### How to Make Big Data Analytics More Interactive?

#### Dominik Slezak

#### Abstract

Big data applications need scalable methods for data exploration and knowledge discovery. The solutions of fundamental KDD tasks which work fine for more standard cases, require to be revised for truly huge and complex data sources. With a growing complexity of the corresponding computational problems, there is also a growing need to interact with the domain experts, to better specify exploration goals which can get narrowed down basing on results obtained so far. With that in regard, there is an ongoing research on how to decompose the workflows of complex data mining processes onto smaller pieces whose outcomes can be iteratively browsed by the users. In this talk, we report some examples of feature selection techniques aimed at the analysis of high dimensional data sets and discuss how user interaction can help to improve them. We also refer to one of our recent projects concerning risk management in coal mines in order to illustrate how modern feature selection algorithms help the end-users 1. to work with big data exploration systems.

The concept of big data has been around for years; most organizations now understand that if they capture all the data that streams into their businesses, they can apply analytics and get significant value from it. But even in 2. the 1950s, decades before anyone uttered the term "big data," businesses were using basic analytics (essentially numbers in a spreadsheet that were manually examined) to uncover insights and trends.

The new benefits that big data analytics brings to the table, however, are speed and efficiency. Whereas a few years ago a business would have gathered information, run analytics and unearthed information that could be used for future decisions, today that business can identify insights for immediate decisions. The ability to work faster – and stay agile – gives organizations a competitive edge they didn't have before.

Big data analytics helps organizations harness their data and use it to identify new opportunities. That, in turn, leads to smarter business moves, more efficient operations, higher profits and happier customers. In his report *Big Data in Big Companies*, IIA Director of Research Tom Davenport interviewed more than 50 businesses to understand how they used big data. He found they got value in the following ways:

**Cost reduction.** Big data technologies such as Hadoop and cloud-based analytics bring significant cost advantages when it comes to storing large amounts of data – plus they can identify more efficient ways of doing business.

- **Faster, better decision making.** With the speed of Hadoop and in-memory analytics, combined with the ability to analyze new sources of data, businesses are able to analyze information immediately and make decisions based on what they've learned.
- New products and services. With the ability to gauge customer needs and satisfaction through analytics comes the power to give customers what they want. Davenport points out that with big data analytics, more companies are creating new products to meet customers' needs.

Machine learning, a specific subset of AI that trains a machine how to learn, makes it possible to quickly and automatically produce models that can analyze bigger, more complex data and deliver faster, more accurate results – even on a very large scale. And by building precise models, an organization has a better chance of identifying profitable opportunities – or avoiding unknown risks.

Dominik Slezak University of Warsaw, Poland, E-mail: slezak@mimuw.edu.pl

## **Biomedical Data Mining**

Data needs to be high quality and well-governed before it can be reliably analyzed. With data constantly flowing in and out of an organization, it's important to establish repeatable processes to build and maintain standards for data quality. Once data is reliable, organizations should establish a master data management program that gets the entire enterprise on the same page.

Data mining technology helps you examine large amounts of data to discover patterns in the data – and this information can be used for further analysis to help answer complex business questions. With data mining software, you can sift through all the chaotic and repetitive noise in data, pinpoint what's relevant, use that information to assess likely outcomes, and then accelerate the pace of making informed decisions.

Hadoop an open source software framework can store large amounts of data and run applications on clusters of commodity hardware. It has become a key technology to doing business due to the constant increase of data volumes and varieties, and its distributed computing model processes big data fast. An additional benefit is that Hadoop's open source framework is free and uses commodity hardware to store large quantities of data.

By analyzing data from system memory (instead of from your hard disk drive), you can derive immediate insights from your data and act on them quickly. This technology is able to remove data prep and analytical processing latencies to test new scenarios and create models; it's not only an easy way for organizations to stay agile and make better business decisions, it also enables them to run iterative and interactive analytics scenarios.

Predictive analytics technology uses data, statistical algorithms and machine-learning techniques to identify the likelihood of future outcomes based on historical data. It's all about providing a best assessment on what will happen in the future, so organizations can feel more confident that they're making the best possible business decision. Some of the most common applications of predictive analytics include fraud detection, risk, operations and marketing. With text mining technology, you can analyze text data from the web, comment fields, books and other textbased sources to uncover insights you hadn't noticed before. Text mining uses machine learning or natural language processing technology to comb through documents – emails, blogs, Twitter feeds, surveys, competitive intelligence and more – to help you analyze large amounts of information and discover new topics and term relationships.

Big Data Analytics takes this a step further, as the technology can access a variety of both structured and unstructured datasets (such as user behaviour or images). Big data analytics tools can bring this data together with the historical information to determine what the probability of an event were to happen based on past experiences.

