Machine Learning for Medical Decision Support
and Individualized Treatment Assignment

Finn Kuusisto

Department of Computer Sciences
Doctoral Defense
August 14, 2015
Health Care Expenditure as % of GDP

*World Health Statistics 2015, World Health Organization (WHO)*
“Tonight, I'm launching a new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes — and to give all of us access to the personalized information we need to keep ourselves and our families healthier.”

-President Barack Obama, State of the Union Address, January 20, 2015
Precision Medicine

- Tailoring medical treatment to individual characteristics of each patient
- Classify individuals into subpopulations that differ in:
  - Susceptibility to particular diseases
  - Biology and/or prognosis of diseases they develop
  - Response to specific treatments
Supervised Learning

**Given:** Values of the input features and the output feature (response, class) for many patients

**Do:** Build a model that can accurately predict the unknown value of the output class for new (previously unseen) patients whose values of the input features are known

Classical methods: linear and logistic regression

Other methods: decision trees, random forests, support vector machines, Bayesian networks, artificial neural networks, etc.
Machine learning results can be made more clinically-relevant by tailoring current approaches to meet clinical objectives through the development of new algorithms to model individual response to treatment, and by incorporating clinical expertise into model development and refinement.
Publications

Clinical Collaboration


Individualized Treatment Effects


Clinical Collaboration


Individualized Treatment Effects


Outline

• Introduction
• Advice-Based Learning Framework
• Support Vector Machines for Uplift Modeling
• Conclusions
Outline

• Introduction
• Advice-Based Learning Framework
• Support Vector Machines for Uplift Modeling
• Conclusions
Decision Support

Great opportunities for machine-learned decision support systems

But...

Standardized, complete, and sufficient training data is rarely available
Comprises two parts

1) Categories of advice sources

2) Iterative process for model refinement
ABLe - Advice Categories

Task
- What is the problem and scope?
- What predictor variables are important?
- How should the problem be modeled?

Relationships Among Variables
- What combinations of variables are important to the task?

Parameter Values
- What is the clinical objective?
- What model parameters best represent that objective?
ABLle - Iterative Process

Repeated iterations to optimize performance

Start

MDE and CSE
define/refine advice

Accept model

Build model

Evaluate model
Upgrade Prediction

1. Mammogram
2. Needle Biopsy
3. Radiologic-Histologic Correlation
4. Excision

Abnormality
Benign Tissue
Non-definitive Diagnosis
Final Diagnosis

Malignant
= “Upgrade”

Image Sources:
1. NIH - wikimedia.org/wiki/File:Woman_receives_mammogram.jpg
2. Itayba - wikimedia.org/wiki/File:Normal.jpg
3. UW Hospital and Clinics
4. NIH - wikimedia.org/wiki/File:Surgical_breast_biopsy.jpg
• 5-15% of core needle biopsies non-definitive
• Approximately 35,000-105,000* per year
• 80-90% of non-definitive biopsies are benign

* Based on 2010 annual breast biopsy utilization rate in the United States
Upgrade Prediction

1. Mammogram
2. Needle Biopsy
3. Radiologic-Histologic Correlation
4. Excision

Abnormality
Benign Tissue
Non-definitive Diagnosis

Malignant
= “Upgrade”
Phase 1

Task
- Simple probabilistic model (Naïve Bayes)
- Standardized BI-RADS descriptor features
- Some non-standard pathology features and demographics
- Predict probability of malignancy
- Assume excision at $\geq 0.02$ model score (to balance risk)

Relationships Among Variables
- Rules predicting increase/decrease risk of malignancy

Parameter Values
- None
Relationships Among Variables

If-Then rules from domain expert (Beth) that suggest increase/decrease risk of upgrade.

High-risk mass rule:

IF
Irregular mass shape is present OR
Spiculated mass margin is present OR
High density mass is present OR
Increasing mass size
THEN
Risk of upgrade increases
Biopsies in Practice (2006-11)

Core Needle Biopsies
- 2,808

Core Needle Biopsies + Dx Mammogram
- 1,910
  - Malignant Biopsy: 601
  - Benign Biopsy: 1,309
    - Non-definitive: 157
      - Malignant (upgrade): 29
      - Benign (non-upgrade): 128
Phase 1 Results

- Naïve Bayes to predict *malignancy*
- Assume excision at $\geq 0.02$ model score
- Experiments with and without expert rule features

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Rules</th>
<th>Data + Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Malignant Excisions</strong></td>
<td>8 (27.6%)</td>
<td>1 (3.4%)</td>
<td>9 (31.0%)</td>
</tr>
<tr>
<td>Missed (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Benign Excisions</strong></td>
<td>46 (35.9%)</td>
<td>5 (3.9%)</td>
<td>63 (49.2%)</td>
</tr>
<tr>
<td>Avoided (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Observations & Refinements

