Enlightened by Randomness

By Phanish Puranam , the Roland Berger Chaired Professor of Strategy and Organisation Design at INSEAD, and Prothit Sen , Assistant Professor of Strategy at the Indian School of Business

To train AI properly, organisations may have to take a drastic step: start making decisions with no rhyme or reason.

Randomness is often the enemy in managerial discourse. We are fooled by it, puzzled by it and must suffer the consequences when the best laid plans of managers and CEOs are disrupted by random events. Yet, randomness can be a powerful antidote to a problem well recognised by adopters of machine learning (ML) and AI – the problem of "algorithmic bias".

Let's take the example of an organisational practice that we are all familiar with – hiring (see **Capelli, Tambe and Yakubovich (2019)** for an excellent overview of AI applications in HR). Text analysis algorithms can be trained to automatically filter applicants based on, say, words (institution names, college degrees, etc.) on their CVs[1]. However, the real power of AI lies in its potential ability to radically transform the hiring practice by accurately predicting the answer to the question, "Would we benefit if we hired this candidate (over others)?" Equivalent predictions in other domains would be about the value of investing in a particular project, targeting a particular customer through a marketing campaign, engaging a particular supplier or

choosing a particular site for a store.

Past data can be helpful in making such predictions, and that is indeed the promise of predictive analytics through ML. But we must watch out for an important managerial blind spot. The blind spot is the danger of bias in the data used to "train" ML models to predict the future performance of a candidate. Consider the illustration in Figure 1. Let's say the firm starts out hiring mostly blue candidates and rejects most red ones based on some criteria (which could be the incorrect view that only blue skills are valuable). Over time, some blue candidates come to wear crowns (i.e. perform well). These are "good hires", the type the firm wants to maximise. Based on data on current employees, one can set up an ML algorithm to predict whether future candidates being evaluated for recruitment will eventually wear a crown or not. This sounds great, but we missed something called "selection bias".
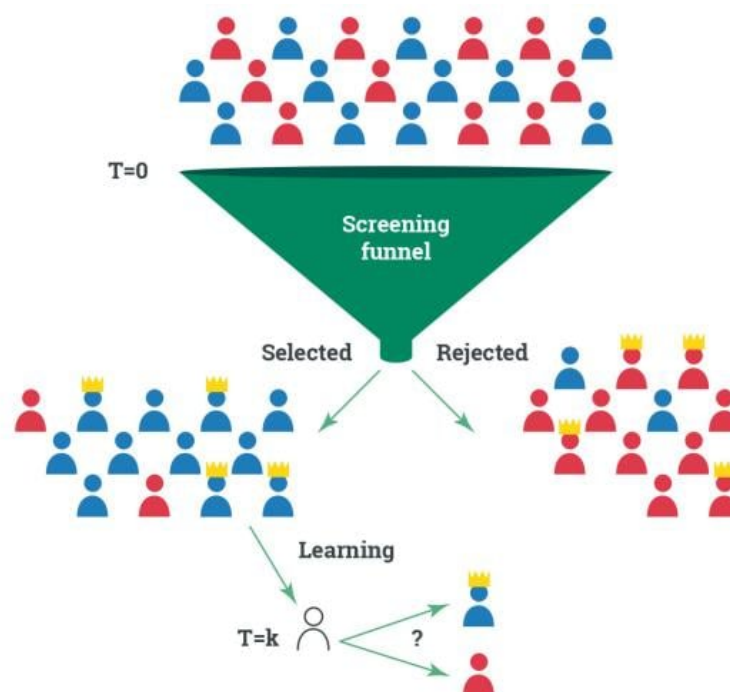

Figure 1: How a biased selection process (at T=0) is being used
to train the AI algorithm for future hiring decisions (at T=k)

We trained the algorithm to form associations between attributes (including colour) and outcome (crown or not) *conditional* on selection into the firm. In other words, we did not expose the algorithm to the counterfactual, i.e. the

rejected (red and blue) candidates. The algorithm did a great job of answering the question, "Will this candidate perform well if selected?" but not, "Should we hire this candidate over somebody else? [2]" Even if your existing non-algorithmic HR processes were perfectly accurate and never made hiring mistakes, you would be ill-advised to create an ML "digital twin" of those processes. Why? What makes a star employee may not be the same as what gets you hired. Even if it were, there might not be enough variation on this dimension among your employees for the algorithm to learn about it.

