Flip01 Presentation

Baobao Song
Hunan University

January 15, 2021

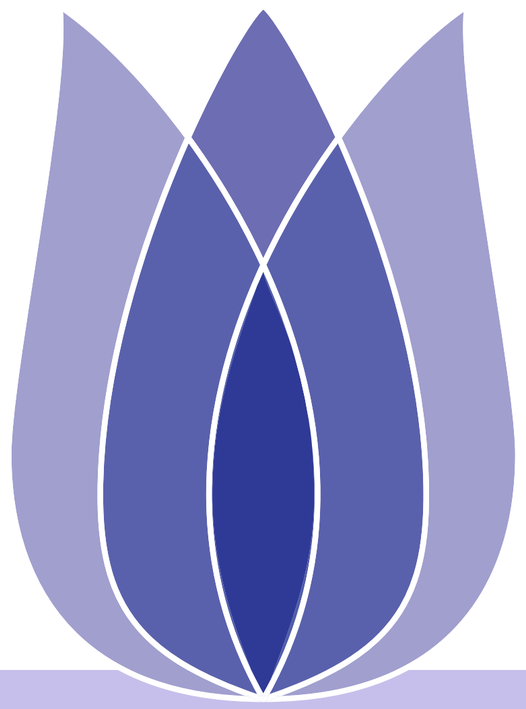

Table of Content

Problem

Text Preprocessing

Text features

Modeling and Forecasting

Problem

- Description and Evalution
- Basic Information of Data

Text Preprocessing

- Remove punctuation and stopwords
- Stemming

Text features

- Fundamental Feature
- Levenshtein Distance
- TF-IDF
- Word2Vec

Modeling and Forecasting

- Model Construction
- Model Comparison
- Score

Problem
Description and Evalution
Basic Information of Data
Text Preprocessing
Text features
Modeling and Forecasting

Problem

Description and Evalution

Problem
Description and Evalution
Basic Information of Data
Text Preprocessing
Text features
Modeling and Forecasting

Description	Predict the relevance of search results on homedepot.com
Evalution	Root mean squared error (RMSE) Predict a relevance in [1,3] for each id in the test set

Basic Information of Data

Problem
Description and Evalution
Basic Information of Data
Text Preprocessing
Text features
Modeling and Forecasting

Table 1:Data

Name	Attribute or description
train.csv	id, product_id, product_title,search_term, relevance
test.csv	id, product_id, product_title,search_term
product_descriptions.csv	product_uid,product_description
attributes.csv	product_uid,name,value
relevance_instructions.doc	rate the relevance of a search result
sample_submission.csv	id,relevance

- id - a unique Id field which represents a (search_term, product_uid) pair
- product_uid - an id for the products
- product_title - the product title
- product_description - the text description of the product (may contain HTML content)
- search_term - the search query
- relevance - the average of the relevance ratings for a given id

Problem

Text Preprocessing

Remove punctuation and stopwords

Stemming

Text features

Modeling and Forecasting

Text Preprocessing

Remove punctuation and stopwords

Problem

Text Preprocessing

Remove punctuation and stopwords

Stemming

Text features

Modeling and Forecasting

- Merge test.csv and train.csv datasets
- Concat product description in product_descriptions.csv
- Convert to lowercase
- Remove punctuation and stopwords

