Smart Grid: Something about electricity markets and predicting energy use so we can reduce waste

Paragraph overview - kind of like an abstract

Related Work

Aim of this Research

“Quote” - probably related to motivation

A one-liner to finish off.

References

“Quote” - probably related to motivation

Reference Image

Industry Links and Research Impact

Statistic
Smart Grid: Something about electricity markets and predicting energy use so we can reduce waste

Paragraph overview - kind of like an abstract

Related Work

“Quote” - probably related to motivation

Aim of this Research

“Quote” - probably related to motivation

References

Industry Links and Research Impact

Statistic

Photo Caption
**Motivations and Challenges**

The primary aim of real-estate value prediction differs significantly from one stakeholder to another: (1) A developer is looking to maximise their returns; (2) Lenders need to be aware of the risk of default; (3) Home buyers have non-return related priorities, such as lifestyle suitability and location. Varying motives, macro-environmental factors and data sparsity are just some of the reasons why real-estate valuation remains a challenging task. Additionally, unprecedented leverage is reported to be so valuable.

Some of the reasons why real-estate value prediction can be so valuable:

1. **Home buyers**: have non-return related priorities, such as lifestyle suitability and location. 
2. **Lenders**: need to be aware of the risk of default. 
3. **Developers**: are looking to maximise their returns.

**Background Reading**

From the perspective of lagged returns, the most popular valuation methodologies employ constant-quality indices, notably repeat sale regression (RSR) [1]. In a comparison study of methodologies, the RSR algorithm was tested on 53,000 properties in and around Chicago and only gained an R² of 0.34 [2]. [3] introduced a RIPPER regression on five spatial regression models with an R² of 0.34. [4] introduced a RIPPER regression on 5,359 properties in and around Chicago and only gained an R² of 0.34 [2]. [5] introduced a RIPPER regression on 5,359 properties in and around Chicago and only gained an R² of 0.34 [2]. [6] introduced a RIPPER regression on 5,359 properties in and around Chicago and only gained an R² of 0.34 [2]. [7] introduced a RIPPER regression on 5,359 properties in and around Chicago and only gained an R² of 0.34 [2].

**Methodology**

Our method introduces a novel, four-stage, methodology for real-estate valuation (see figure 1):

1. **Stage 1 (Temporal Interpolation)**: A space-time interpolation was put forward to provide a time singular dataset (Dt). The mean value of each area was calculated and then extended on each property in the land registry’s sales dataset. The interpolation was tested on yielding an R² value of 0.71.

2. **Stage 2 (Spatial Dependency Identification)**: Universal kriging, an interpolation based on Gaussian processes set by some prior covariance function was utilized. This method uniquely assumes non stationarity. UnK considers the spatial correlation between the points that need to be interpolated and their neighbouring points [12]. Four covariance functions (kernels) were tested: Epanechnikov, Gaussian, Polynomial and Exponential. The best performing employed a fifth order polynomial kernel function with an R² of 0.839.

3. **Stage 3 (property, network and economic features)**: Manual feature selection was undertaken. The remaining features include building size and height, title size, property type (detached, terraced, apartment, etc.), freehold status and old/new build status, proximity to schools and train stations, traffic flow, population density, variable mortgage interest rates, total number of houses sold each month, inflation and GBP-USD exchange rate.

4. **Stage 4 (Gaussian Process Regression)**: All of the previous findings were merged into a single dataset: and then a Gaussian Process (GP) was trained. A Gaussian Process (GP) is a powerful non-parametric Bayesian model, specified by a mean and a covariance (kernel) function. 16,000 location-stratified instances were trained. The covariance function (kernel) chosen was a "Radial Basis" (Gaussian) function.

**Stage 1 (Temporal Interpolation)**: A space-time interpolation was put forward to provide a time singular dataset (Dt). The mean value of each area was calculated and then extended on each property in the land registry’s sales dataset. The interpolation was tested on yielding an R² value of 0.71.

**Stage 2 (Spatial Dependency Identification)**: Universal kriging, an interpolation based on Gaussian processes set by some prior covariance function was utilised. This method uniquely assumes non stationarity. UnK considers the spatial correlation between the points that need to be interpolated and their neighbouring points [12]. Four covariance functions (kernels) were tested: Epanechnikov, Gaussian, Polynomial and Exponential. The best performing employed a fifth order polynomial kernel function with an R² of 0.839.

**Stage 3 (property, network and economic features)**: Manual feature selection was undertaken. The remaining features include building size and height, title size, property type (detached, terraced, apartment, etc.), freehold status and old/new build status, proximity to schools and train stations, traffic flow, population density, variable mortgage interest rates, total number of houses sold each month, inflation and GBP-USD exchange rate.

**Stage 4 (Gaussian Process Regression)**: All of the previous findings were merged into a single dataset: and then a Gaussian Process (GP) was trained. A Gaussian Process (GP) is a powerful non-parametric Bayesian model, specified by a mean and a covariance (kernel) function. 16,000 location-stratified instances were trained. The covariance function (kernel) chosen was a "Radial Basis" (Gaussian) function.

**Results**

For comparison, two performance metrics were applied at each stage: Coefficient of Determination (R²) and Root Mean Squared Error (RMSE) providing an absolute and relative measure for validation. Table 1 shows the results for each stage in the analytics pipeline. Figure 2 shows the actual versus predicted values of a small set of samples. In the final two stages, a ten-fold stratified sampling technique was implemented and the average result for each fold was calculated; the standard deviation between each fold was 4.483. Figure 2 visualises the GPR’s prediction versus actual price for all properties trained and tested on. The models t-value and p-value were reported to be 27.9178 and ≤ 2.2e−16 respectively.

**Impact**

NimbusMaps embeds the techniques outlined in this paper. The interface is powered by GoogleMaps with polygon overlays. A customer is able to search by postcode, current location or title number. A title number selection produces information on ownership, site size, number of buildings, flood risk, estimated residential value and traffic flow are returned.

**References**

A Spatio-Temporal, Gaussian Process Regression, Real-Estate Price Predictor

Paragraph overview - kind of like an abstract

Related Work

"Quote" - probably related to motivation

Aim of this Research

"Quote" - probably related to motivation

A one-liner to finish off.

References

"Quote" - probably related to motivation

Industry Links and Research Impact

Statistic

Photo Caption