Keywords Optimization via Learning Text Representations

Feng Gui (Sophomore)
Runliang Li (Sophomore)
Fred Zhang (Sophomore)
Wilson Zhang (Freshman)
Wuming Zhang (Sophomore)

Expansion is Easy; Prediction is Hard

Expansion is Easy; Prediction is Hard

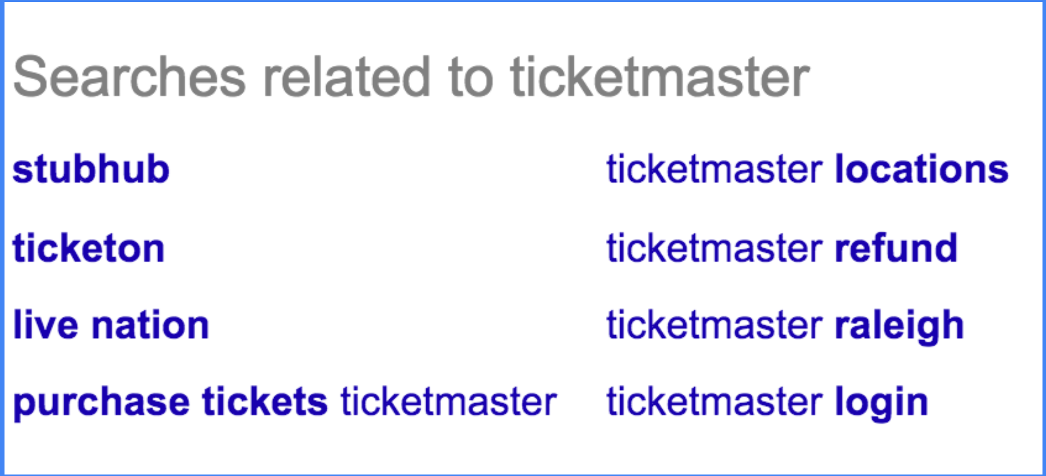
❑ Combinations of existing keywords

❑ Data-driven methods (e.g., “related search”)

❑ (Computational) Linguistics (e.g., thesaurus, WordNet)

Expansion is Easy; Prediction is Hard

- Combinations of existing keywords
- Data-driven methods (e.g., “related search”)
- (Computational) Linguistics (e.g., thesaurus, WordNet)



Searches related to ticketmaster

stubhub

ticketon

live nation

purchase tickets ticketmaster

ticketmaster **locations**

ticketmaster **refund**

ticketmaster **raleigh**

ticketmaster **login**

Expansion is Easy; Prediction is Hard

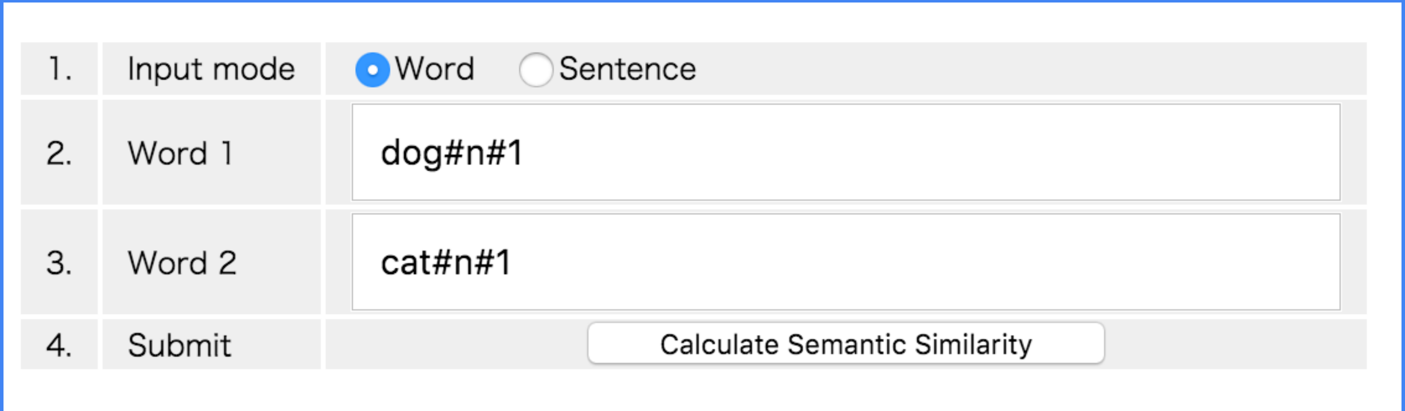
❑ Combinations of existing keywords

❑ Data-driven methods (e.g., “related search”)

