Bioinformatics: Biology X

Bud Mishra

Room 1002, 715 Broadway, Courant Institute, NYU, New York, USA

Model Building/Checking, Reverse Engineering, Causality

Outline

Main theses

"There are seven main propositions in the text. These are:

1. "The world is everything that is the case.
2. "What is the case (a fact) is the existence of states of affairs.
3. "A logical picture of facts is a thought.
4. "A thought is a proposition with sense.
5. "A proposition is a truth-function of elementary propositions.
6. "The general form of a proposition is the general form of a truth function, which is: $\langle \bar{p}, \bar{\xi}, \neg\bar{\xi} \rangle$
7. "Where (or of what) one cannot speak, one must pass over in silence."

–Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, 1921.

Outline

Administrivia

- **Instructor:** Bud Mishra
- Room 1002, 715 Broadway
- email: mishra@nyu.edu
- phone: 212-998-3464
- Office Hours: Tuesdays, 1:30 pm

Administrivia

- **Course Details:** G22.3033-010
  || Bioinformatics
- **Time and Place:** Tuesdays, 5:00-6:50 pm EST
  || Room 1221, 719 Broadway
- **Number of Credits:** 3 credits
- **Course Work:** Software Project, Analyzing Gene Expression Data
- Diseases (Cancer, Diabetes, Autism, CFS, Mortality)
- **Languages of Choice:** Python, Matlab, Mathematica, R (No Perl please)

Text Books

- Uri Alon || An Introduction to Systems Biology: Design Principles of Biological Circuits || Chapman & Hall/CRC, 2006.
- Edmund M Clarke, Orna Grumberg, Doron A Peled || Model Checking, The MIT Press, 2001.
- Patrick Suppes || A probabilistic theory of causality, 1970. (Available at: `http://suppes-corpus.stanford.edu/article.html?id=106-1`)
- Jon Williamson || Bayesian Nets and Causality: Philosophical and Computational Foundations, Oxford University Press, 2005.

Outline

Automating Biology

- **Main Thesis**
- A computational biologist is one who thinks about how biologists think about what they think is biology.
- Biology of the future will be done by a biologist and his dog: The biologist to ensure that large-scale high-throughput experiments and computational analysis of the resulting data are carried out properly, and the dog to bite him if he ever touches the experiments or the computers.
  1. *How should time-course biological data be anlayzed? Mathematical Models?*
  2. *How can we find causal connections among biological processes?*
  3. *What are the most important applications of such data?*

Areas we wish to touch on...

- Systems Biology
- Statistical Algorithms for Data Analysis
- Model Selection
- Model Checking

Let us think about these inter-connected questions from a single global perspective... CAUSALITY

Why is the study of biology intrinsically a causal pursuit?

Part I

- Overview and known biology (Genome, etc)
- What do biologists mean by A causes B?
- Pathways, metabolic processes, computational biology
- Gene expression and regulatory networks
- The search for biological mechanisms (Knockout experiments, association studies, etc.)

What does it mean for one thing to cause another?

Part II

- Overview of causality (Regularities, Counterfactuals, and Processes)
- Probabilistic Causality and the common cause principle
- Review of probability, statistics and information theory background.
- Type and Token causality

How can we identify causes?

Part III

- Overview of graphical models
- Temporal Logic and Model Checking
- Statistical testing and Empirical Bayes Methods
- Causality and time
- Temporal logic and causality
- Back to biology: experiment design, clinical trials, personal genomics

Questions???

Heated Discussions on the Suggested Topics...
Resulting in a New and Better Syllabus...
That EVERYONE Loves!

Some Reading

1. Nir Friedman, Michal Linial, Iftach Nachman, and Dana Pe'er. *Using Bayesian Networks to Analyze Expression Data*. Journal of Computational Biology, 7(3-4): 601–620, 2000.
2. Stuart Glennan. *Rethinking mechanistic explanation*. Philosophy of Science, 69(3):S342–S353, 2002. ISSN 00318248. URL `http://www.jstor.org/stable/3081105`.
3. D. Deutscher, I. Meilijson, S. Schuster, and E. Ruppin. *Can single knockouts accurately single out gene functions?* BMC Systems Biology, 2(1):50, 2008.

1. Stuart S. Glennan. *Mechanisms and the nature of causation*. Erkenntnis (1975-), 44(1):49–71, 1996. ISSN 01650106. URL http://www.jstor.org/stable/20012673.
2. Peter Machamer, Lindley Darden, and Carl F. Craver. *Thinking about mechanisms*. Philosophy of Science, 67(1): 1–25, 2000. ISSN 00318248. URL http://www.jstor.org/stable/188611.
3. John Leslie Mackie. *The Cement of the Universe*. Clarendon Press, 1974.
4. David Lewis. *Causation*. The Journal of Philosophy, 70(17):556–567, oct 1973. ISSN 0022-362X

1. David Lewis. *Causation as influence*. The Journal of Philosophy, 97(4):182–197, apr 2000. ISSN 0022-362X
2. David Hume. *An Enquiry Concerning Human Understanding*. Dover Publications, 2004.
3. Phil Dowe. *Physical causation*. Cambridge University Press, 2000.
4. Patrick Suppes. *A probabilistic theory of causality*. North-Holland Amsterdam, 1970.
5. H. Reichenbach. *The direction of time*. Courier Dover Publications, 2000.
6. N. Cartwright. *Causation: One word, many things.* Philosophy of Science, 71(5):805–819, 2004.

1. Ellery Eells. *Probabilistic Causality*. Cambridge University Press, 1991.
2. Ellery Eells. *Probabilistic Causality*. Cambridge University Press, 1991.
3. Elliott Sober and David Papineau. *Causal factors, causal inference, causal explanation*. Proceedings of the Aristotelian Society, Supplementary Volumes, 60:97–136, 1986. ISSN 03097013. URL `http://www.jstor.org/stable/4106899`.
4. Christopher Read Hitchcock. *The mishap at reichenbach fall: Singular vs. general causation*. Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 78(3):257–291, 1995.

1. Jon Williamson. *Bayesian nets and causality: philosophical and computational foundations*. Oxford University Press, 2005.
2. Hans Hansson and Bengt Jonsson. *A logic for reasoning about time and reliability*. Formal Aspects of Computing, 6(5): 512–535, 1994.

1. E.M. Clarke, O. Grumberg, and D.A. Peled. *Model checking*. Springer, 1999.
2. Bradley Efron. *Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis*. Journal of the American Statistical Association, 99(465):96–105, 2004.
3. Clive W.J. Granger. *Testing for Causality: A Personal Viewpoint*. Journal of Economic Dynamics and Control, 2: 329–352, 1980.
4. C. J. Langmead. *Towards inference and learning in dynamic bayesian networks using generalized evidence*. Technical Report CMU-CS-08-151, Carnegie Mellon University, 2008.

[End of Lecture #1]

See you next week!