Big Data and Healthcare Analytics

Nikhlesh Lakhmani
R.V. College of Engineering, Bangalore, India.

Abstract: In recent years, limited concepts have captured the imagination of health-care practitioners as much as the advent of “big data” and the advanced analytical methods and technologies used to interpret it—it is a trend seen as having the potential to revolutionize biology, medicine, and health care [10]. As new types of data and tools become available, a unique opportunity is emerging for smarter and more effective discovery, development, and commercialization of innovative biopharmaceutical drugs [2].

1. Introduction

Health care systems have leveraged Big Data for quality and performance improvements using analytics—the systematic use of data combined with quantitative as well as qualitative analysis to make decisions [11]. Analytics have been utilized in various aspects of health care including predictive risk assessment, clinical decision support, home health monitoring, finance, and resource allocation [3]. Visual analytics is one example of an analytics technique with an array of health care and research applications that are well described in the literature [11].

2. Terminology of Analytics

The jargon surrounding the use of large and varied types of data in healthcare is evolving, but the term analytics is achieving wide use—its systematic use of data combined with statistical and quantitative analysis, explanatory, predictive models, and fact-based management to drive decisions & actions” [1]. IBM defines analytics as “the systematic use of data and related business insights developed through applied analytical disciplines (e.g. statistical, contextual, quantitative, predictive, cognitive, other [including emerging] models) to drive fact-based decision making for planning, management, measurement and learning. Analytics may be descriptive, predictive or prescriptive” [4].

• Descriptive – standard types of reporting that describe current situations and problems
• Predictive – simulation and modeling techniques that identify trends and portend outcomes of actions taken
• Prescriptive – optimizing clinical, financial, and other outcomes

Much work is focusing now on predictive analytics, especially in clinical settings attempting to optimize health and financial outcomes [5].

There are a number of terms related to data analytics. A core methodology in data analytics is machine learning, which is the area of computer science that aims to build systems and algorithms that learn from data [6]. One of the major techniques of machine learning is data mining, which is defined as the processing and modeling of large amounts of data to discover previously unknown patterns or relationships [7]. A subarea of data mining is text mining, which applies data mining techniques to mostly unstructured textual data. Another close but more recent term in the vernacular is big data, which describes large and ever-increasing volumes of data that adhere to the following attributes [1]:

• Volume – ever-increasing amounts
• Velocity – quickly generated
• Variety – many different types
• Veracity – from trustable sources

3. Challenges in Data Analytics

The data generated in the routine care of patients may be limited in its use for analytical purposes. For example, such data may be inaccurate or incomplete. It may be transformed in ways that undermine its meaning (e.g., coding for billing priorities) [8]. It may exhibit the well-known statistical phenomenon of censoring, i.e., the first instance of disease in record may not be when it was first manifested (left censoring) or the data source may not cover a sufficiently long time interval (right censoring) [1]. Data may also incompletely adhere to well-known standards, which makes combining it from different sources more difficult. Finally, clinical data mostly only allows observational and not experimental studies, thus raising issues of cause-and-effect of findings discovered.

Others have noted larger challenges around analytics and big data. With data analytics the research questions asked tend to be driven by what can be answered, as opposed to prospective hypotheses [1]. Data is not always as objective as
one might like, and that “bigger” is not necessarily better. Finally, ethical concerns over how the data of individuals is used, the means by which it is collected, and the possible divide between those who have access to data and those who do not [2]. Similar concerns are associated with myriad of technical, financial, and ethical issues that must be addressed before one will be able to make use of big data routinely for clinical practice and other health-related purposes [3]. These challenges also create ethical issues, such as who owns data and who has privileges to use it.

4. Disrupting Health Care through Big Data and Predictive Analytics

Predictive analytics for crucial areas of need (e.g., disease burden confirmation for risk score accuracy, medication adherence, and monitoring of evidence-based guidelines for prevention screening) determines patient behavior and physician treatment plan harmony whereby the health care gap necessitates a timely care intervention [9]. The patient-specific, gap-specific insight enables the provider to engage in patient dialogue regarding the importance of managing their chronic conditions. Results of the interventions combined with available data should be systematically integrated back into refreshed analytics to evaluate the need for future interventions [3]. Further, the intervention feedback loop creates a continuous and more insightful clinical analytics process that drives improvement and ultimately results in sustainable quality care.

Creating a robust data analytics engine to perform at optimal levels that commands maximum gap closure and drives meaningful changes requires superior data quality and a comprehensive approach to data collection and data management [2, 9]. Massive data sets from traditionally disparate data sources must be integrated into a single system where predictive clinical algorithms may be applied on a continuous basis. Strong data governance, through vigorous data integrity checks, ongoing baseline assessments, and timely data issue resolution are keys to success [9].

5. Conclusion

As the volume of clinical data and the need for analytics continues to accelerate, systematic approaches will be required for sustained success. The growing quantity of clinical and research data, along with methods to analyze and put it to use, can lead to improve personal health, healthcare delivery, and biomedical research. However, there is also a continued need to improve the completeness and quality of data as well as conduct research to demonstrate how to best apply it to solve real world problems. In addition, human expertise, including in informatics, will be required to optimally carry out business intelligence and advanced analytics.

6. References


