Assignment 11- Data Collection

Jack Wright

11/4/2020

I have selected the reccomender system for zillow.com, and will look at how they choose which homes to show the customer.

Scenario Design

Who are your target users

Zillow serves multiple groups, prospective home buyers, home sellers, real-estate agents and groups in real estate investment. All of which interact with the site in different ways, but for this I will focus on how home buyers interact with the reccomender system.

What are their key goals?

The key goals of the customer are to identify homes based on a variety of criteria, such as price location and home type. They are not looking for a "match" like they might be when shopping on amazon for toilet paper. They are trying to find something that meets a wide set of requirements they might have, or maybe ones they don't know they have.

How can you help them achieve those goals?

One key element to avoiding ranking reccomendations is to display the results on a map. Since a home is a unique product (everyone can't buy or rent the same home over and over). You also cannot describe the quality of the home by customer review. The map allows you to avoid this type of soting.

We can help the customer achieve their goals by allowing them to select search criteria that is relevant to their needs.

For example, if a customer knows they want to live in a certain location but is less sensitive to the size of the house, results should be tailored such.

A final step would be to reccomend other homes similar, but maybe not exact matches to their criteria.

Does it make sense for the reccomender system to perform scenario design twice, once for the organization and once for the organization's customers?

I do think it would be useful to perform scenario design for multiple use cases. The real estate market faces in many different directions, with many different goals in mind.

Reverse engineering the reccomender system

One obstacle that Zillow faces as opposed to other shopping websites is that they have a large number of users who do not have accounts. (Sites like amazon utilize information on a users previous searches and purchases to inform their results.) Zillow overcomes this deficit of information by employing a "twin neural network" sometimes called a "siamese neural network."[1]

A twin neural network uses the same weights while working on two different input vectors to compute comparable output vectors [2]. In this case, I believe that zillow is taking all of the parameters of the selected home, as a vector and then computes the geometric distance to the other homes in its database. The "closest" (not physical distance, but distance in the vector space) results are returned on a carousel next to the listing of the home the user has selected.

A problem encountered when utilizing a neural network for categorical data (in one case zip code) is creating a numerical representation. Zillow attacked this issue with "skip gram modeling."[1]

The skip gram model is a classification model that, instead of trying to predict an element, tries to predict the CONTEXT of an element[3]. An easy to understand example of the skip gram model is if the data is "the quick brown fox jumps over the lazy dog." we would return [quick,fox] given the target "brown."[3]

Zillow uses the skip gram model to create a sentence like "the quick brown fox jumps over the lazy dog" out of which zip codes the user has clicked and then the "target" as the current zip code they are cliked on.

Reccomendations

One thing that was touched upon in the references is the difference between a "hot" listing and a "cold" listing and the problems that this presents to their modeling. a "hot" listing is one that is off the market quickly, i.e. purchased or rented fast. So the user interactions are low. For other shopping reccomender systems. user interaction is one of the primary ways they might rank a product. User sentiment might be deemed high on a product that has lots of interaction and lots of "likes."

A "good" property, or one lots of people want to rent or buy will have lower interaction than one that is "bad" and may be on the market for a year collecting interactions. building the reccomender system to identify hot and cold properties would prove useful to consumers.

I know from personal experience that I have visited apartments that looked good in the pictures and had a low price, only to go there and realize it did not meet my needs (or the needs I was searching for.) I learned through experience to start looking at the length of the listing and how many people had viewed the property to rule out homes that were too good to be true.

Zillow has noted that a large portion of their user base is people who are looking for a home for the first time. Making it clear that a property has warning signs for why it might not meet the search terms would be a very valuable addition.

References

1. Sangdi Lin **Home Embeddings for Similar Home Reccomendations** https://www.zillow.com/tech/embedding-similar-home-recommendation/

2. **Siamese neural network** https://en.wikipedia.org/wiki/Siamese_neural_network

3. Dipanjan Sarkar **Implementing Deep Learning Methods and Feature Engineering for Text Data: The Skip-gram Model** https://www.kdnuggets.com/2018/04/implementing-deep-learning-methods-feature-engineering-text-data-skip-gram.html