Bill's 1970 Predictions
(emphasis added)

- "the computer as an intellectual tool can reshape the present system of health care, fundamentally alter the role of the physician, and profoundly change the nature of medical manpower recruitment and medical education — in short, the possibility that the health-care system by the year 2000 will be basically different from what it is today."
- "familiar projections envision the computer performing a wide variety of functions such as the scheduling of hospital admissions, the keeping of medical records and the operation of laboratory and pharmacy ... in the area of "housekeeping" activities."
- "the increasing shortage of physician manpower and ... geographic maldistribution."
- "difficult challenge of maintaining a high level of physician competence in the face of a continued expansion of medical knowledge that tends to widen progressively the gap between what a doctor should know and what he can retain and utilize."
- "exploitation of the computer as an "intellectual," "deductive" instrument — a consultant that is built into the very structure of the medical-care system and that augments or replaces many traditional activities of the physician."

- "in the not too distant future the physician and the computer will engage in frequent dialogue, the computer continuously taking note of history, physical findings, laboratory data, and the like, alerting the physician to the most probable diagnoses and suggesting the appropriate, safest course of action."
- "help free the physician to concentrate on the tasks that are uniquely human such as the application of bedside skills, the management of the emotional aspects of disease, and the exercise of good judgment in the nonquantifiable areas of clinical care."
- "Computer-supported "health-care specialists," aided by a variety of automated devices for history taking, blood analysis and other procedures, and trained to perform a careful physical examination, might take over a large segment of the responsibility for the delivery of primary medical care."
- "it is conceivable that the computer could also take over a variety of specialized functions that are now performed by highly skilled physicians. It is entirely possible, for example, that the administration of anesthesia — a function now uniquely human — could be largely or fully automated if new monitoring technics were combined with the capacity of the computer instantaneously to analyze and respond to large volumes of physiologic data."
“Artificial Intelligence in Medicine”

- Simulate the behavior of human expert clinicians
  - Flowcharts
  - Rules
  - Pattern Matching
  - NOT machine learning

Flowchart

Bi/Lincoln Labs
Clinical Protocols
1978

Mycin

- Task: Diagnosis and prescription for bacterial infections of the blood (and later meningitis)
- Method:
  - Collection of modular rules
  - Certainty factors
  - Backward chaining from goal to given facts
    - 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

RULE037

IF the organism
1) stains grampos
2) has coccus shape
3) grows in chains

THEN There is suggestive evidence (.7) that the identity of the organism is streptococcus.

Matching as Basis for Reasoning

The “Present Illness” Program

- From initial complaints, guess suitable hypothesis.
- Use current active hypotheses to guide questioning
- 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
- Failure to satisfy expectations is the strongest clue to a better hypothesis; differential diagnosis

Active Pattern Matching in PIP

Inexorable Growth of Healthcare Spending

Life Expectancy

<table>
<thead>
<tr>
<th>Country</th>
<th>Life Expectancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>South Africa</td>
<td>62.34</td>
</tr>
<tr>
<td>Brazil</td>
<td>73.53</td>
</tr>
<tr>
<td>Russia</td>
<td>70.47</td>
</tr>
<tr>
<td>Albania</td>
<td>78.13</td>
</tr>
<tr>
<td>USA</td>
<td>79.68</td>
</tr>
<tr>
<td>Taiwan</td>
<td>79.98</td>
</tr>
<tr>
<td>UK</td>
<td>80.54</td>
</tr>
<tr>
<td>Germany</td>
<td>80.57</td>
</tr>
<tr>
<td>France</td>
<td>81.75</td>
</tr>
<tr>
<td>Sweden</td>
<td>81.98</td>
</tr>
<tr>
<td>Japan</td>
<td>84.74</td>
</tr>
</tbody>
</table>
Responses

- Health Maintenance Organizations — risk sharing with providers
- Pay for performance/outcomes, not effort
- Universal health insurance
- Doc-in-a-box
- Electronic medical records
- Personal health information systems
- Improved public health, immunizations, sanitation, ...
- See Clay Christensen’s analysis

- Science and engineering
  - Biological/genetic understanding of disease
  - Instrumentation
  - “Big Data” analysis

“The Learning Health Care System”

