Predictive Modelling for Clinical Applications

Harry Freitas da Cruz
Data Management for Digital Health
Summer 2017
Recap Use Case Nephrology

- The urinary system
- How the kidneys work
- Kidney diseases
- Bayesian networks for AKI
- Machine learning in Nephrology
Agenda

- Predictive analytics in healthcare
- Clinical Decision Support Systems (CDSS)
- Clinical Data Repository (CDR)
- Establishing Clinical Prediction Models
Your heart attack will happen tomorrow.
Predictive Analytics in Healthcare
Opportunities and Challenges

**OPPORTUNITIES**

Patient safety

Precision medicine

Risk stratification

**CHALLENGES**

Physician and patient trust

Data standardization

Novel algorithms and tools

**Predictive Modelling for Clinical Applications**

Data Management for Digital Health, Summer 2017

https://www.healthcatalyst.com/three-approaches-to-predictive-analytics-in-healthcare
Predictive Analytics in Healthcare Clinical Decision Making

- Huge variety of clinical (increasingly genomic) data
- Questions to be answered:
  - What disease does this patient have?
  - Should this patient be treated?
  - Should testing be done?
- Accurate, complete, relevant data
- Reliance on pattern recognition and customary practices
- Evidence-based medicine (clinical guidelines)
- Medical errors are frequent

Clinical Decision Support Systems Definition

According to Kabari et al.:

“Clinical decision support systems (CDSS) provide clinicians, staff, patients, and other individuals with **knowledge** and **person-specific information**, intelligently filtered and presented at **appropriate times**, to enhance health and health care”¹

---

Clinical Decision Support Systems
Typical Application Scenarios

- Usual scenarios
  - Diagnostic support
  - Preventive care
  - Treatment planning / recommendations

- How can this be achieved?
  - Contextual retrieval of highly relevant information
  - Patient-specific reminders and recommendations
  - Organization and presentation of information

- Information logistics / 5 „rights“
  - Information, person, format, channel, time

Clinical Decision Support Systems
Contextual Retrieval of Highly Relevant Information

Predictive Modelling for Clinical Applications
Data Management for Digital Health, Summer 2017
12

https://www.visualdx.com/benefits/recognize-drug-reactions
Clinical Decision Support Systems
Patient-specific Reminders and Recommendations

Predictive Modelling for Clinical Applications
Data Management for Digital Health, Summer 2017
Clinical Decision Support Systems
Organization and Presentation of Information

Predictive Modelling for Clinical Applications
Data Management for Digital Health, Summer 2017
14
Clinical Data Repository Heterogeneous Landscape

Clinical Data Repository Challenges and opportunities

- Difficult and costly to develop a general purpose CDR
- Integration of different data sources is a daunting task
- Data inconsistency, redundancies are frequent
- Difficult to adapt to changing needs of users

- 360° patient view
- Explore patterns in disease progression and management
- Discover unknown patterns in the data
- Faster hypothesis testing -> Clinical studies


Predictive Modelling for Clinical Applications
Data Management for Digital Health, Summer 2017
16
Clinical Data Repository Nephrology Use Case (AKI): Patient Dashboard

Kidney dashboard

Patient dashboard

Age: 77
Sex: M
Ethnicity: WHITE

Acute Kidney Injury Risk

Serum creatinine (mg/dL)

Gromerular Filtration Rate (mL/min/1.73 m²)
Clinical Prediction Models
Occam’s Razor

- Models are an abstraction of reality
- May still be useful depending on the purpose
- Start simple: robust and difficult to break
- As per William of Occam (1287): ontological parsimony
- In diagnostics, not always the case:
  - Hickam’s dictum
  - “Patients can have as many diseases as they damn well please”
Clinical Prediction Models
Supervised or Unsupervised?

- **Supervised learning**
  - Labeled data is available
  - Categorical or numerical responses
  - Decision trees, Bayesian nets, ridge regression, etc.

- **Unsupervised learning**
  - No labeled data
  - Finding hidden patterns in data
  - Hierarchical clustering, k-means, etc.

