The System is the Sensor... and it's Closed-loop

Robert Nowak
University of Wisconsin-Madison

Overview: Science is arguably the pinnacle of human intellectual achievement, yet the scientific discovery process itself remains an art. Human intuition and experience are still the driving forces of the high-level discovery process: we determine which hypotheses and theories to entertain, which experiments to conduct, how data should be interpreted, when hypotheses should be abandoned, and so on. Meanwhile machines are limited to low-level tasks such as gathering and processing data. While this arrangement may have sufficed in the past century, today large-scale networked systems are a major scientific focus. The complexity and high-dimensionality of these systems makes it incredibly difficult for humans alone to manage and optimize the process of science. A grand challenge for scientific discovery in the 21st century is to devise machines that directly participate in the high-level discovery process.

New Theory and Methods in Information and Decision Systems: There are two problems with the mainstream approach information and decision systems, both owing to the fact that it is based on a separation between data collection (experiment, sensing, sampling) and data analysis (statistical inference):

1) the analysis of fundamental limits of information processing and statistical inference based on this separation can be misleading.
2) it is often contrary to the process of scientific discovery.

Regarding the second problem, it is clear to me, after collaborating with scientists from many fields (and by watching my child's development), that the discovery process is anything but an open-loop system. Feedback is the rule. Data selection is often more central to the process than is data analysis. Deciding which question to ask next, based on past experience, is the key. Our theoretical understanding of this issue is far from complete; reliable, automated, computational tools for addressing it are almost non-existent.

To illustrate the first problem, let me mention a few examples that we have studied. The upshot is that feedback in sensing, experiment and data selection drastically changes the so-called "fundamental" limits of statistical inference.

1) *Learning binary classifiers.* Consider the simple problem of learning a linear classifier in d dimensions. Given a random set of labeled examples, to learn a classifier with an error rate within ϵ of the optimal rule requires $O(d/\epsilon)$ examples. In contrast, if unlabeled examples can be adaptively selected in a sequential fashion for labeling, then the same performance is possible with only $O(d \log 1/\epsilon)$ labeled examples [1].

2) *Reconstructing piecewise smooth images.* Consider the space of images consisting of Lipschitz smooth regions separated by Lipschitz smooth boundaries. Conventional imaging systems sample the imaging field on a uniform grid of points. For the class of images above, the error of a reconstruction based on n samples is $O(n^{-1/2})$. If the samples are selected adaptively, one after the other, then the reconstruction error is $O(n^{-1})$ [2].
3) *Sparse signal inference.* Suppose we observe a sparse signal of length n in noise. Most of the n elements of the signal are zero, but a small, sublinear number are non-zero. How well can we identify the locations of the non-zero elements? Suppose we have a fixed sensing budget that is proportional to n. If this budget is allocated uniformly across the signal, then the non-zero locations can be reliably determined if and only if their magnitudes are $O(\sqrt{\log n})$ in the high-dimensional limit. On the other hand, if the sensing budget is allocated in an adaptive sequential fashion, then it is possible to gradually focus on the relevant signal dimensions. Through adaptation, it is possible to reliable determine the non-zero locations as long as their magnitudes exceed any arbitrarily growing function of n [3]. In effect, adaptation boosts the SNR by an $O(\log n)$ factor.

Application — Network Science: The engineering and scientific study of large-scale networked systems will be a major focus of 21st century, in technology, biology, sociology, and cognitive science. Deciding where, when and what to sense or measure is a crucial problem in the science of such systems. In the Internet, for example, it is impossible to monitor everywhere and all the time. Even network operators can only monitor a small portion of their networks. Similar problems arise in modern systems biology. Biological systems are not defined by the independent functions of individual genes, but rather they depend on the complex interactions of thousands of genes, proteins, and small molecules. The dynamic interplay of these elements is precisely coordinated by signaling networks that orchestrate their interactions. High-throughput experimental techniques now provide biologists with incredibly rich and potentially revealing datasets, but the flexibility of these methods renders it impossible to exhaustively explore the experimental space. Traditional approaches to network inference are passive, in the sense that all data are collected prior to analysis in a non-adaptive fashion. One can envision, however more active strategies in which information gleaned from previously collected data is used to guide the selection of new experiments and data [4]. In the parlance of machine learning, such feedback-driven data selection is called active learning. Very little is known about active learning methods for network inference.

Interrelationships:

1) *Information Theory.* Rényi was the first person to carefully address the problem of sequential data selection in experimentation [5], [6]. He pointed out the relationship between this problem and the classic problems in Shannon's information theory. Horstein's seminal work on feedback in channel coding is the first rigorous analysis of feedback in a learning system (the channel coding problem is equivalent to binary search with noise). Feedback in communications and network coding are two areas of renewed interest in the information

theory community and both are closely related to the challenges outlined above.

2) *Optimization and Control.* Over the past decade we have witnessed an explosion of activity integrating optimization and statistical inference. This has enabled practical methods for very high-dimensional data analysis. One of the distinguishing features of these methods is that they are adaptive to underlying low-dimensional structure hidden in the data. Active learning attempts to extend this sort of adaptivity to the selection of experiments and data. However, the fact that learning systems are inherently noisy makes the analysis of feedback between data analysis and selection incredibly complicated. Undoubtedly, ideas from optimization theory and feedback control will play a central role in the develop of new theory and methods in active learning. After all, the largest learning system ever built — the Internet — is based on feedback and adaptation [7].

References

[1] R. Nowak. Noisy Generalized Binary Search. *Advances in Neural Information Processing Systems (NIPS)*, 22:1366–1374, 2009.

[2] R. Castro, R. Willett, and R. Nowak. Faster Rates in Regression via Active Learning. *Advances in Neural Information Processing Systems (NIPS)*, 18:179–186, 2005.

[3] J. Haupt, R. Castro, and R. Nowak. Distilled Sensing: Selective Sampling for Sparse Signal Recovery. *Proc. 12th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 216–223, 2009.

[4] S. Zehetmayer, P. Bauer, and M. Posch. Optimized multi-stage designs controlling the false discovery or the family-wise error rate. *Statistics in Medicine*, 27:4145–4160, 2008.

[5] A. Rényi, “On a problem in information theory,” *MTA Mat. Kut. Int. Kozl.*, p. 505516, 1961, reprinted in *Selected Papers of Alfred Rényi*, vol. 2, P. Turan, ed., pp. 631-638. Akademiai Kiado, Budapest, 1976.

[6] ——, “On the foundations information theory,” *Review of the International Statistical Institute*, vol. 33, no. 1, 1965.

[7] J. Doyle, D. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger. The ”robust yet fragile” nature of the Internet. Proc Natl Acad Sci U S A. 2005 Oct 11;102(41):14497-502.