Psagent- A Machine Learning Approach to Real Estate

Vishwal Pejakkala Kakrannaya, Vivek Kumar, Sangmesh, Vishal Jaiswal, Sridevi G M
Department of ISE, SJBIT, Bengaluru, Karnataka, India

***Abstract*—The need for a higher degree of accuracy with respect to real estate guidance provided by agents is constantly growing with the growth of the real estate industry. However, human real estate agents, when making suggestions related to buying or selling of property, may be subjected to bias or a decrease in accuracy resulting from limited computational potential. Several attempts have been made to replace human agents with computers for a higher degree of accuracy but such attempts fail to replace human agents completely in the sense that such attempts do not perform all tasks performed by real estate agents. We propose a system that behaves like a human real estate agent in that it attempts to perform every task a human agent performs with a higher degree of accuracy without any room for bias.**

Keywords - Learning, Hypothesis, Regression, Classification

I. INTRODUCTION

Real estate agents primarily perform three main tasks- predict the price of houses based on the selling price of houses with similar parameters previously sold, provide intuition for bargaining based on the selling price of houses previously sold and suggest alternatives to the given option- for a potential buyer. For a seller, however, agents just perform the first two of the aforementioned tasks. Attempts so far have not focussed on performing all three tasks and each attempt does not perform one or more of these tasks.

Human real estate agents process information related to houses that were previously sold to extract knowledge and understanding in a way similar to how machines learn from historic data. Based on this knowledge, we can arrive at the understanding that machine learning techniques can be employed to perform the tasks that real estate agents perform but to do so, we need to understand which of the several parameters of houses affect the selling price of houses. This understanding is used by the human real estate agents to perform the three primary tasks. We can use this information from past sale records to understand the way said parameters affect the price of the house. The relationship we obtain can finally be used to predict the selling price of a house based on the parameters of said house. Multivariate Linear Regression on a data set is performed in a similar style and thus, we can use multivariate linear regression to perform the first of three tasks that a human real estate agent performs.

The second task human real estate agents perform is that they provide the intuition necessary for bargaining to a customer by telling the customer if a house can be bought or sold at a given price. The way a real estate agent does this is by looking at the real estate situation at that given instance of time, at that locality. The agent looks at data related to the parameters and selling price of the previously sold houses with similar parameters and tries to establish whether the house to be bought or sold can be bought or sold at a given price. We can make use of K-Nearest Neighbor's (KNN) algorithm to perform this classification.

The third task that real estate agents do is find alternative houses for a buyer and alternatives presented are generally similar in parameters. We can make use of Euclidean Distance to find alternatives based on object distance between the house the buyer or seller is interested in and other houses available for sale.

We propose a system that performs all three tasks in the ways specified before- prediction with the use of multivariate linear regression, providing intuition for bargaining by making use of KNN algorithm and providing alternatives by making use of Euclidean Distance.

The rest of the paper reviews related work by evaluating the functions of the systems, explains the proposed system in terms of the techniques used and ends with a conclusion of the paper.

II. RELATED WORK

As discussed in the previous sections, several attempts have been made to replace the human real estate agents but the attempts fail to perform one or more of the tasks human agents perform. We review some of those attempts in the rest of this section and juxtapose them with the human agents.

Iain Pardoe[2] proposed that the realtor's experience and local knowledge could be used to subjectively value a home based on its characteristics and the prices of similar homes nearby and that the same could be achieved by regression analysis[7]. This performs the first of three primary tasks that real estate agents perform in a way similar to our proposed system but does not attempt to perform the other two tasks. Nihar Bhagat et al.,[1] felt the need to predict efficient house prices for real estate customers and used linear regression to achieve it. This system used previous market trends and price

ranges have been used. The system was buyer-focused and like Iain Pardoe's attempt [2], did not address the need to perform the second and third tasks mentioned in the previous sections. In another attempt made by Aaron Ng[3] in 2005, a mobile application for house price prediction was made to predict the prices of houses in London. Here the prediction was made based on previous data and a range was given both on the client and the server side. This approach, much like the two aforementioned attempts, did not attempt to completely replace real estate agents in that it did not focus on attempting to perform all three functions.

