**From:** Schrembeck, Steven (CAI - Atlanta - CON)   
**Sent:** Friday, January 4, 2019 12:33 PM  
**To:** Flusberg, Benjamin (CAI - Atlanta) <[Benjamin.Flusberg@coxautoinc.com](mailto:Benjamin.Flusberg@coxautoinc.com)>  
**Subject:** Hey Ben -- ideas from zillow

Hey Ben,

I was reading the Zillow data science blog, and found this really cool technique for improving "Similar homes" recommendations on their platform. I thought you might be interested in the ideas for our own use, so here it is!

<https://www.zillow.com/data-science/embedding-similar-home-recommendation/>

In case you don't have the time to read it, here's my highlights:

They are trying to learn a semantic representation of homes that home shoppers use to decide what they want to look at next. (i.e. "I like craftsman style", "Near the highway", etc.) And hopefully learn stuff that isn't obvious to computers.

Here's how they do it:

1. They are using user click stream data (order of clicks on homes in a 10 minute session) to create positive labels. Home clicks in other sessions are negative labels. **Do we capture this timestamp browsing info for our VDPs?**
2. Using that data, they trained two identical neural networks to create meaningful vector representations of two different homes (there's another variation on this described) and then they measure the distance between those two vectors to create a final similarity score. The neural nets are incentivized to generate representations useful for calculating similarity score, omitting all other data.
3. To create a final list of similar homes, they measure the distance (k-nearest neighbor) between all home vectors and a single home, then simply order the results from smallest to largest.
4. To test this they ranked how high their unseen recommendations appeared in an ordered list of real user clicks -- it outperformed their existing approach handily

This is even more exciting for our recommendation systems than Zillow, because we can replicate a "similar vehicles" usecase from browsing data, but we can also use this same approach on*transaction* information to predict which cars users would like to buy as well.