

به نام خدا

گزارش کار پروژه‌ی درس پردازش زبان های طبیعی

موضوع:

تحلیل نظرات کاربران به منظور یافتن محبوب ترین برند موبایل
با استفاده از رویکردهای SVM, مبتنی بر لغتنامه و ترکیبی

سعیده عباس زاده کشکا

۹۷۱۳۱۱۱۳

رویکرد اول: پیاده سازی با استفاده از SVM

مجموعه داده: مجموعه داده مورد استفاده قرار گرفته برای آموزش دستهبند، در ادرس زیر قرار دارد:

/dataset/mobile_digikala.csv

این مجموعه داده شامل ۲۷۲۳ نظر برچسبگذاری شده درباره محصول موبایل از شرکت دیجیکالا می باشد، که حدوداً از ۸۰۰ نظر منفی و ۲۰۰۰ نظر مثبت تشکیل شده است. و از سایت www.dataheart.ir قابل خریداری است. به منظور آموزش دسته بند: حدود ۰/۹ از تعداد نظرات برای آموزش و ۰/۱ برای تست (به صورت رندوم) در نظر گرفته شده اند.

نحوه ی پیاده سازی:

پیاده سازی در ادرس

/sentiment/tf-idf_SVM.py

قرار دارد.

۱- آموزش مدل و آزمایش: در این مرحله ابتدا بردار tf-idf هر نظر استخراج شده و سپس باتوجه به بردار داده آموزشی، مدل مورد نظر استخراج شده (از تابع GridSearchCV() به منظور استخراج بهترین پارامترها و در نهایت بهترین تخمین زننده استفاده شده است). دقت تخمین برای داده تست به شرح زیر است:

svm score: 0.7875457875457875

precision of positives: 0.8837209302325582

recall of positives: 0.8

precision of negatives: 0.6237623762376238

recall of negatives: 0.7590361445783133

همچنین به منظور مقایسه، از دستهبند logistic regression نیز استفاده شد که دقت این روش نیز به شرح زیر است:

logreg score: 0.7472527472527473

نتیجه گیری ۱: با توجه به رندوم انتخاب شدن داده تست و آموزش، دقت الگوریتمها به ازای هربار اجرا متفاوت بوده، اما به طور کلی دقت در بازه ۰/۸۳-۰/۷۵ قرار داشته است. همچنین باتوجه به داده آموزش و تست، گاه دقت الگوریتم SVM بیشتر بوده است و گاه Logistic Regression.

۲- تعیین محبوبترین برند موبایل: تعداد ۱۴۵۴۹ نظر با استفاده از خزنده وب از سایت mobile.ir استخراج شد. کد خزنده در پوشه crawler قرار دارد. داده استخراج شده از سایت mobile.ir در ادرس زیر قرار دارد:

/dataset/mobile.ir.final.json

نظرات متعلق به یک سال اخیر (از تاریخ تیرماه سال ۹۷ تا تاریخ تیرماه ۹۸) می باشند. اطلاعات استخراج شده برای هر نظر به شرح زیر است:

برند و مدل (model_brand)

نویسنده (holder)

تاریخ (date)

متن نظر (commnt)

پس از استخراج نظرات، بردار tf-idf هر نظر ساخته شد و به دسته‌بند SVM داده شد و کلاس هر نظر پیش‌بینی شد. سپس به منظور تعیین محبوب‌ترین برند موبایل، ماژول items_ranking در ادرس

/sentiment/items_ranking.py

نوشته شد. این ماژول با در نظر گرفتن تعداد برجسب‌های مثبت و منفی برای هر برند، امتیاز هر برند را استخراج و براساس آن مرتب می‌کند.

خروجی نهایی به همراه امتیازها به صورت زیر است:

('سامسونگ', ۰.۰۹۹۰۶۵۱۵۲۰۱۶۶۰۵۱۵)
('شیائومی', ۰.۰۷۵۵۱۶۴۱۰۶۸۸۶۲۱۶۱)
('هوای', ۰.۰۵۰۵۸۲۴۷۵۷۱۶۳۱۹۴۸)
('انر', ۰.۰۳۶۹۸۱۰۱۹۱۴۱۶۴۳۶)
('سونی', ۰.۰۲۷۳۱۳۶۶۵۵۴۰۲۰۴۳۲۷)
('نوکیا', ۰.۰۲۱۲۳۵۳۴۲۵۲۳۷۹۹۱۴۴)
('اچ تی سی', ۰.۰۱۳۱۸۴۸۱۰۳۹۵۹۸۶۰۰۴)
('اپل', ۰.۰۱۳۱۵۹۳۶۰۴۶۳۹۳۹۲۵۲)
('موتورولا', ۰.۰۰۸۳۱۴۸۹۳۳۹۶۷۹۴۱۶۴)
('بلک بری', ۰.۰۰۷۳۱۳۲۱۷۴۰۳۲۲۰۴۵۳)
('لنوو', ۰.۰۰۷۲۷۳۵۴۳۵۲۹۴۱۴۸۶۲)
('ایسوس', ۰.۰۰۶۹۲۵۱۲۳۵۹۵۲۹۳۰۸۳)
('ال جی', ۰.۰۰۶۶۱۹۵۸۹۷۴۶۰۱۵۰۷)
('اوپو', ۰.۰۰۶۱۴۴۷۵۲۲۱۶۵۹۵۹۸۶)
('وان پلاس', ۰.۰۰۲۷۰۱۶۱۴۲۰۸۷۸۷۰۰۲۶)
('سونی اریکسون', ۰.۰۰۰۹۵۵۰۳۰۳۶۹۱۹۱۹۶۹۳)
('مایکروسافت', ۰.۰۰۰۷۳۶۴۲۷۶۳۸۵۸۵۱۶۱۲)
('ویوو', ۰.۰۰۰۵۹۲۱۶۳۳۵۲۴۴۹۷۱۷)
('گوگل', ۰.۰۰۰۴۸۱۱۳۲۷۲۳۸۷۹۳۰۴۳۷)
('جی ال ایکس', ۰.۰۰۰۳۶۴۸۱۴۹۲۲۴۸۸۴۲۰۱)
('میزو', ۰.۰۰۰۲۹۷۸۴۴۰۶۷۱۲۵۶۵۲)
('زد تی ای', ۰.۰۰۰۲۷۴۹۳۲۹۸۴۹۹۳۲۴۱۲)
('الکاتل', e-05۶/۸۷۳۳۲۴۶۲۰۲۴۸۸۱۳)
('آی میت', ۰/۰)
('آمازون', ۰.۰۰۰۲۷۴۹۳۲۹۸۴۸۰۹۹۵۲۵۳-)

