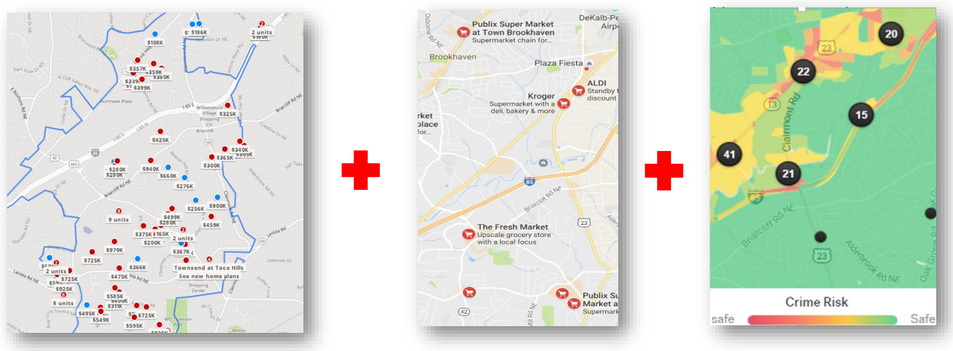
Real Estate Search and Recommendation System

Yunhe Song; Yichao Yao; Jin Yan; Qinlan Zhang; Xu Zhang

Georgia Tech College of Computing
Computational Science and Engineering

Introduction

Our project focuses on the development of interactive, map-based real estate recommendation websites for homebuyers in Atlanta including environmental factors, which is important in making decision, besides price and physical factors in existing search website such as Zillow. To provide valuable information on environmental factors, a real estate Geographic Information System (GIS) will be created by quantifying data from Google map, Zillow, Police Department, City of Atlanta and so on, and then a recommendation system will be created accordingly.



✚ Environmental factors ✚ A control panel to set preference ➡ Find dream home after one search

Approaches

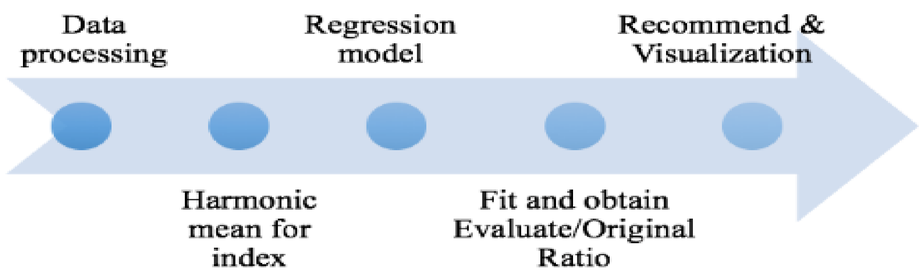
1) Using Harmonic mean of distance to neighbor school/hospital to generate related environmental index.

$$s = \frac{1}{\sum \frac{1}{s_i}}$$

2) Select appropriate variables and function of model to regress with trained data. (Compare R^2 and relative error graph)

$$\ln(y) = f(physical\ property, environmental\ factor)$$

3) Using model obtained in trained data and run it in test data(selling house)
4) Changing the priority of different environmental factor will alter the coefficients of selected variable in proportion
5) Houses accord to customers' demands with high estimated/selling price ratio will be recommended to customers.



Why does it work?

The price of a read estate is determined by a combination of different factors, a good model can be found by regression.

What's new?

Besides physical properties, environmental factors are included in regression model. Recommendations are based on user priority setting.

Data

All dataset contains longitude and latitude for GIS, includes:

Property house Dataset	Downloaded from RealtyTrac.
#House sold	9736
#House on sale	6604
Duration	Jan.2016 – Oct 2016
Crime rate:	Downloaded from Atlanta Police Department
#Crime data	261976
Duration	2009 - 2016
Environmental factors	Downloaded from Atlanta Regional Commission

