

Key Points for Hotel Recognition

Oliver Broadrick and Marshall Thompson



Sex Trafficking Investigations Using Images from Hotels



Europol 
@Europol

...

This is a hotel room. Do U know which one? This info can help police solve a child sexual abuse case. Tell us on europol.europa.eu/stopchildabuse...



8:00 AM · Aug 11, 2017 · Twitter Web Client

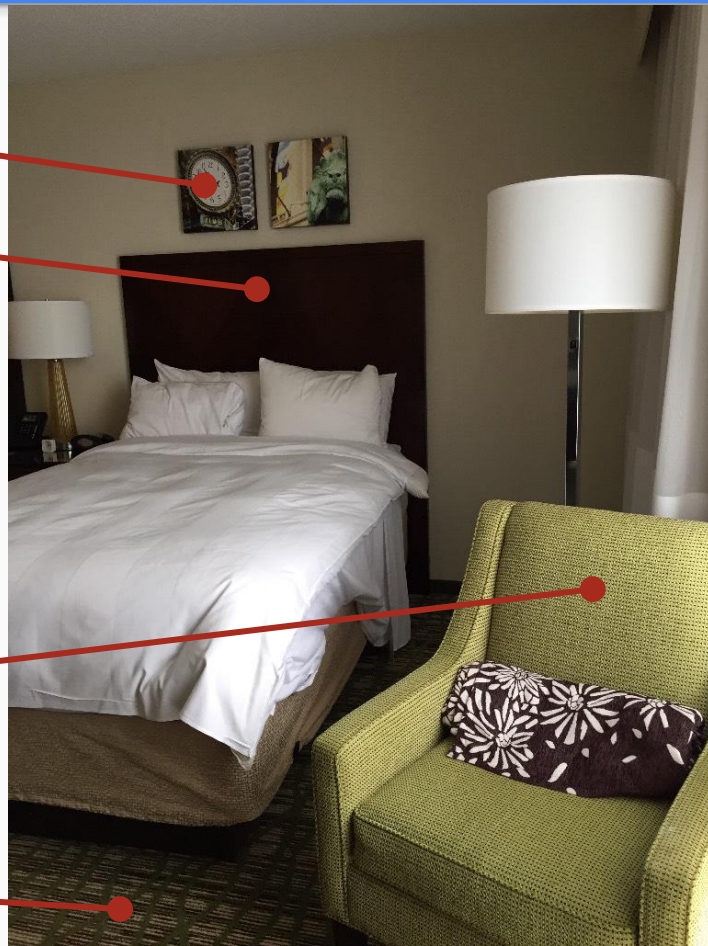
4,548 Retweets **285** Quote Tweets **541** Likes



