World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:11, No:02, 2017

Optimal Pricing Based on Real Estate Demand Data

Authors : Vanessa Kummer, Maik Meusel

Abstract : Real estate demand estimates are typically derived from transaction data. However, in regions with excess demand, transactions are driven by supply and therefore do not indicate what people are actually looking for. To estimate the demand for housing in Switzerland, search subscriptions from all important Swiss real estate platforms are used. These data do, however, suffer from missing information—for example, many users do not specify how many rooms they would like or what price they would be willing to pay. In economic analyses, it is often the case that only complete data is used. Usually, however, the proportion of complete data is rather small which leads to most information being neglected. Also, the data might have a strong distortion if it is complete. In addition, the reason that data is missing might itself also contain information, which is however ignored with that approach. An interesting issue is, therefore, if for economic analyses such as the one at hand, there is an added value by using the whole data set with the imputed missing values compared to using the usually small percentage of complete data (baseline). Also, it is interesting to see how different algorithms affect that result. The imputation of the missing data is done using unsupervised learning. Out of the numerous unsupervised learning approaches, the most common ones, such as clustering, principal component analysis, or neural networks techniques are applied. By training the model iteratively on the imputed data and, thereby, including the information of all data into the model, the distortion of the first training set—the complete data—vanishes. In a next step, the performances of the algorithms are measured. This is done by randomly creating missing values in subsets of the data, estimating those values with the relevant algorithms and several parameter combinations, and comparing the estimates to the actual data. After having found the optimal parameter set for each algorithm, the missing values are being imputed. Using the resulting data sets, the next step is to estimate the willingness to pay for real estate. This is done by fitting price distributions for real estate properties with certain characteristics, such as the region or the number of rooms. Based on these distributions, survival functions are computed to obtain the functional relationship between characteristics and selling probabilities. Comparing the survival functions shows that estimates which are based on imputed data sets do not differ significantly from each other; however, the demand estimate that is derived from the baseline data does. This indicates that the baseline data set does not include all available information and is therefore not representative for the entire sample. Also, demand estimates derived from the whole data set are much more accurate than the baseline estimation. Thus, in order to obtain optimal results, it is important to make use of all available data, even though it involves additional procedures such as data imputation.

Keywords : demand estimate, missing-data imputation, real estate, unsupervised learning
Conference Title : ICMLDA 2017 : International Conference on Machine Learning and Data Analysis
Conference Location : Paris, France
Conference Dates : February 23-24, 2017