Data Mining for Translation to Practice

Chih-Lin Chi, Ph.D., M.B.A.
Assistant Professor, School of Nursing
Core Faculty, Institute for Health Informatics
University of Minnesota

Second International Conference on Research Methods for Standard Terminologies
April 15, 2015
DISCLOSURES

There are no conflicts of interest or relevant financial interests that have been disclosed by this presenter or the rest of the planners and presenters of this activity that apply to this learning session.
Steps for Translating from Big Data to Practical Use

1. **Computational**
   A. Develop research question and data-mining approaches
   B. Demonstrate preliminary results of these approaches for a single Problem
   C. Standardize the process and develop data-mining pipeline for other Problems
   D. Validate with world-wide structured nursing data
   E. Simulated clinical trail using client randomization

2. **Practical**
   A. Test on home-visiting care scenarios
   B. Integrate with current workflow and develop software and guidelines to facilitate the use in practical settings (e.g., identify patients, identify personalized interventions)
   C. Implementation
The IHI Triple Aim

Population Health

Experience of Care

Per Capita Cost

Predict need for intervention

Think about a difficult problem in a population. Regardless of outcome, who will need more interventions?
Predict responsiveness to interventions

Within the population, which individuals will be responsive to more interventions for this problem, compared to those who are less responsive?
Predict type of intervention that will be efficient and effective for an individual

Understand different intervention patterns in order to personalize care planning based on an individual’s characteristics
Benefits of Standardized Terminologies in Data Mining

Big Data + Data Mining = Progress to Triple Aim

Why use a Standardized Terminology for Big Data?

– Pre-classification of clinical knowledge
– Outcome metrics
– Relational database structure
Benefit: Pre-classification of Clinical Knowledge

Problem representation
- Domains
- Problems
- Signs/symptoms

Intervention representation
- Categories

Outcomes
- Knowledge
- Behavior
- Status
Benefit 2: Outcome Metrics

Explicit outcome measurement for all problems
Not looking for surrogate or proxy measures
  E.g. claims data, laboratory results
Less chance of missing values
Benefit 3: Relational database structure

All data relate to a central concept (Problem) Improves clinical and theoretical management of information
Data Mining for Translation to Practice: Oral health

1. **Computational**
   A. Develop research question and data-mining approaches
   B. Demonstrate preliminary results of these approaches for a single Problem
   C. Standardize the process and develop data-mining pipeline for other Problems
   D. Validate with world-wide structured nursing data
   E. Simulated clinical trial using client randomization

2. **Practical**
   A. Test on home-visiting care scenarios
   B. Integrate with current workflow and develop software and guidelines to facilitate the use in practical settings (e.g., identify patients, identify personalized interventions)
   C. Implementation
Our Translational Project Starting from Data-Mining Approaches: Oral health problem

The health of the mouth and surrounding craniofacial (skull and face) structures is central to a person’s overall health and well-being.

Social determinants affect oral health. In general, people with lower levels of education and income, and people from specific racial/ethnic groups, have higher rates of disease.

People with disabilities and other health conditions, like diabetes, are more likely to have poor oral health.

https://www.healthypeople.gov/2020/topics-objectives/topic/oral-health
Data Set

Clients (N=1,618 or subset)
Characteristics (demographic and signs/symptoms)

Interventions (113,989 or subset)
Teaching, Guidance, and Counseling
Treatments and Procedures
Case Management
Surveillance

Outcomes
Knowledge
Behavior
Status
Steps for Translating from Big Data to Practical Use

1. **Computational**
   - A. Develop research question and data-mining approaches
   - B. Demonstrate preliminary results of these approaches for a single problem
   - C. Standardize the process and develop data-mining pipeline for other problems
   - D. Validate with world-wide structured nursing data
   - E. Simulated clinical trial using client randomization

2. **Practical**
   - A. Test on home-visiting care scenarios
   - B. Integrate with current workflow and develop software and guidelines to facilitate the use in practical settings (e.g., identify patients, identify personalized interventions)
   - C. Implementation
Our research question as an example: A small percentage of clients consume a high percentage of service resources (80-20 rule in Oral health problem)

