Multi-view Active Learning

Ion Muslea

University of Southern California
Outline

- Multi-view active learning
- Robust multi-view learning
- View validation as meta-learning

- Related Work
- Contributions
- Future work
Background & Terminology

• Inductive machine learning
  – algorithms that learn concepts from labeled examples
• Active learning: minimize need for training data
  – detect & ask-user-to-label only most informative exs.
• Multi-view learning (MVL)
  – disjoint sets of features that are sufficient for learning
    • Speech recognition: sound vs. lip motion
  – previous multi-view learners are semi-supervised
    • exploit distribution of the unlabeled examples
    • boost accuracy by bootstrapping views from each other
Multi-view active learning maximizes the accuracy of the learned hypotheses while minimizing the amount of labeled training data.
Outline

• Multi-view active learning
  – The intuition
  – The Co-Testing family of algorithms
  – Empirical evaluation

• Robust multi-view learning

• View validation as meta-learning

• Related Work

• Contributions

• Future work
A Simple Multi-View Problem

- **Features:**
  - salary
  - office number

- **Concept:** Is Faculty?
  - View-1: salary > 50 K
  - View-2: office < 300

**GOAL:** minimize amount of labeled data
Co-Testing

Labeled Examples

Unlabeled Examples
Co-Testing

Salary

Office

Labeled Examples

Unlabeled Examples
Co-Testing

Labeled Examples

Unlabeled Examples
The **Co-Testing** Family of Algorithms

- **REPEAT**
  - Learn one hypothesis in each view
  - Query one of the *contention points* *(CP)*

- Algorithms differ by:
  - output hypothesis: *winner-takes-all, majority/weighted vote*
  - query selection strategy:
    - **Naïve:** randomly chosen *(CP)*
    - **Conservative:** equal confidence *(CP)*
    - **Aggressive:** maximum confidence *(CP)*
When does Co-Testing work?

• **Assumptions:**

1. **Uncorrelated views**
   - for any \(<x_1, x_2, L>\): given \(L\), \(x_1\) and \(x_2\) are uncorrelated
   - views *unlikely* to make same mistakes \(\Rightarrow\) contention points

2. **Compatible views**
   - perfect learning in both views
   - contention points are *fixable mistakes*

• **under these assumptions**, there are classes of learning problems for which Co-Testing converges faster than single-view active learners
Experiments: *four real-world domains*

- Random Sampling
- Uncertainty Sampling
- Query-by-Committee
- Query-by-Boosting
- Query-by-Bagging
- Naïve Co-Testing
- Conservative Co-Testing
- Aggressive Co-Testing

<table>
<thead>
<tr>
<th>Ad</th>
<th>Parse</th>
<th>Courses</th>
<th>Wrapper</th>
</tr>
</thead>
<tbody>
<tr>
<td>IB</td>
<td>C4.5</td>
<td>Naïve-Bayes</td>
<td>Stalker</td>
</tr>
</tbody>
</table>

- remove advertisements
- "is this image an ad?"

- learns shift-reduce parser that converts Japanese discourse tree into an equivalent English one
- discriminates between course homepages and other pages
- extract relevant data from Web pages

- **wins**
- **works**
- **cannot-be-applied**
Main Application: **Wrapper Induction**

- Extract *phone number*: find its **start** & **end**

---

... Hilton <p> Phone: <b> (211) 111-1111 </b> </p> Fax: (211) 121-1...

SkipTo( Phone : <b> ) SkipTo( </b>)

---

... Phone (toll free) : <i> (800) 171-1771 </i> </i> Fax: (800) 777-1...

SkipTo( Phone ) SkipTo( Html ) SkipTo( Html )
Co-Testing for Wrapper Induction

- **Views:** tokens *before* & *after* extract. point

... Hilton <p> Phone: <b> (211) 111-1111 </b> Fax: <b> (211) ...</p>

SkipTo(Phone) SkipTo(<b>) BackTo( Fax ) BackTo( ( Nmb )

... Motel 6 <p> **Phone**: <b> (311) 101-1110 </b> Fax: <b> (311) ...</p>

... **Phone** (tool free): <i> (800) 171-1771 </i> Fax: <b> (111) ...</i>

...
Results on 33 tasks: 2 rnd exs + queries

Tasks

Queries until 100% accuracy

Random sampling

18+
Results on 33 tasks: 2 rnd exs + queries

![Bar chart showing queries until 100% accuracy for Naïve Co-Testing and Random sampling across 33 tasks. The chart indicates that tasks 1, 3, 5, 9, and 17 require more than 18 queries for 100% accuracy.](chart.png)
Results on 33 tasks: 2 rnd exs + queries
Co-Testing vs. Single-View Sampling

