Brief Summary on Big Data

S. Sharma

Abstract. This paper provides a brief summary on Big Data research.

Summary

The age of data science has just started and is experiencing the tactical evolution by leaps and bound in all dimensions of science. The digital footprints and sizeable traces of the data are being produced and left to deposit in massive quantities by various data outlet harbingers such as Facebook, Twitter, Wikipedia, Google to name a few. The scientists from almost every community are expressing significant interest and eagerness to access the spoors of the raw data to facilitate the optimal decision makings. Disparate scientific communities unilaterally cherish the gr from the existing database models. In the last few years, due to the internet sites, the sources of data collection have been increased. The data is inundated from many potential sources such as sensors - gathering climatic information, power meter readings, social media sites, videos and digital pictures, traffic data, etc. These sources are responsible for the rapid growth of voluminous data - about 2.5 quintillion bytes every day. Because of the growth of the data with such a fast velocity, the size of data is a central moving issue today. In 2012, in a single dataset, the size of the data likely to increase from few dozen terabytes to many petabytes and about 90% of the whole data in world today is produced in the last two years only. Not only for storage capacity, but also from processing aspects, the existing database management systems fall short to handle them. In this era of data science, in its entirety, the term “data” is further redefined, which is popularly known as “Big Data” today. This may help to convincingly reformulate the operational-oriented definition of “Big Science” into data-centric definition and can be rephrased as – “The science that deals with Big Data is Big Science.” The concept of Big Data is relatively new and has a large scope of further research. This section highlights some important aspects about the Big Data. Primarily, the Big Data is described to have five concerns: volume of the data, velocity of the data, variety of the data, veracity of the data, and the value of the data. All of the concerns are briefly described here. Today, the industries are inundated with voluminous data of all types. The data is growing easily from terabytes to even petabytes. Tweets from the social network sites alone are responsible for up to 12 terabytes of data every day. On an annual basis the power meters contribute to a huge amount of data – 350 billion readings. The data, getting collected at the enterprises is at a very fast pace. Sometimes even the delay of a minute may cause the discrepancy in the analyzed output. For especially time-sensitive systems such as fraud detection, the Big Data must be analyzed as it comes into the enterprises to obtain the best results such as scanning though the 5 million trade transactions, produced every day, to capture the potential frauds. As per the description about Big Data, it can be of any type, regardless of its structured or unstructured nature. The data type majorly includes – text, audio and video data, sensor data, log files, etc. One of the main reasons that the enterprises prefer to work with Big Data is the finding of new insights, when analyzing all such data types together. The examples include – monitoring live video feeds from surveillance cameras in order to narrow down to the point of interest; to analyze the bulk data growth in the videos, documents and images to enhance the customer satisfaction. In the business world, the leaders do not completely trust on the information extracted from Big Data. Though, they exploit this information to obtain better decision making. If someone does not trust on the input, how she/ he can trust on the output and changes that one will act of such outputs are next to negligible. As the Big Data is growing in size and variety every day, instating the confidence in Big Data poses greater challenges. Regardless the engineering arena, enormous efforts are invested for the curation of Big Data with the expectations of being it content-rich and valuable. Though, the richness of the data may be highly valuable for certain communities and may be of no use for others that do not see the apt outcome from the extraction. So, in the latter case, the efforts investment does not warrant any return. So, a primary careful analysis about the value of the Big Data before its curation may help building the competitive advantage.
Conclusion. This paper was a one pager on Big Data research.

References


