Artificial Intelligence in Medicine

Peter Szolovits
Prof. of EECS, HST
CSAIL Clinical Decision Making Group
http://medg.csail.mit.edu

6.034
April 7, 2010

Personal History

- Interest in Knowledge Representation
  - How do you write down in a computer real-world knowledge that is useful for
    - inference
    - learning
    - consistency checking
  - Special types of knowledge
    - Likelihood
    - Time
    - Space
    - Belief

What do Doctors do?

- Three classical medical tasks
  - diagnosis
  - abduction: reasoning from effects to causes
  - prognosis
  - predictive models
  - therapy
    - choose actions, in light of diagnosis and prognosis
- Need to choose diagnostic tests makes it more like therapy
- Additional contemporary tasks
  - monitoring
  - prevention
  - public health and epidemiology
  - biomedical research

The Medical Cycle

- initial presentation
- observe/measure
- plan
- interpret
- classify
- diagnosis(-es)
  - patient data
  - information
The Meta-level Cycle

How can we emulate diagnosis?

- Flowcharts
- Pattern matching
- Rules
- Probabilistic Networks

The “Learning Health Care System”

Flowchart

BI/Lincoln Labs
Clinical Protocols
1978
Simple Representation of Disease/Symptom Associations

Diagnosis by Edge-Punched Cards

- Dx is intersection of sets of diseases that *may cause* all the observed symptoms
- Difficulties:
  - Uncertainty
  - Multiple diseases
  - "Problem-Knowledge Coupler" of Weed

Diagnosis by Card Selection

Taking the Present Illness: Diagnosis by Pattern Directed Matching
Hypotheses about the Patient

The Short-Term Memory

From initial complaints, guess suitable hypothesis.
Use current active hypotheses to guide questioning
Failure to satisfy expectations is the strongest clue to a better hypothesis; differential diagnosis
Hypotheses are activated, de-activated, confirmed or rejected based on
(1) logical criteria
(2) probabilities based on:
   findings local to hypothesis
   causal relations to other hypotheses

PIPs's Theory of Diagnosis

The Long-Term Memory: Medical Knowledge of Diseases and Symptoms

Memory Structure in PIP

Triggers
Logical Criteria
Probabilistic Scoring Function
Differential Diagnosis Heuristics
Causally and Associationally Related Hyp's
Manifestations
Hypothesis
PIP's Model of Nephrotic Syndrome

NEPHROTIC SYNDROME, a clinical state

FINDINGS:
1. Low serum albumin concentration
2. Heavy proteinuria
3. >5 gm/day proteinuria
4. Massive symmetrical edema
5. Facial or peri-orbital symmetric edema
6. High serum cholesterol
7. Urine lipids present

IS-SUFFICIENT: Massive pedal edema & >5 gm/day proteinuria
MUST-NOT-HAVE: Proteinuria absent

SCORING . . .

MAY-BE- CAUSED-BY: AGN, CGN, nephrotoxic drugs, insect bite, idiopathic nephrotic syndrome, lupus, diabetes mellitus
MAY-BE-COMPLICATED-BY: hypovolemia, cellulitis

DIFFERENTIAL DIAGNOSIS:
- neck veins elevated → constrictive pericarditis
- ascites present → cirrhosis
- pulmonary emboli present → renal vein thrombosis