Observations

- No output threshold with acceptable performance
- Non-definitive biopsies broken into 3 categories at diagnosis
  - Atypical/Radial Scar (ARS)
  - Insufficient (I)
  - Discordant (D)
- ARS and I cases consistently mislabeled

Refinements

- Focus exclusively on discordant cases
Discordant Biopsies (2006-11)

Discordant Biopsy

60

Malignant (upgrade) 10

Benign (non-upgrade) 50
Phase 2 Results

- Naïve Bayes to predict *malignancy* of discordants
- Assume excision at \( \geq 0.02 \) model score
- Experiments with and without expert rule features

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Rules</th>
<th>Data + Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Malignant Excisions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missed (%)</td>
<td>3 (30.0%)</td>
<td>1 (10.0%)</td>
<td>3 (30.0%)</td>
</tr>
<tr>
<td><strong>Benign Excisions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avoided (%)</td>
<td>29 (58.0%)</td>
<td>17 (34.0%)</td>
<td>27 (54.0%)</td>
</tr>
</tbody>
</table>
Observations & Refinements

Observations

- Good ranking of cases by output model scores
- Most cases assigned less than 0.02 risk

Refinements

- Make model conservative
  - Different costs for false negatives (FN) versus false positives (FP)
  - Take from utility analysis literature in mammography
Phase 3 Results

- Naïve Bayes to predict *malignancy* of discordants
- Cost ratio of 150:1 for FN:FP
- Assume excision at $\geq 0.02$ model score
- Experiments with and without expert rule features

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Rules</th>
<th>Data + Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malignant Excisions</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>Missed (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benign Excisions</td>
<td>5 (10.0%)</td>
<td>5 (10.0%)</td>
<td>12 (24.0%)</td>
</tr>
<tr>
<td>Avoided (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Outline

• Introduction
• Advice-Based Learning Framework
• Support Vector Machines for Uplift Modeling
• Conclusions
Clinical Trial

Clinical experiment to determine the average effect of some treatment for:

- Safety
- Efficacy
Clinical Trial

Treatment Group

Control Group

Pretrial

Outcome

28.6%

57.1%
Clinical Trial

ATE = 28.6% - 57.1% = -28.5 percentage points
Clinical Trial

<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretrial</td>
<td></td>
</tr>
<tr>
<td>Outcome</td>
<td></td>
</tr>
</tbody>
</table>

28.6% 57.1%
Clinical Trial

<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretrial</td>
<td></td>
</tr>
<tr>
<td>Outcome</td>
<td></td>
</tr>
</tbody>
</table>

ITE =  

28.6%  57.1%
ITE Challenge

- Cannot observe both treatment and control outcomes for any one individual

- Need a lot of data to model ITE for even a moderate number of individual features
How do we choose which customers to target with some marketing activity?

<table>
<thead>
<tr>
<th><strong>Persuadables</strong></th>
<th>Customers who respond positively to marketing activity.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sure Things</strong></td>
<td>Customers who respond positively regardless.</td>
</tr>
<tr>
<td><strong>Lost Causes</strong></td>
<td>Customers who respond negatively regardless.</td>
</tr>
<tr>
<td><strong>Sleeping Dogs</strong></td>
<td>Customers who respond negatively to marketing activity.</td>
</tr>
</tbody>
</table>
True customer groups are **unknown**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td>No Response</td>
</tr>
<tr>
<td>Persuadables,</td>
<td>Sleeping Dogs,</td>
</tr>
<tr>
<td>Sure Things</td>
<td>Lost Causes</td>
</tr>
</tbody>
</table>
## Standard Model

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Response</strong></td>
<td><strong>No Response</strong></td>
</tr>
<tr>
<td>Persuadables, Sure Things</td>
<td>Sleeping Dogs, Lost Causes</td>
</tr>
</tbody>
</table>

**Positive**
- Persuadables
- Sleeping Dogs
- Sure Things

**Negative**
- Persuadables
- Sleeping Dogs
- Lost Causes
# Response Model

## Treatment vs. Control

<table>
<thead>
<tr>
<th>Response</th>
<th>No Response</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persuadables, Sure Things</td>
<td>Sleeping Dogs, Lost Causes</td>
<td>Sleeping Dogs, Sure Things</td>
</tr>
<tr>
<td>Persuadables, Sure Things</td>
<td>Sleeping Dogs, Lost Causes</td>
<td>Persuadables, Sure Things</td>
</tr>
</tbody>
</table>

## Positive

- Persuadables
- Sure Things

## Negative

- Sleeping Dogs
- Lost Causes
Uplift Modeling
(RADCLIFFE & SIMPSON, 2008)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Response</th>
<th>No Response</th>
<th>Control</th>
<th>Response</th>
<th>No Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persuadables, Sure Things</td>
<td>Sleeping Dogs, Lost Causes</td>
<td>Sleeping Dogs, Sure Things</td>
<td>Persuadables, Lost Causes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

