CSE 847: Machine Learning—Project Proposal

An Exploration and Implementation of Automated Valuation Models to Learn and Predict the Value of Real Estate

Mick Langford
langfo37@msu.edu

Mike Lingg
linggmic@egr.msu.edu

Jordi Lucero
luceroj2@msu.edu

February 17, 2017

1 Problem Description

Automated Valuation Models (AVM) have become increasingly popular as the real estate market has embraced the World Wide Web.[2] Banks have also shown great interest in using AVMs to help mitigate fraud by human appraisal.[3] Our goal is to explore various machine learning techniques to implement an AVM and predict the true value of a house based on features commonly found on real estate listings. Our data will be drawn from the Nashville, TN housing market, using a dataset posted on Kaggle[1].

We will begin by exploring linear regression models that take into account physical attributes of each house and location, but we will also explore nonlinear models, such as deep learning with neural networks and decision trees .

2 Related Work

An obvious and popular example is Zillow's proprietary Zestimate®. Zillow uses a closed source AVM that takes into account special features of the home, location, and market conditions. Zillow admits to using features such as physical attributes, tax assessments, and both prior transactions. Zillow claims to have data on 110 million homes and estimates on approximately 100 million homes.[4]

Relevant papers include the doctoral dissertation of Lowrance which explores and compares various linear models on housing data for the Los Angeles County.[5] Park and Bae explore machine learning algorithms such as C4.5, RIPPER, Nave Bayesian, and AdaBoost.[6] Bin performed a study that estimates a hedonic price function using a semi-parametric regression.[7] This may be particularly useful for real estate listings that are incomplete or for data that is entered erroneously. Bourassa et al. consider the spatial dependence of house prices, which is intuitively an important factor.[8][9] Kauko et al. research neural network models to help investigate segmentation in the housing market of Helsinki, Finland.[10] Azadeh et al. present an algorithm based on fuzzy linear regression and a fuzzy cognitive map to handle uncertainty in the housing market and improve the analysis of housing price fluctuations.[11] Fan et al. introduce a decision tree approach for modeling and predicting house prices.[12]

3 Project Milestones

To do.

References

[1] *Nashville Housing Data: Home value data for the booming Nashville Market* Retrieved from `https://www.kaggle.com/tmthyjames/nashville-housing-data/`

[2] *Data-driven property valuations: the real deal?* Retrieved from `http://blog.kaggle.com/2010/06/21/data-inc-are-avms-soothsayers-or-the-real-deal/`

[3] Schroeder, Steve. *Fighting Fraud: A combination of collateral assessment and AVMs can maximize mortgage-fraud management* Retrieved from `http://www.scotsmanguide.com/Residential/Articles/2005/10/Fighting-Fraud/`

[4] *What is a Zestimate? Zillow's Home Value Forecast.* Retrieved from `http://www.zillow.com/zestimate/`

[5] Lowrance, R. E. (2015). *Predicting the Market Value of Single-Family Residential Real Estate* (Doctoral Dissertation). New York University. Retrieved from `http://gradworks.umi.com/36/85/3685886.html`

[6] Park, B., & Bae, J. K. (2015). *Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data.* Expert Systems with Applications, 42(6), 2928-2934. doi:10.1016/j.eswa.2014.11.040

[7] Bin, O. (2004). *A prediction comparison of housing sales prices by parametric versus semi-parametric regressions.* Journal of Housing Economics, 13(1), 68-84. doi:10.1016/j.jhe.2004.01.001

[8] Bourassa, S. C., Cantoni, E., & Hoesli, M. (2010). *Predicting House Prices with Spatial Dependence: Impacts of Alternative Submarket Definitions.* SSRN Electronic Journal. doi:10.2139/ssrn.1090147

[9] Bourassa, S. C., Cantoni, E., & Hoesli, M. (2007). *Spatial Dependence, Housing Submarkets, and House Prices.* SSRN Electronic Journal. doi:10.2139/ssrn.771867

[10] Kauko, T., Hooimeijer, P., & Hakfoort, J. (2002). *Capturing Housing Market Segmentation: An Alternative Approach based on Neural Network Modelling.* Housing Studies, 17(6), 875-894. doi:10.1080/02673030215999

[11] Azadeh, A., Ziaei, B., & Moghaddam, M. (2012). *A hybrid fuzzy regression-fuzzy cognitive map algorithm for forecasting and optimization of housing market fluctuations.* Expert Systems with Applications, 39(1), 298-315. doi:10.1016/j.eswa.2011.07.020

[12] Fan, G., Ong, S. E., & Koh, H. C. (2006). *Determinants of House Price: A Decision Tree Approach.* Urban Studies, 43(12), 2301-2315. doi:10.1080/00420980600990928