❑ (Computational) Linguistics (e.g., thesaurus, WordNet)

Expansion is Easy; Prediction is Hard

- Combinations of existing keywords
- Data-driven methods (e.g., “related search”)
- (Computational) Linguistics (e.g., thesaurus, WordNet)



1.	Input mode	● Word ○ Sentence
2.	Word 1	dog#n#1
3.	Word 2	cat#n#1
4.	Submit	Calculate Semantic Similarity

Expansion is Easy; Prediction is Hard

❑ Combinations of existing keywords

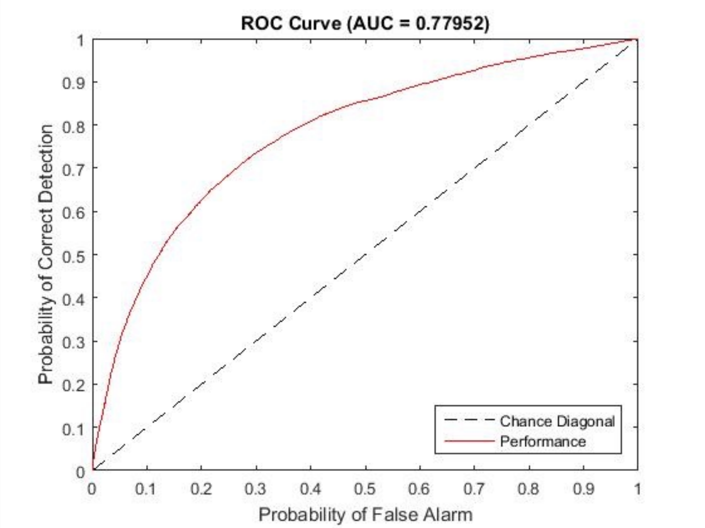
❑ Data-driven methods (e.g., “related search”)

❑ (Computational) Linguistics (e.g., thesaurus, WordNet)

Predict What Keywords are Good

1. Label the data—Good or Bad.
2. Construct text representation —$\mathbb{R}^{2857}$!
   - Bag-of-Words (BoW)
   - tf-idf vectors
3. Dimensionality Reduction (e.g., PCA)
4. Train Machine Learning classifier(s)— SVM, Random Forests, etc.

Predict What Keywords are Good



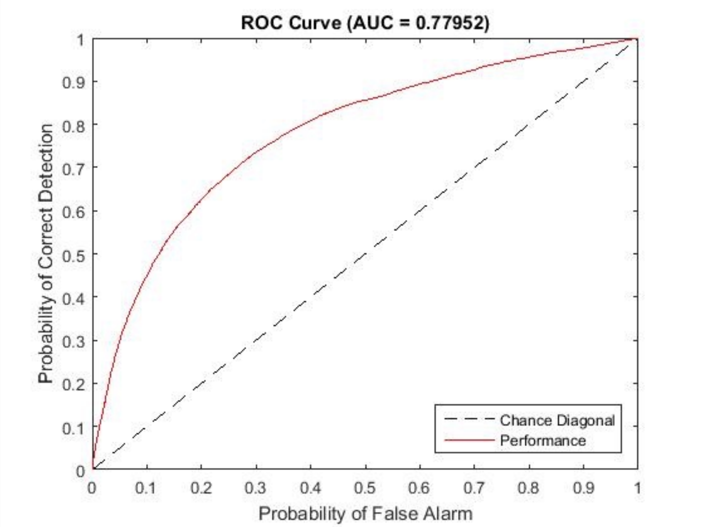
Bad.

tation —$\mathbb{R}^{2857}$!

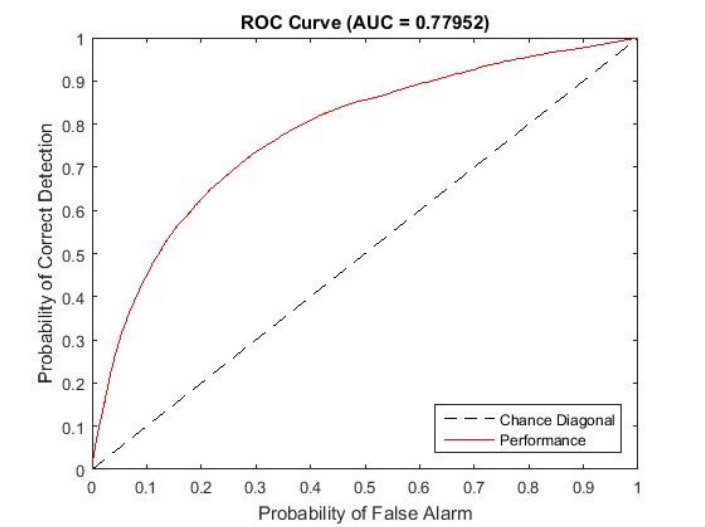
on (e.g., PCA)

classifier(s)— SVM, Random Forests, etc.

