Active Learning: Recent Advances
Charles Ling
PhD (Upenn), Professor
University of Western Ontario, Canada
cling@csd.uwo.ca; clinging.csd.uwo.ca
Joint work with Jun Du et al
KDD08, ICDM09, TKDE10, ECML10, ICDM10, 11
Outline

• Introduction
• Active Learning with Generalized Queries (AGQ)
• Experimental Comparison
• Discussions
• Recent Advances
• Conclusions and Future Works
Introduction: Why

- Traditional ML: large training set; passive
- Human learning: small training set…
  - But we learn actively, in many ways…
  - ML: active: acquire missing values, new examples, ask for labels of unlabeled examples, …
  - Reduce # of labelled examples significantly
- Which unlabeled examples to ask??
  - Correct labels provided by users or “Oracle”
  - More powerful than semi-supervised learning
Introduction: Previous Works

• Pool-based Active Learning (Tong & Koller 2002, Roy & McCallum 2001; Baram, El-Yaniv, & Luz 2004)
  • Choose one with most uncertain predicted label
• Direct Query Construction (Ling & Du, KDD 2008)
  • Construct one with most uncertain predicted label
• Exponential speed-up proved for threshold concept
• On real data, speed-up is often small
Introduction – Limitations

• Previous works always ask specific queries
• Example: Predicting heart disease based on a patient dataset with 30 attributes
  • Answer: Yes/No (for this specific patient).
Limitations of Specific Queries

• Many of the attributes may not be relevant;
  • Example: name, temperature, ... etc. are not relevant
• Too many attribute confuse human experts;
• Labels returned are also specific (only for the specific queries).
Outline

• Introduction
• Active Learning with Generalized Queries (AGQ)
• Experimental Comparison
• Discussions
• Recent Advances
• Conclusions and Future Works
An Extension: Generalized Queries

• Active learners can ask generalized queries.
• Example: Query: “are people over 50 with chest pain likely to have heart disease?” (only 2 relevant attributes, age and chest pain)
• Advantages
  • More natural and relevant
  • One generalized query = many specific queries
  • Answer apply to all these examples
• Need to identify irrelevant features
Can Feature Selection Do It?
Feature Selection + Pool_AL < AGQ

Suppose the target is:

All attributes are **relevant**, thus FS could do **nothing**.

Generalized query for L6: [x1=0, x3=0, x5=0] ->?, should obtain **certain answer** from oracle.
Difficulties of Generalized Queries

• Answer from the oracle can often be uncertain.
  • Query: “are people over 50 with chest pain likely to have heart disease?” Yes with a 90% probability
  • Query: “are people over 50 likely to have heart disease?” Yes with a 60% probability
• More general, more powerful, but more uncertain
• Answer could be inaccurate (90% or 92%?)
• How to ask good generalized queries based on limited info (small training set)?
Our Task...

Assuming oracle can answer GQ with prob label, design a robust active learner that attempts to ask few generalized queries (large speed-up)

End results:

• AGQ (Active learning with Generalized Queries)
  • Query: over 50 with chest pain [ICDM 2009]
• AGQ+ [IEEE TKDE 2010]
  • Query: age 50-60, mild or severe chest pain
• Applications in medical domain, text mining...
• More recent advances in AL
AGQ Process

Step 1: [1, 0, 1, 0, 0, 0, 1,...] (most uncertain example)

Step 2: [1, *, *, 0, *, 0, *,...] (generalized query)

Step 3: Oracle

Step 4: 1, 90%

0, 10%

[1, 0, 1, 0, 0, 0, 1,...]=(1,90%)(0, 10%)
[1, 1, 1, 0, 0, 0, 1,...]=(1,90%)(0, 10%)
[1, 0, 0, 0, 0, 0, 1,...]=(1,90%)(0, 10%)
[1, 1, 0, 0, 0, 0, 1,...]=(1,90%)(0, 10%)

......
Key Steps

1. Finding the Most Uncertain Example;
2. Constructing the Generalized Query;
3. Asking Generalized Queries to Oracle;
4. Updating the Training Dataset;
1. Finding Most Uncertain Example

- Build model on current training examples;
- Find **most uncertain** example
  - From a pool of unlabeled examples (pool-based)
  - Direct query construction (KDD 2008)
- Example: \([1, 0, 1, 1, 0, 1]\) (from the pool)
  - Prediction: 52% for class 1; 48% for class 0;
  - Most uncertain: prob closest to 50%
2. Constructing Generalized Query

- Find irrelevant attributes in the chosen example (i.e., the most uncertain example)

- Example:
  - Input: the chosen example \([1, 0, 1, 1, 0, 1]\)
  - Output: generalize to \([1, *, 1, *, 0, 1]\)
How to Generalize?