- “one in which progress in science, informatics, and care culture align to generate new knowledge as an ongoing, natural by-product of the care experience, and seamlessly refine and deliver best practices for continuous improvement in health and health care” — IOM
- Needs not currently met:
  - Comprehensive collation of all clinical, social, demographic, behavioral, ... data that are now captured in the health care system
  - Routine capture of novel data sources:
    - genomes, gene expression, etc.
    - environmental factors (e.g., metagenomics)
    - physiological response to life situations
      - (related to fitness and wellness)
  - Technical infrastructure
    - Storage and analysis of truly “big data”
    - Incentives and demonstrations of utility

What is Evidence-Based Medicine?
- RCTs, Meta-Analysis, Systematic Reviews

- Randomized Controlled Clinical Trials
  - E.g., is drug A more effective than drug B for condition X?
  - Narrow selection of patient cases and controls
  - Careful collection of systematically organized data
  - Statistical analysis of outcomes
  - => Statistically significant conclusions
- But:
  - Heterogeneity: Most cases to which RCT results are applied do not fit trial criteria
  - Short Follow-Up: Trials run for limited times, but use is longer
  - Small Samples: Some effects are rare but devastating
- Instead: consider every patient’s experience as a source of knowledge by which to improve health care

What if we had all these data?

⇒ Decision Support

- For patients
  - Manage ongoing care of chronic conditions, titrate medications, change behaviors, notice actionable conditions, ...
- For providers
  - Double-check data interpretation, improve diagnostic acumen, optimize therapeutic decisions, predict outcomes, help to follow up, ...
- For institutions and public health
  - Focus on critical needs, rationalize priorities, organize public education, ...
- For researchers
  - Provide insights into the science and practice of medicine
- Key: ability to predict future events in individuals or population
  - “It’s tough to make predictions, especially about the future” — Yogi Berra
  - How?
Deep Pathophysiological Models


What is the Space of Models to Learn?

- Classification vs. Regression
  - Classification chooses one of a discrete set of answers, or a probability distribution over such a set
    - e.g., diagnosis
  - Regression predicts some dependent variable, typically continuous
    - e.g., predict a lab value, time to some event
- Probabilistic inference vs. decision analysis
  - i.e., are decisions formally modeled?
- "Hidden" states vs. all explicit
  - Hidden states: HMM, BN, MDP, etc.
  - Explicit: autoregressive models, covariance, interpolation, logistic regression, etc.

Bayesian Network Models: Alarm


Bayesian Network Models: Pathfinder


100+ nodes, but structure is nearly Naive Bayes!
**Causal Discovery**

**Causality**

- Suppes, 1950's
- Statistical association
- Temporal succession
- No confounders (!)
- Hidden variables

- A node, X, is conditionally independent of all other nodes in the network given its Markov blanket: its parents, U, children, Y, and children's parents, Z.
- Focus on effect size

**Pearl: Drawing Causal Conclusions from Observational Data**

- In an ordinary experiment, we control something (X) to observe its effects
- Can we approximate such control by relying on observations of “natural experiments”?
- Possible problem: What if X is influenced by (something that is influenced by …) one of the effects we measure?
  - E.g., A doctor chooses a treatment based on some lab test, but we are interested in the effect of that treatment on the test.
- Pearl's theorem: Causal interpretation of observational data requires making independence assumptions.

**Markovian Models**

- Limited “history” encapsulates all past influences
- Decisions based on maximizing (discounted) expected value of future events
  - Partially Observable Markov Decision Processes (POMDP)
  - For n possible states, complexity \( \sim O(2^n) \)
  - For k (binary) variables, \( n=O(2^k) \), thus overall \( 2^{2k} \)
    - \( k=2, 16 \)
    - \( k=3, 256 \)
    - \( k=4, 65,536 \)
    - \( k=5, 4,294,967,296 \)
    - \( \ldots \)
    - \( k=160, 2^{10^{109}} \) (number of protons in universe \( \sim 136 \times 2^{256} \), or about \( 1.57 \times 10^{79} \))
- Need for many simplifications, approximations
Google’s Lessons

