$	37.8	58.1	61.6	75.0
	AUC	81.4	87.1	85.5	93.2
LR	P	65.8	71.3	58.0	81.7
	R	**77.4**	81.3	**78.8**	**88.4**
	$F1$	71.1	75.8	66.8	84.8
	AUC	85.9	91.6	89.2	95.4
SVM	P	73.7	82.8	77.9	86.7
	R	64.7	69.9	50.3	81.4
	$F1$	68.6	75.5	61.1	83.9
	AUC	85.0	91.3	88.8	95.1
RF	P	**80.7**	**95.0**	**85.2**	**90.2**
	R	73.3	**84.7**	59.3	87.1
	$F1$	**76.7**	**89.5**	**69.9**	**88.6**
	AUC	**92.2**	**97.8**	**93.9**	**97.5**

Table 1: Results (%) for concept prerequisite relation classification.

Classification Results (cont.)

Classifier	Metric	Data Mining	Geometry	Physics	Precalulus
NB	P	71.5	84.4	54.3	85.7
	R	28.5	44.3	71.9	66.9
	$F1$	37.8	58.1	61.6	75.0
	AUC	81.4	87.1	85.5	93.2
LR	P	65.8	71.3	58.0	81.7
	R	**77.4**	81.3	**78.8**	**88.4**
	$F1$	71.1	75.8	66.8	84.8
	AUC	85.9	91.6	89.2	95.4
SVM	P	73.7	82.8	77.9	86.7
	R	64.7	69.9	50.3	81.4
	$F1$	68.6	75.5	61.1	83.9
	AUC	85.0	91.3	88.8	95.1
RF	P	**80.7**	**95.0**	**85.2**	**90.2**
	R	73.3	**84.7**	59.3	87.1
	$F1$	**76.7**	**89.5**	**69.9**	**88.6**
	AUC	**92.2**	**97.8**	**93.9**	**97.5**

Table 1: Results (%) for concept prerequisite relation classification.

- Worst F1 and AUC
- The strong independence assumption does not hold for the proposed feature set.

Classification Results (cont.)

Classifier	Metric	Data Mining	Geometry	Physics	Precalulus
NB	P	71.5	84.4	54.3	85.7
	R	28.5	44.3	71.9	66.9
	$F1$	37.8	58.1	61.6	75.0
	AUC	81.4	87.1	85.5	93.2
LR	P	65.8	71.3	58.0	81.7
	R	**77.4**	81.3	**78.8**	**88.4**
	$F1$	71.1	75.8	66.8	84.8
	AUC	85.9	91.6	89.2	95.4
SVM	P	73.7	82.8	77.9	86.7
	R	64.7	69.9	50.3	81.4
	$F1$	68.6	75.5	61.1	83.9
	AUC	85.0	91.3	88.8	95.1
RF	P	**80.7**	**95.0**	**85.2**	**90.2**
	R	73.3	**84.7**	59.3	87.1
	$F1$	**76.7**	**89.5**	**69.9**	**88.6**
	AUC	**92.2**	**97.8**	**93.9**	**97.5**

Table 1: Results (%) for concept prerequisite relation classification.

- Similar F1 and AUC scores
- Higher recall for LR
- Higher precision for SVM

Classification Results (cont.)

Classifier	Metric	Data Mining	Geometry	Physics	Precalulus
NB	P	71.5	84.4	54.3	85.7
	R	28.5	44.3	71.9	66.9
	$F1$	37.8	58.1	61.6	75.0
	AUC	81.4	87.1	85.5	93.2
LR	P	65.8	71.3	58.0	81.7
	R	**77.4**	81.3	**78.8**	**88.4**
	$F1$	71.1	75.8	66.8	84.8
	AUC	85.9	91.6	89.2	95.4
SVM	P	73.7	82.8	77.9	86.7
	R	64.7	69.9	50.3	81.4
	$F1$	68.6	75.5	61.1	83.9
	AUC	85.0	91.3	88.8	95.1
RF	P	**80.7**	**95.0**	**85.2**	**90.2**
	R	73.3	**84.7**	59.3	87.1
	$F1$	**76.7**	**89.5**	**69.9**	**88.6**
	AUC	**92.2**	**97.8**	**93.9**	**97.5**

Table 1: Results (%) for concept prerequisite relation classification.