	id	product_uid	product_title	search_term	relevance	is_train	product_description	search_term_transform	product_title_transform	pro_des_trans
0	2	100001	Simpson Strong-Tie 12-Gauge Angle	angle bracket	3.00	1	Not only do angles make joints stronger, they ...	angle bracket	simpson strong - tie 12-gauge angle	angles make joints stronger also provide consi...
1	3	100001	Simpson Strong-Tie 12-Gauge Angle	l bracket	2.50	1	Not only do angles make joints stronger, they ...	l bracket	simpson strong - tie 12-gauge angle	angles make joints stronger also provide consi...
2	9	100002	BEHR Premium Textured DeckOver 1-gal. #SC-141 ...	deck over	3.00	1	BEHR Premium Textured DECKOVER is an innovativ...	deck	behr premium textured deck 1- gal . sc -141 tu...	behr premium textured deckover innovative soli...
3	16	100005	Delta Vero 1-Handle Shower Only Faucet Trim Ki...	rain shower head	2.33	1	Update your bathroom with the Delta Vero Singl...	rain shower head	delta vero 1- handle shower faucet trim kit ch...	update bathroom delta vero single - handle sho...
4	17	100005	Delta Vero 1-Handle Shower Only Faucet Trim Ki...	shower only faucet	2.67	1	Update your bathroom with the Delta Vero Singl...	shower faucet	delta vero 1- handle shower faucet trim kit ch...	update bathroom delta vero single - handle sho...
...	...	...	...	...	...	...	...	...	...	...
240755	240756	224424	stufurhome Norma 24 in. W x 16 in. D x 34 in. ...	24 whtie storage cabinet	NaN	0	Create a neat yet stylish storage space for or...	24 whtie storage cabinet	stufurhome norma 24 . w x 16 . x 34 . h linen ...	create neat yet stylish storage space organizi...
240756	240757	224425	Home Decorators Collection 49 in. D Alessandro...	adirondeck cusion	NaN	0	Our Bullnose Adirondack Chair Cushions fit Adi...	adirondeck cusion	home decorators collection 49 . alessandro spi...	bullnose adirondack chair cushions fit adirond...
240757	240758	224426	Simpson Strong-Tie HB 3-1/2 x 14 in. Top Flang...	hb	NaN	0	Joist hangers are designed to provide support ...	hb	simpson strong - tie hb 3-1/2 x 14 . top flang...	joist hangers designed provide support underne...
240758	240759	224427	1/4 in. -20 tpi x 1-1/2 in. Stainless Steel Bu...	hex sockets	NaN	0	These socket cap screws are ideal for applicat...	hex sockets	1/4 . -20 tpi x 1-1/2 . stainless steel button...	socket cap screws ideal applications require w...
240759	240760	224428	Bosch 4 in. Bi-Metal Hole Saw	4 inch hole saw	NaN	0	The Bosch quick change bi-metal hole saws feat...	4 inch hole saw	bosch 4 . bi - metal hole saw	bosch quick change bi - metal hole saws featur...

240760 rows × 10 columns

Figure 1: Remove punctuation and stopwords

Stemming

Problem

Text Preprocessing

Remove punctuation and stopwords

Stemming

Text features

Modeling and Forecasting

Stemming by SnowballStemmer

	id	product_uid	product_title	search_term	relevance	product_description	search_term_transform	product_title_transform	pro_des_trans
0	2	100001	simpson strong - tie 12- gaug angl	angl bracket	3.00	angl make joint stronger also provid consist s...	angle bracket	simpson strong - tie 12-gauge angle	angles make joints stronger also provide consi...
1	3	100001	simpson strong - tie 12- gaug angl	l bracket	2.50	angl make joint stronger also provid consist s...	l bracket	simpson strong - tie 12-gauge angle	angles make joints stronger also provide consi...
2	9	100002	behr premium textur deck 1- gal . sc -141 tugb...	deck	3.00	behr premium textur deckov innov solid color c...	deck	behr premium textured deck 1- gal . sc -141 tu...	behr premium textured deckover innovative soli...
3	16	100005	delta vero 1- handl shower faucet trim kit chr...	rain shower head	2.33	updat bathroom delta vero singl - handl shower...	rain shower head	delta vero 1- handle shower faucet trim kit ch...	update bathroom delta vero single - handle sho...
4	17	100005	delta vero 1- handl shower faucet trim kit chr...	shower faucet	2.67	updat bathroom delta vero singl - handl shower...	shower faucet	delta vero 1- handle shower faucet trim kit ch...	update bathroom delta vero single - handle sho...
...	...	...	...	...	...	...	...	...	...
240755	240756	224424	stufurhom norma 24 . w x 16 . x 34 . h linen s...	24 whtie storag cabinet	NaN	creat neat yet stylish storag space organ bath...	24 whtie storage cabinet	stufurhome norma 24 . w x 16 . x 34 . h linen ...	create neat yet stylish storage space organizi...
240756	240757	224425	home decor collect 49 . alessandro spiceberri ...	adirondeck cusion	NaN	bullnos adirondack chair cushion fit adirondac...	adirondeck cusion	home decorators collection 49 . alessandro spi...	bullnose adirondack chair cushions fit adirond...
240757	240758	224426	simpson strong - tie hb 3-1/2 x 14 . top flang...	hb	NaN	joist hanger design provid support underneath ...	hb	simpson strong - tie hb 3-1/2 x 14 . top flang...	joist hangers designed provide support underne...
240758	240759	224427	1/4 . -20 tpi x 1-1/2 . stainless steel button...	hex socket	NaN	socket cap screw ideal applic requir well tool...	hex sockets	1/4 . -20 tpi x 1-1/2 . stainless steel button...	socket cap screws ideal applications require w...
240759	240760	224428	bosch 4 . bi - metal hole saw	4 inch hole saw	NaN	bosch quick chang bi - metal hole saw featur p...	4 inch hole saw	bosch 4 . bi - metal hole saw	bosch quick change bi - metal hole saws featur...

