Data / algorithms: where is the value?

Marcin Detyniecki
Head of Research and Development
& Group Chief Data Scientist

« Data is the black gold of the 21st century ». This popular metaphor is frequently used by experts and media alike. It considers data as the most precious asset of our time, not unlike oil in the last century. One could take the metaphor one step further, comparing algorithms to refineries since they transform raw data into usable material, just as a refinery transforms oil into gasoline. So, where is the value? In the data or in the algorithm?

A s is often the case, reality is more complex than it seems. There is no hard and fast rule: depending on the circumstances and the context, the value may reside on the side of the data or the algorithm. It may even transition from one to the other over time. Data and algorithms are also part of a larger ecosystem that encompasses them. If we do compare data to crude oil and algorithms to refineries, we need to add a few parameters for the metaphor to be complete such as the driver who will drive car (i.e. human intelligence) and the infrastructure that allows automotive circulation (i.e. human intelligence). This combination is required for value to emerge.

Even though data and value often go hand in hand, this is not always the case. In reality, data does not hold intrinsic value in itself: the value depends on another set of elements. More precisely, data creates value whenever it allows a company to gain an advantage over its competitors. This is the matter of competitive advantage.

When data enables value

In order to find out whether there is in fact a competitive advantage, one should ask a series of questions. One of the most important ones is the degree of exclusivity of the data: Am I the only one who owns this specific dataset,

or is it shared? Whenever it is accessible to all, data has little intrinsic value, even if its quality is excellent. For example, although extremely rich, Wikipedia cannot allow a company to differentiate itself from its competitors. Conversely, when a company is the sole owner of data, it can create a unique algorithmic model. By unlocking this competitive advantage over its competitors, the company can generate value.

Other parameters also come into play, such as the direct contribution of the data to the value proposition of the company or product; or the speed at which the collected data becomes obsolete. In an article dedicated to the issue of competitive advantage, the Harvard Business Review notes that the faster this obsolescence occurs, the easier it is for a competitor to enter the market. The article provides two examples: the first is from the collision avoidance company Mobileye, whose data accumulated from car manufacturers over the years is still valid to this day, and for which value is maintained. The second is from the social gaming company Zynga. It aggregates data about its users, but the information collected from one game do not transfer well to the next because of changing player preferences. This data is therefore difficult to value.

The nature of the project also plays an important role in the equation. For example,

data will have a strong value if the project is about automating a process, especially when interacting with a customer.

When the algorithm enables value

However, things are different in the context of a competitive market based on shared data. In this situation, it is the search for a particular algorithm or a more efficient configuration that will create the advantage. All efforts must then be focused on optimizing the algorithm. This is particularly the case when processing very large volumes of data. Let's take the example of claim validation, on a basis of millions of e-mails. For each file, a decision must be made on whether or not to validate the claim. In such a case, an algorithm whose performance is superior to others – even marginally – can generate value.

Injecting intelligence

Injecting human intelligence and knowledge into the algorithm is an important element to improve competitiveness and therefore its potential to generate value. Let's take the example of the CrashZam accident detection tool, developed by AXA. It uses the sound of a car to detect accidents in real time. Whenever the sound reveals braking followed by an impact, the tool determines that an accident has occurred. In this case, the tool relies heavily on public data collected on the web, but it is the expertise of our AI engineers that enables the algorithm to correctly identify the sound sequence of an accident.

More than just retrieving this data, the work consisted of exploiting it in an intelligent way by creating features to model the sounds and detect clues to determine whether an accident had indeed occurred. It is the way the algorithm analyzes the situation and provides an accurate diagnosis that makes the solution useful. The intelligence that was injected into it and that enables this analysis is central to the success of the tool. Therefore, the value lies in the algorithm.

When the value fluctuates

Things, however, are not set in stone. The value maybe in the algorithm at a certain point in time, but also reside in the data at another point down the line.

Let's go back to the CrashZam example. By collecting a sufficient amount of data, it could in theory become possible to achieve the same results using a standard machine learning algorithm. The value would then move to the data. After some time, one can imagine that all market players would own similar data. In that scenario, the important remaining parameter would lie in the accuracy of the tool. The most reliable solution – the one that limits the number of false positives – would be the most sought-after. In order to make the difference, one would have to look for the best algorithm. The value would have moved again.

Transferring value through data

In a system that allows unlimited queries to the algorithm, it is possible to extract the value

through the data. Let's take another example: that of a company offering an online medical prevention service. This service allows users to fill in their health data and provides them with a check-up based on these elements and on expert medical knowledge. Since this knowledge has been injected into the system, the value is well within the algorithm.

But as soon as a third-party company has the possibility to launch a large number of queries to this service, it can build up a database. Through machine learning, it will be able to reproduce the behavior of the original algorithm. The value is then passed into the data, which will allow creating a new algorithm to achieve similar, reliable results.

Contacts

Marcin Detyniecki

Head of Research and Development & Group Chief Data Scientist

marcin.detyniecki @axa.com

And tomorrow?

In the future, a new way of extracting value will become a reality. The next challenge will be the interpretability of algorithms built from data. Today, many algorithms are "black boxes". As we have just seen, they manage to exploit data and create value without human intervention, but no one is in a position to know why – the value lies in the algorithm or the data that made it possible to build it. This is why tomorrow, algorithms should be able to provide explanations. It will then be possible to extract not only the prediction, but also the knowledge that results from it. This will be a formidable competitive lever.

3 questions to...

Amit Jain
Head of Data Science and Lead
Data Innovation and Data Strategy
for AXA Switzerland

How can we deliver value through data and algorithms?

Business Value is only generated when your solution is being used. In this sense, data and algorithms are part of an artificial intelligence solution, but are not the only ingredients. The secret sauce is to co-create the solution together with the business experts on a specific business case. It needs to be a joint ambition.

What are the challenges to overcome to deliver this value?

If we say that data is the new oil, it is also true that data quality is key. Without quality data, it is very challenging to deliver value. An organization not only needs a solid data foundation, but also a strong data culture so that everyone uses data properly. In order to build an AI solution, not only do you need a lot of raw data, you also must have high-quality labels. Depending on the problem, labelling must be done by business experts. So ideally this is already integrated into their daily workflow. This is why motivating people to embrace new ways of working is extremely important.

How would you recommend encouraging this change?

Innovation is managed in phases. The starting point is a joint vision of a new way of working defined in a human centric way supported by technology. In the exploration phase, the innovation team then needs to build a first solution to demonstrate it can technically solve the problem. In the scaling phase (or pilot), the financial proof has to be provided. At the end there is the need for a very disciplined approach to operationalize the solution for all users. In the beginning of the innovation project, the teams need freedom to experiment and learn. Co-creation engages people. In later stages, innovation requires strong managers and support from the hierarchy so the solution is integrated into the business processes. All along the way, the key is to prove that you are constantly providing value to the business in an ethical way by using data and analytics.

AXA