Case, continued

- PRESENTING PROBLEM: A MIDDLE AGED WOMAN WITH PEDAL EDEMA.
- THE CASE CAN BE SUMMARIZED AS FOLLOWS:
  - THIS IS A MIDDLE-AGED WOMAN, WHO HAS PEDAL EDEMA, WHICH IS NOT-PAINFUL, NOT-ERYTHEMATOUS, PITTING, SYMMETRICAL. 4+, WITHOUT-TEMPORAL-PATTERN, OCCASIONAL AND FOR-WEEKS. SHE DOES NOT HAVE DYSPNEA. SHE HAS HEAVY ALCOHOL CONSUMPTION. SHE HAS JAUNDICE. SHE HAS PAINFUL HEPATOMEGALY. SHE HAS SPLENOMEGALY. SHE HAS ASCITES. SHE HAS PALMAR ERYTHEMA, SHE HAS SPIDER ANGIOMATA. SHE DOES NOT HAVE PAROTID ENLARGEMENT. SHE HAS MODERATELY-ELEVATED DIRECT-AND-INDIRECT BILIRUBIN. SHE HAS PROLONGED PROTHROMBIN TIME. SHE HAS MODERATELY-ELEVATED SGPT. SHE HAS MODERATELY-ELEVATED SGOT. SHE HAS MODERATELY-ELEVATED LDH. SHE HAS NOT-RECEIVED BLOOD TRANSFUSIONS. SHE HAS NOT-EATEN CLAMS. SHE DOES NOT HAVE ANOREXIA. SHE HAS MELENA. SHE DOES NOT HAVE HEMATEMESIS. SHE HAS LOW SERUM IRON. SHE HAS ESOPHAGEAL VARICES.
- DIAGNOSES THAT HAVE BEEN ACCEPTED ARE: ALCOHOLISM AND GI BLEEDING.
- THE LEADING HYPOTHESIS IS CIRRHOSIS.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>fit of case to hypothesis</th>
<th>fraction of finding explained</th>
<th>average score</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIRRHOSIS</td>
<td>0.72</td>
<td>0.78</td>
<td>0.75</td>
</tr>
<tr>
<td>HEPATITIS PORTAL</td>
<td>0.75</td>
<td>0.30</td>
<td>0.53</td>
</tr>
<tr>
<td>HYPERTENSION</td>
<td>0.72</td>
<td>0.17</td>
<td>0.45</td>
</tr>
<tr>
<td>CONSTRICTIVE PERICARDITIS</td>
<td>0.17</td>
<td>0.13</td>
<td>0.15</td>
</tr>
</tbody>
</table>

A Case of a Middle-Aged Woman with Pedal Edema (swollen feet)

QMR Partitioning
Multi-Hypothesis Diagnosis

- Set aside complementary hypotheses
- … and manifestations predicted by them
- Solve diagnostic problem among competitors
- Eliminate confirmed hypotheses and manifestations explained by them
- Repeat as long as there are coherent problems among the remaining data
Knowledge Base:
- 956 hypotheses
- 4090 manifestations (about 75/hypothesis)
- Evocation like P(H|M)
- Frequency like P(M|H)
- Importance of each M
- Causal relations between H's

Diagnostic Strategy:
- Scoring function
- Partitioning
- Several questioning strategies

QMR Scoring
- Positive Factors
  - Evoking strength of observed Manifestations
  - Scaled Frequency of causal links from confirmed Hypotheses
- Negative Factors
  - Frequency of predicted but absent Manifestations
  - Importance of unexplained Manifestations
- Various scaling parameters (roughly exponential)

Example Case
Assume a bipartite graph representation of diseases/symptoms

Given a set of symptoms, how to proceed?

If we could “guess” an appropriate clustering of the symptoms so that each cluster has a single cause …

… then the solution is \((d_5, d_6) \times (d_3, d_7, d_8, d_9) \times (d_1, d_2, d_4)\)

Symptoms

- High urine osmolality (Osm)
- High urine specific gravity (Sg)
- Low urine sodium (Na)
- Low urine pH (pH)

**Synopsis** in Renal Disease

- **Diseases**
  - Hypertension (HTN)
  - Acute glomerulonephritis (AGN)
  - IgA nephropathy (IgA)
  - Prerenal azotemia (PRA)
  - Hepatorenal syndrome (HRS)
  - Renal vasculitis (RV)
  - Congestive heart failure (CHF)
  - Aldosteronism (Aldo)
  - Constrictive pericarditis (Peri)
  - Diabetic ketoacidosis (DKA)
  - Analgesic nephropathy (AN)
  - Hypokalemic nephropathy (HKN)
  - Chronic renal failure (CRF)
  - Renal tubular acidosis (RTA)

