ISE 599
Special Topics Applied Predictive Analytics

Mayank Kejriwal

Metrics

- Can’t live without them
- “Measurement is the first step that leads to control and eventually to improvement.
  If you can’t measure something, you can’t understand it.
  If you can’t understand it, you can’t control it.
  If you can’t control it, you can’t improve it.”
  — H. James Harrington

Examples

- We saw several IR metrics: NDCG, MAP for ranked list, precision and recall for or unranked list (set)
- TOEFL synonymy test for word embeddings, language models
- In general, precision, recall and F-measure apply in a whole series of settings
- Others: AUC, ROC curves, accuracy…
- Curves and metrics like the F-Measure are useful for quantifying tradeoffs but must be used carefully

Metrics must be carefully chosen to reflect your actual problem

- Recall difference between in vitro vs. in vivo evaluations
- Both have their own sets of (distinct) metrics, since they are asking different evaluation questions

Campbell's law

- "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."

- Related:
- "When a measure becomes a target, it ceases to be a good measure." — Goodhart's Law

So what happens?

- When first formulated, the measure appears effective and highly correlated with your original goal
- Over time, people learn to game it (human beings are creative)
  - Soviet Russia story
- May be a very real and subtle problem in Artificial Intelligence and NLP!
  - Turing test
  - Chinese room
- Can appear in a variety of guises in science: p-hacking, model selection bias, data bias…

- *…over time, it's the measure and not the original goal that keeps appearing in reports, messages, communications, and discussions. Revisiting the goal often isn't done. As such, slowly, the group changes. The measurement takes over.*

Examples from real life: what could do wrong?

- Grocery store cashiers are sometimes set a metric (have to bag a certain number of items per hour)
- Sales
  - a typical compensation structure can look like:
  - If the salesman sold less than $50k worth of vacuum cleaners per month, he'd net a 3% commission; $50-99k, and the commission would increase to 5%; $100k+, and he'd get 7%
  - What happened?

Online advertising

- Suppose click through rate (CTR) is the metric you have to optimize
- Some clients demand that the ad agency or marketer achieve a CTR of (say 15%)
- Suppose Ferrari wants you to get a 15% CTR on Facebook (16x the industry average)…what do you do?

Journalism

- You have to write an article that hits 2 million pageviews per month
- What do you do?

Computer programming

- Lines of code!
- Still very common, but getting superseded by tests
- What can go wrong in either case?

Is there any field that is immune?

- Law? *His worth would be quantified in billable hours, not impact*...
- Academia? *Publications and clout, not the creation of knowledge...*
- Teaching? *Standardized test score targets, not the enrichment of minds...*
- Law Enforcement? *Monthly ticket quotas...*
- Politics? *He'd sooner be fed to wolves...*

General lessons

- Improper targets can:
  - Encourage “gaming” the system (e.g., bagging free groceries)
  - Incentivize the wrong aspects of work (e.g., writing trivial code)
  - Erode morale (e.g., writing clickbait)
  - Harm customers (e.g., turning away critical surgery patients)

Some suggestions

- OKR approach: Objectives and Key Results
- Objectives should be quantitative and inspirational, time bound and actionable by the team independently
- Key results should be difficult but not impossible
- Do not change OKRs halfway through your bounded time period!
- Other methodologies for dealing with these issues are typically covered in a project management course

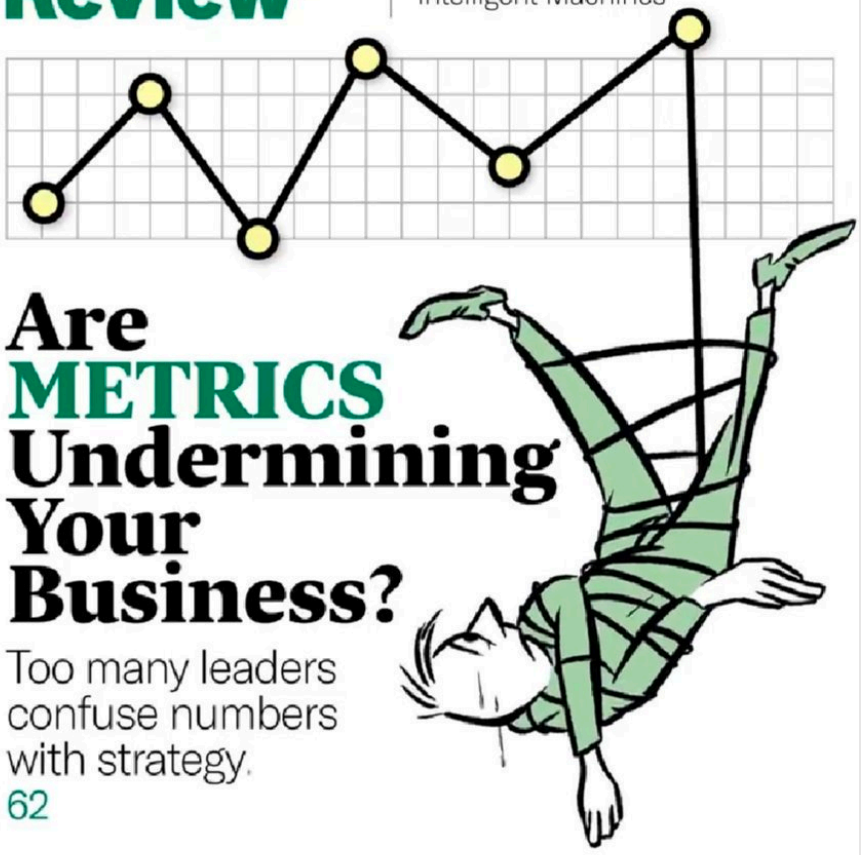
Harvard Business Review

44 How Dual-Career Couples Make It Work

70 Put Purpose at the Core of Your Strategy

116 Learning to Work with Intelligent Machines

HBR.ORG
September–October 2019



Are METRICS Undermining Your Business?

Too many leaders confuse numbers with strategy.
62