Tasks

Queries until 100% accuracy

Aggressive Co-Testing  Query-by-Bagging
First Contribution

**Co-Testing**: multi-view active learning

- Querying contention points
- Converges faster than single-view
  - variety of domains & base learners
Outline

• Multi-view active learning

• Robust multi-view learning
  – motivation
  – Co-EMT = active + semi-supervised learning
  – robustness to assumption violations

• View validation as meta-learning

• Related Work

• Contributions

• Future work
Motivation

- **Active learning:**
  - queries *only* the most informative examples
  - ignores all remaining (unlabeled) examples

- **Semi-supervised learning (previous MVL):**
  - few labeled + many unlabeled examples
    - *unlabeled examples:* model examples’ distribution
    - use this model to boost accuracy of small training set

- **Best of both worlds:**
  1. **Active:** make queries
  2. **Semi-supervised:** use remaining (unlabeled) exs.
Co-EMT = Co-Testing + Co-EM

• Given:
  – views $V_1$ & $V_2$
  – $L$ & $U$, sets of labeled & unlabeled examples

• Co-Testing

  REPEAT
  - use Co-EM($L, U$) to learn $h_1$ and $h_2$
  - use labeled examples in $L$ to learn $h_1$ and $h_2$
  – query contention point: $h_1(u) \neq h_2(u)$

Semi-supervised MVL
- few labeled + many unlabeled exs
- uses unlabeled exs to bootstrap views from each other
The Co-EMT Synergy

1. Co-Testing boosts Co-EM:  
   - *stand-alone Co-EM* uses random examples  
   - Co-Testing provides more informative examples

2. Co-EM helps Co-Testing:  
   - *stand-alone Co-Testing* uses only labeled exs  
   - Co-EM also exploits unlabeled examples
Two real-world domains

COURSES

error rate (%)

ADS

error rate (%)

Co-EMT
Co-EM
Co-Training
semi-supervised EM

Co-Testing
Semi-supervised **MVL**: bootstrapping views

**Task:** is Web page *course homepage* (+) or *not* (-) ?

**V2**: words in hyperlinks

... Spring teaching ...

... favorite class ...

... my favorite class ...

**V1**: words in pages

![Image of course homepage]
**Assumption:** compatible, independent views
Incompatible views

CS-511: Neural Nets

Neural nets papers:

...neural nets ...

...neural nets ...

Neural nets papers:...
Correlated views: *domain clumpiness*

- **Theory clump**
- **A.I. clump**
- **Systems clump**
- **Faculty clump**
- **Admin clump**
- **Students clump**
A Controlled Experiment

<table>
<thead>
<tr>
<th>Clumps per class</th>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Co-EM
- Co-Training
- EM

Incompatibility (%)

Legend:
- Co-EM
- Co-Training
- EM
Co-EMT is robust!
Second Contribution

**Co-EMT**: robust multi-view learning

- interleave *active* & *semi-supervised MVL*
Outline

• Multi-view active learning
• Robust multi-view learning

• View validation as meta-learning
  – Motivation
  – Adaptive view validation
  – Empirical results

• Related Work
• Contributions
• Future work
Motivation: Wrapper Induction

In **MVL**, the *same* views may be:

- *adequate* for some tasks
- *inadequate* for other tasks
The Need for View Validation

• **Not only** for wrapper induction:
  • **Speech recognition:** sound *vs.* lip motion
    • Task-1: recognize *Tom Brokaw*’s speech
    • Task-2: recognize *Ozzy Osbourne*’s speech
    • ...
  • **Web page classification:** hyperlink *vs.* page words
    • Task-1: terrorism / economics news
    • Task-2: faculty / student homepage
    • ...