- Best F1 and AUC
- Better than linear combination of features
- Chosen for active learning experiments

Feature Analysis

Data Mining	Geometry	Physics	Precalculus
Authority diff	PageRank diff	PageRank diff	PageRank diff
LDA entropy of A	In degree of A	RefD	Authority diff
PageRank diff	Out degree of A	# mentions of A in B	RefD
In degree of A	RefD	In degree of B	# mentions of A in B
RefD	# mentions of A in B	Authority diff	Out degree of A
LDA entropy of B	LDA entropy of A	Link proportion of B	A in B's 1st sentence
In degree of B	A in B's 1st sentence	Out degree of A	In degree of A
LDA cross entropy (A;B)	Length of A	In degree of A	Hub diff
Link proportion of A	# NPs in A	LDA entropy of A	# NPs in A
LDA cross entropy (B;A)	# mentions of B in A	# NPs in B	# mentions of A in B

Table 1: Top 10 important features for each domain.

Ranked by "mean decrease impurity" of Random Forest

Feature Analysis

- Common top features: PageRank diff, Authority diff, RefD, …
- Graph-based > Text-based
- Symmetric pairwise features such as similarity and PMI are NOT important in current in-domain classification setting

Data Mining	**Geometry**	**Physics**	**Precalculus**
Authority diff	PageRank diff	PageRank diff	PageRank diff
LDA entropy of A	In degree of A	RefD	Authority diff
PageRank diff	Out degree of A	# mentions of A in B	RefD
In degree of A	RefD	In degree of B	# mentions of A in B
RefD	# mentions of A in B	Authority diff	Out degree of A
LDA entropy of B	LDA entropy of A	Link proportion of B	A in B's 1st sentence
In degree of B	A in B's 1st sentence	Out degree of A	In degree of A
LDA cross entropy (A;B)	Length of A	In degree of A	Hub diff
Link proportion of A	# NPs in A	LDA entropy of A	# NPs in A
LDA cross entropy (B;A)	# mentions of B in A	# NPs in B	# mentions of A in B

Table 1: Top 10 important features for each domain.

