Title: Machine Intelligence For Housing Finance

Presented By: Prof. Kay Giesecke

Abstract

The housing market is massive yet the housing finance technology is stuck in the 20th century. For example, 2000 pages per application are used, the process is slow and takes around 45 days and is very costly. Further, plenty of bad decisions are made in terms of grid-based underwriting. This talk described how machine intelligence can transform the housing market by providing a means to replace the grid and also generate societal benefits. The key idea is to leverage data to understand borrower behavior. The results demonstrate the efficacy of this idea in better predictions to avoid foreclosures. Using deep learning, significant betterment in the prediction performance has been shown.

Motivation

The housing market is massive yet the housing finance technology is stuck in the 20th century. For example, 2000 pages per application are used, the process is slow and takes around 45 days and is very costly. Further, plenty of bad decisions are made in terms of grid-based underwriting.

Approach

This talk described how machine intelligence can transform the housing market by providing a means to replace the grid and also generate societal benefits. The key idea was to leverage data to understand borrower behavior. Using deep learning, significant betterment in the prediction performance was shown. All the stakeholders were identified ranging from lenders originators, insurers, services and regulators.

Methodology

The approach followed was machine learning accessing the unprecedented amounts of data that are now available. There was data of around 200 million loans in the 1994-2015 time window from sources such as Freddie Mac and Fannie Mae. There was data of property level sales transactions and tax records in the time window 2000-2014 (approximately 100 million sales). Further, there was zip code level data on the economic, financial and demographic attributes of the participants of the financial transactions. In addition to these high level data, there was also more granular, borrower level transaction data such as credit history, social media activity, and property level images.

The transaction was modeled as markov decision process with the states being current, 30 days late, 60 days late, 90+ days late, foreclosure, REO, and paid off. Based on the data from 120 million loans, a transition matrix was formed using the frequentist approach of maximum likelihood estimation.

However, representing such a massive transition matrix is intractable and difficult to work with to do predictive modeling. Therefore, a deep neural network was trained to map inputs to the conditional state transition probabilities. These inputs were around 300 risk factors that were identified based on the collected data. The key idea here was to find a relative importance ranking between different risk factors with the goal of predicting delinquency and defaulting. A deep neural network was the representation of choice because of the ease of capturing non-linearities in the data that could otherwise not be modeled.

Results

The results demonstrate the efficacy of this idea in better predictions. Investment portfolio was designed. Predictions were used to address feature volatility. Delinquency rate was fitted. The regulatory agencies was explanations for all decisions there significant work was performed on assessing the statistical significance in terms of risk factor importance and ranking.

The most interesting out of these ideas was the assessment of explainability in terms sensitivity analysis. This led to risk factor importance ranking in terms of which features that are input to the neural network contribute the most to the terms in the transition distribution matrix. As expected, this methodology successfully extracted some of the obvious indicators of risk such as FICO score, it also yielded some surprising results. For instance, zillow zip code housing price change since origination was an important factor in determining the risk. Another such example is the state unemployment rate. Using these importance factors of risk, regulators will now be more likely to adopt this methodology in the housing market since decisions made can be better explained based on the risk importance factors.

Future Directions

This was only the tip of the iceberg in terms of potential applications of these methodologies. Some other domains to apply these ideas would be student loans, credit cards, and SME credit.

References

[1] Sirignano, Justin, Apaar Sadhwani, and Kay Giesecke. "Deep learning for mortgage risk." *arXiv preprint arXiv:1607.02470* (2016).
[2] Horel, E., Mison, V., Xiong, T., Giesecke, K., & Mangu, L. “Sensitivity based Neural Networks Explanations.” *arXiv preprint arXiv:1812.01029* (2018).
[3] Horel, Enguerrand, and Kay Giesecke. "Towards explainable ai: Significance tests for neural networks." *arXiv preprint arXiv:1902.06021* (2019).