- **Symptoms**
  - High urine osmolality (Osm)
  - High urine specific gravity (Sg)
  - Low urine sodium (Na)
  - Low urine pH (pH)

<table>
<thead>
<tr>
<th>Osm</th>
<th>AGN</th>
<th>IgA</th>
<th>PRA</th>
<th>HRS</th>
<th>RV</th>
<th>CHF</th>
<th>Aldo</th>
<th>Peri</th>
<th>DKA</th>
<th>AN</th>
<th>HKN</th>
<th>CRF</th>
<th>RTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Symptom Clustering is Efficient

- Like in any “planning island” approach, reducing an exponential problem to several smaller exponential problems vastly improves efficiency, if it captures some insight into the problem.

- Wu's algorithm (SYNOPSIS) will keep a compact encoding even if it overgenerates slightly.

E.g., suppose that of the set of diseases represented by \((d_5, d_6) \times (d_3, d_7, d_8, d_9) \times (d_1, d_2, d_4)\), \(d_6 \times d_8 \times d_1\) is not a candidate. To represent this precisely would require enumerating the 23 valid candidates. Instead, the factored representation is kept.

In a diagnostic problem drawn from a small subset of the Internist database, it is a power of 3 faster and a power of 5 more compact than standard symptom clustering.

Guide search via probabilities, if we have a reasonable model(!)

Reasoning Using Rules

- Mycin used backward chaining (from conclusions back to facts) with a collection of <1000 rules
- Domain: bacterial infections
- E.g., RULE037
  - If the organism
    - stains grampos
    - has coccus shape
    - grows in chains
  - Then there is suggestive evidence (0.7) that it is streptococcus.

How Mycin Works

- To find out a fact
  - If there are rules that can conclude it, try them
  - Ask the user
- To “run” a rule
  - Try to find out if the facts in the premises are true
  - If they all are, then assert the conclusion(s), with a suitable certainty
- Backward chaining from goal to given facts

\[ \text{Naive Bayes or Bayesian Classifier} \]

\[
\begin{align*}
H \mid M_1, \ldots, M_5 & = \frac{P(H \mid M_1, \ldots, M_5)}{P(H)}, \\
L(M_i \mid H) & = \frac{P(H \mid M_i)}{P(H \mid \overline{M}_i)} \\
O(H \mid M_1, \ldots, M_5) & = O(H) \times L(M_1 \mid H) \times \cdots \times L(M_5 \mid H) \\
\log O(H \mid M_1, \ldots, M_5) & = \log O(H) + \log L(M_1 \mid H) + \cdots + \log L(M_5 \mid H) \\
W(H \mid M_1, \ldots, M_5) & = W(H) + W(M_1 \mid H) + \cdots + W(M_5 \mid H)
\end{align*}
\]

\[ \text{Single expression of any piece of knowledge} \]

\[ \text{Information used everywhere appropriate} \]

\[ \text{Dynamically traces out behavior of (what might be) a flowchart} \]
**Bipartite Graph with Probabilities**

- Bayesian Network
- Computation cost exponential in number of “loops”

| $H_1$ | $H_2$ | $H_3$ | $P(M_1|H_1,H_2,H_3)$ | $P(\neg M_1|H_1,H_2,H_3)$ |
|-------|-------|-------|---------------------|---------------------|
| $F$   | $F$   | $F$   | 0.001               | 0.999               |
| $F$   | $F$   | $T$   | 0.01                | 0.99                |
| $F$   | $T$   | $F$   | 0.007               | 0.993               |
| $F$   | $T$   | $T$   | 0.2                 | 0.8                 |
| $T$   | $F$   | $F$   | 0.1                 | 0.9                 |
| $T$   | $F$   | $T$   | 0.15                | 0.85                |
| $T$   | $T$   | $F$   | 0.2                 | 0.8                 |
| $T$   | $T$   | $T$   | 0.7                 | 0.3                 |

