Finding Your Dream Home: REX
House Recommendations

Cecilia Yang, Nikhil Vanderklauuw, Weiru Chen, Selina Wu
Agenda

1. Introduction and Preparation
   a. Problem Statement and Motivation
   b. Datasets

2. Modeling Approaches

3. Evaluation, Interpretation and User study:
   What works? Why choose it?

4. Conclusions & Future Work
Problem Statement and Motivation
How do you find your dream house?

How can we go through all of REX’s listings to find the perfect home for each user?
Problem Statement

REX wants to improve users’ experience in buying homes.

Our Goal: Develop a model that serves open minded house-hunters with personalized matches for discovering their perfect home.

Metric for Success: receiving positive engagement with house listings suggested by our recommender.

How many of the listings we recommend end up being visited by users?
Project Trajectory

**Data Exploration**
- Understand the features in selecting an ideal house
- Explore the datasets provided
- Identify the data that would be helpful to us
- Assess quality/missingness

**Baseline Model**
- Create collaborative filtering and content filtering models features as baselines
- Request for more data from REX

**Advanced models**
- Incorporate co-occurrence information to model user behavior
- Develop Siamese Neural Network Random Forest models
- Fine-tune the models and add image features by image modeling

**Evaluation and user study**
- Evaluate and compare the results from different models
- Design a user study to test our assumptions
Description of Data
Listing Data: Numeric

Dataset that included **descriptive features** about **LA house listings**

**Main Features**

- **Location**: 6145 Longridge Ave, Van Nuys, CA 91401
- **Price**: $829,000 ($466 / sf)
- **Bedrooms**: 3
- **Bathrooms**: 2
- **Square Footage**: 1,778 sf, 7,381 lot sf

**Additional Features**

<table>
<thead>
<tr>
<th>Details</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>House Type</strong></td>
<td>Single-Family</td>
</tr>
<tr>
<td><strong>Stories</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Year Built</strong></td>
<td>1954</td>
</tr>
<tr>
<td><strong>A/C</strong></td>
<td>Central</td>
</tr>
<tr>
<td><strong>Heating</strong></td>
<td>Forced Air</td>
</tr>
<tr>
<td><strong>Pool</strong></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Fireplace</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Parking Spaces</strong></td>
<td>3 spaces</td>
</tr>
<tr>
<td><strong>Parking Type</strong></td>
<td>Carport, Driveway</td>
</tr>
<tr>
<td><strong>Basement Sqft</strong></td>
<td>—</td>
</tr>
<tr>
<td><strong>HOA Fee</strong></td>
<td>$ —</td>
</tr>
</tbody>
</table>

**Sources**

- Data provided by listers to REX
- MLS broker database
- Zipcode-level market surveys to encode location information
Listing Data: Images

We extracted features from images of 9.5k listings as supplementary features.
Web Interaction Data

Features

Listings viewed by user
- Learn about specific interests of each user
  - What kinds of houses were they browsing?

Create user-listing “score” from click data
- High intent: Heart, Request Info...
- Low intent: scrolls/clicks...

Timestamp

Items Clicked
- Image carousel, Favorite/Heart, Payment calculator ...

92k users viewed LA listings
533/5.5k listings had interactions (9%)
Co-Occurrences

**Goal:** Learn from browsing behavior to supervise recommendations
- Assume that co-clicked homes within a window during a session are relatively similar
- Incorporate collaborative information by aggregating across users

**We learned:** REX browsing is *exploratory*
- Users may be interested in a *variety* of houses
- We want to be able to *capture diversity* of viewings in our recos

---

**ANCHOR**
- 2 bd | 1 ba
- 812 sqft
- $600k

**30x**
- 2 bd | 1 ba
- 972 sqft
- $450k

**13x**
- 3 bd | 2 ba
- 1,126 sqft
- $635k

**8x**
- 2 bd | 3 ba
- 1,300 sqft
- $470k

**8x**
- 3 bd | 3 ba
- 2,500 sqft
- $3.55m

---

User 1
- Browsing Journey
- Total: 504 *co-occurring* listings

- 1
- 2
- 3
- 4
- 5
- 6
- 7

User 2
- Browsing Journey

- 7
- 9
- 5
- 8
- 3
- 2
- 4

Co-Occurrences for listing 3: 2, 4, 5, 8
Modeling
Baseline: Traditional Recommendation Models