Clinical Prediction Models

Types of Models

- Classification-based
  - Binary or multi-class
  - Diagnosis, risk stratification, hospital readmission
- Regression-based
  - Optimal drug dosage
  - Treatment plan adjustment
- Survival analysis
  - Time-to-event models
  - Cancer mortality


Predictive Modelling for Clinical Applications

Data Management for Digital Health, Summer 2017

21
Clinical Prediction Models
Establishing a CPM: Preparation and Dataset Selection

- **Step 1: Preparation**
  - What is the target outcome?
  - What is the target patient?
  - What is the target user?

- **Step 2: Dataset selection**
  - Is the data needed available?
  - Is the data representative?
  - What is the validation strategy?

---

Clinical Prediction Models
Establishing a CPM: Preparation and Dataset Selection

- Step 1: Preparation
  - What is the target outcome?
  - What is the target patient?
  - What is the target user?

- Step 2: Dataset selection
  - Is the data needed available?
  - Is the data representative?
  - What is the validation strategy?


Predictive Modelling for Clinical Applications
Data Management for Digital Health, Summer 2017
23
Clinical Prediction Models
Establishing a CPM: Variable Handling

- Expert judgement often necessary
- Previous significant factors should be used
- Avoid predictors that are possible correlated
- Merging categorical variables should be considered
- Scale matters in continuous variables: consistency
- Nominalization often needed for continuous variables
- Consider scale transformation (e.g. log)
- Normalization of values (e.g. from 0-1)
- Identify and handle outliers

Clinical Prediction Models
Establishing a CPM: Missing Data

- Leaving out the missing ones: complete case analysis
- Single imputation x multiple imputation
  - Using “other” or “unknown”
  - Averaging occurrences, median or mean
- MICE (Multiple Imputation using Chained Equations)
- Regression model from the existing variables

Clinical Prediction Models
Establishing a CPM: Model Generation

- Selecting the proper algorithm
- Performing feature selection
  - Backward elimination
  - Stepwise selection
- Trade-offs between goodness of fit and complexity
  - Akeike Information Criterion
  - Bayes Information Criterion
- Perform parameter tuning
  - Optimization of hyperparameters

http://doi.org/10.3803/EnM.2016.31.1.38

Predictive Modelling for Clinical Applications
Data Management for Digital Health, Summer 2017
26
Clinical Prediction Models
Establishing a CPM: Evaluation and Validation

- Internal validation
  - Cross-validation
  - Bootstrapping
- External validation
  - Using a different data source
  - Ensure transportability and generalizability
- Measures of performance
  - ROC Curve
  - R², p-values
- True positive rate (TPR) / true negative rate
Clinical Prediction Models
Establishing a CPM: Measures of Performance

<table>
<thead>
<tr>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity and specificity</td>
</tr>
<tr>
<td>Discrimination (ROC/AUC)</td>
</tr>
<tr>
<td>Predictive values: positive, negative</td>
</tr>
<tr>
<td>Likelihood ratio: positive, negative</td>
</tr>
<tr>
<td>Accuracy: Youden index, Brier score</td>
</tr>
<tr>
<td>Number needed to treat or screen</td>
</tr>
<tr>
<td>Calibration: Calibration plot, Hosmer-Lemeshow test</td>
</tr>
<tr>
<td>$R^2$ statistical significance: p-value (e.g. likelihood ratio test)</td>
</tr>
<tr>
<td>Magnitude of association, e.g., $\beta$ coefficients, odds ratio</td>
</tr>
<tr>
<td>Model quality: Akeike IC/ Bayes IC</td>
</tr>
<tr>
<td>Net reclassification index and integrated discrimination improvement</td>
</tr>
<tr>
<td>Net benefit</td>
</tr>
<tr>
<td>Cost-effectiveness</td>
</tr>
</tbody>
</table>

Measure of model performance (Lee 2016)

---


Predictive Modelling for Clinical Applications
Data Management for Digital Health, Summer 2017
28
What to Take Home?

- There is lots of potential for health care analytics
- Clinical Decision Support Systems offer a way to tap into that
- Step-by-step process to establishing clinical prediction models
  - Preparation
  - Data selection
  - Model generation
  - Evaluation and validation
To Know More

Predictive Modelling for Clinical Applications
Data Management for Digital Health, Summer 2017
33
What’s Coming Next?

- Applying the step-by-step process with an example and tool (Rapid Miner)
- Instructions for the exercise