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.

BIO: Dimitris Bertsimas is currently the Boeing Professor of Operations Research and the co-director of the Operations Research Center at the Massachusetts Institute of Technology. He has received a BS in Electrical Engineering and Computer Science at the National Technical University of Athens, Greece in 1985, a MS in Operations Research at MIT in 1987, and a Ph.D in Applied Mathematics and Operations Research at MIT in 1988. Since 1988, he has been with MIT's Sloan School of Management. His research interests include optimization, stochastic systems, data mining and their application. He is a member of the National Academy of Engineering, and he has received numerous research awards including the Erlang prize (1996), the SIAM prize in optimization (1996), the Bodossaki prize (1998) and the Presidential Young Investigator award (1991-1996).