• Main Idea
  • For all irrelevant attributes, examples with any combination of their values have same predicted label and probability.

• Example:
  • Given: \(x_1=[1, 0, 1, 1, 0, 1]\), and \(P(1|x_1) = 52\%\)
    - 2\(^{nd}\) and 4\(^{th}\) attributes (red ones) are irrelevant
  • Then: \(P(1|x_2) = 52\%, x_2=[1, 0, 1, 0, 0, 1]\);
    \(P(1|x_3) = 52\%, x_3=[1, 1, 1, 0, 0, 1]\);
    \(P(1|x_4) = 52\%, x_4=[1, 1, 1, 1, 0, 1]\);
Greedy Hill-Climbing Search

Given a small threshold $p$:

- For each attribute, construct a fix number of examples with different attribute values;
- Estimate the probability of these examples by the current classifier;
- Choose the attribute with minimum probability variance $v$; if $v < p$, add it to the subset of irrelevant attributes.
- Repeat, until $v > p$

$P$ represents how “conservative” the generalization would be.

In AGQ, $p = 0.0001$

(More details: see paper)
3. Asking Generalized Queries

• Asking generalized queries to the oracle, and obtain an answer with probability.
• If generalized queries are “conservative”, answers should be certain
• On UCI datasets, we first construct the “target concept” on the whole datasets to simulate the oracle

• Example:
  • Query: How to classify [1, *, 1, *, 0, 1]?
  • Answer: 90% for positive
4. Updating Training Set

- Update training set, according to the generalized queries and probability answers

- Example:
  - Given: [1, *, 1, *, 0, 1]; 90% for 1 (10% for 0)
  - Add four specific examples into training set, with probability labels (90% for 1 and 10% for 0)
    - [1, 0, 1, 0, 0, 1], [1, 0, 1, 1, 0, 1], [1, 1, 1, 0, 0, 1], [1, 1, 1, 1, 0, 1].
  - Limit the number of examples added
Outline

• Introduction
• Active Learning with Generalized Queries (AGQ)
• Experimental Comparison
• Discussions
• Recent Advances
• Conclusions and Future Works
Artificial Data

Target:
5 relevant + 5 irrelevant att.

