Chapter 11
Enhancement of Data Quality in Health Care Industry:
A Promising Data Quality Approach

Asmaa S. Abdo
Menoufia University, Egypt

Rashed K. Salem
Menoufia University, Egypt

Hatem M. Abdul-Kader
Menoufia University, Egypt

ABSTRACT

Ensuring data quality is a growing challenge, particularly when emerging big data applications. This chapter highlights data quality concepts, terminologies, techniques, as well as research issues. Recent studies have shown that databases are often suffered from inconsistent data, which ought to be resolved in the cleaning process. Data mining techniques can play key role for ensuring data quality, which can be reutilized efficiently in data cleaning process. In this chapter, we introduce an approach for dependably generating rules from databases themselves autonomously, in order to detect data inconsistency problems from large databases. The proposed approach employs confidence and lift measures with integrity constraints to guarantee that generated rules are minimal, non-redundant and precise. Since healthcare applications are critical, and managing healthcare environments efficiently results in patient care improvement. The proposed approach is validated against several datasets from healthcare environment. It provides clinicians with automated approach for enhancing quality of electronic medical records. We experimentally demonstrate that the proposed approach achieves significant enhancement over existing approaches.

DOI: 10.4018/978-1-5225-2229-4.ch011

INTRODUCTION

With massive and vast amount of generated data from health care applications domain, which consider a great asset to healthcare organizations in today's economy. The assessment of data extremely depends on its degree of quality (Saha et al., 2014).

Quality is applied to data by meaning is it "fitness for use" or "potential for use" (Martín et al., 2010). Data quality is essential characteristic that determines the reliability of data for data management purposes (Wang et al., 2014). The quality of data is an increasingly pervasive problem, as data in real world databases quickly degenerates over time and effects the results of the mining. This result in what is called "Dirty Data" (Chiang et al., 2008; Li et al., 2014). Such dirty data often emerges due to violations of integrity constraints, which results in incorrect statistics, and ultimately wasting of time and money (Yakout et al., 2010). It has been estimated that erroneous data leading to lose billions of dollars annually due to poor data quality in decision making that negatively effect on achieving customer satisfaction (Fan et al., 2012; Yakout et al., 2010). As a result, detecting inconsistent data is very important task in the data cleaning process. Doubtless, ensuring high quality dependable data is a competitive advantage to all businesses, which requires accurate data cleaning solutions (Fan et al., 2012; Wang et al., 2014).

We need to verify four attributes of data quality as shown in Figure 1. Attributes of data quality is detailed as follows: Complete, Accurate, Available and Timely. Completeness of data denote as all related data about one entity is linked. Accuracy refer to data free from common problems such spelling mistake, typographical error, and random abbreviations. Availability of data mean required data accessible on demand that make customer not need to search manually for information. Timely consider data up to dated available for management purposes.

Indeed, to ensure quality of data there is need to data cleaning. Data cleaning also called (data cleansing or scrubbing) refers to the process of maintaining corrupted and/or inaccurate records in order to enhancing the quality of data. This process is mandatory in data management cycle before mining and analyzing data (Mezzanzanica et al., 2013).

A manual process of data cleansing is also exhausting process, time consuming, and itself prone to errors (Li et al., 2014). Thus, make us search for automate solutions as powerful tools, which automate or greatly assist in the data cleansing process to achieve a reasonable quality level in existing data. As this

Figure 1. Four attributes of data quality