A Probabilistic Look into the Semantics of Medicine

Peter Haug, MD

International Conference on Semantic Computing 2014
June 16-19, Newport Beach, CA
Agenda

• Medical Models
• “Big” Healthcare Data
  – Capture as a part of the care process
  – Strong Temporal Component
  – Inherently messy
• Extracting Models from Data
  – A place for Ontologies
  – Semi-Automatic extraction of Diagnostic Models
• Bayesian Modeling
  – Variations on the theme
Agenda

• Medical Models
• “Big” Healthcare Data
  – Capture as a part of the care process
  – Strong Temporal Component
  – Inherently messy
• Extracting Models from Data
  – A place for Ontologies
  – Semi-Automatic extraction of Diagnostic Models
• Bayesian Modeling
  – Variations on the theme
Models for Healthcare Delivery

A foundation for clinical research and care delivery

Modeling and Model Implementation
- Medical Data Models
- Diagnostic Models
- Therapeutic Models
- Workflow/Business Process Models
- Clinical Trials Models
- Natural Language Models
- Research Models
- Physiologic Models
- Ontologies
- Temporal Models
- Predictive Models
- Translational Research
- Data Visualization
Bayes Equation

$$P(D | F) = \frac{P(F \text{ and } D)}{P(F)}$$

- Probability of Disease When Finding exists
- Probability of Both The Disease and Finding
- Probability of Finding
Bayes Equation-A Graphical View

Rendered as a Bayesian Network

<table>
<thead>
<tr>
<th>Disease</th>
<th>Present</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease Present</td>
<td>34.0</td>
<td></td>
</tr>
<tr>
<td>Disease Absent</td>
<td>66.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Finding</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finding Positive</td>
<td>38.9</td>
<td></td>
</tr>
<tr>
<td>Finding Negative</td>
<td>61.1</td>
<td></td>
</tr>
</tbody>
</table>
Bayes Equation – Various Variables

Rendered as a Bayesian Network
Agenda

• Medical Models
• “Big” Healthcare Data
  – Capture as a part of the care process
  – Strong Temporal Component
  – Inherently messy
• Extracting Models from Data
  – A place for Ontologies
  – Semi-Automatic extraction of Diagnostic Models
• Bayesian Modeling
  – Variations on the theme
Sources for Building Medical Models

What are the sources of knowledge for Clinical Decision Support

- Clinical Trials
  - Prospective Trials
  - Observational Studies
- Physiologic Principals/Models
- Expert Opinion
- Common Sense
- CLINICAL DATA
Clinical Data Serves Us Twice

Multiple Episodes of Care

Supporting Care/Collecting Data

Clinical Database

Data Warehouse: Re-Organizing the Data

Research and Analysis

Medical Decision Support

Intermountain Healthcare
# EDW Interesting Facts...

<table>
<thead>
<tr>
<th>Fact</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td># of EDW login accounts</td>
<td>2500</td>
</tr>
<tr>
<td># of EDW “Consumers”</td>
<td>Thousands</td>
</tr>
<tr>
<td># Records in EDW</td>
<td>70,000,000,000</td>
</tr>
<tr>
<td>Size (bytes)</td>
<td>~ 10 Terabytes (production)</td>
</tr>
<tr>
<td>Avg monthly queries</td>
<td>~ 150,000,000</td>
</tr>
<tr>
<td>Most queried table</td>
<td>LKUP.PATIENT_MASTER</td>
</tr>
<tr>
<td>Largest table (records)</td>
<td>HELP.PT_DATA (Patient Data Table from HELP1) 19.5 Billion Records</td>
</tr>
<tr>
<td># queriable tables</td>
<td>12,500</td>
</tr>
<tr>
<td>Avg query run time</td>
<td>6 seconds</td>
</tr>
<tr>
<td>Avg rows returned per query</td>
<td>3500</td>
</tr>
<tr>
<td># of years of data (for major data sets)</td>
<td>10 – 15 years (25+ for others)</td>
</tr>
</tbody>
</table>
Agenda