240760 rows × 9 columns

Figure 2: Stemming Result

Problem

Text Preprocessing

Text features

Fundamental Feature

Levenshtein Distance

TF-IDF

Word2Vec

Modeling and Forecasting

Text features

Fundamental Feature

Problem
Text Preprocessing
Text features
Fundamental Feature
Levenshtein Distance
TF-IDF
Word2Vec
Modeling and Forecasting

- the length of search term
- the number of same keywords between search terms and product title
- the number of same keywords between search terms and product description

	id	product_uid	product_title	search_term	relevance	pro_des_trans	len_of_query	commons_in_title	commons_in_desc
0	2	100001	simpson strong - tie 12-gaug angl	angl bracket	3.00	angles make joints stronger also provide consi...	2	1	1
1	3	100001	simpson strong - tie 12-gaug angl	l bracket	2.50	angles make joints stronger also provide consi...	2	1	1
2	9	100002	behr premium textur deck 1- gal . sc -141 tugb...	deck	3.00	behr premium textured deckover innovative soli...	1	1	1
3	16	100005	delta vero 1- handl shower faucet trim kit chr...	rain shower head	2.33	update bathroom delta vero single - handle sho...	3	1	1
4	17	100005	delta vero 1- handl shower faucet trim kit chr...	shower faucet	2.67	update bathroom delta vero single - handle sho...	2	2	2
...	...	...	...	...	...	...	...	...	...
240755	240756	224424	stufurhom norma 24 . w x 16 . x 34 . h linen s...	24 whtie storag cabinet	NaN	create neat yet stylish storage space organizi...	4	3	3
240756	240757	224425	home decor collect 49 . alessandro spiceberri ...	adirondeck cusion	NaN	bullnose adirondack chair cushions fit adirond...	2	0	0
240757	240758	224426	simpson strong - tie hb 3-1/2 x 14 . top flang...	hb	NaN	joist hangers designed provide support underne...	1	1	1
240758	240759	224427	1/4 . -20 tpi x 1-1/2 . stainless steel button...	hex socket	NaN	socket cap screws ideal applications require w...	2	2	2
240759	240760	224428	bosch 4 . bi - metal hole saw	4 inch hole saw	NaN	bosch quick change bi - metal hole saws featur...	4	3	2

240760 rows × 9 columns

Figure 3: Fundamental Feature of database

Levenshtein Distance

Problem
Text Preprocessing
Text features
Fundamental Feature
Levenshtein Distance
TF-IDF
Word2Vec
Modeling and Forecasting