Comparison:
<table>
<thead>
<tr>
<th>Query 1</th>
<th>AGQ</th>
<th>Pool</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[1, 1, 1, 0, *, *, *, *, *, *]</td>
<td>[1, 1, 1, 0, 1, 1, 1, 1, 0, 0]</td>
</tr>
<tr>
<td></td>
<td>L2</td>
<td>L2</td>
</tr>
<tr>
<td>Ideal Query</td>
<td>[* , 1, 1, 0, *, *, *, *, *, *, *]</td>
<td>[* , 1, 1, 0, *, *, *, *, *, *, *]</td>
</tr>
<tr>
<td>Answer</td>
<td>0, 100%</td>
<td>0</td>
</tr>
<tr>
<td>No. of Examples</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Error Rate</td>
<td>0.18</td>
<td>0.27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 2</th>
<th>AGQ</th>
<th>Pool</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[0, *, 0, 1, *, *, *, *, *, *, *]</td>
<td>[1, 0, 1, 1, 0, 0, 1, 0, 0, 1]</td>
</tr>
<tr>
<td></td>
<td>L5, L6</td>
<td>L3</td>
</tr>
<tr>
<td>Ideal Query</td>
<td>-</td>
<td>[* , 0, 1, *, *, *, *, *, *, *, *]</td>
</tr>
<tr>
<td>Answer</td>
<td>0, 54%</td>
<td>0</td>
</tr>
<tr>
<td>No. of Examples</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Error Rate</td>
<td>0.21</td>
<td>0.22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 3</th>
<th>AGQ</th>
<th>Pool</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[0, 1, 0, 1, 1, 0, 0, *, 1, *]</td>
<td>[1, 1, 1, 1, 0, 1, 1, 1, 0, 1]</td>
</tr>
<tr>
<td></td>
<td>L5</td>
<td>L1</td>
</tr>
<tr>
<td>Ideal Query</td>
<td>[0, *, 0, *, 1, *, *, *, *, *, *, *]</td>
<td>[* , 1, 1, 1, *, *, *, *, *, *, *]</td>
</tr>
<tr>
<td>Answer</td>
<td>1, 100%</td>
<td>1</td>
</tr>
<tr>
<td>No. of Examples</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Error Rate</td>
<td>0.16</td>
<td>0.26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 4</th>
<th>AGQ</th>
<th>Pool</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[0, 1, 0, 1, 0, 1, *, *, 0, *]</td>
<td>[1, 0, 1, 1, 0, 1, 0, 0, 1, 1]</td>
</tr>
<tr>
<td></td>
<td>L6</td>
<td>L3</td>
</tr>
<tr>
<td>Ideal Query</td>
<td>[0, *, 0, *, 0, *, *, *, *, *, *, *]</td>
<td>[* , 0, 1, *, *, *, *, *, *, *, *]</td>
</tr>
<tr>
<td>Error Rate</td>
<td>0, 100%</td>
<td>0</td>
</tr>
<tr>
<td>No. of Examples</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Error Rate</td>
<td>0.17</td>
<td>0.26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 5</th>
<th>AGQ</th>
<th>Pool</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[1, *, 0, *, 0, *, 1, *, *, *]</td>
<td>[1, 1, 1, 0, 0, 1, 0, 0, 1, 1]</td>
</tr>
<tr>
<td></td>
<td>L4</td>
<td>L2</td>
</tr>
<tr>
<td>Ideal Query</td>
<td>[1, *, 0, *, *, *, *, *, *, *, *, *]</td>
<td>[* , 1, 1, 0, *, *, *, *, *, *, *]</td>
</tr>
<tr>
<td>Answer</td>
<td>1, 100%</td>
<td>0</td>
</tr>
<tr>
<td>No. of Examples</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Error Rate</td>
<td>0.13</td>
<td>0.2</td>
</tr>
</tbody>
</table>
### UCI Data

14 UCI datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th># Att</th>
<th># Inst</th>
<th>Class Dist.</th>
<th>Train</th>
</tr>
</thead>
<tbody>
<tr>
<td>breast-cancer</td>
<td>9</td>
<td>277</td>
<td>196/81</td>
<td>1/5</td>
</tr>
<tr>
<td>breast-w</td>
<td>9</td>
<td>699</td>
<td>458/241</td>
<td>1/10</td>
</tr>
<tr>
<td>colic</td>
<td>22</td>
<td>368</td>
<td>232/136</td>
<td>1/5</td>
</tr>
<tr>
<td>credit-a</td>
<td>15</td>
<td>690</td>
<td>307/383</td>
<td>1/20</td>
</tr>
<tr>
<td>credit-g</td>
<td>20</td>
<td>1000</td>
<td>700/300</td>
<td>1/100</td>
</tr>
<tr>
<td>diabetes</td>
<td>8</td>
<td>768</td>
<td>500/268</td>
<td>1/10</td>
</tr>
</tbody>
</table>

...
## Statistics on UCI Data

<table>
<thead>
<tr>
<th>Dataset</th>
<th># Don’t-care Att.</th>
<th>#. Inst</th>
<th>Certainty of Oracle</th>
<th>Iteration of “Pool”</th>
<th>Iteration of AGQ</th>
<th>% Reduction of Iteration</th>
<th>AGQ (w/t/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>breast-cancer</td>
<td>2.7 (30%)</td>
<td>14.54</td>
<td>95%</td>
<td>35</td>
<td>18</td>
<td>49%</td>
<td>W</td>
</tr>
<tr>
<td>breast-w</td>
<td>5.35 (59%)</td>
<td>32.31</td>
<td>87%</td>
<td>18</td>
<td>18</td>
<td>0%</td>
<td>T</td>
</tr>
<tr>
<td>colic</td>
<td>13.15 (60%)</td>
<td>35.68</td>
<td>91%</td>
<td>15</td>
<td>8</td>
<td>47%</td>
<td>W</td>
</tr>
<tr>
<td>credit-a</td>
<td>6.38 (43%)</td>
<td>16.43</td>
<td>88%</td>
<td>12</td>
<td>5</td>
<td>58%</td>
<td>W</td>
</tr>
<tr>
<td>credit-g</td>
<td>8.54 (43%)</td>
<td>4.97</td>
<td>87%</td>
<td>50</td>
<td>12</td>
<td>76%</td>
<td>W</td>
</tr>
<tr>
<td>diabetes</td>
<td>3.02 (38%)</td>
<td>27.31</td>
<td>89%</td>
<td>50</td>
<td>16</td>
<td>68%</td>
<td>W</td>
</tr>
<tr>
<td>heart-statlog</td>
<td>5.92 (46%)</td>
<td>12.52</td>
<td>89%</td>
<td>50</td>
<td>25</td>
<td>50%</td>
<td>W</td>
</tr>
<tr>
<td>Avg.</td>
<td>12.53 (51.36%)</td>
<td>16.49</td>
<td>90.21%</td>
<td>35.14</td>
<td>24.21</td>
<td>36%</td>
<td>9/4/1</td>
</tr>
</tbody>
</table>
Outline