• Medical Models
• “Big” Healthcare Data
  – Capture as a part of the care process
  – Strong Temporal Component
  – Inherently messy
• Extracting Models from Data
  – A place for Ontologies
  – Semi-Automatic extraction of Diagnostic Models
• Bayesian Modeling
  – Variations on the theme
Using Ontologies to Extract Build Diagnostic Models

*There is a need for predictive (diagnostic) models for care.*

- Predictive models have value in clinical care environments
  - Have successfully deployed pneumonia and sepsis diagnostic models
- Building models from data is resource intensive.
- Ontologies can support diagnostic modeling
  - Requires a database containing data collected during routine care.
  - An **Ontology** to capture clinical relationships among data elements.
  - An application (the Ontology-driven Diagnostic Modeling System) is used to automate the initial analysis.
Ontologies Describe How Clinical Data are Related to Diseases

- Pneumonia
  - Bacterial Pneumonia
    - Pneumococcal pneumonia
      - Pneumococcal pneumonia ICD9: 481
    - Other Bacterial Pneumonia
      - Other bacterial pneumonia
        - Other bacterial pneumonia ICD9: 482
  - Pneumonia, Organism unspecified ICD9: 486
  - More Pneumonias
    - has_X-ray_Manifestation
    - has_Micro_Manifestation
  - More Bacterial Pneumonias
    - has_??_Manifestation
  - More Manifestations
    - has_Altered_VS
    - has_Altered_Lab_Value
    - has_Sign
    - Localize Infiltrate
      - X-ray Finding: Localized Infiltrate
        - SNOMED: 128309002
      - + Sputum Culture
        - Sputum Culture: Positive
        - SNOMED: 442773002
    - Vital Signs: Temperature
      - LOINC: 8310-5
    - White Blood Count
      - Hematology: White Blood Count
      - LOINC: 62239-9
    - Pulmonary Rales
      - Signs: Chest Auscultation-Rales
      - PTXT: 28.1.3.22.34.2.1.32
Ontology-Driven Model Discovery

Using knowledge embedded in ontologies to automate research?

- Disease Ontology
  - Concept Retrieval (from Ontology)
  - Concept Translation to EDW Representation
  - Data Warehouse/Analytic Health Repository
  - Natural Language Processing Subsystem

- Structural Knowledge Retrieval from the Ontology
  - Data Retrieval from the Analytic Health Repository

- Analytic Workbench
  - Diagnostic Models
  - Model Comparisons
  - Model Explanation (by reference to the Ontology)

Output
- Analysis Results
- Analytic Data
- Prediction Algorithm
- Relevant Ontologic Concepts

Intermountain Healthcare
A Care Delivery Framework
(*multi-factor screening and treatment*)

**Example:** Community-Acquired Pneumonia

- **Chest Xray Reports**
- **Chest Xray Report Processing (Structured Data Extraction)**
- **Data Supporting Pneumonia Assessment**
- **Pneumonia Screening Tool**
- **Clinical Data Repository**

**Computable Medical Knowledge Repository**

Does the patient have pneumonia?

Should we used the protocol?

**Pneumonia Protocol Enrollment**

**Pneumonia Treatment Protocol**

Apply **Pneumonia Care Protocol**.
Alerting for Pneumonia in the Patient Tracking System

- System Watches the Data Flow in the ED
- Identifies Possible Pneumonia Patients
Hospital Ward Admission recommended. The patient has Moderate CAP based on no HCAP factors, a 30 day mortality risk of 8.2%, and a 33% likelihood of needing ICU treatment, based on 2 Severe CAP criteria.