- Much of human knowledge is not like physics!
- "... invariably, simple models and a lot of data trump more elaborate models based on less data"
- "... simple n-gram models or linear classifiers based on millions of specific features perform better than elaborate models that try to discover general rules"
- "... all the experimental evidence from the last decade suggests that throwing away rare events is almost always a bad idea, because much Web data consists of individually rare but collectively frequent events"


Simple Predictive Models

- Naive Bayes
- (Logistic) Regression
- Support Vector Machine
- Classification Trees (or CART)
- Random Forest
- Artificial Neural Nets (today “deep learning”)
- Ensemble methods
- …

Inferential Models:
Naïve Bayes

- \( y \) (unobserved) is the diagnosis
- \( x \) (observed) are the symptoms

Bipartite Graph Models

- \( y \) are diseases, unobserved (no longer exhaustive & mutually exclusive)
- \( x \) are symptoms
  - each \( x \) depends on all diseases, hence \( 2^m \) conditional probabilities
  - further assumptions reduce this complexity (e.g. "noisy or")
Markov Model

- Transition model among time sequences
  - Observed: y’s are observable
  - Hidden: y’s are unobservable

(Deep) Neural Networks

- Every node gets a logistic regression function of its inputs
- Number of nodes in each layer may vary
- Number of layers is another hyper parameter
- Training by back propagation
  - change weights in proportion to error signal
- 1-layer network used in word2vec

Influence Diagram

- Models not only hidden and observable variables but also:
  - Tests and interventions
  - Utilities of various states
  - Compact representation of complex decision tree
- Issues
  - Complexity of fitting, inference
  - Discounting utilities
  - For recurring problems, policy vs. optimal choice (e.g., chronic treatment)

Logistic Regression

- Simple, fast, unsophisticated, but often works well
- Given a number of cases \(i=1, \ldots, k\)
- For each case, we have an outcome \(y\) and a vector of features \(x = \{x_1, \ldots, x_n\}\)
  \[\hat{y} = \text{logit}^{-1}(\beta_0 + \sum \beta_i x_i) = \frac{1}{1 + \exp(- (\beta_0 + \sum \beta_i x_i))}\]
- Estimate \(\beta\)s by least squares fit
  - Minimize \(\sum_{i=1, \ldots, k} (y - \hat{y})^2 + \sum_{i=0, \ldots, n} ||\beta||\)
  - 2nd (regularization) term penalizes model complexity
    - L1 norm minimizes number of non-0 \(\beta\)s
      - LASSO
    - L2 norm minimizes prediction error
      - Ridge regression
Use All Possible Data

Model Prediction = f(Inputs)

Heterogeneous Sources of Clinical Data

- **Tabular**, standardized data
  - Billing codes: encounter, most important conditions, severity, interventions, …
  - Demographics
  - Laboratory measurements
  - Medication prescriptions
  - Discrete measurements from monitors, wearable instruments, …
- Continuously recorded **signals**
  - ECG/ECG monitors, wearables, …
- **Narrative** reports
  - Doctors’ and nurses’ notes
  - Specialist reports: pathology, radiology, …
  - Patient self-observations, blogs, email exchanges, …
- **Questionnaires**: smoking, drugs, exercise habits, travel, …
- **Imaging**: x-rays, CT scans, MRI, PET, ultrasound, …
- **Environmental** conditions: pollution, epidemics, …
- **Genetics**: gene expression, SNP, CNV, exome, genome, epigenetics, metagenetics, …

Using MIMIC Data to Build Predictive Models

- Mortality
  - Comparison to SAPS II
  - Daily Acuity Scores
  - Real-time Acuity Scores (real-time risk assessment)
- Other clinical events
  - Pressor weaning, intra-aortic balloon pump weaning
  - Onset of septic shock, acute kidney injury
- Data set (MIMIC 2, earlier snapshot, today ~5-6x)
  - 10,066 patients: 7,048 development, 3,018 validation
  - Selected cases with adequate data
  - Excluded neurological and trauma cases
- **only tabular data**
- Derived variables can summarize essential contributions of dynamic variation
  - Integrals, slopes, ranges, frequencies, etc.
  - Transformed variables: inverse, abs, square, square root, log-abs, abs deviation from mean, log abs deviation, …

Summary of Mortality Models

![Graph 1](http://dspace.mit.edu/handle/1721.1/46690)

