



Housing Price Prediction Using Search Engine Query Data

Qian Dong*

National Bureau of Statistics of China , Beijing, P.R.China– queenad@hotmail.com

Abstract

The real estate industry is one of the economics drivers of the Chinese economy, and the housing price has been earning constant attention ever since. But the data published by government statistical agencies are usually delayed, thus cannot fulfill the public demand. This article provides an optimal modal of predicting the price trends in new and secondary housing market in eight cities including those in Beijing, Shanghai, Tianjin, Chongqing, and some other relatively developed cities in China. Based on the Baidu Search Index (BSI), we picked 12 keywords that influence the second-hand housing price most and 8 keywords that influence the price of new houses most. With the cross-validation technique, we fitted and forecasted the housing prices in both markets by using 6 analytical models including Linear Regression, Regression Tree, Random Forests, Bagging, m-Boosting and Support Vector Machine (SVM). Among the six models we used, the SVM and Random Forest models predicted the best, while the Regression Tree model predicted the worst one. Most of the public attention is on the aspects of the transaction and housing policy among the key factors that influence the price of secondary housing; while the price trends and real estate policy are the focus of public attention among the key factors that influence the price of new houses. We concluded that the data collected through website searching could not only predict the housing prices, but it could also derive some specific patterns and trends of economic behavior among major society. Besides, this prediction model is highly timely since it could predict the price trends of the real estate industry two weeks prior to the data published by official statistic agencies.

Key words: Searching Data; Real Estate Price Prediction; Cross Validation; SVM; Random Forest