

PRINCETON UNIVERSITY
OPERATIONS RESEARCH & FINANCIAL ENGINEERING SEMINAR

PREDICTION OF HEALTH CARE COSTS VIA DATA-MINING AND ALGORITHMIC DISCOVERY OF MEDICAL KNOWLEDGE

DIMITRIS BERTSIMAS
MIT

Abstract

Rising health care costs are one of the world's most important problems. Correspondingly, predicting such costs with accuracy is a significant first step in addressing this problem. Since the 1980s, there have been research efforts for predictive modeling of medical costs based on claims data that utilize heuristic rules and classical regression methods that have not been appropriately validated in populations that the methods have not seen. In this study, we utilize modern data mining methods, specifically classification trees and clustering algorithms, and claims data from close to four hundred thousand members over three years to provide a) predictions of health care costs in the third year, based on medical and cost data from the first two years, which we rigorously validate, and b) an illustration through two examples that our methods can lead to discovery of medical knowledge. We quantify the accuracy of our predictions using out of sample data from over one hundred thousand members. The key insights we obtain are: a) our data mining methods provide accurate predictions of medical costs and represent a powerful tool for prediction of health care costs, b) the pattern of past cost data are strong predictors of future costs, c) medical information is an accurate prediction of medical costs particularly on high risk members, and d) new medical knowledge can be obtained through our methods.

Tuesday, February 12, 2008
E-219, E-Quad
4:30 PM