Health Care Acquired Pneumonia Factors
A single positive HCAP factor means the patient has HCAP.
- Hospitalization \( \geq 2 \) days: Yes \( \text{or} \) No
  - within 90 days:
- Nursing Home Resident: Yes \( \text{or} \) No
- Wound Care or Infusion Therapy within 30 days:
- Chronic Dialysis within 30 days: Yes \( \text{or} \) No

Vital Factors
- Age: 78 Years
- Confusion (patient not oriented to person, place, or time):
  - Temperature: 37.5°C
  - Respiratory Rate: 26 BPM
  - Systolic Blood Pressure: 102 mm/Hg

Radiology
- Pleural Effusion:
- Infiltrates:
  - Single Lobe
  - Multilobar

Labs
- BUN: 21 mg/dL
- WBC: 29300 cells/mm³
- Platelet Count: 268000 cells/mm³
- \( \text{PaO}_2/\text{FiO}_2 \) Ratio: 252.3 mm/Hg

\[
P/F \text{ Ratio} = \frac{52.974}{0.210 - 252.255} \\
\text{SpO}_2 \rightarrow \text{PaO}_2 \text{ - conversion by Ellis equation} = 52.974 \\
\text{Calculated FiO}_2 = 0.210 \times (0.000 \times 0.03) = 0.210 \\
\text{SpO}_2 \text{ on room air} = 87.000 \\
\text{FiO}_2 \text{ at room air} = 0.21
\]
Agenda

• Medical Models
• “Big” Healthcare Data
  – Capture as a part of the care process
  – Strong Temporal Component
  – Inherently messy
• Extracting Models from Data
  – The classic research paradigm
  – Semi-Automatic extraction of Diagnostic Models
• Bayesian Modeling
  – Variations on the theme
Intermountain Sepsis Bundle

To be successful, 11 components of the Sepsis Bundle should be accomplished.

Intermountain Healthcare
Screening Sepsis Model

Goal: Identify Sepsis Patients within 2 hours of admission.

Simple Sepsis Model: Designed for Early Detection
Sepsis Model: Extending Its Coverage

• Initial Model: Designed to function within the first 2 hours
• Revised Model: Designed to function any time within the first 24 hours

Complex Sepsis Model: Effective at Multiple Times?
Adding Semantics to the Model

*Saying “Findings are consistent with Diagnosis”*

---

**Pneumonia**

- Present: 0.49
- Absent: 99.5

**Sepsis**

- Present: 1.75
- Absent: 98.2

---

**CXR_Consistent**

- False: 94.5
- True: 5.47

**Orientation_Consistent**

- True: 11.3
- False: 88.7

---

**Alert_and_oriented_x_3**

- Absent: 11.3
- Present: 88.7

---

**Not_Oriented_x_3**

- Present: 0.49
- Absent: 99.5

---

**Not_Oriented_to_time**

- Absent: 94.1
- Present: 5.92

**Not_Oriented_to_situation**

- Absent: 99.9
- Present: 0.052

---

**Not_Oriented_to_place**

- Absent: 97.0
- Present: 3.04

**Not_Oriented_to_person**

- Absent: 99.6
- Present: 0.44

---

**Not_Oriented_to_S_O**

- Absent: 100
- Present: 0

---

**CXR_Ordered**

- Present: 0
- Absent: 100

---

**Multi_Lobar_Infiltrates**

- Present: 33.3
- Absent: 33.3
- Unknown: 33.3

---

**Single_Lobe_Infiltrate**

- Absent: 33.3
- Present: 33.3
- Unknown: 33.3

---

"Chest X-ray Results consistent with Pneumonia"

---

"Mental Status consistent with Sepsis"
Conclusion

“All models are wrong; some models are useful.”
Attributed to statistician George Box

• Probabilistic Models have a role in clinical care
• Enterprise Data Warehouses can contribute to Model development
• Stored Medical Knowledge (in the form of Ontologies) can accelerate Modeling
• Bayesian Models can flexibly represent a variety of clinical conditions.
Comments and Questions

Questions???