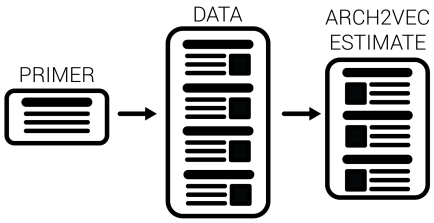
John Scheeler, jas875
Jordan Stout, jds459
Nov 21, 2017
CS 5306

Crowdsourcing A Better Estimate

Crowdsourcing Effort

For this project we explored the creation of a data set for training a Real Estate evaluation neural network tool for improved recommendation services and price evaluations. This will allow for immediate classification on an online tool to augment the data available on traditional real estate websites. We used crowdsourcing as our primary tool for dataset collections and evaluation training. To propagate the information needed for populating our crowdsourcing effort we scraped existing sources. This project will build on Jordan's advanced knowledge in machine learning with John's studies in architecture and real estate as a B.Arch.



Key Questions

Based on a variable size set of factors, how plausible is a human intelligence based, real estate research model? Currently real estate research and appraisals are done by individuals working to price houses based on a set of factors. These numbers are often used as a baseline for price, but very rarely shared publicly. This allows an owner of property to price independently of appraised value. This can lead to markets in which some houses are inaccurately priced if the owner is willing to spend more time on the market.

To purchase these houses, one must then determine what they are willing to pay for a property based on other comparable properties on the market and real estate cost estimate tools, such as Zillows Zestimate. We propose a solution where one is able to use the "wisdom of the crowd" to develop cost effective estimates, with higher fidelity, responsiveness to market fluctuations, and accurate comparative properties. This is something that current models, such as Zillow's Zestimate, are known to struggle with.

Zillow is aware of the problems in their algorithm and as of May of this year, offered up a one million dollar prize to the team that increased the accuracy of the algorithm most significantly.[1]

To help better understand common fallacies in pricing estimates, such as Zestimate, and other problems inherent to real estate market encounters we referred to home buying forums such as the financial Stack Exchange site and r/RealEstate. Users described being able to come up with more accurate costs by focusing on recently sold comparable houses in the region of interest. These metrics must be analyzed to compare amenities, area, and location, factors which in combination increase the difficulty of the problem significantly.

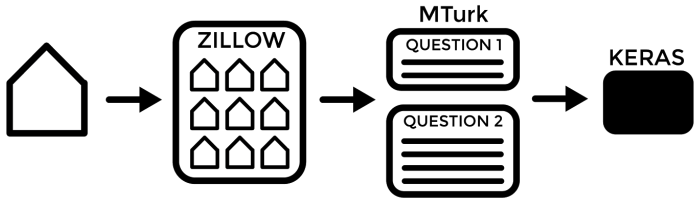
One user documented their process of creating and aggregating comparable properties to create a better estimate than a Zestimate, which is where our work began. The user utilized Zillow's recommendation tool to find houses they deemed comparable, eliminating many of the commonly used metrics like cost and images, and focused strictly on known features (the specifics of which were excluded from the article). This information was then re-combined with price and square footage information to calculate an allowable price per square foot, which was then used to calculate a final estimate for the target house.

Data Needed

There will be two primary data needs for this project. The first being real estate data gathered by scraping both existing data sets and real estate websites. The second being training data for our neural network gathered by crowdsourcing price estimates for existing home listings. This will allow us to check the guesses against existing data and better train our users to create better guesses.

Once that data is gathered, it can be placed into our “Arch2Vec” model, which will break down a house into its composite factors (style, size, location, year built etc.) and then generate an approximate price point. This can be done fairly simply using multi-dimensional geometry to generate clusters of similar vectors. Based on distance from known price points we can then generate the approximation.