• Introduction
• Active Learning with Generalized Queries (AGQ)
• Experimental Comparison
• Discussions (skip)
• Recent Advances
• Conclusions and Future Works
Outline

• Introduction
• Active Learning with Generalized Queries (AGQ)
• Experimental Comparison
• Discussions
• Recent Advances
• Conclusions and Future Works
Recent Advances in AL

• A New Paradigm of Active Learning
• Hierarchical Classification for Next Search Engine
• Making reasonable assumptions on oracle
  – Cost-sensitive oracle
  – Ambiguous oracle
  – Oracle with explanation
  – ...

(Skip)
A New Paradigm of Active Learning

• Most previous works of AL (pool-based, AGQ, ...) are based on asking near-boundary examples

• In scientific discovery, scientists are active learners... They do perform near-boundary exp

• What else?

• When anomaly (surprise) is found, they “repeat” experiments to confirm or disapprove it!
The Most Famous Failed Exp

- Newtonian physics predicted that the speed of light will change in the “aether wind”
- In 1887, Michelson–Morley experiment

Anomaly is found Exp. repeatedly: AL
A new theory in 1905 Einstein Special Relativity
Predicted as certain, positive example. Ask Oracle (do experiment) Negative! Anomaly!! Actively repeat exp (itself, neighbours)

For noisy, probabilistic concepts, and unexplored space in complex concepts
AL in Hierarchical Classification
For the Next Search Engines
Current: Flat list, only limited structure
Amazon.com: Big Brain Academy: Wii Degree: Video Games
The Wii sequel to Big Brain Academy for Nintendo DS includes three multiplayer modes for up to eight players. Players also can exchange student-record books ...
www.amazon.com › Video Games › Wii › Puzzle - Cached - Similar

Big Brain Academy: Wii Degree - Wikipedia, the free encyclopedia
Big Brain Academy: Wii Degree includes a single player mode whereby the player uses a brain to effectively answer questions correctly. The game also ...
en.wikipedia.org/wiki/Big_Brain_Academy:_Wii_Degree - Cached

Big Brain Academy: Wii Degree for Nintendo Wii | EBGames
EBGames: Buy Big Brain Academy: Wii Degree, Nintendo of America, Nintendo Wii, Find release dates, customer reviews, previews and screenshots.
www.ebgames.com/Catalog/ProductDetails.aspx?product_id... - Cached - Similar

Big Brain Academy: Wii Degree Video Game | Reviews, Trailers ...
View Big Brain Academy: Wii Degree video game trailers, exclusive features, and online reviews. View exclusive interviews, actual Big Brain Academy: Wii ...
www.gametrailers.com/game/big-brain-academy-wii-degree/3424 - Cached
Integrating search and browsing together
– Search alone: filters left at top level (“all pages”)
– Browsing alone: search keyword left blank
– Search within browse; browse then search

Multi-label, Multi-instance Hierarchical Classification with Active Learning
Outline

• Introduction
• Active Learning with Generalized Queries (AGQ)
• Experimental Comparison
• Discussions
• Recent Advances
• Conclusions and Future Works
Conclusions

• Active Learning can be very powerful
• Many different ways to be “active”
  • Generalized queries reduces queries significantly
  • Different assumptions on oracle
    • Cost-sensitive, ambiguous, with explanation, ...
  • Anomaly-based AL (vs boundary-based AL)
    • ... ...
• Widespread applications (search engines, text, ...)
• ...
Current & Future Works

• Theoretical Guarantee
  • Some results already

• Active learning in human and machines

• Real-World Applications
Thanks for Active Listening!