20% patients use 70% of intervention resource
Data Mining (Visualization) Method Used to Show Intervention Usage

Excel

Sort and rank clients based on percentage of interventions received for the episode of care
Create line graph of cumulative percentage of interventions for the entire sample
Steps for Translating from Big Data to Practical Use

1. **Computational**
   A. Develop research question and data-mining approaches
   B. Demonstrate preliminary results of these approaches for a single Problem
   C. Standardize the process and develop data-mining pipeline for other Problems
   D. Validate with world-wide structured nursing data
   E. Simulated clinical trail using client randomization

2. **Practical**
   A. Test on home-visiting care scenarios
   B. Integrate with current workflow and develop software and guidelines to facilitate the use in practical settings (e.g., identify patients, identify personalized interventions)
   C. Implementation
Detail Research Question 1 of 3: Predict Intervention Usage

Regardless of outcome, who will need more interventions?

For 50% threshold
Maximal accuracy ~ 60%
Maximal AUC ~ 75%

For 75% threshold
Maximal accuracy ~ 74%
Maximal AUC ~ 77%

Prediction measured using receiver operating curves and area under the curve (AUC).
Data Mining Method Used to Predict Intervention Usage

Support vector machines in Matlab software
- Input: Client characteristics (demographics and signs/symptoms from first encounter)
- Output: Interventions across all clients (compared to 50th and 75th percentiles)
Detail Research Question 2 of 3: Predict Personalized Responsiveness to Interventions

Within the population, which individuals will be responsive to more interventions for this problem, compared to those who are less responsive?
Data Mining Method Used to Predict Personalized Responsiveness to Interventions

Support vector machines in Matlab software with sensitivity analysis

Input: Client characteristics (demographics and signs/symptoms from first encounter), interventions, and any KBS improvement from admission to discharge

Output: Responsive score based on personal characteristics
Detail Research Question 3 of 3: Predict Personalized Nursing Intervention

How to personalize care planning based on an individual’s characteristics and what intervention patterns can be used to help personalization?

Intervention patterns typically used in Oral health

<table>
<thead>
<tr>
<th></th>
<th>Teaching, guidance, and counseling</th>
<th>Treatments and procedures</th>
<th>Case management</th>
<th>Surveillance</th>
<th>Number of clients</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>100.00%</td>
<td>24</td>
</tr>
<tr>
<td>B</td>
<td>0.00%</td>
<td>10.00%</td>
<td>0.00%</td>
<td>90.00%</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>0.00%</td>
<td>20.00%</td>
<td>0.00%</td>
<td>80.00%</td>
<td>285</td>
</tr>
<tr>
<td>D</td>
<td>30.00%</td>
<td>0.00%</td>
<td>30.00%</td>
<td>40.00%</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>30.00%</td>
<td>10.00%</td>
<td>10.00%</td>
<td>50.00%</td>
<td>1</td>
</tr>
<tr>
<td>F</td>
<td>40.00%</td>
<td>0.00%</td>
<td>10.00%</td>
<td>50.00%</td>
<td>210</td>
</tr>
<tr>
<td>G</td>
<td>50.00%</td>
<td>0.00%</td>
<td>10.00%</td>
<td>40.00%</td>
<td>234</td>
</tr>
<tr>
<td>H</td>
<td>60.00%</td>
<td>0.00%</td>
<td>10.00%</td>
<td>30.00%</td>
<td>1</td>
</tr>
</tbody>
</table>
Data Mining Method Used to Summarize Intervention Patterns

Simple cluster analysis in Excel using round-up or round-down technique

Proportion of interventions by category observed in the data
Relative Improvement of Predicted Personalized Nursing Intervention

Relative improvement is 51% (compared to maximum possible improvement for all clients)

Choosing the right pattern can improve care (efficiency and effectiveness)
Maximum possible improvement: 1.18% (the largest possible improvement space) Standard deviation: 0.64%
Comparison Baseline: Random Assignment of Interventions