• **Solution:** meta-learning
  • from past experiences, learn to …
  • … predict whether **MVL** is adequate for new, unseen task
Meta-learner: *Adaptive View Validation*

- **GIVEN**
  - labeled tasks \([\text{Task}_1, L_1], [\text{Task}_2, L_2], \ldots, [\text{Task}_n, L_n]\)

- **FOR EACH** \(\text{Task}_i\) **DO**
  - generate *view validation example*
    \[e_i = \langle \text{Meta-F1}, \text{Meta-F2}, \ldots, L_i \rangle\]

- train **C4.5** on \(e_1, e_2, \ldots, e_n\)

For each *new, unseen task* use learned decision tree to predict whether **MVL** is adequate for task.
View Validation *Meta-Features*

- use labeled examples to learn $h_1$ & $h_2$
- **The meta-features:**
  - F1: agreement of $h_1$ & $h_2$ on unlabeled examples
  - F2: $\min(\text{TrainError}(h_1), \text{TrainError}(h_2))$
  - F3: $\max(\text{TrainError}(h_1), \text{TrainError}(h_2))$
  - F4: $F3 - F2$
  - F5: $\min(\text{Complexity}(h_1), \text{Complexity}(h_2))$
  - F6: $\max(\text{Complexity}(h_1), \text{Complexity}(h_2))$
  - F7: $F6 - F5$

**Illustrative View Validation Rule:**

**IF**

$h_1$ & $h_2$ agree on at least 62% unlabeled exs & $|\text{TrainError}(h_1) - \text{TrainError}(h_2)| < 10$

**THEN**

task’s views are adequate for $\text{MVL}$
Empirical Results

- **WI**: wrapper induction (33 tasks)
- **TC**: text classification (60 tasks)
Third Contribution

View validation:

meta-learner that uses past experiences to predict whether or not MVL is appropriate for new, unseen task
Related Work: Active Learning

• counterexamples [Angluin 88], query generation [Lang ‘92]

• Selective Sampling
  – uncertainty reduction [Lewis 94, Schohn 01, Thompson 99]
  – version space reduction [Seung 92, Cohn 94, Abe 98]
  – expected-error minimization [Lindenbaum 99, Tong 00, Roy 01]

• Co-Testing vs. existing selective samplers
  – multi-view vs. single-view active learning
  – “domain” oriented vs. “base learner” oriented

• Co-EMT vs. “EM + Query-by-Committee” [McCallum+ ‘98]
Related Work: \textit{Multi-view Learning}

- **Theory of Co-Training:**
  - [Blum+Mitchell 98] formalization of multi-view learning
  - [Dasgupta+ 01] Co-Training’s proof of convergence
  - [Abney 02] allowing (some) view correlation
- **Extensions:**
  - algorithmic [Collins 99] [Nigam 00] [Pierce 01] [Ghani 02]
  - applicability [Nigam 00] [Goldman 00] [Raskutti 02]

- **Co-Testing vs. existing multi-view learners**
  - all other \textit{MVL} are “passive” & semi-supervised
Related Work: *Meta-learning*

- **Meta-features**
  - general features [Aha 92][Brazdil+ 95][Todorovski+ 99]
    - simple features: number of classes, features, examples, …
    - statistical: default accuracy, std.-dev., skewness, kurtosis, …
    - information theoretic: class, attribute, and joint entropy, …
  - classifier-based [Bensusan 99]: max-depth & shape of DT, …
  - landmarking [Pfaringer 00]: accuracies of simple, fast learners
- **Adaptive View Validation** vs. existing approaches:
  - single- vs. multi-view learning
  - few labeled + many unlabeled examples
  - landmarking (**training error**) + classifier-based (**complexity**)
Contributions

1. **Co-Testing**: multi-view active learning
   - Querying contention points
   - Converges faster than single-view learners …
     - … on a variety of domains & base learners

2. **Co-EMT**: novel multi-view learner
   - Interleaving active & semi-supervised learning
   - Robust behavior on large spectrum of tasks

3. **View Validation**: is task appropriate for MVL?
   - Meta-learning algorithm that uses past experiences to predict whether or not MVL is appropriate for new, unseen task.
Future Work

• **View Detection**
  – propose feature split into views
    • INPUT: learning task (features + examples)
    • OUTPUT: split of features into several views (*if possible*)

• **Co-Testing**
  – myopic *vs.* look-ahead queries
    • select optimal *sequence of queries*
  – **Co-Testing** for regression & semi-supervised clustering

• **Adaptive View Validation**
  – “general purpose” *vs.* “per multi-view problem”
    • train on tasks from a variety of multi-view problems