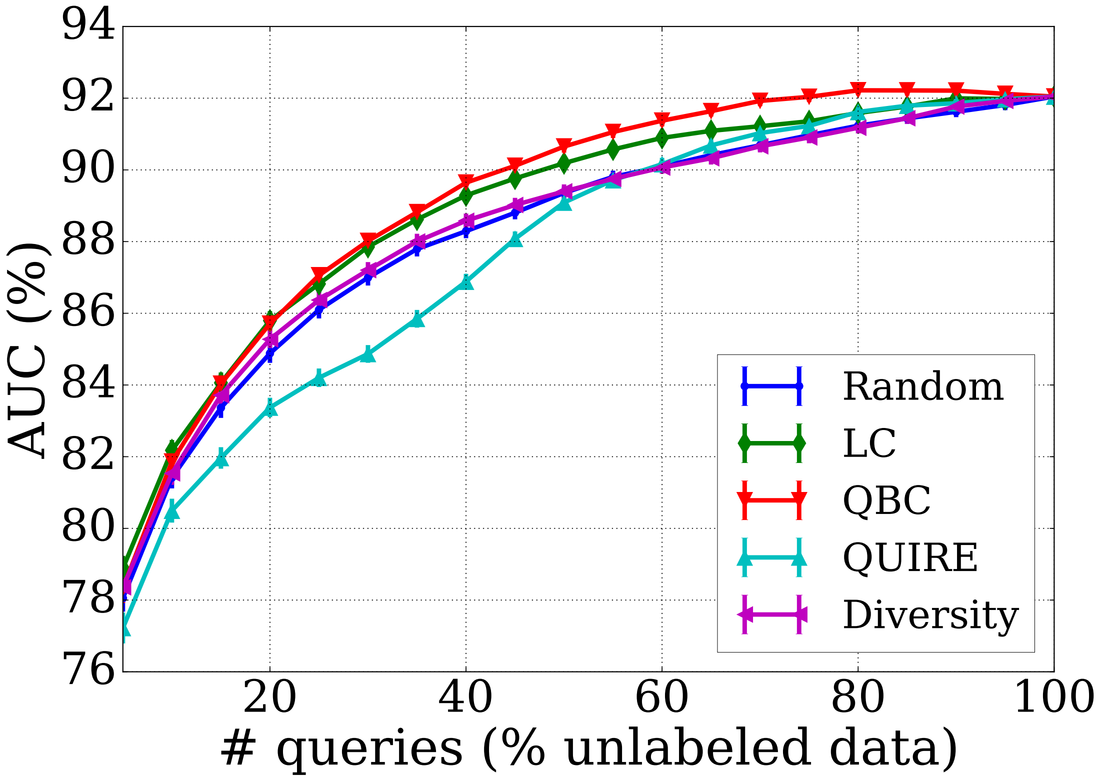
Active Learning – Settings

- Randomly split a dataset into a training set $\mathcal{D}$ and a test set $\mathcal{D}_{test}$ with a ratio of 2:1
- Randomly select 20 samples from the training set as the seed set $\mathcal{D}_l$
- The active learning is repeated 300 times with distinct random seeds.
- Report average scores and confidence intervals ($\alpha = 0.05$)

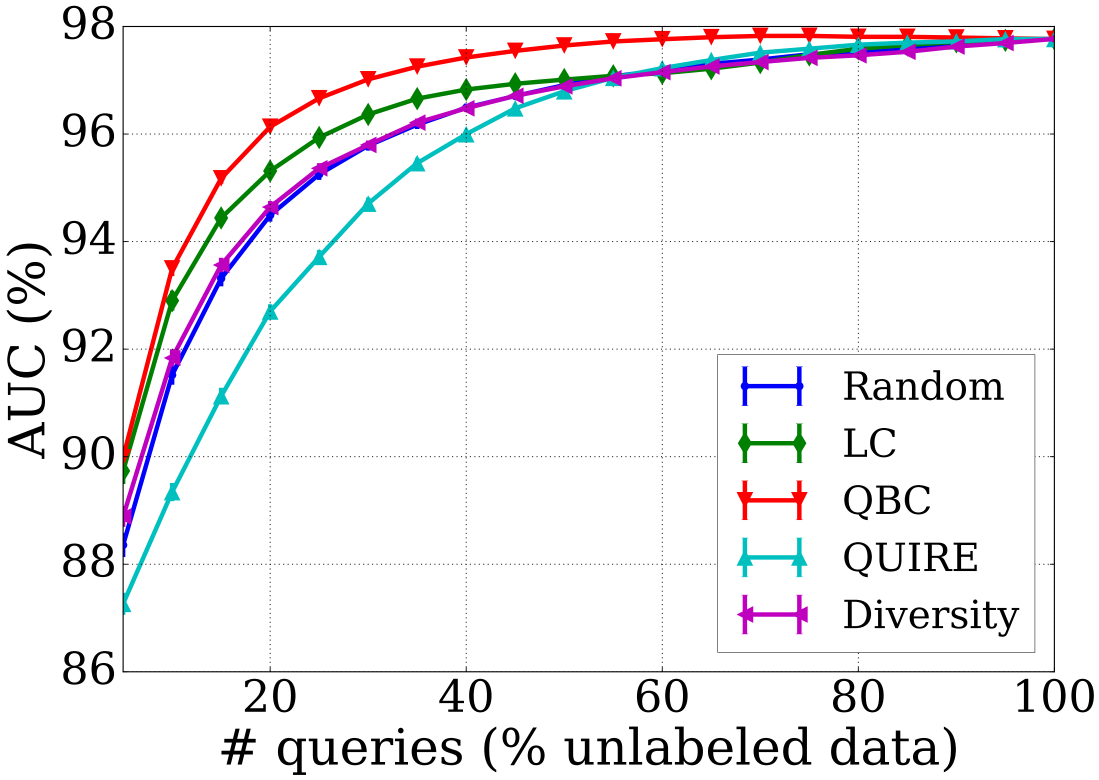
Active Learning – Strategies (cont.)

