PERGAMON

Expert Systems with Applications 24 (2003) 167–172

Expert Systems with Applications

www.elsevier.com/locate/eswa

Analysis of healthcare quality indicator using data mining and decision support system

Young M. Chae[a,*], Hye S. Kim[a], Kwan C. Tark[b], Hyun J. Park[b], Seung H. Ho[a]

[a]*Graduate School of Health Science and Management, Yonsei University, C.P.O. Box 8044, Seoul, South Korea*

[b]*Department of Quality Improvement, Severance Hospital, Yonsei University Medical Center, C.P.O. Box 8044, Seoul, South Korea*

Abstract

This study presents an analysis of healthcare quality indicators using data mining for developing quality improvement strategies. Specifically, important factors influencing the inpatient mortality were identified using a decision tree method for data mining based on 8405 patients who were discharged from the study hospital during the period of December 1, 2000 and January 31, 2001. Important factors for the inpatient mortality were length of stay, disease classes, discharge departments, and age groups. The optimum range of target group in inpatient healthcare quality indicators were identified from the gains chart. In addition, a decision support system (DSS) was developed to analyze and monitor trends of quality indicators using Visual Basic 6.0. Guidelines and tutorial for quality improvement activities were also included in the system. In the future, other quality indicators should be analyzed to effectively support a hospital-wide continuous quality improvement (CQI) activity and the DSS should be well integrated with the hospital order communication system (OCS) to support concurrent review.

Keywords: Continuous quality improvement; Data mining; Decision support system

1. Introduction

An increasing concern with improving the quality of care in various components of the health care system has led to the adoption of quality improvement approaches originally developed for industry. These include 'total quality management' (TQM) (Deming, 1986), an approach which employs process control measures to ensure attainment of defined quality standards, and 'continuous quality improvement' (CQI) (Juran, 1988), a strategy to engage all personnel in an organization in continuously improving quality of products and services.

CQI was originally based on the quality assurance (QA) paradigm, which emphasizes monitoring of incidents, mortality and morbidity audits, and hospital infection audits. However, manufacturing industry experiences have shown that QA programs, which focus on end-product evaluation/audit, have little effect on improving quality or decreasing costs (Sakofsky, 1996). In manufacturing industries, process quality improvement strategies have been proven to be far effective than product oriented quality control programs. At the beginning of the nineties, the emphasis was shifted from the QA paradigm to that of process oriented TQM/CQI, concurrently with the realization of the advantages of the latter throughout the industry.

Process improvement strategies operationalize the plan-do-check-act (PDCA) process quality management cycle (Deming, 1986). The outcome targets from the continuous and final QA criteria to be used at quality evaluation-and-improvement checkpoints. The key inputs to the PDCA process are patient assessment/outcome data that are compared to the expected outcome targets and best practice guidelines or protocols. Each set of evaluation results can be used as part of the decision support information for revising the care plan and improving the intervention strategies.

In Korea, QA activity has been launched in 1981 as a part of the Hospital Standardization Project organized by the Korean Hospital Association. Since the Korean Society of Healthcare QA was established in 1994, more comprehensive quality management, evaluation, and research have been implemented. The Hospital Service Evaluation System began in 1995 for an evaluation of CQI activities at the tertiary hospitals initially, but was later expanded to the hospitals with less than 200 beds. Recently, for a more systematic and practical evaluation of hospital quality, researches on reconceptualization of CQI, development of QI standard and QI indicators, establishment of a QI

* Corresponding author. Tel.: +82-2-361-5048; fax: +82-2-392-8996/7734.

E-mail address: ymchae@yumc.yonsei.ac.kr (Y.M. Chae).

0957-4174/03/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0957-4174(02)00139-2

department, and development of a QI manual are actively in progress.

