

1. For classifying product names to categories:
  1. What precision (P@1) were you able to achieve?
    - i. 0.605
  2. What fastText parameters did you use?
    - i. Lr =1
    - ii. Epoch =25
    - iii. Ngram = 2
  3. How did you transform the product names?
    - i. Using given normalization; cat  
 /workspace/datasets/fasttext/training\_lite.txt |sed -e "s/([.!?,'/()])/\1 /g" |  
 tr "[:upper:]" "[:lower:]" | sed "s/^[[:alnum:]]\_/\_ /g" | tr -s ' ' >  
 /workspace/datasets/fasttext/normalized\_training\_lite.txt
  4. How did you prune infrequent category labels, and how did that affect your precision?
    - i. I used pandas to remove less frequent labels. Had some issues getting the model set up with the new data, will try to implement later this week.
  5. How did you prune the category tree, and how did that affect your precision?
    - i. Wasn't able to get to this point, but looking to start later this week.
2. For deriving synonyms from content:
  1. What were the results for your best model in the tokens used for evaluation?

~/fastText-0.9.2/fasttext nn /workspace/datasets/fasttext/title\_model.bin

Query word? iphone

4s 0.84206

apple 0.820812

ipod 0.747951

3gs 0.68661

ipad 0.672341

4th 0.606912

mophie 0.562621

shell 0.557209

macbeth 0.556639

earbud 0.553249

Query word? nintendo

ds 0.964093

wii 0.932778

3ds 0.794615

gamecube 0.775196

psp 0.761968

advance 0.739912

xbox 0.73516

playstation 0.733872

boy 0.733771

360 0.70983

Query word? razr  
motorola 0.779999

droid 0.706869

nokia 0.651545

atrix 0.647557

phones 0.616268

treo 0.60941

8530 0.592447

sph 0.587375

cell 0.585448

9700 0.581893

Query word? leather

armless 0.6399

recliner 0.637171

berkline 0.614312

sofa 0.591253

case 0.562563

seating 0.560304

curved 0.558419

executive 0.557078

theaterseatstore 0.533088

jivo 0.531959

2. What fastText parameters did you use?
  - i. Epoch 25
  - ii. minCount 20
3. How did you transform the product names?
  - i. 

```
cat /workspace/datasets/fasttext/titles.txt | sed -e "s/([.!?/,/()])/\1 /g" | tr "[:upper:]" "[:lower:]" | sed "s/^[[:alnum:]]/ /g" | tr -s ' ' > /workspace/datasets/fasttext/normalized_titles.txt
```
3. For integrating synonyms with search:
  1. How did you transform the product names (if different than previously)?
    - i. 

```
cat /workspace/datasets/fasttext/normalized_titles.txt | tr " " "\n" | grep "...." | sort | uniq -c | sort -nr | head -1000 | grep -oE '^[^ ]+$' > /workspace/datasets/fasttext/top_words.txt
```
  2. What threshold score did you use?
    - i. 0.75
  3. Were you able to find the additional results by matching synonyms?
    - i. I had trouble getting final synonyms to work in my analyzer, but submitting now to finish up later