An Overview of Biomedical Informatics

Ira J. Kalet, Ph.D. Professor Emeritus, Radiation Oncology joint with Biomedical Informatics and Medical Education

University of Washington Seattle, Washington USA

July, 2014

Outline

The transformation of biology and medicine

An Introduction to Logic

Logic and public health Biochemical logic

Ontologically Computing the Spread of Tumor Cells The UW Foundational Model of Anatomy (FMA) A Model of Tumor Spread

The Rest of Biomedical Informatics

Biology and Medicine: a new era

- Biology is experiencing an explosion of data and chunks of knowledge.
- Medical practice is rapidly generating huge amounts of clinical data, laboratory data, images, and more.
- Three approaches to this are possible:
 - 1. organizing the data with tags and structure to make it widely accessible and enable integration (the Semantic Web),
 - 2. searching the data for patterns (data mining, knowledge discovery), and
 - 3. constructing computation based theories from what is known, to predict what is yet to be observed.

What is Biomedical Informatics?

concerned with

- representation
- retrieval
- manipulation

of biomedical data, information and knowledge in context.

- uses computer systems and software to implement and test ideas
- studies such uses, problems and solutions

From the National Library of Medicine Long Range Plan:

"Medical informatics attempts to provide the theoretical and scientific basis for the application of computer and automated information systems to biomedicine and health affairs."

From Milt Corn, Deputy Director for Research and Education, NLM

"We are debating where and how to draw the line between biology as a handy domain for the informatics problem (good), and biology projects for which some computation is needed (bad)."

Biomedical Informatics

Core Knowledge Areas:

- biomedical data and knowledge representation,
- biomedical information access,
- biomedical decision making, and
- information and technology use in biomedical contexts.

Domain Areas:

- Biology (e.g., molecular biology, biological structure, physiology)
- Patient care (e.g., clinical medicine, nursing, pharmacy, dentistry, patient centered health care)
- Public Health (e.g., health care organization and delivery, surveillance and epidemiology)

Distinctions and contrasts

- Engineering (building artifacts) vs. science (discovering some new knowledge)
- Biomedical Informatics vs. Biomedicine
- Theory vs. Experiment
- All are important!

Core technical knowledge

- Computer programming
- Mathematical logic
- Probability and statistics
- Information systems
- Networks computers (Internet) and abstract

Logics

Logics are formal languages for representing information and knowledge such that conclusions can be drawn.

- syntax defines the structure (sentence vs. non-sentence)
- semantics defines the "meaning" of a sentence, i.e., what is a true statement or a false statement

A logic consists of:

- A set of primitives
- Ways of building formulas (connectives, quantifiers, punctuation)
- Inference rules (how to derive new formulas)

A public health example

- If particulate material is being moved and it is dry it will become airborne.
- If particles are toxic and airborne, then there will be an increase of incidence of asthma in the vicinity.
- If there is an increase of asthma incidence, there will be an increase of clinic visits in the area.

How can this be formalized? We abbreviate the things being described using symbols and use logic connectors to represent (causal) relations. Thus:

- moved $\wedge dry \rightarrow airborne$
- ▶ airborne \land toxic \rightarrow more-asthma
- ▶ more-asthma \rightarrow more-clinic-visits

Inference rules

Given a set of formulas (the *axioms*), a new formula can be inferred by applying an *inference rule*, to derive a new formula. Here are some inference rules:

"and" elimination:

From a conjunction, $\alpha \wedge \beta$, we can infer α and β separately.

"and" introduction:

Given two sentences, α and β , we can infer the conjunction, $\alpha \wedge \beta$.

double negative elimination:

The double negative, $\neg \neg \alpha$ is equivalent to the formula α .

Modus ponens (ancient and very important): From the two formulas, α and α → β, we can infer β

Theorems and Proof

- A proof is a sequence of formulas, where the next formula is derived from previous ones using the inference rules.
- A theorem is a formula that can be derived (proved) from a set of axioms.
- A *tautology* is a formula that is true regardless of the interpretation. The formula p → (p ∨ q) is an example of a tautology, but the formula p → q is not a tautology. (A tautology is sometimes called a *valid* formula.)

The public health example revisited

The knowledge:

The facts for today:

D: moved E: dry F: toxic

What is entailed by these sentences? Can the entailments be proved?