- Random: baseline
- LC: least confident sampling, a widely used uncertainty sampling variant
- QBC: query-by-committee algorithm. Use query-by-bagging [Mamitsuka, ICML'98] with a committee of three decision trees
- QUIRE: follow authors' experimental approach to use an RBF kernel with \lambda = 1
- Diversity: diversity sampling

Active Learning – Results

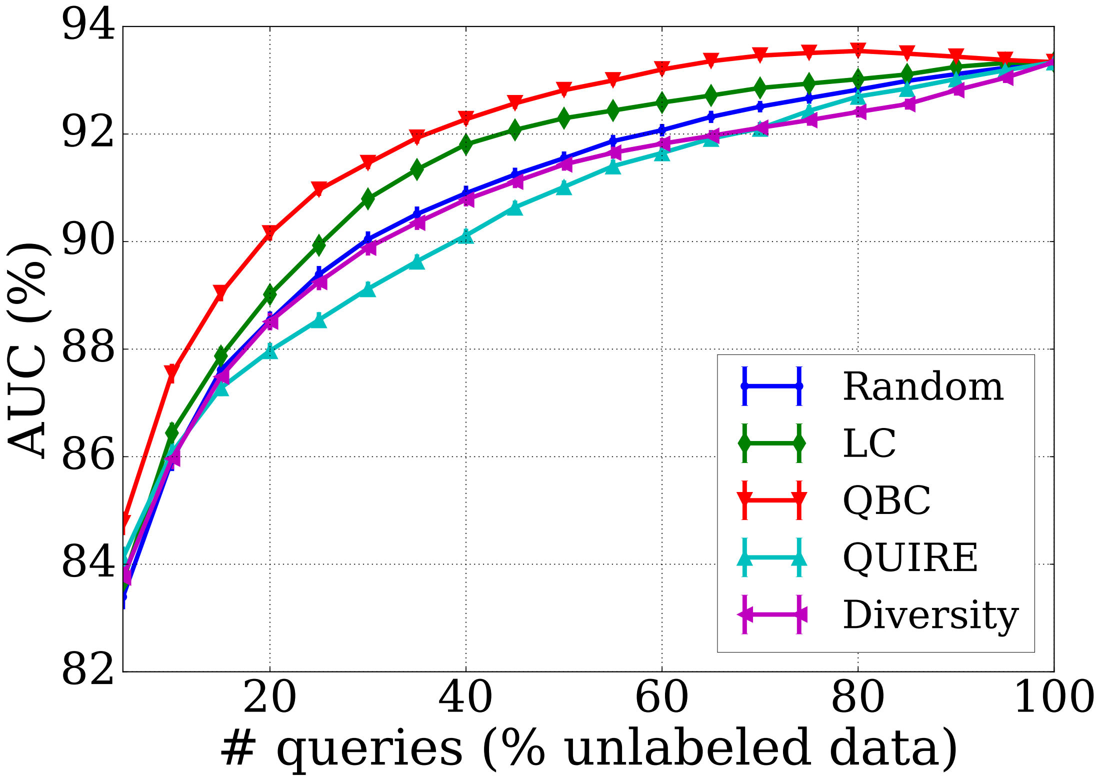


(a) Data Mining

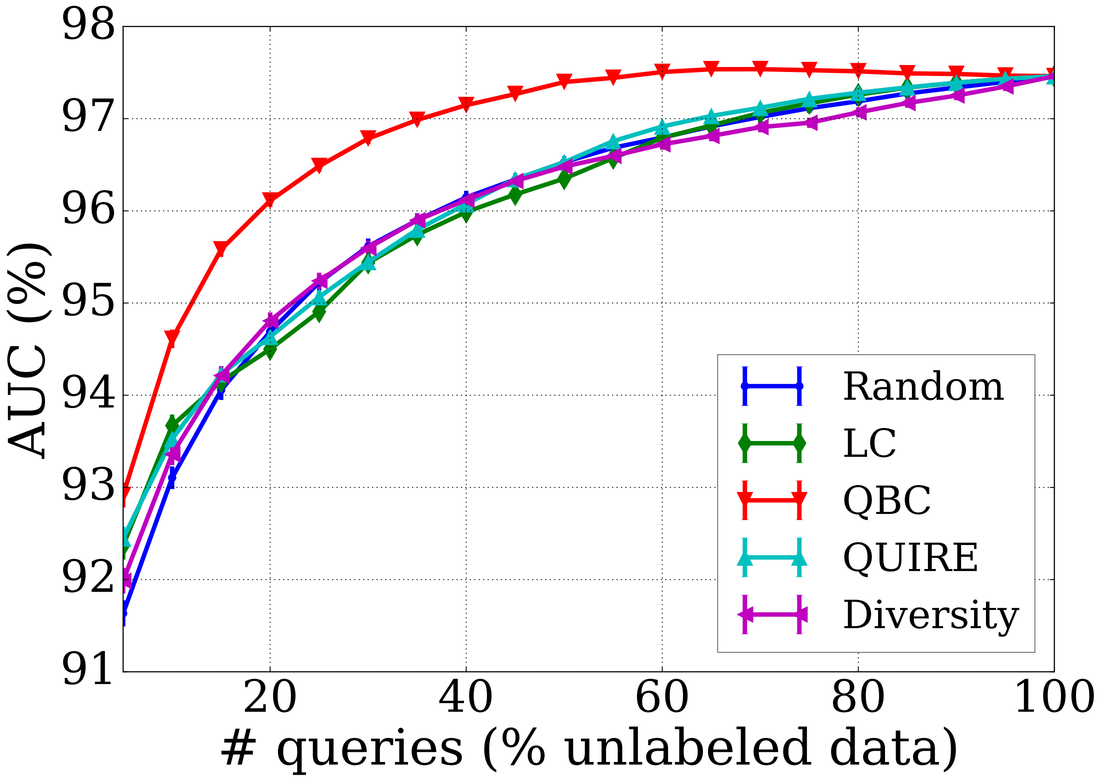


(b) Geometry

Active Learning – Results (cont.)



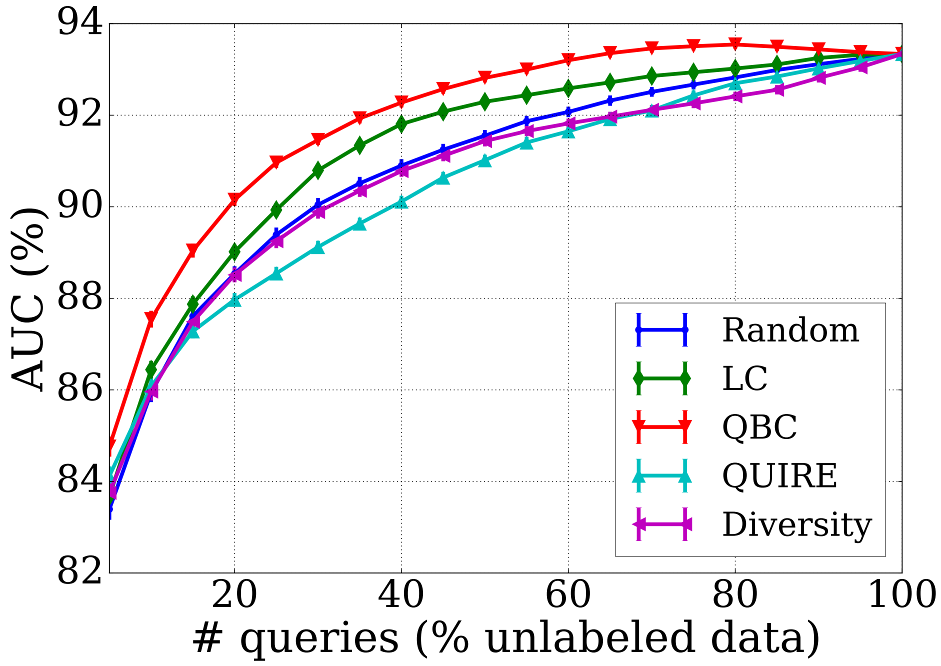
(c) Physics



(d) Precalculus

Active Learning – Results (cont.)

- LC, QBC > Random
- Diversity is not significantly different from Random
- QUIRE performs worse than Random
- QBC is better than other compared methods



Summary

- Investigated the applicability of active learning to concept prerequisite learning.
- Found that QBC outperforms uncertainty sampling, QUIRE, and diversity sampling.
- Proposed a novel set of feature for the task. Top features hopefully will be helpful for other prerequisite learning methods.

Future Work

- Design active learning query strategies tailored to the concept prerequisite learning problem.
  - Active Learning of Strict Partial Orders: A Case Study on Concept Prerequisite Relations (*Liang et al., arXiv:1801.06481 [cs.LG], January 2018)*
- Design more complex and effective concept pair representation and study their effect on the classification performance.

Thanks! Questions?

Chen Liang cul226@ist.psu.edu
http://www.personal.psu.edu/cul226/