**A Realistic Bipartite Graph**

- Symptoms of prerenal azotemia, and their alternative causes (Tom Wu PhD 1992)

**What to do with Tons of Data?**

- Partners Healthcare has ~4M records of previously-treated patients
- Mayo Clinic has ~60M discharge summaries
- We have 30K detailed records on patients in the ICU (Intensive Care Unit)
- Surely, these must be useful for something!
- $State_i(patient) = f_i(\text{features}_{patient})$
  - We can learn $f_i$ from data
  - Use it to predict $State_i$ for future patients
- What are useful $State_i$?
  - Death
  - Specific diseases
  - Effectiveness of particular therapies
  - Optimal timing of various interventions
  - ...

**Using MIMIC data to build predictive models**

- Mortality
- Comparison to SAPS II
- Stationary Daily Acuity Score
- Daily Acuity Scores (one for each day $n$ of ICU stay)
- Real-time Acuity Scores
- Secondary Outcomes
  - Weaning from Vasopressors
  - Weaning from Intraortic Balloon Pump
  - Onset of Septic Shock
  - Acute Kidney Injury
  - Weaning from Mechanical Ventilator
  - Tracheotomy Insertion
  - First response to Vasopressor Reduction

Caleb Hug, PhD 2009
http://dspace.mit.edu/handle/1721.1/46690
Cleaning the data—half the research time

- Missing values
- Some values are not measured for some clinical situations
- Failures in data capture process
- Episodically measured variables
- Unclear/undefined clinical states
- Imprecise timing of meds, ...
- Partially measured i/o
- Proxies: e.g., which ICU=what disease
- Derived variables: integrals, slopes, ranges, frequencies, etc.
- Transformed variables: square root, log, etc.
- Select subset of data with enough data!

SAPS II

Table 4.1: SAPS II Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Max Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>15</td>
</tr>
<tr>
<td>Heart rate</td>
<td>11</td>
</tr>
<tr>
<td>Systolic BP</td>
<td>13</td>
</tr>
<tr>
<td>Body temperature</td>
<td>3</td>
</tr>
<tr>
<td>PaO2/FIO2 (if ventilated or continuous)</td>
<td>11 positive airway pressure</td>
</tr>
<tr>
<td>Urinary output</td>
<td>11</td>
</tr>
<tr>
<td>Serum urea nitrogen level</td>
<td>10</td>
</tr>
<tr>
<td>WBC count</td>
<td>12</td>
</tr>
<tr>
<td>Serum potassium</td>
<td>3</td>
</tr>
<tr>
<td>Serum sodium level</td>
<td>5</td>
</tr>
<tr>
<td>Serum bicarbonate level</td>
<td>6</td>
</tr>
<tr>
<td>Bilirubin level</td>
<td>9</td>
</tr>
<tr>
<td>Glasgow Coma Score*</td>
<td>26</td>
</tr>
<tr>
<td>Chronic diseases</td>
<td>17</td>
</tr>
<tr>
<td>Type of admission</td>
<td>8</td>
</tr>
</tbody>
</table>

*If the patient is sedated, the estimated GCS prior to sedation

Outcomes

Many univariate analyses

Table 3.15: Preprocessed Data

<table>
<thead>
<tr>
<th>Number of Patients</th>
<th>Number of Rows</th>
<th>Number of Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,066</td>
<td>1,044,982</td>
<td>438</td>
</tr>
</tbody>
</table>
Evaluating the models

Selected features for each day of ICU stay

Evaluation of Survival Models

“There’s Gold in Them Thar Hills”

- Federal investment of ~$45B in healthcare IT
- Strong pressures to make healthcare more efficient