Translating logic into code

The goal is to be able to write programs that automatically do the proofs, so we need a way to express propositional formulas in a form that a computer program can process. We use *prefix* notation, writing the logical connectives as *operators* and the terms as symbols. Here is how to represent the \rightarrow connective:

(<- fact expr)

This means if expr can be proved or is asserted, then fact is true also. The expr can use the connectives, and, or and not. An example would be

(<- airborne (and moved dry))</pre>

Public health knowledge in code

Here are the rules

- > (<- more-asthma (and airborne toxic))
- > (<- more-clinic-visits more-asthma)

and here are the assertions for the day of the incident.

```
> (<- moved)
> (<- dry)
> (<- toxic)
```

We need a function, prove, that takes an assertion and determines if it is a theorem. So, in our example,

```
(prove 'more-clinic-visits)
```

Will succeed.

The prove function in action

```
> (prove 'more-clinic-visits)
 0[2]: (PROVE MORE-CLINIC-VISITS)
   1[2]: (PROVE MORE-ASTHMA)
     2[2]: (PROVE (AND AIRBORNE TOXIC))
       3[2]: (PROVE TOXIC)
       3[2]: returned T
       3[2]: (PROVE AIRBORNE)
         4[2]: (PROVE (AND MOVED DRY))
           5[2]: (PROVE DRY)
           5[2]: returned T
           5[2]: (PROVE MOVED)
           5[2]: returned T
         4[2]: returned T
       3[2]: returned T
     2[2]: returned T
   1[2]: returned T
0[2]: returned T
т
```

The Fundamental Dogma of Cell Biology



Biochemical pathways



Biochemical reachability logic

Metabolic reactions can be represented as logic formulas:

	interestates a
(<- f1p fru)	ment.
(<- glh f1p)	
(<- dap f1p)	(<- atp)
(<- g3p glh)	(<- adp)
(<- g6p (and glu atp))	(<- gdp)
(<- f6p g6p)	(<- gtp)
(<- fbp (and f6p atp))	(<- fru)
	(<- glu)
(<- gtp (and sca gdp P))	(<- NAD+)
(<- FADH2 (and suc FAD))	(<- FAD)
(<- NADH (and mal NAD+))	(<- P)
$(\leq H + (and mal NAD +))$	

A series of assertions define which molecules are in the initial environment.

Can we produce pyruvate (pyr) from the above system? It is a simple call to the same prove function.

```
> (prove 'pyr)
T
```

What is an ontology?

- Definition: a compact description of reality (more at Wikipedia page "Ontology"), or alternatively, "A specification of a conceptualization" (Tom Gruber)
- Usually refers to a set of descriptions of entities and their relationships, a class hierarchy.
- Many formalisms: First order logic, frames, description logics.
- May be only terminology systems, or may include axioms.
- In general, a set of differential equations such as Maxwell's Equations can be considered an ontology, but the term is never used in this way.

Frame representations

A frame system generalizes the idea of a class (e.g. in object-oriented programming). A frame can describe a class or an individual. Here is an example of an individual:

```
(heart-of-joe
 (instance-of heart)
 (belongs-to joe)
 (beat-rate 72)
 (volume 34)
 (systolic-pressure 125)
 ...)
```

Frame representations, continued

And here is an example of a class meta-object:

```
(heart
  (own-slots
    (part-of cardiovascular-system)
    (contained-in middle-mediastinal-space)
    (has-parts (right-atrium left-atrium
                right-ventricle left-ventricle
                mitral-valve tricuspid-valve ...))
    (arterial-supply (right-coronary-artery
                      left-coronary-artery))
    (nerve-supply ...))
  (template-slots (belongs-to beat-rate volume
                   systolic-pressure ...))
  ...)
```

Defining the GTV, CTV and PTV



Involved lymph nodes

Predicting lymphatic dissemination of tumor

Case description: Adenoid cystic carcinoma in the nasal cavity, behind the nasal vestibule, between the left upper turbinate and the nasal bone.

Problem: Which lymph nodes are likely to have tumor? Where are they?

Possible solution: Can the Foundational Model of Anatomy (FMA) knowledge resource tell us which are the primary lymph nodes and paths in this location, and then can we predict how far the cells will go? Lymphatic chains and nodes in the head and neck



The theory is tested by comparing its predictions with surgical data.