In a different approach [4] used by Vishal Raman et al., the goal was to find houses and suggest them based on the budget of the user. It mines the web to extract information about houses for sale based on the budget of the user and the requirements of the user. This, in a way, performs the third task performed by real estate agents but does not focus on the first two tasks mentioned in the previous sections.

III. PROPOSED WORK

The system we propose performs the three primary tasks a human real estate agent does, thereby attempting to replace human agents, with a higher degree of accuracy. The three primary tasks we make use of to build proposed system are explained in detail in this section.

Multivariate linear regression helps understand data by understanding the relationship between the input variables (the parameters of a house that affect the selling price of the house) which are the independent variables and the output variable (the selling price of the house) which is the dependent variable since its value depends on the values of the independent variables. The relationship, between the parameters of the house that affect the selling price and the selling price, thus determined becomes the hypothesis which behaves like a black box in predicting the house prices. The following equation is made use of to perform multivariate linear regression and can be written as Y=XB+U,where Y is a matrix with series of multivariate measurements[5] (each column being a set of measurements on one of the dependent variables), X is a matrix of observations on independent variables that might be a design matrix (each column being a set of observations on one of the independent variables), B is a matrix containing parameters that are usually to be estimated and U is a matrix containing errors (noise). The errors are usually assumed to be uncorrelated across measurements and follow a multivariate normal distribution. If the errors do not follow a multivariate normal distribution, generalized linear models may be used to relax assumptions about Y and U. Fig.1 provides an example for a simple linear regression model with only one independent variable. The model when there are multiple variables can be represented using a multidimensional space.

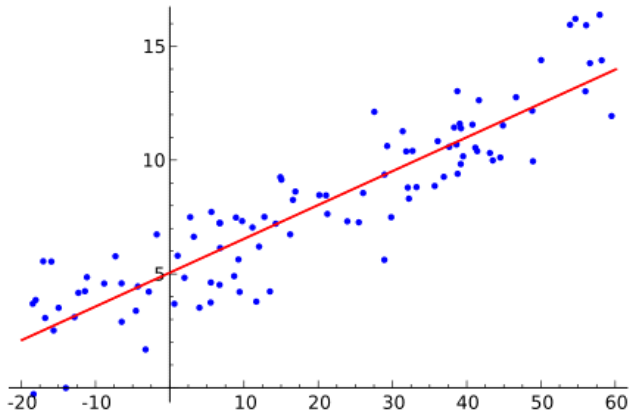

Fig.1 Regression analysis

We use Euclidean distance to determine which alternatives are to be presented to the user. For this, we consider the house that the customer is interested in and determine the nearest alternatives in terms of object distance and present a select number of those to the user. The equation for Euclidean Distance to find the straight-line distance between two points is as follows,

[6].

In a plane with p1 at (x1, y1) and

p2 at (x2, y2), it is $\sqrt{((x1 - x2)^2 + (y1 - y2)^2)}$.

We make use of K-Nearest Neighbor's algorithm to determine whether or not a house can be bought at a given price. We make use of Euclidean distance as a part of KNN to determine the K nearest neighbors, where K is the number of neighbors to be considered for classification. We classify the house to be bought or sold on the basis of whether or not a house can be bought or sold at a given price.

IV. CONCLUSION

The system we propose performs all three primary tasks- house price prediction, provide intuition necessary for bargaining and suggest alternatives- that a real estate agent performs, thereby attempting to replace human real estate agents whose predictions may have reduced accuracy due to limited computational power or potential bias. The system performs the three tasks in ways similar to how the human real estate agents perform the tasks by making use of multivariate linear regression, K Nearest Neighbor's algorithm and Euclidean distance.

REFERENCES

[1] Nihar Bhagat, Ankit Mohokar, Shreyash Mane "House Price Forecasting using Data Mining", October 2016.
[2] Iain Pardoe,"Modeling Home Prices Using Realtor Data", 2008.
[3] Aaron Ng, "Machine Learning for a London house price prediction mobile application", 2015.

[4] Vishal Venkat Raman, Swapnil Vijay, Sharmilabanu K "Identifying customer interest in real estate using data mining", 2014.
[5] https://en.wikipedia.org/wiki/General_linear_model
[6] https://en.wikipedia.org/wiki/Euclidean_distance
[7] Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, "Introduction to Linear Regression Analysis", 2015.