Expand Data Analysis and Insights with Space-Time Benchmarking

In sports, "most improved" designations are insightful and motivating. For example, Google reports over half a million hits on the query "*most improved player*". Such designations imply not just a comparison to other players, but also to the same player in the recent past.

Outside of sports, peer comparison is prominent in business, where it's known as benchmarking. Since the benchmarking task has traditionally been carried out manually (and mentally), the inability to deal with too much data has led to a focus on peer comparisons during a fixed measurement period, e.g., last month, quarter, or fiscal year, as well as other shortcuts. Mixing in the past is a lot of extra work.

Now here's a general rule for innovation: Given the presence of lots of data and metrics, hard mental work, few experts, and necessary shortcuts due to these and other factors, think Artificial Intelligence! Finding information on the web before there were automated search engines (remember that Yahoo started as a browsable directory of web pages!) is a good historical example, since web search engines are infused with artificial-intelligence thinking.

Automation through artificial intelligence can give rise to an expansive role for business benchmarking, which we'll call **Space-Time Benchmarking**, because it enables comparing to peers (who occupy other spaces) AND back in time, AND not just considering improvements but changes in general. After all, to promote change and improvement, why not benchmark on changes and improvements?

Space-time benchmarking is impractical without automation, because the scope for benchmarking insights increases by more than double. For example, consider that any metric X can be combined with *change-in*-X to enable insights such as "*Acme Inc. had the smallest quarterly rise in revenue of all companies that have at least $1B in revenue*."

Let's do a simple model of the effects. Suppose there are N metrics and M possible peer groups. For a given target entity, simple benchmarks imply NxM amount of work. Now let's throw in *change-in-*X metrics. The work doubles to NxM + NxM = 2NxM. Further allowing pairings of X with *change-in-*X*,* the work goes up to 2NxM + NxM = 3NxM. If we consider both quarterly changes and, e.g., annual changes, then it's 4NxM. If change-in-X can be paired with change-in-Y, as in *"the smallest annual revenue increase of anyone whose labor force grew by at least 10%",* then the work expands still more. The total impact is actually greater in reality, due to combinatorial effects.

We have carried out automated space-time benchmarking experiments on several public data sets. The most recent application benchmarks the tax systems of 195 countries using the USAID Collecting Taxes Database. The same source makes historical data available, so that metrics such as corporate income tax productivity has values both for the last measurement period and for three years earlier. Here are three example outputs, discovered and written by software, on the United States that involve *changes over time*:

1. USA has the 6th-biggest rise over 3 years in tax collection overhead (+38.0%) among all the 195 nations. That +38.0% compares to an average of +16.2% and standard deviation of +87.5% across the 195 nations. Also, that +38.0% represents a rise from 0.45% to 0.62%.

2. USA has the 7th-biggest rise over 3 years in corporate income tax collection as a percentage of GDP (+190.0%) among all the 195 nations. That +190.0% compares to a median of 0.00% across the 195 nations. Also, that +190.0% represents a rise from 0.90% to 2.6%.

3. USA has the biggest rise over 3 years in corporate income tax productivity (+130.0%) among the 28 nations with at least a 35.0% corporate income tax rate. That +130.0% compares to an average of -1.7% and standard deviation of +46.6% across the 28 nations. Also, that +130.0% represents a rise from 0.03 to 0.07.

Readers may enter a country and browse the insights themselves at taxes.onlyboth.com.

Automated benchmarking relies on a classical AI approach: (1) Study the potentially-large space of acceptable solutions, (2) Figure out what makes one solution better than another, (3) Devise an algorithm that will search this space while using knowledge to guide its search in favorable directions, (4) Rank the outputs according to other domain knowledge, and finally (5) Write up conclusions as a human consultant would.

Space-time benchmarking across peers and time greatly expands the potential to find valuable and motivating insights. It places the task even further beyond the reach of traditional, largely-manual methods, just as full-text search, rather than just search over titles, took search engines far beyond what people could accomplish without automation.

To promote change, let's benchmark for it.

Raul Valdes-Perez