Chapter 8
Integration of BI in Healthcare: From Data and Information to Decisions

Xue Ning
University of Colorado, Denver, USA

ABSTRACT

In the digital age, the healthcare industry is generating a huge amount of data and information. Although there are structured data such as EHRs, the major data type is unstructured data such as clinical text. The sources of health data are also diversified, including medical data, clinical data, patient-generated data, and social media data. Different methods are applied to analyze the variety of data and obtain health information. When the various types of information are generated, information retrieval and extraction techniques can be used for further decision-making. Data and information-enabled decision-making is a complex process. Many tools and methods are developed to support decision-making in healthcare. Along with the benefits of integrating business intelligence in healthcare, issues and challenges exist. This chapter discusses the health data and information and how they support decision-making in healthcare.

DOI: 10.4018/978-1-7998-2310-0.ch008
DATA AND INFORMATION IN HEALTHCARE

Big Data and Text Data Analysis

In the big data era, healthcare industry also generates huge amount of data in different formats, including sensor data, biomedical signal and image data, and various text data (Wang, Kung, & Byrd, 2018). Text data in healthcare can be structured format such as part of the EHR text, but most clinical text data is unstructured, for example, free-style doctor notes (Dong et al., 2016). The clinical text refers to reports written by the clinicians for describing patients, their pathologies, personal, social, and medical histories, findings made during interviews or procedures, etc.

It is critical to converting the unstructured clinical text into a structured format for user convenience. The traditional manual coding method is costly and time-consuming. Currently, one of the most popular and powerful methods is natural language processing (NLP). NLP is an interdisciplinary effort between linguistics and computer science domain, which is considered as a subfield of artificial intelligence (AI). The first step of NLP approach in healthcare is generating and understanding the natural language, then creating syntactic and semantic language knowledge on top of domain knowledge. Data mining from unstructured text with NLP approaches can process large volume of clinical text to automatically encode clinical information in a timely manner (Kreimeyer et al., 2017).

NLP methodologies are useful for the clinical text preprocessing such as text segmentation, text irregularities handling, identification of domain specific abbreviation and missing punctuations. More advanced processing includes the application of syntactic and semantic interpreter, algorithms, database handler to convert the clinical text into a format that is compatible for database storage (Friedman & Johnson, 2006).

The NLP methodology consists of some text analysis methods as the core components. For example, morphological analysis, lexical analysis, syntactic analysis, and semantic analysis (Chopra, Prashar, & Sain, 2013). Morphological analysis refers to the process of converting a sentence into a sequence of tokens that are mapped to their canonical base form. For instance, go is the canonical base form of goes and went. Lexical analysis is based on lexicon (i.e., a special dictionary), which provides the necessary rules and data for carrying out linguistic mapping. Syntactic analysis requires grammatical knowledge and parsing techniques. This method can identify the formal relationship between words in the text. The semantic analysis determines the words or phrases in the text that are clinically relevant and extracts their semantic relations. By applying these text analysis techniques, the analyzed data is classified and standardized for database storage.
Related Content

MaxDiff Choice Probability Estimations on Aggregate and Individual Level
[www.igi-global.com/article/maxdiff-choice-probability-estimations-on-aggregate-and-individual-level/192168?camid=4v1a](www.igi-global.com/article/maxdiff-choice-probability-estimations-on-aggregate-and-individual-level/192168?camid=4v1a)

Prototyping
[www.igi-global.com/chapter/prototyping/199664?camid=4v1a](www.igi-global.com/chapter/prototyping/199664?camid=4v1a)

A Data Mining Approach for Risk Assessment in Car Insurance: Evidence from Montenegro
Ljiljana Kašelan, Vladimir Kašelan and Milijana Novovi-Buri (2014). *International Journal of Business Intelligence Research* (pp. 11-28).
[www.igi-global.com/article/a-data-mining-approach-for-risk-assessment-in-car-insurance/122449?camid=4v1a](www.igi-global.com/article/a-data-mining-approach-for-risk-assessment-in-car-insurance/122449?camid=4v1a)
Utility based Tool to Assess Overall Effectiveness of HRD Instruments

www.igi-global.com/article/utility-based-tool-to-assess-overall-effectiveness-of-hrd-instruments/176925?camid=4v1a