- Levenshtein text similarity between search terms and product title
- Levenshtein text similarity between search terms and product description

id	product_uid	product_title	search_term	relevance	product_description	len_of_query	commons_in_title	commons_in_desc	dist_in_title	dist_in_desc
2	100001	simpson strong - tie 12- gaug angl	angl bracket	3.00	angl make joint stronger also provid consist s...	2	1	1	0.173913	0.039539
3	100001	simpson strong - tie 12- gaug angl	l bracket	2.50	angl make joint stronger also provid consist s...	2	1	1	0.139535	0.029801
9	100002	behr premium textur deck 1-gal . sc -141 tugb...	deck	3.00	behr premium textur deckov innov solid color c...	1	1	1	0.112676	0.010323
16	100005	delta vero 1-handl shower faucet trim kit chr...	rain shower head	2.33	updat bathroom delta vero singl - handl shower...	3	1	1	0.389610	0.064257
17	100005	delta vero 1-handl shower faucet trim kit chr...	shower faucet	2.67	updat bathroom delta vero singl - handl shower...	2	2	2	0.351351	0.052525
...	...	...	...	...	...	...	...	...	...	...
240756	224424	stufurhom norma 24 . w x 16 . x 34 . h linen s...	24 whtie storag cabinet	NaN	creat neat yet stylish storag space organ bath...	4	3	3	0.468085	0.076795
240757	224425	home decor collect 49 . alessandro spiceberri ...	adirondeck cusion	NaN	bullnos adirondack chair cushion fit adirondac...	2	0	0	0.268908	0.056478

Figure 4: Levenshtein text similarity

TF-IDF

Problem

Text Preprocessing

Text features

Fundamental Feature

Levenshtein Distance

TF-IDF

Word2Vec

Modeling and Forecasting

- Cosine similarity between search terms and product title
- Cosine similarity between search terms and product description

_train	product_description	len_of_query	commons_in_title	commons_in_desc	dist_in_title	dist_in_desc	all_texts	tfidf_cos_sim_in_title	tfidf_cos_sim_in_desc
1	angl make joint stronger also provid consist s...	2	1	1	0.173913	0.039539	simpson strong - tie 12- gaug angl.angl make j...	0.270386	0.191815
1	angl make joint stronger also provid consist s...	2	1	1	0.139535	0.029801	simpson strong - tie 12- gaug angl.angl make j...	0.000000	0.000000
1	behr premium textur deckov innov solid color c...	1	1	1	0.112676	0.010323	behr premium textur deck 1- gal . sc -141 tugb...	0.194095	0.212764
1	updat bathroom delta vero singl - handl shower...	3	1	1	0.389610	0.064257	delta vero 1- handl shower faucet trim kit chr...	0.137359	0.051747
1	updat bathroom delta vero singl - handl shower...	2	2	2	0.351351	0.052525	delta vero 1- handl shower faucet trim kit chr...	0.355233	0.133826
...	...	...	...	...	...	...	...	...	...
0	creat neat yet stylish storag space organ bath...	4	3	3	0.468085	0.076795	stufurhom norma 24 . w x 16 . x 34 . h linen s...	0.214326	0.136015

Figure 5: TF-IDF method

Word2Vec

Problem
Text Preprocessing
Text features
Fundamental Feature
Levenshtein Distance
TF-IDF
Word2Vec
Modeling and Forecasting

- Cosine similarity between search terms and product title
- Cosine similarity between search terms and product description

ımons_in_title	commons_in_desc	dist_in_title	dist_in_desc	all_texts	tfidf_cos_sim_in_title	tfidf_cos_sim_in_desc	w2v_cos_sim_in_title	w2v_cos_sim_in_desc
1	1	0.173913	0.039539	simpson strong - tie 12- gaug angl.angl make j...	0.270386	0.191815	0.393470	0.476581
1	1	0.139535	0.029801	simpson strong - tie 12- gaug angl.angl make j...	0.000000	0.000000	0.280904	0.275490
1	1	0.112676	0.010323	behr premium textur deck 1- gal . sc -141 tugb...	0.194095	0.212764	0.493544	0.358379
1	1	0.389610	0.064257	delta vero 1- handl shower faucet trim kit chr...	0.137359	0.051747	0.543286	0.367758
2	2	0.351351	0.052525	delta vero 1- handl shower faucet trim kit chr...	0.355233	0.133826	0.799418	0.565387
...	...	...	...	...	...	...	...	...
3	3	0.468085	0.076795	stufurhom norma 24 . w x 16 . x 34 . h linen s...	0.214326	0.136015	0.604527	0.580506

