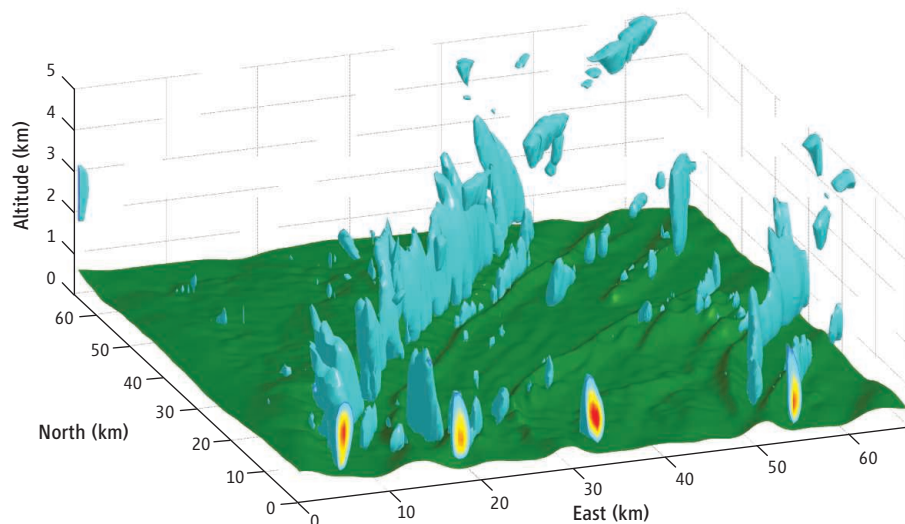


been demonstrated on a small UAV (15), and there is ongoing research in control strategies to improve cruise performance of small and mini-UAVs by gust energy harvesting.

The improved perception provided by the sensing and processing systems, coupled with the improved persistence provided by atmospheric energy harvesting, should enable long-endurance missions that are far beyond the capabilities of current robotic aircraft. Eventually a soaring-capable, autonomous, mini-UAV that is equipped with a sophisticated sensing system will be able to follow a migrating bird and provide close-up in-flight video, as well as in situ atmospheric measurements. Successful completion of such a challenging mission would demonstrate that flight by human-built robotic aircraft could rival that of birds.

References and Notes

1. The Aerosonde UAV has flown meteorology missions into hurricanes (www.aerosonde.com/), but it is still too large for many military and scientific missions.
2. The Defense Advanced Research Projects Agency (DARPA) definition of MAV specifies a wingspan of 15 cm or less. Here we use the term more loosely to include all vehicles that are hand-launchable and can be operated by a single person. An example is the Aerovironment Wasp MAV (www.avinc.com/uas/small_uas/wasp/).
3. S. Griffiths *et al.*, in *Advances in Unmanned Aerial Vehicles: State of the Art and the Road to Autonomy*, K. P. Valavanis, Ed. (Springer-Verlag, Dordrecht, Netherlands, 2007), pp. 213–244.
4. J. J. Leonard *et al.*, Dynamic Map Building for an Autonomous Mobile Robot. *Int. J. Robot. Res.* **11**, 286 (1992).
5. S. Thrun, Learning metric-topological maps for indoor mobile robot navigation. *Artif. Intell.* **99**, 21 (1998).
6. M. Achtelik, A. Bachrach, R. He, S. Prentice, N. Roy, *Proc.*



Wind-assisted outdoor flight. The potential for extracting energy for flight from natural winds created by mountain “wave”—long-period oscillations of the atmosphere—over central Pennsylvania (Allegheny Plateau, Bald Eagle Ridge, and Tussey Ridge). The cyan isosurfaces bound the regions where soaring can occur—vertical wind velocity exceeds the sink rate of the vehicle. Nighttime wind-field changes are shown in video S2.

SPIE **7332**, 733219 (2009).

7. A. Torralba, K. P. Murphy, W. T. Freeman, M. A. Rubin, in *Proceedings of the Ninth IEEE International Conference on Computer Vision* (IEEE, Piscataway, NJ, 2003), pp. 273–280.
8. K. Lai, D. Fox, *Proceedings of Robotics: Science and Systems* (MIT Press, Cambridge, MA, 2009).
9. Y. Wei, E. Brunskill, T. Kollar, N. Roy, in *Proceedings of the IEEE International Conference on Robotics and Automation* (IEEE, Piscataway, NJ, 2009), pp. 3761–3767.
10. J. W. S. Rayleigh, *Nature* **27**, 534 (1883).
11. M. J. Allen, V. Lin, AIAA Aerospace Sciences Meeting and Exhibit (AIAA Paper 2007-867, American Institute of Aeronautics and Astronautics, Reno, NV, 2007).
12. A. Chakrabarty, J. W. Langelaan, AIAA Guidance, Naviga-

tion and Control Conference, AIAA Paper 2009-6113 (American Institute of Aeronautics and Astronautics, Chicago, IL, 2009).