Random baseline improvement: -0.03% (randomly choose 1 of the 8 patterns for each patient)

Showing importance of personalized interventions
Data Mining Method Used to Predict Personalized Nursing Intervention

Support vector machines in Matlab software with optimization

Input: Client characteristics (demographics and signs/symptoms from first encounter) and intervention patterns for each client

Output: Any KBS improvement from admission to discharge
Steps for Translating from Big Data to Practical Use

1. **Computational**
   A. Develop research question and data-mining approaches
   B. Demonstrate preliminary results of these approaches for a single Problem
   C. Standardize the process and develop data-mining pipeline for other Problems
   D. Validate with world-wide structured nursing data
   E. Simulated clinical trial using client randomization

2. **Practical**
   A. Test on home-visiting care scenarios
   B. Integrate with current workflow and develop software and guidelines to facilitate the use in practical settings (e.g., identify patients, identify personalized interventions)
   C. Implementation
## Building Evidence (your study here)

### Evidence matrix

<table>
<thead>
<tr>
<th>Problems Data Source</th>
<th>Oral health Problem</th>
<th>Caretaking/parenting</th>
<th>Growth and development</th>
<th>Pregnancy</th>
<th>Other Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dakota County, MN</td>
<td>Personalized intervention evidence</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
</tr>
<tr>
<td>Other County, MN</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
</tr>
<tr>
<td>Other County, Other State</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
</tr>
<tr>
<td>Other Country</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
<td>To be done</td>
</tr>
</tbody>
</table>

Goal: Standardize the data-mining process and develop a pipeline to automate and document evidence.
More Data = More Confidence in Findings and More External Generalizability

Data Mining Methods can increase confidence when using observational data

Personalized method helps us understand and validate predictions (remember netflix)
Typical Data Mining Validation Process: Compare Predictive with True Outcomes

Patient Data 1

Patient Data 2

Training

Prediction model

Independent Variable

Dependent Variable (Outcome)

Validation (compare predicted to true outcomes)

Independent Variable

Dependent Variable (outcome)

Predictive performance: e.g., AUC, accuracy, confusion matrix, F-measure, … etc

More complicated method to further reduce variation, such as (multiple) 10-fold cross validation, leave one out, etc
Steps for Translating from Big Data to Practical Use

1. Computational
   A. Develop research question and data-mining approaches
   B. Demonstrate preliminary results of these approaches for a single Problem
   C. Standardize the process and develop data-mining pipeline for other Problems
   D. Validate with world-wide structured nursing data
   E. Simulated clinical trail using client randomization

2. Practical
   A. Test on home-visiting care scenarios
   B. Integrate with current workflow and develop software and guidelines to facilitate the use in practical settings (e.g., identify patients, identify personalized interventions)
   C. Implementation
Using clinical trial simulations to provide baseline information to learn efficacy improvement

- Arm 1: Non-personalized approach
- Arm 2: Personalized approach

Predicted outcome for personalized approach → Predicted outcome for non-personalized approach → Compare efficacy improvement
Steps for Translating from Big Data to Practical Use

1. Computational
   A. Develop research question and data-mining approaches
   B. Demonstrate preliminary results of these approaches for a single Problem
   C. Standardize the process and develop data-mining pipeline for other Problems
   D. Validate with world-wide structured nursing data
   E. Simulated clinical trail using client randomization

2. Practical
   A. Test on home-visiting care scenarios
   B. Integrate with current workflow and develop software and guidelines to facilitate the use in practical settings (e.g., identify patients, identify personalized interventions)
   C. Implementation
20% patients use 70% of intervention resource.
Take home message for data mining techniques and the use for translational research

- Triple Aim
- Standardized Terminology
- Methods for Translation
  - Data mining to practical implementation
- Oral health example
  - Small percentage of clients receive most interventions
  - Personalization of care
- Collaboration
Martin KS. The Omaha System: A key to practice, documentation, and information management (Reprinted 2nd ed.). Omaha, NE: Health Connections Press; 2005.