Big data applications need adaptable techniques for information investigation and information disclosure. The arrangements of crucial KDD assignments which work fine for progressively standard cases, require to be overhauled for really immense and complex information sources. With a developing unpredictability of the relating computational issues, there is likewise a developing need to interface with the space specialists, to more readily determine investigation objectives which can get limited basing on results got up until this point. With that in respect, there is a continuous exploration on the most proficient method to disintegrate the work processes of complex information mining forms onto littler pieces whose results can be iteratively perused by the clients. In this discussion, we report a few instances of highlight choice strategies focused on the examination of high dimensional informational indexes and talk about how client cooperation can assist with improving them. We likewise allude to one of our ongoing undertakings concerning hazard the executives in coal mineshafts so as to represent how current component determination calculations help the end-clients to work with enormous information investigation frameworks.

The idea of enormous information has been around for a considerable length of time; most associations presently get that on the off chance that they catch all the information that streams into their organizations, they can apply investigation and get noteworthy incentive from it. In any case, even during the 1950s,

2nd International Conference on Big Data Analysis and Data Mining November 30-December 01, 2015 San Antonio, USA

# **Biomedical Data Mining**

decades before anybody expressed the expression "huge information," organizations were utilizing fundamental investigation (basically numbers in a spreadsheet that were physically analyzed) to reveal bits of knowledge and patterns.

The new advantages that big data investigation brings to the table, notwithstanding, are speed and effectiveness. While a couple of years prior a business would have accumulated data, run investigation and uncovered data that could be utilized for future choices, today that business can distinguish experiences for guaranteed choices. The capacity to work quicker – and remain coordinated – gives associations a serious edge they didn't have previously.

Enormous information investigation assists associations with tackling their information and use it to recognize new chances. That, thus, prompts more brilliant business moves, increasingly proficient tasks, higher benefits and more joyful clients. In his report Big Data in Big Companies, IIA Director of Research Tom Davenport talked with in excess of 50 organizations to see how they utilized large information. He discovered they got an incentive in the accompanying manners:

1. Cost decrease. Huge information advances, for example, Hadoop and cloud-based examination bring critical cost points of interest with regards to putting away a lot of information – in addition to they can recognize increasingly proficient methods of working together.

2. Faster, better dynamic. With the speed of Hadoop and in-memory examination, joined with the capacity to break down new wellsprings of information, organizations can investigate data promptly – and settle on choices dependent on what they've realized.

3. New items and administrations. With the capacity to check client needs and fulfillment through examination comes the ability to give clients what they need. Davenport calls attention to that with huge information investigation, more organizations are making new items to address clients' issues.

AI, a particular subset of AI that prepares a machine how to learn, makes it conceivable to rapidly and naturally produce models that can break down greater, Extended Abstract

progressively complex information and convey quicker, increasingly precise outcomes – even for a huge scope. What's more, by building exact models, an association has a superior possibility of recognizing gainful chances – or maintaining a strategic distance from obscure dangers.

Data should be high caliber and very much represented before it tends to be dependably broke down. With information continually streaming all through an association, it's essential to set up repeatable procedures to manufacture and keep up guidelines for information quality. When information is solid, associations ought to set up an ace information the board program that gets the whole endeavor in the same spot.

Data mining innovation causes you analyze a lot of information to find designs in the information – and this data can be utilized for additional investigation to help answer complex business questions. With information mining programming, you can filter through all the turbulent and tedious clamor in information, pinpoint what's important, utilize that data to survey likely results, and afterward quicken the pace of settling on educated choices.

Hadoop an open source programming system can store a lot of information and run applications on bunches of product equipment. It has become a key innovation to working together because of the steady increment of information volumes and assortments, and its conveyed figuring model procedures enormous information quick. An extra advantage is that Hadoop's open source structure is free and uses ware equipment to store huge amounts of information.

By breaking down information from framework memory (rather than from your hard plate drive), you can get prompt bits of knowledge from your information and follow up on them rapidly. This innovation can expel information prep and systematic preparing latencies to test new situations and make models; it's not just a simple path for associations to remain coordinated and settle on better business choices, it likewise empowers them to run iterative and intuitive examination situations.

Prescient investigation innovation utilizes information, factual calculations and AI methods to recognize the

## **Biomedical Data Mining**

probability of future results dependent on verifiable information. It's everything about giving a best appraisal on what will occur later on, so associations can feel progressively sure that they're settling on the most ideal business choice. Probably the most widely recognized uses of prescient examination incorporate extortion identification, hazard, tasks and advertising.

With text mining innovation, you can examine text information from the web, remark fields, books and other content based sources to reveal bits of knowledge you hadn't saw previously. Text mining utilizes AI or normal language preparing innovation to search over reports – messages, web journals, Twitter channels, studies, serious knowledge and the sky is the limit from there – to assist you with investigating a lot of data and find new subjects and term connections.

Large Data Analytics makes this a stride further, as the innovation can get to an assortment of both organized and unstructured datasets, (for example, client conduct or pictures). Enormous information examination devices can unite this information with the recorded data to figure out what the likelihood of an occasion were to happen dependent on past encounters.