- Starting from tumor location, compute lymphatic drainage paths from the FMA.
- ► For each path, compute probabilities of downstream travel.
- For each lymph node group, compute the aggregate probability of metastasis according to the distance.

Conceptual Overview of the FMA

The FMA describes anatomical (structural) entities and relationships from subcelluar (macromolecular) structure to body parts and organ systems. Its components are:

- Anatomical taxonomy (At): classification hierarchy, with shared characteristics and distinctive ones
- Anatomical Structural Abstraction (ASA): part-whole and spatial relationships
- Anatomical Transformation Abstraction (ATA): morphological transformation during prenatal development and postnatal life cycle
- Metaknowledge (Mk): principles, rules and definitions

The code for the metastatic path model

```
(defun find-all-paths (start)
  (labels ((successors (path)
             (get-children (first path) "efferent to")))
    (path-search start
                 #'(lambda (current) ;; stop when no more
                     (null (funcall successors current)))
                 nil ;; get all the paths
                 #'successors #'cons #'append)))
(defun lymphatic-paths (site)
  (apply #'append
         (mapcar #'(lambda (start)
                       (find-all-paths start))
                 (get-children site "lymphatic drainage"))))
```

Paths for the Soft Palate

```
> (fms-connect))
> (lymphatic-paths "Soft palate")
(("Jugular lymphatic trunk"
  "Inferior deep lateral cervical lymphatic chain"
  "Superior deep lateral cervical lymphatic chain")
. . .
 ("Right lymphatic duct" "Right jugular lymphatic trunk"
  "Right inferior deep lateral cervical lymphatic chain"
  "Right superior deep lateral cervical lymphatic chain"
  "Right retropharyngeal lymphatic chain")
. . .
 ("Thoracic duct" "Left jugular lymphatic trunk"
  "Left inferior deep lateral cervical lymphatic chain"
  "Left superior deep lateral cervical lymphatic chain"
  "Left retropharyngeal lymphatic chain")
 ("Thoracic duct" "Left jugular lymphatic trunk"
  "Left superior deep lateral cervical lymphatic chain"
  "Left retropharyngeal lymphatic chain"))
```

A Markov Model for Tumor Growth and Metastasis



The mathematical formulas for the model

The state is an array of probabilities, indexed by the T-stage (s) and the farthest involved lymphatic chain (c). Here, g represents the probability of primary tumor growth by one T-stage, and m represents the probability of metastasis one step further down the lymphatic path. The indices s1, c1 are the initial state specifiers and s2, c2 are the final state specifiers.

$$\begin{aligned} \mathcal{M}_{s2,c2,s1,c1} &= m\delta_{c1+1,c2}\delta_{s1,s2} + mg\delta_{c1+1,c2}\delta_{s1+1,s2} \\ &+ g\delta_{c1,c2}\delta_{s1+1,s2} + (1-m-g+mg)\delta_{c1,c2}\delta_{s1,s2} \end{aligned}$$

To compute the probability for each successive step, multiply the current state vector by the transition matrix to get the new state vector.

$$P_{s,c}(t+1) = \sum_{i=0,j=0}^{i=s,j=c} M_{s,c,i,j} P_{i,j}(t)$$

A few other areas that are also part of BHI

This only touches on some of the foundations with a few examples. The field is very broad and includes many other areas of activity and methodologies such as:

- Natural language processing of biomedical text,
- Discovery of relationships in large data sets (data mining, machine learning),
- Invention of new ideas and methods for use of computers and information organization and access to improve health,
- Social and behavioral aspects of information seeking and use,
- Security and privacy of medical information, and
- Medical device safety.

Publications for more information

- Ira Kalet, "Principles of Biomedical Informatics", second edition, Academic Press (Elsevier), 2013.
- Edward H. Shortliffe and James J. Cimino, "Biomedical Informatics: Computer Applications in Health Care and Biomedicine", fourth edition, Springer, 2014.
- George Starkschall and R. Alfredo Siochi, "Informatics in Radiation Oncology (Imaging in Medical Diagnosis and Therapy)", CRC Press, 2013.

Thanks

- to many students and colleagues,
- to the National Library of Medicine for grant support for the FMA, Biomedical Informatics Research Training and a Publication Grant for my book, "Principles of Biomedical Informatics", and
- to you all for listening.