POSITIVE
Persuadables
Sure Things
Sleeping Dogs
NEGATIVE
Sleeping Dogs
Lost Causes
Persuadables

Treatment ↔ Control
Uplift Modeling
(RADCLIFFE & SIMPSON, 2008)

**Lift**
The number of responders that a classifier identifies at a given proportion of the population targeted.

**Uplift**
The difference in lift produced by a classifier between treatment and control subgroups.

\[ AUU = AUL_T - AUL_C \]
COX-2 Inhibitors

- Non-steroidal anti-inflammatory drug (NSAID)
- Significantly reduced occurrence of adverse gastrointestinal effects common to other NSAIDs (e.g. ibuprofen)
- Wide use for treatment of ailments such as arthritis
- Later clinical trials showed increased risk of myocardial infarction (MI), or “heart attack”
COX-2 Inhibitors

Main Assumption
Patients with an increased risk of MI due to treatment with COX-2 inhibitors are directly analogous to **Persuadables**.
Support Vector Machines

Find maximum-margin separating plane between positive and negative examples.
Extend previous SVM work maximizing AUC (Joachims, 2005) to maximize AUU instead.
ROC and AUC
SVM for Uplift

Let the positive skew of data be:

\[ \pi = \frac{P}{P + N} \]

Then (Tuffery, 2011):

\[ AUL = P \times \left( \frac{\pi}{2} + (1 - \pi)AUC \right) \]
SVM for Uplift

\[ AUU = AUL_T - AUL_C = P_T \times \left( \frac{\pi_T}{2} + (1 - \pi_T)AUC_T \right) - P_C \times \left( \frac{\pi_C}{2} + (1 - \pi_C)AUC_C \right) \]

\[ \max(AUU) \equiv \max(P_T \times (1 - \pi_T)AUC_T - P_C \times (1 - \pi_C)AUC_C) \]

\[ \propto \max \left( AUC_T - \frac{P_C \times (1 - \pi_C)}{P_T \times (1 - \pi_T)} AUC_C \right) \]

\[ \max(AUU) \equiv \max(AUC_T - \lambda AUC_C) \]
Uplift Modeling Simulation: Persuadable ROC

- Generated synthetic customer population
- Subjected customer population randomly to simulated marketing activity
- Measured ROC with **Persuadables** as the positive class, others as negative
Uplift Modeling Simulation: Persuadable ROC
COX-2 Inhibitor Results
## COX-2 Inhibitor Results

<table>
<thead>
<tr>
<th>Model</th>
<th>AUU</th>
<th>COX-2 AUL</th>
<th>No COX-2 AUL</th>
<th>AUU p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM&lt;sup&gt;Upl&lt;/sup&gt;</td>
<td>50.7</td>
<td>123.4</td>
<td>72.7</td>
<td>-</td>
</tr>
<tr>
<td>COX-2-Only</td>
<td>13.8</td>
<td>151.5</td>
<td>137.7</td>
<td>0.002*</td>
</tr>
<tr>
<td>Standard</td>
<td>1.2</td>
<td>147.7</td>
<td>146.5</td>
<td>0.002*</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.002*</td>
</tr>
</tbody>
</table>
• Introduction
• Advice-Based Learning Framework
• Support Vector Machines for Uplift Modeling
• Conclusions
Contributions

In This Presentation

• Developed framework for collaboration between clinicians and machine learning experts to address challenges in decision support (Kuusisto et al., 2015)

• Developed support vector machine for uplift modeling to address COX-2 inhibitor treatment and understand indolent breast cancer in older patients (Kuusisto et al., 2014)
Contributions

In This Presentation

• Developed framework for collaboration between clinicians and machine learning experts to address challenges in decision support (Kuusisto et al., 2015)

• Developed support vector machine for uplift modeling to address COX-2 inhibitor treatment and understand indolent breast cancer in older patients (Kuusisto et al., 2014)

Other Contributions

• Investigated use of machine learning for accurately estimating individualized treatment effects versus traditional approaches with RCT and observational data (Weiss et al., 2015)

• Developed statistical relational uplift modeling algorithm to understand factors contributing to indolent breast cancer in older patients (Nassif et al., 2013)

• Applied inductive logic programming with rule evaluation function tailored to meet clinical objective (Kuusisto et al., 2013)
Overall Conclusions

• Close collaboration with clinicians is essential to develop models to meet clinical objectives
• Leveraging clinical expertise in model-building can alleviate challenges of gathering sufficient data for rare diseases
• Machine learning and uplift modeling have potential applications in treatment assignment and knowledge discovery
Acknowledgements