Figure 6: Word2Vec method

Problem

Text Preprocessing

Text features

Modeling and Forecasting

Model Construction

Model Comparison

Score

Modeling and Forecasting

Model Construction

Problem
Text Preprocessing
Text features
Modeling and Forecasting
Model Construction
Model Comparison
Score

- Gradient Boosting Regressor
- Use cross_val_score to adjust parameters

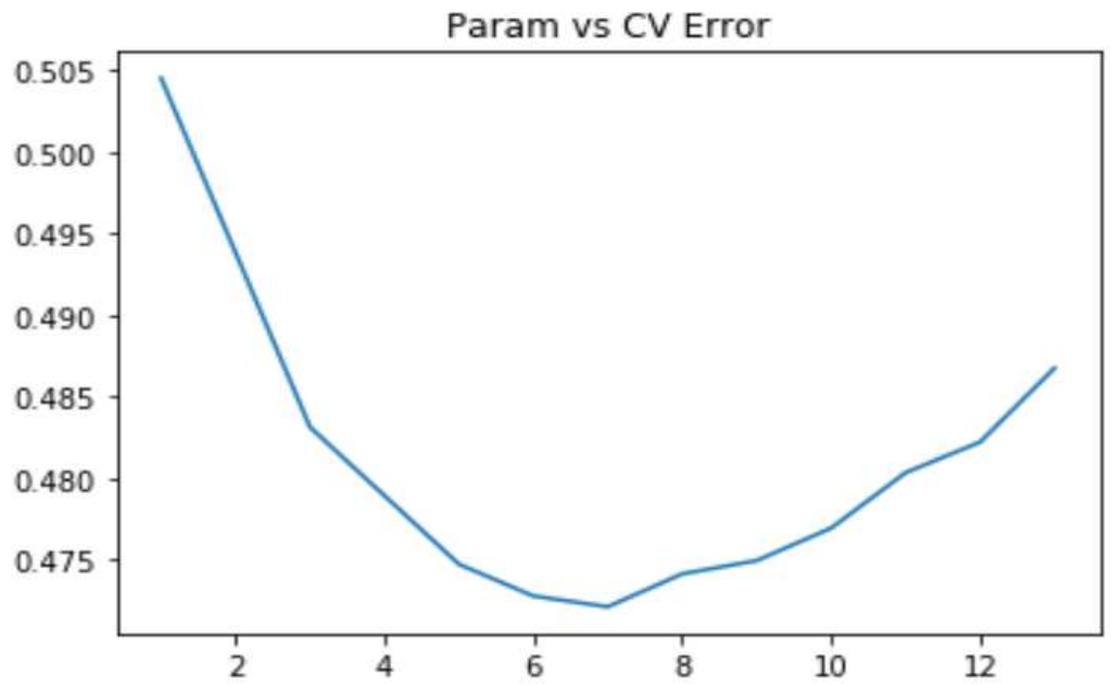

Figure 7: Adjusting max depth of Gradient Boosting Regressor

Model Comparison

Problem
Text Preprocessing
Text features
Modeling and Forecasting
Model Construction
Model Comparison
Score

- Methods
  - Random Forest Regressor
  - Gradient Boosting Regressor
  - XGboost
- Comparison:Gradient Boosting Regressor > Gradient Boosting Regressor > XGboost

Problem

Text Preprocessing

Text features

Modeling and Forecasting

Model Construction

Model Comparison

Score

- rank:347/2123

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
submissionm.csv	just now	0 seconds	1 seconds	0.48032

Complete

Jump to your position on the leaderboard

Figure 8: Random Forest Regressor Score

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
submissiongbrr.csv	5 hours ago	1 seconds	1 seconds	0.47101

Complete

Jump to your position on the leaderboard

Figure 9: Gradient Boosting Regressor Score