Review
Beyond the hype: Big data concepts, methods, and analytics

Auer Thomas

October 18, 2022

The paper "Beyond the hype: Big data concepts, methods, and analytics" describes fundamental terms, paradigms, and approaches in the context of big data. The authors define the key aspects of 'big data' by more distinct criteria than just 'size' and show the relevancy of other characteristics such as data types, frequency of used items, and methods & tools used to analyze data.

Due to the different interpretations of 'Big Data', the 3 V's have emerged as a common definition framework: Volume, Variety, and Velocity.

○ *Volume* refers to the size of the saved data.

○ *Variety* defines the structure of the data - such as structured (Tabular), semi-structured (XML) or unstructured (Text, Video) data.

○ *Velocity* is the rate at which data is generated and the speed with which the analysis should be carried out.

There are also some more definitions such as *Veracity*, the irregularity of some data sources; *Variability and complexity*, the variation in data flow & the complexity of the fundamental system and *Value*, the value of the data in accordance to its density and approximate returns on investment. The paper then describes how different data types, such as text, audio, video, or social media can be processed in analytics to acquire intelligence - *text mining*, *information extraction*, *text summarization*, *Question answering*, *sentiment analysis*, *speech analytics* using LVCSR or Phonetics, *video content analytics* via compressed data on routed servers or locally on raw data, and, lastly, *social media analytics* using blogs, networks, news, and bookmarking by content or structure. Recent recommender systems target social influence or link prediction to gather data. Furthermore, *predictive analytics* uses aged and current data to predict future outcomes.

The goal of big data is the drive for meaningful decisions, using diverse, high-volume, datasets. The common approach is to use data management systems to gather and store data efficiently for the analytics, which then acquires knowledge from the saved data. Generally, the concrete definition of big data, because of the differences of fundamental technology and layout of definition, depends on the industry itself.