The perils of selection bias

Selection bias can harm you in at least three ways. First, you might hire bad employees, because the algorithm did not learn about the markers that distinguish the good from the bad, since there was too little variation on these markers among those who work for you. Second, there is an opportunity cost of missing out on good employees of the red type because your algorithm cannot recognise their worth (because there isn't enough of them in the data). Third, you have locked yourself into inadvertently reinforcing mistakes in past hiring, because the algorithm will make it even harder to hire red types if the next batch of data come from applying the algorithm to hire.

So, how can we avoid selection bias? One possibility is for you to keep track of the candidates you did not select (e.g. through follow-up interviews or LinkedIn). This can be cumbersome. A far easier way to get around the counterfactual hurdle is to complement AI with **random selection**.

If HR managers are free to conduct a round of hiring purely based on random sampling instead of selection based on *desired* attributes (thus making blue indistinguishable from red), both types would enter the firm. This would allow for an unbiased training set for the ML algorithm.

The removal of the screening funnel (even for one round of fresh hiring, or even a fraction of the intake) may come with short-term costs, but it can significantly improve the prediction accuracy in the evaluation of future candidates. Think of these as the acceptable sacrifices required to obtain the priceless resource of truly reliable data. In fact, one can minimise the cost by restricting randomisation to an optimised candidate pool, culled based on performance on a dummy project or skills test. It is a form of experimentation, admittedly at a cost, that yields fruit in the form of valuable insight.

Randomness in decision making

This is not the only way randomness can make algorithms more useful. For instance, it is well known that all machine learning works on correlations, not necessarily causal relationships. The only guaranteed path to discovering causal effects is through experiments involving randomisation into treatment and control groups.

To distinguish what we are talking about from "A/B testing" (comparing the effectiveness of two variants assigned randomly within a sample), we might call random decision making "A-to-Z testing" because we try all options with equal probability and no screening. The idea is neither new nor startling: To escape biases in one's thinking, it is important to do things that seem counter-intuitive given current beliefs. The challenge of balancing exploration (for better ideas by trying non-intuitive actions) with exploiting the value of current wisdom is a staple feature in learning systems of all kinds, from rats in mazes to self-driving cars[3].

Interestingly, sloppy decision making can produce a form of A-to-Z testing. Bo Cowgill at Columbia University has offered **an elegant argument**: Even if hiring managers are biased, provided they are noisy in how they act on their biases (i.e. occasionally letting in a few red candidates), the training data may still contain enough variation for algorithms to detect the value of red types. Paradoxically, the sloppiness of your hiring practices may protect you to some extent against the biases in the hiring process!

Social impact

While we wrote about hiring, our argument applies to any form of selection (projects, customers, suppliers, sites, etc.). But how does our argument apply to discrimination or other forms of bias against minorities? Everything we said above is relevant. But sadly, it is not enough. A second form of discrimination is often at work against minorities, and that is in evaluation – who gets to "wear the crown" among the hired employees.

Although bias can always extend to both selection and evaluation, the likelihood is much higher when prejudice is present. For example, a manager might, all things being equal, irrationally prefer one form of project over another, but it is hard to justify avoiding the "wrong" kind of project when it is proven to be profitable. However, the inherent subjectivity of the evaluation process can give misogynistic managers a free hand to unfairly

disparage the work of female employees. Even with random hiring, the few "reds" who sneak in may never be allowed to wear the crown. That means that even with A-to-Z testing, the algorithms will learn to perpetuate this bias. Worse, if the bias is widespread across companies, the problem may be hard to even perceive as a business problem in addition to an ethical one, because there will no glaring cases of the "stars that got away" to point to. But one battle at a time: Awareness of the challenge is already a step up the mountain.

Phanish Puranam is the Roland Berger Chaired Professor of Strategy and Organisation Design at INSEAD.

Prothit Sen is an Assistant Professor of Strategy at the Indian School of Business.

Found this article useful? Subscribe to our weekly newsletter.

Follow INSEAD Knowledge on Twitter and Facebook.

[1] See for instance *Prediction Machines* by Ajay Agrawal, Joshua Gans and Avi Goldfarb (Harvard Business Review Press, 2018).

[2] This is a very well-known problem in statistics and in applications in economics. See for instance Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*. 47(1): 153–61.

[3] See for instance John Holland (1989), James March (1991), or Sutton and Barto (2018).

Find article at

https://knowledge.insead.edu/leadership-organisations/enlightened-randomness

About the author(s)

Phanish Puranam is the Roland Berger Chaired Professor of Strategy and Organisation Design at INSEAD.

Prothit Sen is an Assistant Professor of Strategy at the Indian School of Business.