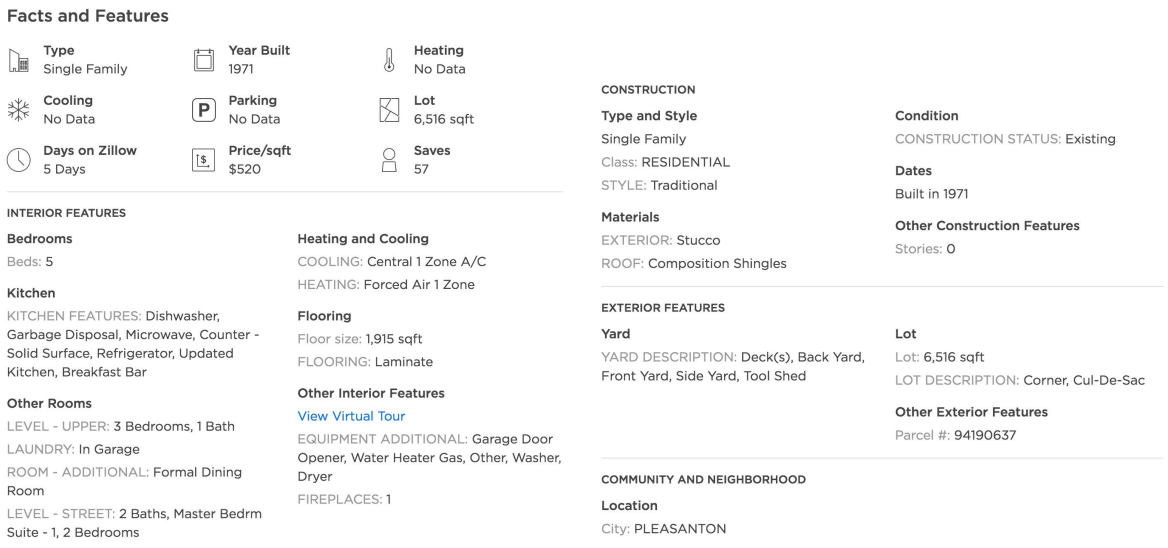
[1]https://www.housingwire.com/articles/40206-have-issues-with-the-zestimate-zillow-is-offering-1-million-to-fix-it



A test set for our crowdsourcing data can be gathered from the same site that we use to gather the other information which will have current house listing prices. Known price points and Zillow Zestimates will provide accuracy and comparables data.

Data Collection

To begin our data collection we need a way of taking information from existing properties listed on the market to help us create our MTurk tasks which would help judge the validity as properties as comparable properties. A simple Zillow scrape allowed us to enter a property based on a specific address or a more generic postal zip code, gain a list of properties nearby, and pull all associated features and amenities.



Facts and Features

Type Single Family
Year Built 1971
Heating No Data
Cooling No Data
Parking No Data
Lot 6,516 sqft
Days on Zillow 5 Days
Price/sqft $520
Saves 57

INTERIOR FEATURES

Bedrooms
Beds: 5

Kitchen
KITCHEN FEATURES: Dishwasher, Garbage Disposal, Microwave, Counter - Solid Surface, Refrigerator, Updated Kitchen, Breakfast Bar

Other Rooms
LEVEL - UPPER: 3 Bedrooms, 1 Bath
LAUNDRY: In Garage
ROOM - ADDITIONAL: Formal Dining Room
LEVEL - STREET: 2 Baths, Master Bedrm Suite - 1, 2 Bedrooms

Heating and Cooling
COOLING: Central 1 Zone A/C
HEATING: Forced Air 1 Zone

Flooring
Floor size: 1,915 sqft
FLOORING: Laminate

Other Interior Features
View Virtual Tour
EQUIPMENT ADDITIONAL: Garage Door Opener, Water Heater Gas, Other, Washer, Dryer
FIREPLACES: 1

CONSTRUCTION

Type and Style
Single Family
Class: RESIDENTIAL
STYLE: Traditional

Materials
EXTERIOR: Stucco
ROOF: Composition Shingles

Condition
CONSTRUCTION STATUS: Existing

Dates
Built in 1971

Other Construction Features
Stories: 0

EXTERIOR FEATURES

Yard
YARD DESCRIPTION: Deck(s), Back Yard, Front Yard, Side Yard, Tool Shed

Lot
Lot: 6,516 sqft
LOT DESCRIPTION: Corner, Cul-De-Sac

Other Exterior Features
Parcel #: 94190637

COMMUNITY AND NEIGHBORHOOD

Location
City: PLEASANTON

To compensate for different housing density based on zip code we also enabled modified geographic fence sizes based on the density of results in the given area to create a similarly sized list of properties.

The first MTurk task created, which served as our training and workforce verification, provided participants with one adress and 10 comparable houses pulled from our initial zillow scrape. To narrow down the initial output of listings to just 10, we manually conducted our own analysis and created a ranking of the results fitment. The turkers were presented with a generic ID for each property to prevent them from doing additional research beyond the metrics provided. For each listing the Turkers were given only information found under zillow's “Facts and features” with images, price, and price per square foot removed to limit visual bias and problems with individual listing prices.

The task then asked each Turker to identify the 3 best comparable properties in no particular order for the sample address provided. Turkers were informed that they would be paid 1 cent for completing the task and 2 cents for every correct answer. This meant that the total possible reward would be 10 cents. We filled 145 hits, with 117 of our initial 145 workers completing the task with at least 2 out of 3 of our correct answers, qualifying them to be able to complete our next task which would serve as our data collection question.