But because of inadequate utilization of QI evaluation results and feedback, heavy workloads, and lack of motivation of this endeavor, CQI has not been successfully implemented in most hospitals in Korea. Moreover, the majority of QI activities heavily relied on manual processes such as chart audit. However, manual QI activity without its connection to underlying clinical information produced by hospital information system had been criticized as contributing nothing to quality improvement (Luttman, 1993). Therefore, there is a need for a decision support system (DSS) that provides patient assessment/outcome information and a clinical pathway to support the PDCA process.

For process quality improvement to be successfully implemented, information on patient care process and the factors influencing quality or treatment outcome must be available at real time for comparison against the desired progress/outcome criteria and development of quality improvement strategies by integrating with the hospital information system. In this study, the factors influencing quality were identified using data mining, and a DSS for process-oriented CQI based on these factors is another key information for the PDCA process. Data mining is a knowledge discovery method from a large-scale information bank such as a data warehouse. Data mining was used in this study in order to identify pattern or rules about various quality problems or indicators from a large-scale data warehouse. While there were several studies on data mining such as identifying significant factors influencing prenatal care (Prather et al., 1997) and automatic detection of hereditary syndromes (Evans, Lemon, Deters, Fusaro, & Lynch, 1997), these systems did not explicitly deal with management issues on CQI activities.

2. Methods

2.1. Subjects and scope

The subjects were 8405 patients who were discharged from the study hospital during the period of December 1, 2000 and January 31, 2001. Of several quality indicators used in the study hospital, this study focused on the inpatient mortality for the decision tree analysis of the influencing factors for quality. Patient characteristics such as age, sex, discharge department, disease classes, and quality indicators were used in the analysis.

2.2. Methods

The decision tree was used in the analysis of the factors influencing inpatient mortality. Decision trees are known as effective classifiers in a variety of domains. In our example, the decision tree categorizes the entire subjects according to whether or not they are likely to have inpatient mortality. Most of the decision tree algorithms use a standard top-down approach to building trees. Chi-squared automatic interaction detection (CHAID) and C5.0 are two popular decision tree inducers, based on the ID3 classification algorithm by Quinlan (1993).

A CHAID tree is a decision tree that is constructed by splitting subsets of the space into two or more child nodes repeatedly, beginning with the entire data set. To determine the best split at any node, any allowable pair of categories of the predictor variables is merged until there is no statistically significant difference within the pair with respect to the target variable. This process is repeated until no insignificant pair is found. The resulting set of categories of the predictor variable is the best split with respect to that predictor variable. In this paper, the CHAID algorithm with growing criteria of the likelihood ratio chi-square statistic was used for building the tree and evaluating splits because most of our variables were ordinal and discrete continuous variables. To identify nodes of interest (that is, nodes with a relatively high probability), a gains chart was used. The gains chart shows the nodes sorted by the number of cases in the target category for each node.

3. Results

3.1. Characteristics of subjects

Among the 8405 patients, 4451 (53.0%) were male and 3954 (47.0%) were female. Patients who were discharged from Internal medicine departments were almost three times (6109) more than those from the surgery departments (2296). Patients in the age group of 41–60 had the highest proportion (31.3%). Among all disease classes, neoplasm had the highest proportion (28.8%). Disease classes with the proportion of less than 5% were grouped under miscellaneous. Complete descriptive statistics for the modifiable risk factors are shown in Table 1.

3.2. Decision tree analysis by CHAID algorithm

The decision tree for inpatient mortality had 17 statistically significant nodes at 5% level (Fig. 1). Among 8405 patients, 170 (2.0%) were inpatient mortality cases. The most significant factor explaining the infant mortality was length of stay (LOS). Mortality rate for the patients with LOS longer than 16 days (6.4%) were almost six times higher than the other two LOS groups. Discharge departments were the next significant factors, followed by the age groups.

Each node depicted in the decision tree can be expressed in terms of an 'if-then' rule, as follows

/*Node 16*/
If $(17 < \text{LOS} < 341$ and Discharge department = Rheumatics Medicine and Age $> 61)$, then inpatient mortality = 23.4%.