The Impact of Prediction Contests

Mu Zhu

Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario N2L 3G1
e-mail: m3zhu@math.uwaterloo.ca

I appreciate this opportunity to discuss issues that are of importance to our community. I work in the area of statistical machine learning. This branch of statistics is more concerned about prediction than about explanation (Shmueli 2010). While not all statisticians would agree with this particular scientific position (see, e.g., Breiman 2001), the choice of research problems — to quote Professor Shing Tung Yau, a famous mathematician and celebrated Fields medalist — is largely "a matter of personal taste" (Yau 2011). My "personal taste" happens to be better aligned with the predictive and not the explanatory aspect of statistical modeling. As such, my discussion below can represent only a limited personal point of view.

Science

A very recent phenomenon in the area of statistical machine learning is the emergence of numerous real-world prediction contests.

In 2006, Netflix shook the world of numerical analysts by announcing a one-million-dollar prediction contest (http://www.netflixprize.com/). The goal was to beat the accuracy of Netflix's Cinematch algorithm for recommending movies by 10% on the root-mean-squared-error (RMSE) scale, a goal that took various teams from all of the world almost three years to accomplish. During the three-year period, numerous conferences were held where researchers working on the contest gathered together to share ideas and compare solutions. In hindsight, the Netflix contest successfully served as a catalyst to generate a great deal of new research (Koren *et al.* 2009). Since then, the web site, http://www.kaggle.com/, has started to sponsor many such similar contests. The site also allows instructors to set up mini-contests for their classes (http://inclass.kaggle.com/).

These prediction contests are changing the landscape for researchers in my area — an area that focuses on making good predictions from finite (albeit sometimes large) amount of data. In my personal opinion, they are creating a new paradigm with distinctive advantages over how research is traditionally conducted in our field.

First, when writing a paper, a researcher often hand-picks the "right" problem so as to illustrate that his/her method is superior to many existing ones.

There is nothing fundamentally wrong or unethical about such a practice. After all, nobody expects a method to work in all situations, so the researcher is merely showing us the kind of situations for which his/her method is particularly well-suited. However, this makes it hard for us to appreciate the value of other researchers' contributions to problems that we are trying to solve ourselves. In reality, it is rare for two researchers to be faced with exactly the same problem; there are always some differences, and sometimes these seemingly small differences matter a lot. Your approach may not work well for my problem, but it doesn't mean your approach is bad. A prediction contest allows researchers from all over the world to work on exactly the same problem. As a result, the advantages and disadvantages of different methods can be more easily compared and understood, and researchers can better learn from each other's (common) experiences.

Second, the kind of mass collaboration afforded by these contests is unprecedented. Never before have so many researchers come together and worked on exactly the same problem on such an immense scale. When many teams arrive at the same conclusion that a certain idea tends to work better for the underlying problem, that conclusion carries a lot of weight. When an algorithm is the top performer against hundreds of competitors — many of which implemented by the top brains in the field — as opposed to a few "rival" algorithms selected specifically by the authors of a paper in order to demonstrate that their algorithm "works," the value and advantages of such an algorithm can be established on much more solid grounds. In other words, there are many valuable lessons for a researcher to learn from such a contest that he/she cannot otherwise learn from reading a journal paper.

For reasons such as these (and many others), I expect to see many more prediction contests in the future, and I expect their influence on our field to grow exponentially. However, I think that, currently, statisticians are slightly disadvantaged and not as well prepared to become dominant players in these contests. For example, everybody says the internet is a giant repository of data and information, but not many statisticians are proficient with tools to crawl the internet themselves and obtain these data for analysis. The very meaning of being a leader means that one has to push the boundary of the current technology, but statisticians traditionally have not been trained to take on such a role. Experiences from the Netflix contest have clearly shown that it usually takes a team of several "hackers" to participate and succeed in these prediction contests, but our graduate students typically lack the necessary computer literacy. It is no surprise that the team winning the Netflix contest was led by industrial researchers from Yahoo Research and AT&T Labs, rather than by statisticians from top academic departments.

Research funding

How does this impact our research funding? As I have mentioned above, it often takes a team to solve a real, large-scale problem and/or to participate in these

prediction contests. As far as funding goes, statistics so far has remained largely a solitary discipline — we collaborate with each other in the sense that many of our papers are co-authored, but we don't establish or operate labs as other scientists do. The current level of funding available to the mathematical sciences does not realistically allow us to operate labs, but I see a need for this to change if we want to stay competitive and useful. While theory cannot be displaced, there is a need to develop the experimental side of statistics. Just like physics, statistics also needs to have both theorists and experimentalists, and we need to fund statistical and analytical labs consisting of post-docs, graduate students, and technical support staffs.

Training

Likewise, these prediction contests have created a demand that we train our graduate students differently. Other than traditional pedagogic elements such as solving problem sets and taking exams, our students must procure hands-on skills not typically taught in classrooms. In order for the discipline of statistics to stay relevant, next-generation statisticians need to become "qualified data hackers," and this kind of training is best acquired in a lab environment.

Institutes

Finally, I will take this opportunity to say a few (very personal) words about our mathematical institutes. I have found institutes such as BIRS to be absolutely invaluable. The thematic workshops and programs that they host and sponsor are infinitely more useful and relevant than a large, all-encompassing professional conference, such as the JSM. Of course, this is not to say that the JSM is useless; it is useful in different ways. I am not exaggerating when I say that more than 50% of my own work has benefited from these institutes in one way or another. There is also something to be said about some of these institutes' relatively isolated locations, such as BIRS and similarly NISS and/or SAMSI in the United States. They are especially conducive to productivity! We need to do whatever we can to keep these institutes up and running.

References

Breiman, L. (2001). Statistical modeling: The two cultures (with discussion). *Statistical Science*, **16**(3), 199–231.

Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, **42**(8), 30–37.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, **25**(3), 289–310.

Yau, S. T. (2011). The shape of inner space. Public lecture delivered at the University of Waterloo, Waterloo, Ontario, Canada on January 18, 2011.