The secondary question cycled through a series of 5 different zip codes, all of which we are personally familiar with and selected for variation in housing markets to better understand how our model would handle different conditions and markets ranging from the 2-5 million range down to 200k-400k. Within these different markets there was also variation in how similar the comparable houses nearby were. To generalize, city suburbs tended to have houses of very similar size and price, while small towns had much larger fluctuations in prices between homes. At random, one result from a selected zip code was picked for each task, and subsequently a list of similar properties was generated.

Out of our 117 approved Turkers, 78 chose to return for a second task. These participants were primed with the same payment information as in our initial study, with the exception that this time a known set of correct answers did not exist. To solve this, we payed all participants as if they answered every question correctly. The participants were then divided into two groups, with the first being asked to identify the 5 best fitting properties of a set of 15, and the second being asked to identify the 10 best from a sample of 20. This deviation between the two groups allowed us to do an initial exploration into the impact of the number of comparable houses used and the effect of increasing the pool of selectable comparable properties.

Once the task was completed, workers were paid and not allowed to return into our HIT pool to minimize workers taking advantage of our inability to better vet the correctness of their answers. This model would prove to be unsuitable if implement on a larger scale where the Turker marketplace wasn't large enough to satisfy our data collection needs without repetition; however,

for this particular test we only needed a relatively small training set of 30+ entries. Our 78 returning workers divided into two groups easily satisfied this requirement, creating a training set with 37 entries and 41 entries respectively, which we deemed acceptable for the scope of this project based prior experience with Keras.

Arch2Vec Implementation

This gave us each of the crowdsourced home's features and associated price points. Using this and the features taken from the home we'd like to estimate, we developed a house to vector to price model. Each of the homes have a position in N-Dimensional space that correlates with a price point. The home that we are trying to estimate has a position in N-Dimensional space as well, but no associated price point.

Using clustering and machine learning models available with Keras (https://keras.io/) we then got an associated price point for our test house. To test the accuracy of this model we took recently sold houses and kept track of the Zillow price estimate vs the actual selling price. Then we used the same house and estimated a price using our model, then used our estimate vs the actual selling price to find a comparison between Zillow's pricing algorithm and our model.

It's worth noting that we implemented a weighting system designed to give greater weight to houses included in the training set multiple times. We felt this was critical as if one Turker thought that the house was similar but none of the others did then it should not be given as much weight in the model.

One interesting addition to our project in future iterations would be to ask Turkers to find houses that are dissimilar to the target home and use houses that are similar and not dissimilar to determine the training set of houses as well as associated weights in the system.

Result

We found that using crowdsourced estimates of real estate data we were able to match the accuracy of existing models of cost estimation in neighborhoods of significant diversity and underperformed the accuracy of existing models in neighborhoods of minimal diversity.

In neighborhoods with a large variety both in home types and price points our model was competitive with Zillow's Zestimate model. Zillow's reacher shows that on average their estimate had an error of 6% nationwide, while we were only able to achieve accuracies ranging from 4% to 11% with an average of 7.3% across our different trail zip codes. The crowdsourcing effort yielded a more exact set of comparable houses than Zillow but failed to take into account

all of the available information about a property including images and previous sale prices, which hint at finish quality and condition. Given this additional data, and properly integrated into our model, we believe we could make our model more competitive with Zillow's Zestimate.

In neighborhoods, especially those in dense suburban areas, that did not have a large variety of price points, we failed to improve upon existing models finding that our average error of 8.6%. This isn't shocking, as our model focused on key features such as lot size, bedrooms, and amenities. This doesn't take into account the state or finish of properties which proved to be one of the key variations in areas with high density of comparable houses. In these locations, previous listing and sales prices help to provide an indication of how a property finishes and condition compare to that of comparable properties. This information could be a starting place for secondary research that is more successful in these suburban areas.

We found little difference in our results when expanding the set of comparable houses for our crowd sourcers to find and allowing for the selection of more comparable properties. We expect that this is because the "Arch2Vec" model will always rely on the small set of houses that most closely resemble the house it is attempting to price. If there are outliers, as we expect there would be in cases of a selected set of size 10, the model will not use that information to inform its final price point.

Work Split

John focused on the early stage project question framing, leveraging his knowledge of real estate to inform what characteristics were most likely to be impactful to market prices excluding known prices and images, and developed the scraping tools used for web data collection. John also took the lead on developing the crowdsourcing effort, with Jordan assisting on implementation. Jordan developed the "Arch2Vec" model and performed testing as crowdsourcing data became available.