Aspects of Big Data in Complex Systems



Stephan Schumann, Karl-Marx-Str.191 Berlin, 01.02.2018

This work aims to structure the different aspects of big data to visualize multiple combinations of paths of a human-system-interaction. The aim of humans to understand better complex systems for various reasons is the underlying motive to do Big-Data-Analysis in the first place. With it comes the goal to use algorithms on big data in automated processes. In this case the borders between the system and Big-Data-Analysis becomes blurr.

The map is the result of a series of lectures by various companies and players in the field of big data at the HPI Potsdam. The arrows represent an information or data stream. An aspect is an area that in itself requires intensive study and often stands for a stage of the data processing pipeline. This map serves the purpose of a discussion of which fields of big-Data will flourish the most (marked red).

	Complex Sys
07	hard to under
	generate



and sometimes is seen as source

> Data stream analysis is often done parallel to storing the stream data on a database for further analysis

Data Stream

Data Preparation



Unifying format and data structure Connecting different data Cleaning from faulty, incomplete data Repairing faulty data Completing incomplete data The results are:

Datawarehouses



tensor-oriented datastructure



DataSelection



Data Stream Analysis



To detect certain events of processes or states of a stream

As the analysis is done on a stream data preperation is mostly included



For fast calculations the data is preferebly not stored on drives but handeled on multiple nodes

Often the results are stored in a Database



Data Mining

Hypothesis-driven searching using algorithms to deduce new information through pattern detection and statistics

Process Mining

subset of data mining with focus on processes, their efficiency and consistency and further aspects

Outlook

Up to now, Big Data is mainly used by companies as an analysis tool, either for internal processes, markets or customer wishes. Many data are generated but for a long time they couldn't entirely be used. Thus, it mainly serves the information gain and the decision making. Due to increased machine learning and greater computing speed on a growing number of nodes, however, it is expected in the future that big data will increasingly serve the economical purposes of giving solutions to desires. Similar to many traffic apps, the end user is no longer interested in a detailed data analysis but simply uses the fastest way to go from A to B.

Discusion

Obviously our Data Input and the generated Data need a lot of preperation. It seems that new systems should also be focusing on a different paradigm of human computer-interaction. Maybe instead of analysing each and every step of a user, a different input would lead to a valuable and more transparent way of data-collection.