13. G. Sachs, *Ibis* **147**, 1 (2005).
14. M. Deittert, A. Richards, C. A. Toomer, A. Pipe, *J. Guid. Control Dyn.* **32**, 1446 (2009).
15. C. K. Patel, *Energy Extraction from Atmospheric Turbulence to Improve Aircraft Performance* (VDM, Saarbrücken, Germany, 2008).

Supporting Online Material

www.sciencemag.org/cgi/content/full/326/5960/1642/DC1
Videos: S1 and S2

10.1126/science.1182497

COMPUTER SCIENCE

Mining Our Reality

Tom M. Mitchell

Something important is changing in how we as a society use computers to mine data. In the past decade, machine-learning algorithms have helped to analyze historical data, often revealing trends and patterns too subtle for humans to detect. Examples include mining credit card data to discover activity patterns that suggest fraud, and mining scientific data to discover new empirical laws (1, 2). Researchers are beginning to apply these algorithms to real-time data that record personal activities, conversations, and movements (3–8) in an attempt to improve

human health, guide traffic, and advance the scientific understanding of human behavior. Meanwhile, new algorithms aim to address privacy concerns arising from data sharing and aggregation (9, 10).

To appreciate both the power and the privacy implications of real-time data mining, consider the data available just to your phone company, based on your phone records and those of millions of other individuals who are going about their daily lives carrying a smart phone—a device that contains a Global Positioning System (GPS) sensor locating you to within a few meters, an accelerometer that detects when you are walking versus stationary, a microphone that detects both conversations and background noises, a camera that records

where each picture was taken, and an interface that observes every incoming and outgoing e-mail and text message. The potential benefits of mining such data are various; examples include reducing traffic congestion and pollution, limiting the spread of disease, and better using public resources such as parks, buses, and ambulance services. But risks to privacy from aggregating these data are on a scale that humans have never before faced.

One line of research is based on watching where people are, where they are heading, and when. Anonymous real-time location data from smart phones are already being used to provide up-to-the-minute reports of traffic congestion in many urban regions through services such as Google Maps (11).

Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA. E-mail: tom.mitchell@cs.cmu.edu

CREDIT: (PLOT) A. CHAKRABARTY/DEPARTMENT OF AEROSPACE ENGINEERING, PENN STATE UNIVERSITY; (DATA) G. S. YOUNG, B. J. GAUDET, N. L. SEAMAN, D. R. STAUFFER/DEPARTMENT OF METEOROLOGY/PENN STATE UNIVERSITY

Downloaded from www.sciencemag.org on December 18, 2009

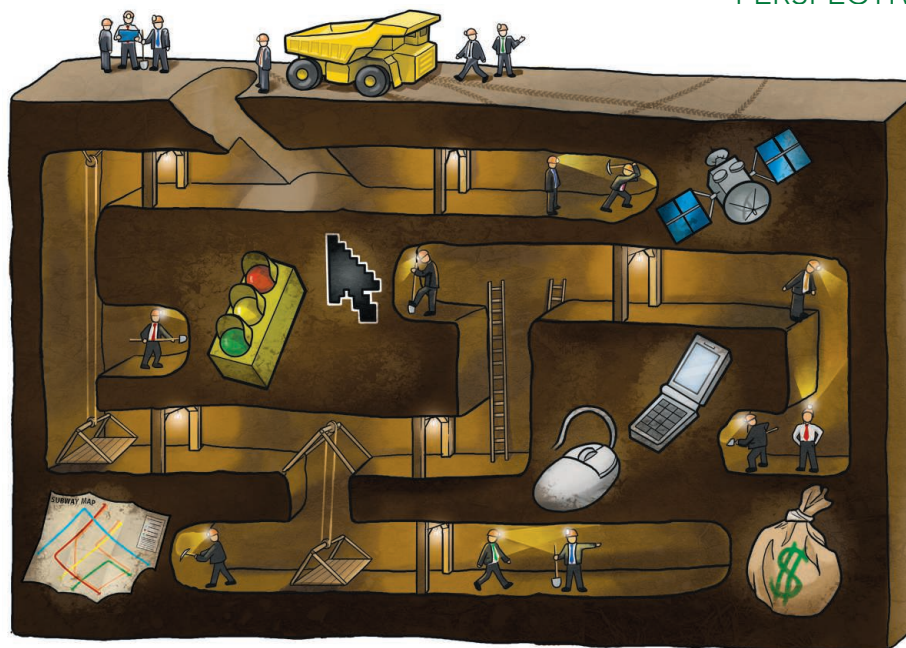
This information could be used to relieve congestion by controlling traffic lights in real time, and to optimize public transport schedules. Pedestrian traffic can also be monitored through GPS smart phones, for example, providing data on which parks are most frequently used and for how long. Even census data might be continuously updated by observing where your smart phone spends most nights.