Schools, Markets hospitals etc

Experiments and Results

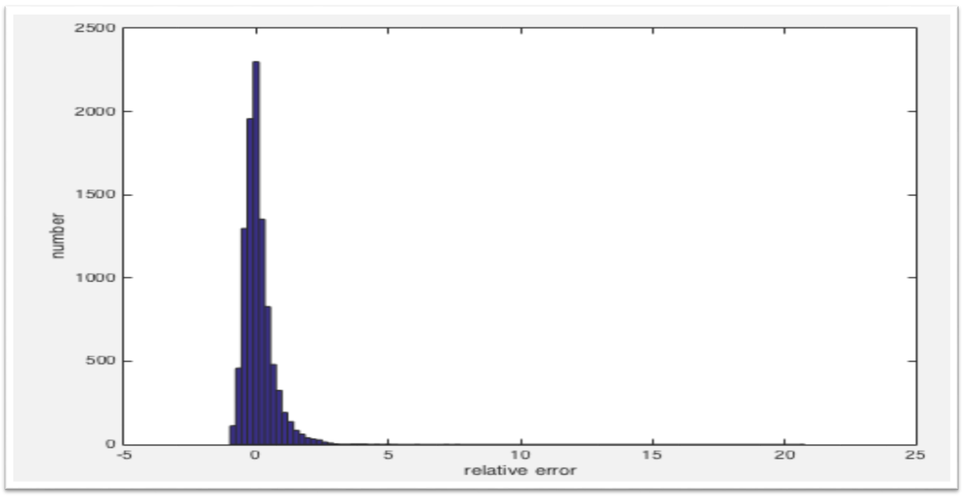


Figure 1 Result for trained data

1. Using sold house data to test correctness of regression model. R^2 = 0.8070 (Relative error figure fig.1)

$$\ln(Price) = 0.9 + 1.4737\ln(size) + 0.1933School_{Index} - 0.0033\ School^2_{index} - 0.2325Hos_{index} + 0.0037Hos^2_{index} + 0.0572market_{index} - 0.0009market^2_{index} - 0.0398crime_{index}$$

2. 97% houses in trained data with relative error less than 20%.
3. This regression model has lower relative error than the regression model with no environmental factors.

Interactive visualization result:

Composition of website: Control Panel + Map

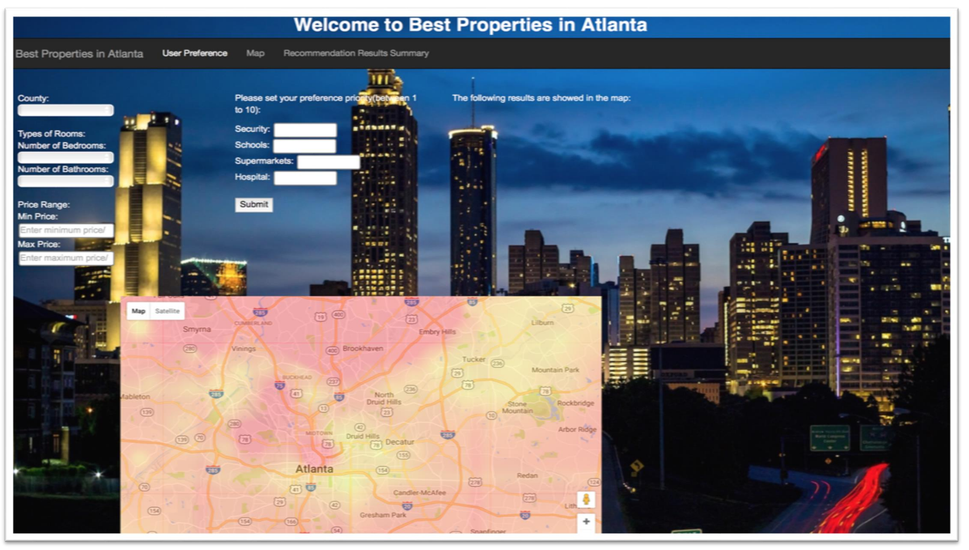


Figure 2 Overview of visualization

Heat Map: Show the level of purchase recommendation of different locations in area under investigation with combination of priority. The redder the higher level of recommendation.

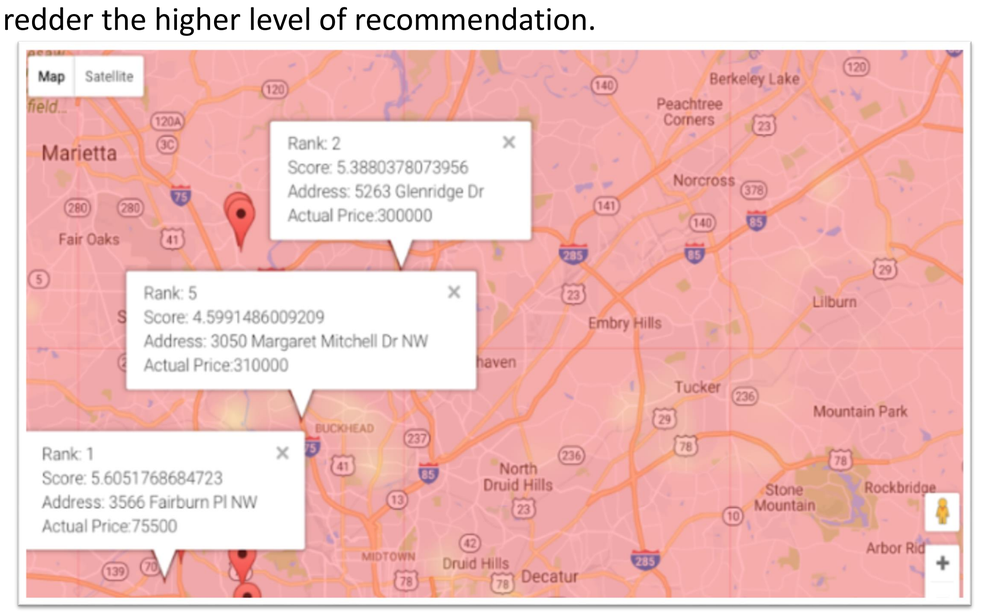


Figure 3 Visualization Result after submit

Markers: Show the top ten recommended properties meeting the user's requirements and preferences. With windows showing information of the property after clicking.