Content-based Filtering

Recommend other listings similar to what the user likes

Data: Listing features
Method: Cosine-Similarity

Collaborative Filtering

Use similarities between users and listings simultaneously to provide recommendations

Data: User-Listing view “scores”
Method: SVD, NMF
Models: **Siamese Network:**

Content-based approaches informed by **collaborative** information
Models: Random Forest
Trained on co-occurrences, predict on User Vectors
Extract Features from the Images

Data – Models (Image Style Clusters) – Evaluation

Indoor

Outdoor

Living Room
Dining Room
Bedroom
Kitchen
Bathroom

Indoor/Outdoor Classification

Indoor Room Type Classification

Places365

REX API
Style Clusters

For each listing:

<table>
<thead>
<tr>
<th>Outdoor</th>
<th>Style Cluster #</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.jpg" alt="Outdoor Image 1" /></td>
<td>0</td>
</tr>
<tr>
<td><img src="image2.jpg" alt="Outdoor Image 2" /></td>
<td>2</td>
</tr>
<tr>
<td><img src="image3.jpg" alt="Outdoor Image 3" /></td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bedroom</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image4.jpg" alt="Bedroom Image 1" /></td>
</tr>
<tr>
<td><img src="image5.jpg" alt="Bedroom Image 2" /></td>
</tr>
</tbody>
</table>
Style Clusters

For each listing

The output from our model

Majority Vote

Outdoor

Style Cluster #

Bedroom

0

2

3

0

3

To Listing Metadata as additional features
Evaluation and Interpretation
Project Workflow

Data
- Listing
  - REX Listing Images
  - REX + MLS Listing Metadata
- User Activity
  - Web Interaction Events
  - User Journey Co-Occurrences

Models
- Style Clustering
- Content-based Filtering
- Collaborative Filtering
- Siamese Neural Network
- Random Forest
- Train on first 75% Temporal split

Predict/Evaluate
- Top k Recommendation
- Precision/Recall Evaluation
- User Study
**Evaluation Metrics:** precision@k, recall@k

- **User Historical Visitings**
  - (train)
  - (test)

- **Model**
  - (recommendations)

- **Intersect**
  - Top k=10
  - Recall @ k
    - Intersect
    - \# test
  - Precision @ k
    - Intersect
    - \# recommendations
Models Summary
Random Forest and Siamese model performed **2x better** than content filter

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision @ 10 (%)</th>
<th>Recall @ 10 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Filtering</td>
<td>0.458</td>
<td>2.77</td>
</tr>
<tr>
<td>Collaborative Filtering</td>
<td>0.54</td>
<td>2.78</td>
</tr>
<tr>
<td>Siamese NN</td>
<td><strong>1.04</strong></td>
<td><strong>5.94</strong></td>
</tr>
<tr>
<td>Random Forest</td>
<td>4.2</td>
<td>25</td>
</tr>
</tbody>
</table>
User Study

We designed a simple user study to see which model produced recommendations that real users would be interested in.

1. Pick 3 houses of interest

2. Select an additional listing of interest
   Which model produced their preferred listing?
User Study Results

Sample Size: 65 (students + REX employees)

Recommendation Chosen

Most Commonly Chosen House
Conclusion & Future Work

- We were able to generate successful recommendations that seem to capture nuances of house hunting behavior
  - Random forest is the best performing model on test set
  - Random forest also produced the recommendations that participants in our user study were most interested in

- Future Work:
  - We are continuing to collect data for the user study
  - Results can be used to convince REX product managers to continue the project - measure increase in user engagement
Poster Video
RESULTS
Style Clustering
- Use clustered labels as features for recommendation models.

For each listing:

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision @ 10 (%)</th>
<th>Recall @ 10 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These clusters are then added to listing metadata.

CONCLUSIONS
- We found that the most versatile model using click journey co-occurrence is the Random Forest.
- Including information from image style clustering improved model performance.
Literature & Sources
- Home Embeddings for Similar Home Recommendations
- Siamese Network Keras for Image and Text Similarity
- Places365-CNNs