Goldman Sachs Asset Management

SPONSORED COMMENTARY

THE BIG DEAL ABOUT BIG DATA

In the present day, we are creating more data than ever before and at an exponential rate. This information can be used for purposes that were unprecedented when data was first collected. Enhancements to technology and computing power have been critical in making sense of the data that is available globally. The growth in distributed databases, where data is stored via a centralised database across several platforms instead of a single platform, allows for highly-scalable parallel processing of vast amounts of data. This can decrease processing time by several orders of magnitude for many applications.

Yet, faster computers and bigger databases do not solve the predicament of digesting the continuous stream of data that we now have access to. As a result, data processing algorithms have evolved from simple data processing to learning how to process. This approach is called Machine Learning.

There are two basic types of Machine Learning algorithms: supervised and unsupervised. Supervised Machine Learning algorithms can make predictions based on historical observations. They analyse historical data ('training data'), model the relationship between input data (defined by its 'features'), and label output data. Unsupervised Machine Learning takes this one step further by analysing a large set of input data in order to create structure around it. As dynamic models have emerged to analyse data that is difficult to quantify, focus has been shifting from structured to unstructured data. We can now extract information from languages, images and speech. Having access to new types of information, and being able to effectively capture and process it, has resulted in entire industries being revolutionised by Big Data.

Active asset management, which has always been about uncovering opportunities before they are priced in by the broader market, is no exception. The Quantitative Investment Strategies (QIS) group at Goldman Sachs Asset Management leverages Natural Language Processing (NLP) by using computers to read and interpret vast amounts of text. This allows the team to process data in multiple languages from multiple sources.

Gauging sentiment between the lines is one simple NLP application and enables analysts to distinguish between positive and negative tones in companies' research reports. An extension to this is Topic Modelling, which attempts to segment a large body of text into topics and themes that can be easily understood and can at the same time be used for systematic analysis in Statistical and Machine Learning applications. An example of this is quarter-to-quarter comparisons of subjects that companies' senior management choose to focus on during earnings calls.

NLP also allows the team to identify subtle relationships between companies that might otherwise go unnoticed. This is known as Intercompany Momentum. While traditional momentum focuses on the persistence of price movements for a single security, Intercompany Momentum seeks to understand how the price movement of one security might impact the price movement of other related securities. These relationships are more obscure but can be assembled from the clustering of companies in text-based data, appearing together in news articles, regulatory filings or research reports.

Leveraging Big Data to inform investment decisions has particular applications in the Environmental, Social, and Governance (ESG) space. In the US and Europe, environmental impact factors have been a key topic for institutional investors. One such factor is Environmental Damage Cost. Companies with a lower Environmental Damage Cost can be subject to increased investor interest, which is believed to potentially lower a company's cost of capital, thereby allowing for price appreciation. This factor helps to strategically tilt towards companies with favourable environmental profiles. The team analyses this information for over 2,000 companies (900 in the US and 1,100 in Europe).

An additional data-driven ESG topic model uses NLP to read through articles on companies' Corporate Social Responsibility activities. This allows the team to identify major environmental themes that are shared by multiple companies. For example, over 24,000 articles have been gathered since 1999, and over 1,000 US companies can now be linked using such information.

At GSAM, we are focused on creating data-driven investment models that can objectively evaluate global, public companies across fundamental and economically-motivated investment themes. These models have historically used a large set of company-specific data such as publicly available financial statements, as well as market data in the form of pricing, returns and volumes. With the growth and availability of non-traditional data sources such as internet traffic, patent filings and satellite imagery, more nuanced and sometimes unconventional data sources can help us gain an informational advantage and make better informed investment decisions.

Research success for us is not about finding a new stock to invest in, but rather finding a new investment factor that can help improve the way we select stocks. Investment factors should be fundamentally-based and economically-motivated, and the data enables us to empirically test our investment hypotheses. By harnessing the power of Big Data, we seek new investment opportunities for our clients. Combined with specialised industry knowledge and technological advancement, the group's culture of research continues to drive evolution in its process as it searches for robust, diversified sources of return.

Find out more about Big Data and our investment capabilities on our website: GSAM.com

JAVIER RODRIGUEZ-ALARCON
Head of the Quantitative Investment Strategies Group for EMEA and AEJ
Goldman Sachs Asset Management

Sources: GSAM; GSAM Perspectives; TED Talk: Kenneth Cukier, "Big data is better data". In the United Kingdom, this material is a financial promotion and has been approved by Goldman Sachs Asset Management International, which is authorized and regulated in the United Kingdom by the Financial Conduct Authority.