**Figure 5.25**: AUC versus day, first 5 ICU days (validation data)

**Figure 5.27**: AUC versus day, patients with ICU stays ≥ 5 days (validation data)
Models for Therapeutic Opportunities and Risks

<table>
<thead>
<tr>
<th>Prediction</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weaning from vasopressors within next 12 hours, remain off for 4 hours</td>
<td>0.809</td>
</tr>
<tr>
<td>Pressor weaning + Survival</td>
<td>0.825</td>
</tr>
<tr>
<td>Weaning from Intra-Aortic Balloon Pump</td>
<td>0.816</td>
</tr>
<tr>
<td>Onset of Septic Shock</td>
<td>0.843</td>
</tr>
<tr>
<td>Acute kidney injury</td>
<td>0.742</td>
</tr>
</tbody>
</table>

• Predictions are not as accurate as mortality models, but still impressive
• Using acuity score instead of such specific models is worse
  • E.g., Vasopressor weaning — 0.679 vs. 0.809

Where do Features Come From?

“Natural Features”

• Coded data
  • lab values
  • demographics (age, gender, ethnicity, socio-economic status, marital status)
  • medications
  • symptoms
  • diagnoses
  • procedures
  • insurance status
  • billing codes
  • death records, hospitalizations, ED visits

• Information extracted from narrative notes
  • <all of the above>
  • physical exam, presenting complaint, vital signs, nursing observations, specialists’ reports, …
  • richer structure: drug given for condition, side-effect of treatment, …

Where do Features Come From?

Abstractions

• Temporal trends
  • average, range, standard deviation, mean-crossings, linear trends, each over various lengths of time
• Knowledge-based
  • data that pertain to specific organs, disease processes
• Clustering-based
  • commonly co-occurring values & trends, patient groupings, linguistic structures
    • unsupervised or supervised
  • bi- or tri-clustering, matrix or tensor factorization
    • mutually constrain clustering among different dimensions
• Topic models
  • bag-of-words
  • bag of more primitive features

Clustering Snapshots

http://groups.csail.mit.edu/medg/ftp/kshetri/Kshetri_MEng.pdf

• ~10,000 patients x ~12,000 possible features/patient = 1M rows (sparse), 30 min snapshots ⇒ 11 clusters

<table>
<thead>
<tr>
<th># in cluster</th>
<th>Survival</th>
<th>GCS</th>
<th>Heart Rate Events</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Prediction using multi-layer abstraction

Plane of Disease Clusters

Plane of Pathophysiological Clusters (Foci)

Two-Stage Abstractions

1. Normalize data
   a. For each focus
      i. Compute magnitude of abnormality of each related measurement
         = abstraction of overall degree of abnormality of this focus
      ii. Cluster the set of directions of abnormality (+, −, 0) of each
          = abstraction of pattern or abnormality of this focus
   b. For the patient’s overall state
      i. Compute aggregate degree of abnormality
         = how severely ill is this patient
      ii. Cluster sets of magnitudes and directions of abnormality of the foci
         = nature of abnormality of this patient

Foci and Features

- Selected ~10,000 patients in MIMIC II database (excluded pediatrics, trauma)
- Over 12 domain foci; many clinical variables per focus
- Over 1 million chart events (after binning into 1 hour windows)

<table>
<thead>
<tr>
<th>Focus</th>
<th>Features in each Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kidney</td>
<td>Creatinine, BUN, BUN2Cr, UrineOut/Hr/Kg, ...</td>
</tr>
<tr>
<td>Liver</td>
<td>AST, Alt, TBili, Dbili, Albumin, tProtein</td>
</tr>
<tr>
<td>Cardio</td>
<td>MAP, HR, CVP, BPSys, BPDis, Cardiac Index, ...</td>
</tr>
<tr>
<td>Respiration</td>
<td>RR, SpO2, FiO2Set, PEEPSet, TidVolSet, SaO2, PIP, MinVent, ...</td>
</tr>
<tr>
<td>Hematology</td>
<td>Hematocrit, Hgb, Platelets, INR, WBC,RBC, PT</td>
</tr>
<tr>
<td>Electrolytes</td>
<td>Na, Mg, K, Ca, Glucose, ...</td>
</tr>
<tr>
<td>Acid-base</td>
<td>Art CO2, Art PaCO2, Art pH, Art BE</td>
</tr>
<tr>
<td>General</td>
<td>GCS, Age, temp</td>
</tr>
<tr>
<td>Medication</td>
<td>Diuretic, Antiarrhythmic, Antiplatelet, Sympathomimetic, ...</td>
</tr>
<tr>
<td>Chronic</td>
<td>AIDS, HematMalig, Metacarcinoma</td>
</tr>
<tr>
<td>Electrocardio (EKG)</td>
<td>PVC, Rhythm types, Ectopic frequency, ...</td>
</tr>
</tbody>
</table>
Disease Severity using “Radial Domain Folding”