Query Image

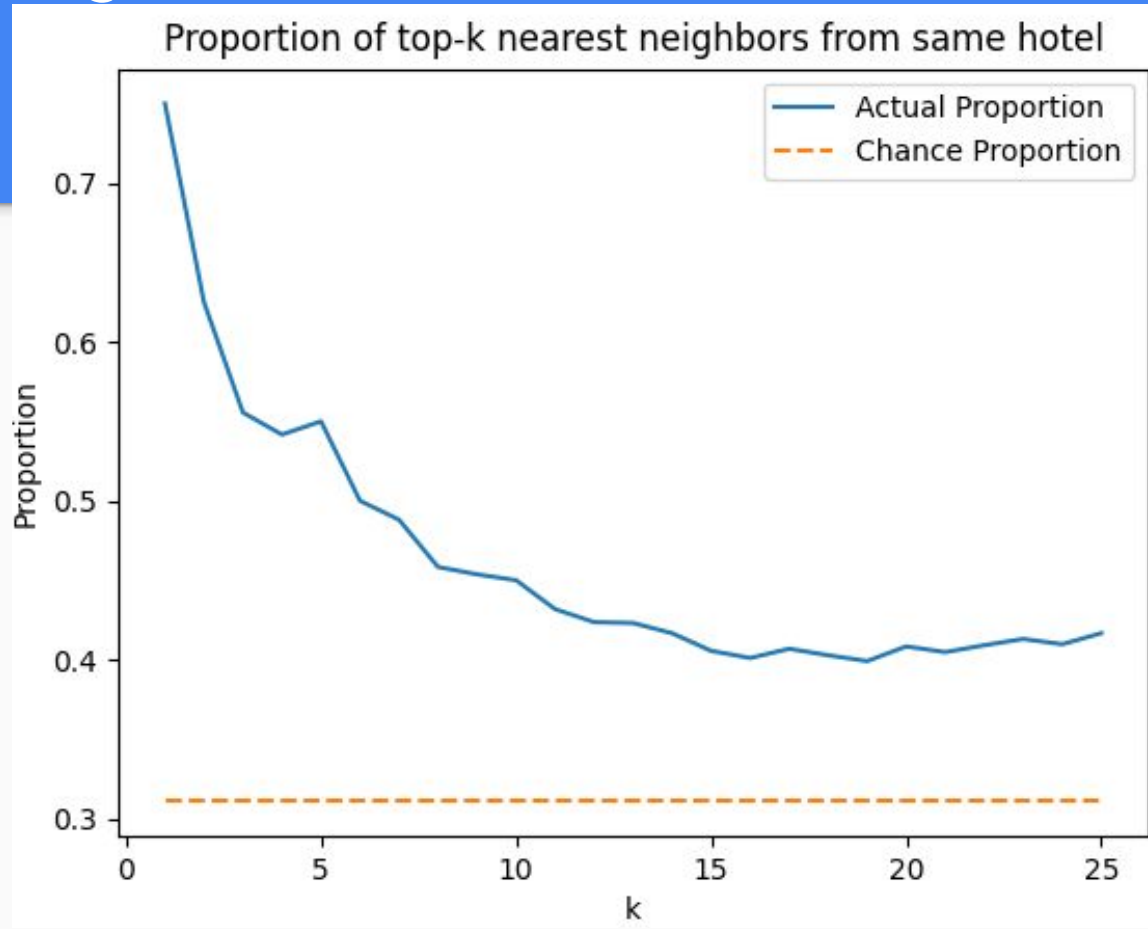


Search Result



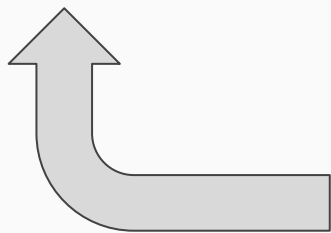
Brute Force Searching Works

- A brute force kNN search will find matches!



Brute force kNN is slow!

```
for point p in query_image:  
    for point p' in database:  
        compute_dist(p, p')
```



$O(mn)$ complexity

$m > 1$ Thousand

is the number of points in a query image

$n = \sim 10$ Billion

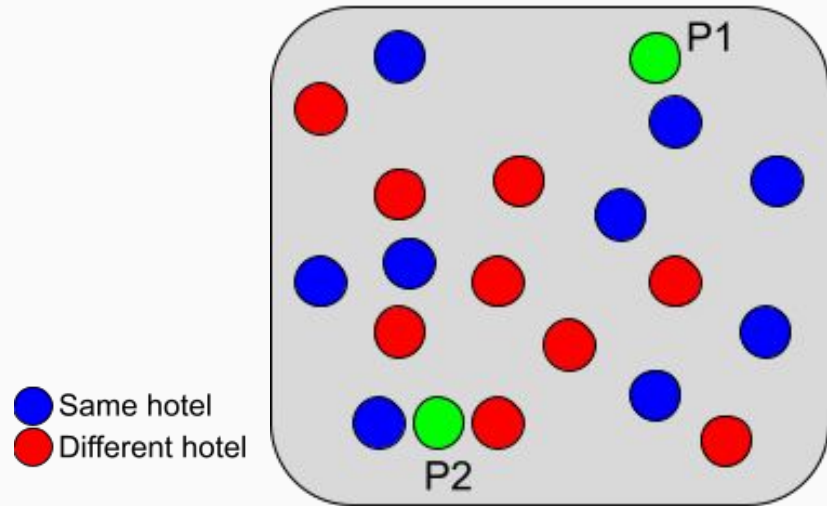
is the number of points in the database

(distance computations on 128 dimensions)