Aggregating such GPS geolocation data with other sources opens up a vast range of new possibilities, as well as new privacy issues. For example, if your phone company and local medical center integrated GPS phone data with up-to-the-minute medical records, they could provide a new kind of medical service: If phone GPS data indicate that you have recently been near a person now diagnosed with a contagious disease, they could automatically phone to warn you.

A second line of research uses real-time sensing of routine behavior to study interpersonal interactions. For example, Pentland (3) has used specially designed work badges—"sociometers"—to study productivity and creativity in the workplace. The sociometers contain infrared sensors, microphones, accelerometers, and location sensors to record the location and duration of conversations between workers, their physical distance apart, who speaks when, speaker intonation, gestures, and upper body motion. Analyzing these data allows Pentland to track various informal and sometimes subconscious cues in conversations, such as the subconscious mimicry by one person of the head nods, gestures, and vocal intonations of the other.

In one study of salary negotiations between individuals monitored by sociometers, this subconscious mimicry was found to be a key indicator of successful negotiations, and was strongly correlated with the feelings both parties reported about their negotiation. Pentland calls such subconscious cues "honest signals" that can indicate how the conversation will turn out—even better than the actual words exchanged. This research, together with the growing collection of individual data on a vast scale, suggests an important new opportunity for behavioral psychology and social science (4): research based on large-scale field data collected as people go about their daily lives, in contrast to laboratory experiments that produce more controlled and more limited data.

A third line of research involves monitoring real-time cyber data to track the ebb and flow of interests and ideas of millions of individuals. For example, the Google Trends



Web site (12) can be used to plot how many people queried Google for a given topic, broken down by year and date, and by country and city of origin. Ginsberg *et al.* have shown (5, 6) that by mining millions of geographically localized health-related search queries, one can estimate the level of influenza-like illnesses in regions of the United States with a reporting lag of just 1 day—faster than the estimates provided by government agencies such as the U.S. Centers for Disease Control and Prevention (CDC).

Analysis of such cyber data allows new types of questions to be asked and answered. For example, Leskovec *et al.* (7) analyzed 10 million postings to 45,000 different blogs over a 1-year period to study the cascading spread of ideas in the blogosphere. The authors developed a computer algorithm to identify the blog routes through which new ideas most often spread, and calculate which handful of blogs out of these 45,000 one should read to maximize the probability of spotting relevant new topics quickly.

As more diverse sensors become pervasive, wireless networking becomes more widespread, and new algorithms are developed, a global sensor network monitoring much of humanity might emerge. The development and use of privacy-preserving data mining algorithms (9) will thus be very important. One promising approach based on secure multiparty computation (10) allows mining data from many different organizations without ever aggregating these data into a central data repository. Each organization performs part of the computation based on its privately held data, and uses cryptography to encode intermediate results that must be communicated to other organizations per-

forming other parts of the computation. Such methods could, for example, be used to mine private medical records held at thousands of individual hospitals to determine which treatments work best for a new flu strain while retaining complete privacy of patient records within each hospital. Other privacy-preserving methods are also being explored, ranging from sharing only statistical summaries of the individual data sets, to inserting random perturbations into individual data records before sharing them.

Perhaps even more important than technical approaches will be a public discussion about how to rewrite the rules of data collection, ownership, and privacy to deal with this sea change in how much of our lives can be observed, and by whom. Until these issues are resolved, they are likely to be the limiting factor in realizing the potential of these new data to advance our scientific understanding of society and human behavior, and to improve our daily lives.

References and Notes

1. D. Waltz, B. G. Buchanan, *Science* **324**, 43 (2009).
2. M. Schmidt, H. Lipson, *Science* **324**, 81 (2009).
3. A. Pentland, *Sloan Manage. Rev.* **50**, 70 (2008).
4. D. Lazer *et al.*, *Science* **323**, 721 (2009).
5. J. Ginsberg *et al.*, *Nature* **457**, 1012 (2009).
6. See www.google.org/flutrends.
7. J. Leskovec *et al.*, *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 420–429 (2007).
8. A. Vespignani, *Science* **325**, 425 (2009).
9. C. C. Aggarwal, P. S. Yu, Eds., *Privacy-Preserving Data Mining: Models and Algorithms* (Springer, New York, 2008).
10. Y. Lindell, B. Pinkas, *J. Privacy Confidentiality* **1**, 59 (2009).
11. For example, real-time traffic in Pittsburgh is tracked at maps.google.com/maps?q=pittsburgh+pa&layer=t.
12. See www.google.com/trends.

10.1126/science.1174459