Mortality Prediction on the Test Data

Our model: AUC of 0.9
SAPS-II: AUC of 0.81;

Our method outperforms customized SAPS-II score

Patient-State Transition: Effects of Past Transitions of Other Foci

\[ \ln(\text{organ\_transition}) \sim \text{current\_state(organ)} + \text{current\_state(otherorgans)} + \text{latest\_transition\_trend(otherorgans)} + \text{duration\_in\_current\_state(organ)} + \text{duration\_since\_last\_change(otherorgans)} \]

Forecasting the transitions of organ systems

Cardiovascular

<table>
<thead>
<tr>
<th>True States</th>
<th>Predicted States</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.017</td>
<td>0.025 0.013 0.032 0.012</td>
</tr>
<tr>
<td>0.009</td>
<td>0.073 0.106 0.16 0.053</td>
</tr>
<tr>
<td>0</td>
<td>0.012 0.161 0.014 0.012</td>
</tr>
<tr>
<td>0.004</td>
<td>0.046 0.127 0.515 0.31</td>
</tr>
<tr>
<td>0</td>
<td>0.018 0.09 0.216 0.676</td>
</tr>
</tbody>
</table>

Respiratory

<table>
<thead>
<tr>
<th>True States</th>
<th>Predicted States</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.103</td>
<td>0.13 0.104 0.194 0.089</td>
</tr>
<tr>
<td>0.011</td>
<td>0.707 0.119 0.127 0.036</td>
</tr>
<tr>
<td>0.003</td>
<td>0.098 0.953 0.012 0.014</td>
</tr>
<tr>
<td>0.003</td>
<td>0.032 0.044 0.719 0.203</td>
</tr>
<tr>
<td>0.003</td>
<td>0.056 0.296 0.318 0.368</td>
</tr>
</tbody>
</table>
What Should We Predict?

- Regularities in the world ⇒ we can predict anything
  - if it occurs often enough
  - at some level of accuracy
- Are these predictions useful?
  - Mortality
  - Deterioration
  - Therapeutic opportunity
  - Activities of daily living
  - Re-hospitalization
- We won’t know until we put them into practice
  - Requires changes in workflow, possible regulatory arrangements, overcoming skepticism
  - Need joint planning with providers, vendors, patient groups, regulators, insurers

Thanks

- Funding from
  - NIH-National Library of Medicine (i2b2)
  - NIH-National Institute of Biomedical Imaging and Bioengineering (MIMIC & ICU Informatics)
  - NIH-National Institute of Human Genomics Research (PIC-SURE BD2K)
  - ONC SHARP 4 project (Secondary Use of Clinical Data)
  - Siemens Research
  - Philips Research
  - Wistron Corp.
  - Quanta Corp.
  - MIT-MGH Partnership
  - Simons Center for the Social Brain
- HMS collaborators:
  - Tianxi Cai, Susanne Churchill, Vivian Gainer, Sergey Goryachev, Elizabeth Karlson, Isaac Kohane, Fina Kurreema, Katherine Luo, Shawn Murphy, Robert Plenge, Soumya Raychaudhuri, Qing Zeng-Treitler, Stanley Shaw, …
- NLP collaborators
  - Özlem Uzuner, Bill Long, Ana Rumshisky, Guergana Savova, Marzyeh Ghassemi, Yuan Luo, Andreea Bodnari, Tawanda Sibanda, Rachel Chasin, …
- My CSAIL research group
  - http://meds.csail.mit.edu