Advisors: Jude Shavlik, David Page

Committee: Elizabeth Burnside, Vitor Santos Costa, Charles Dyer

Collaborators: Ines Dutra, Houssam Nassif, Yirong Wu

Funding: NLM R01LM010921, NIH R01CA165229, NIH R01LM011028, NIGMS R01GM097618

Friends: Eric Lantz, Aubrey Barnard, Alex Cobian, Kendrick Boyd, Jeremy Weiss, Jie Liu, Brandon Smith, Sarah Edlund

Family: Maggie Kuusisto, Larry Kuusisto, Elina Kuusisto
Thank You!
Future Directions
Uplift Bayesian Networks

**Uplift TAN**

\[
I_{DIFF}(A; B|\text{Class}) = I_{treat}(A; B|\text{Class}) - I_{control}(A; B|\text{Class})
\]
Net Benefit Maximization

- Can evaluate treatment assignment model on RCT data (Vickers et al., 2007)
- Could optimize for treatment assignment directly
Model Calibration
Other Work
Breast Cancer States

**In Situ**
- Earlier state
- Cancer localized

**Invasive**
- Later state
- Cancer has invaded surrounding tissue
Breast Cancer Age Differences

**Older**
- Cancer tends to progress *less aggressively*
- Patient has *less* time for progression

**Younger**
- Cancer tends to progress *more aggressively*
- Patient has *more* time for progression
## Uplift SVM Older In Situ Rules

10 = Clinically Interesting  
1 = Clinically Counter-Intuitive

<table>
<thead>
<tr>
<th>Rank</th>
<th>Feature</th>
<th>Older In Situ Correlation</th>
<th>Radiologist Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Linear Calc. Distribution Present</td>
<td>Positive</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>Spiculated Mass Margin Present</td>
<td>Negative</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>Palpable Lump Present</td>
<td>Positive</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>Irregular Mass Shape Present</td>
<td>Negative</td>
<td>9-10</td>
</tr>
<tr>
<td>5</td>
<td>No Family History</td>
<td>Negative</td>
<td>8</td>
</tr>
</tbody>
</table>
Upgrade Rules

Use F-score to learn precise rules to predict benign non-definitive biopsies

**Algorithm**  Rule Learning Procedure

\[
\text{for Train, Test } \in \text{ Folds do}
\]

- Theory ← Aleph(Train, minpos = 2, noise = 0, evalfn = F_\beta);
- Rule* ← argmax F_\beta(Theory, Train);
- Evaluate(Rule*, Test);

\text{end for}
Upgrade Rules

1. The patient did not have a previous surgery, imaging did not present a spiculated mass margin, and the abnormality did not disappear in post-biopsy imaging

<table>
<thead>
<tr>
<th>Benign Avoided</th>
<th>Malignant Missed</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0</td>
</tr>
</tbody>
</table>

2. Imaging did not present an indistinct mass margin, imaging did not present a spiculated mass margin, and the abnormality did not disappear in post-biopsy imaging

<table>
<thead>
<tr>
<th>Benign Avoided</th>
<th>Malignant Missed</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>0</td>
</tr>
</tbody>
</table>

3. Imaging did not present a spiculated mass margin, and the abnormality did not disappear in post-biopsy imaging

<table>
<thead>
<tr>
<th>Benign Avoided</th>
<th>Malignant Missed</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>1</td>
</tr>
</tbody>
</table>

4. Imaging did not present an indistinct mass margin, and the abnormality did not disappear in post-biopsy imaging

<table>
<thead>
<tr>
<th>Benign Avoided</th>
<th>Malignant Missed</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>1</td>
</tr>
</tbody>
</table>

5. The patient has no personal history of breast cancer, and the abnormality did not disappear in post-biopsy imaging

<table>
<thead>
<tr>
<th>Benign Avoided</th>
<th>Malignant Missed</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>0</td>
</tr>
</tbody>
</table>
Use ILP to induce feature set used by BN that maximizes uplift.
SAYL - Older Model

breast category

combined BI-RADS increased up to 3 points over previous mammogram

had previous in situ biopsy at same location

no family history of cancer, and no prior surgery

breast BI-RADS score = 4

breast has mass size ≤ 13 mm
SAYL - Younger Model

- Breast category
- Had previous in situ biopsy at same location
- Breast BI-RADS score = 4
- Combined BI-RADS increased up to 3 points over previous mammogram
- No family history of cancer, and no prior surgery
- Breast has mass size ≤ 13 mm
Individualized Treatment

```
<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs.</th>
<th>RCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>age (older)</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>smoke</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>gender (male)</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>HDL</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>LDL</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>diabetes</td>
<td>39</td>
<td>42</td>
</tr>
<tr>
<td>family history of CVD</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>blood pressure</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>history of angina</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>history of stroke</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>history of depression</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>statin use</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>MI</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>
```
Individualized Treatment