So Let's Reduce Dataset Size

Good vs Bad points:

- 1) Train a Binary Classification Algorithm
- 2) Identify and remove bad points
- 3) Query time and database building time



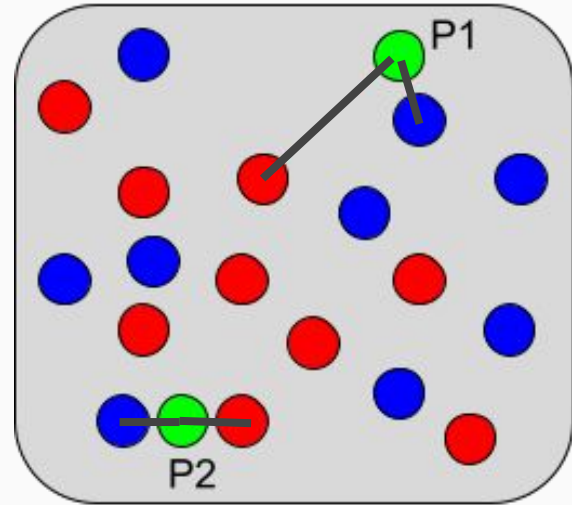
What *exactly* makes a good point?

We consider the *ratio* of distances...

$\text{diff_distance} / \text{same_distance}$

High ratio \Rightarrow good point

● Same hotel
● Different hotel

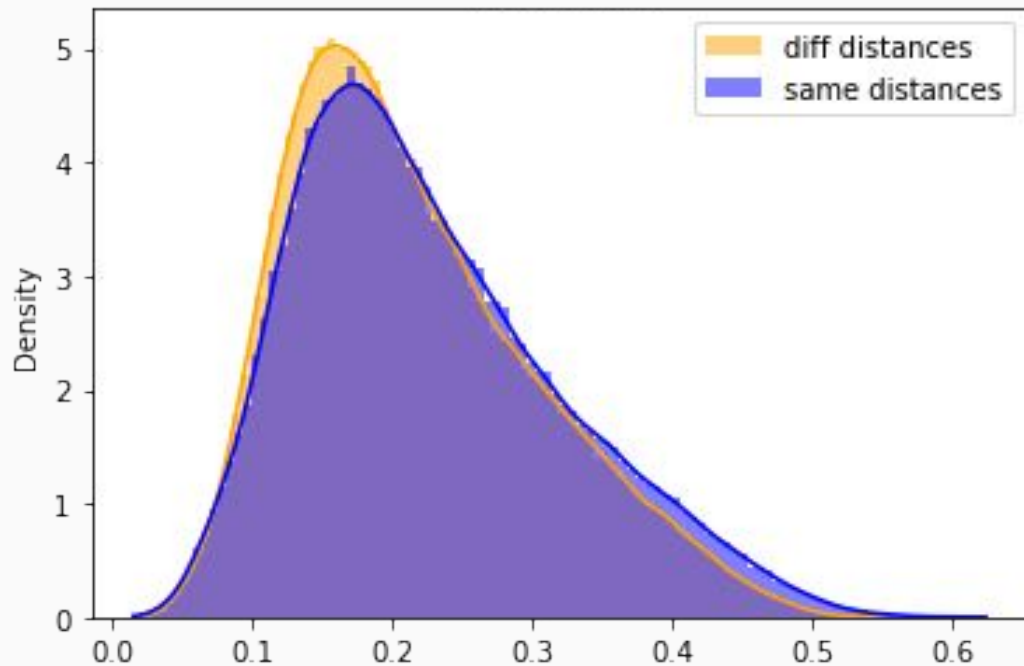


Compute distances for random sample

10 hotels, 25% of the points per hotel

Do brute force distance computations

~175k points for training and testing



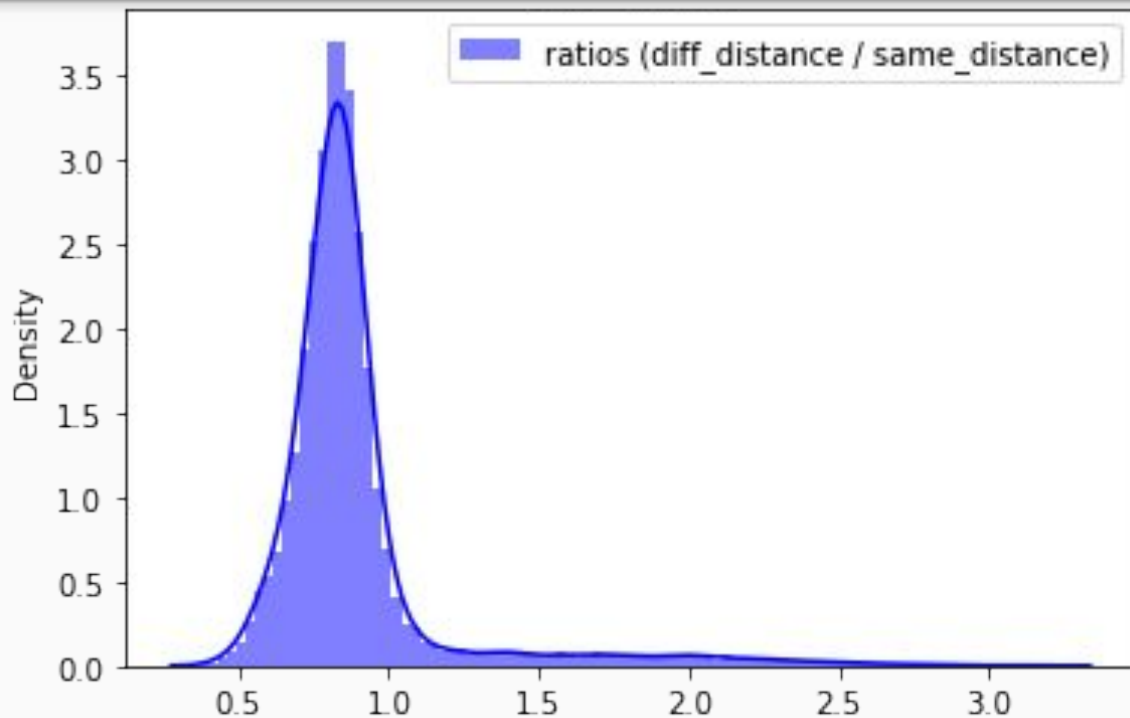
We need to choose a cutoff

1?

Mean?

Median?

Other quantiles?



Building and training a model

Logistic regression model was not sufficient... (we did try!)

Neural network time...

Model Iterations

Models Tried

Model 1: Epochs 30, LR=0.01, Batch Size=32, Loss- BCE, Optimizer- Adam

Results: train/acc=74.46, test/acc=72.9

Model 2: Epochs 30, LR=0.01, Batch Size=64, Loss- BCE, Optimizer- Adam

Results: train/acc=75.51, test/acc=74

Model 3: Epochs 30, LR=0.001, Batch Size=64, Loss- BCE, Optimizer- Adam

Results: train/acc=77.57, test/acc=73

Model 4: Epochs 60, LR=0.001, Batch Size=64, Loss- BCE, Optimizer- Adam ----- Starting to overfit

Results: train/acc=79.10, test/acc=72.7

Model 5: Epochs 50, LR=0.001, Batch Size=128, Loss- BCE, Optimizer- Adam ----- Current Best Model

Results: train/acc=79.10, test/acc=75.5

Conclusion and Future Work

- Decent results

Can try:

- Higher cutoffs
- Experiments with different network architectures
- Advice?