Predict What Keywords are Good



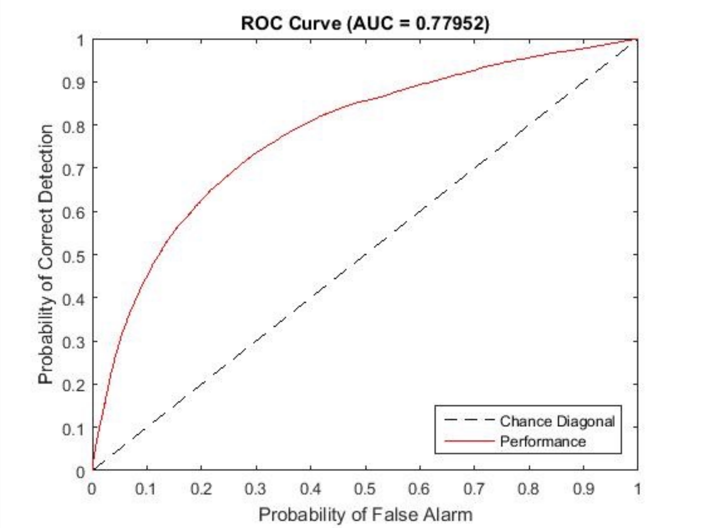
Predict What Keywords are Good



This beats an early result from research community

“Keyword Optimization in Sponsored Search via Feature Selection” (JMLR 08)

Predict What Keywords are Good

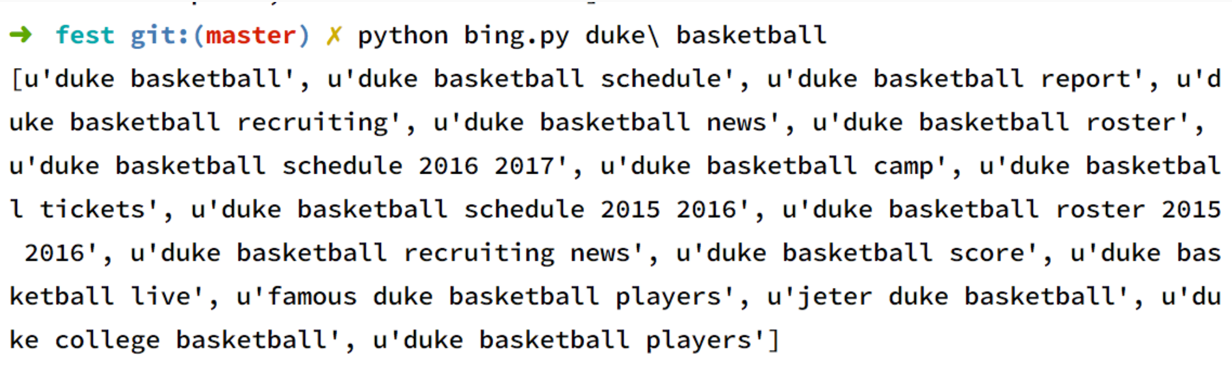


Expand Your Keyword Set

- Fix a word. See what combinations are good.
- (Asynchronizable) script to query Bing’s related search API.

Expand Your Keyword Set

Expand Your Keyword Set



```
➜  fest git:(master) ✗ python bing.py duke\ basketball
[u'duke basketball', u'duke basketball schedule', u'duke basketball report', u'd
uke basketball recruiting', u'duke basketball news', u'duke basketball roster',
u'duke basketball schedule 2016 2017', u'duke basketball camp', u'duke basketbal
l tickets', u'duke basketball schedule 2015 2016', u'duke basketball roster 2015
 2016', u'duke basketball recruiting news', u'duke basketball score', u'duke bas
ketball live', u'famous duke basketball players', u'jeter duke basketball', u'du
ke college basketball', u'duke basketball players']
```

Expand Your Keyword Set

Expand Your Keyword Set

- Fix a word. See what combinations are good.
- (Asynchronizable) script to query Bing's related search API.