رویکرد دوم: پیاده‌سازی با استفاده از رویکرد مبتنی بر لغت‌نامه

مجموعه داده:

به منظور تست دقت این روش از همان داده دارای برچسب mobile_digikala.csv استفاده شده است. همچنین لیست لغات منفی و مثبت نیز در ادرس زیر قرار دارد:

/dataset/dataheart_lexicon/negative_words.txt

/dataset/dataheart_lexicon/positive_words.txt

این داده‌ها نیز از سایت www.dataheart.ir دانلود شدند. لازم به ذکر است که چندین لغت به صورت دستی به لیست این لغات اضافه شدند.

همچنین لیستی از افعال منفی نیز تهیه شده که در ادرس زیر قرار دارد:

/dataset/dataheart_lexicon/negative_verbs.txt

نحوه‌ی پیاده‌سازی:

پیاده‌سازی در ادرس sentiment/lexicon_based(dataheart).py قرار دارد.

در این روش به منظور تعیین قطبیت هر نظر، تعداد کلمات حسی مثبت و منفی در هر نظر استخراج شده و به ازای کلمات حسی که پیش از افعال منفی مشاهده شدند (تا دو کلمه قبل از فعل منفی)، قطبیت کلمه برعکس در نظر گرفته شد. در نهایت برچسب نظر باتوجه به امتیاز به دست آمده تعیین شد. خروجی به ازای تست بر روی داده دارای برچسب شرکت دیجیکالا به صورت زیر است:

accuracy 0.6683804627249358

precision: 0.7042525465400237

recall: 0.6683804627249358

لازم به ذکر است که ابتدا از لغات حسی lexipers استفاده شد که دقت حدود ۰/۴۳ بود. سپس با تغییر لیست لغات منفی و مثبت و استفاده از لغات حسی موجود در سایت dataheart.ir دقت به حدود ۰/۶۳ رسید. سپس با اضافه نمودن چندین لغت به صورت دستی و همچنین در نظر گرفتن افعال منفی‌کننده دقت به حدود ۰/۶۶ رسید.

همچنین در فایل /dataset/founded_negative_verbs.txt افعال منفی موجود در نظرات و کلمات حسی مثبت و منفی که پیش از آن‌ها آمده‌اند به همراه تعداد دفعات مشاهده این کلمات پیش از فعل منفی، آورده شده است. این فایل شامل دو قسمت است: ۱- نظرات مربوط به داده دیجیکالا ۲- نظرات جمع‌آوری شده از سایت mobile.ir

تعیین محبوب‌ترین برند موبایل:

برای این منظور از داده خزیده شده از سایت mobile.ir استفاده شد. خروجی به شرح زیر است:

('سامسونگ', ۰/۱۳۱۷۲۰۴۰۸۳۱۰۲۹۳۶۴)

('شیائومی', ۰/۰۹۴۰۸۲۱۹۲۷۵۰۲۰۱۵۹)

('هواوی', ۰/۰۸۱۴۹۳۰۵۲۴۵۱۸۲۲۱)

('انر', ۰/۰۴۹۱۵۸۹۵۹۰۶۲۱۱۲۱۱۶)

('نوکیا', ۰/۰۳۰۲۴۵۵۳۷۷۰۳۰۴۱۵۸۳)

('سونی', ۰/۰۲۹۲۸۲۸۰۳۱۴۵۰۹۲۴۲۴)

('اپل', ۰/۰۲۸۸۳۴۰۴۶۴۶۷۴۷۴۷۸)

('اچ تی سی', ۰/۰۲۲۴۶۱۵۹۳۳۷۲۲۱۵۳۷)

('ال جی', ۰/۰۲۰۷۹۸۱۰۸۶۲۷۳۳۴۳۶۸)
 ('موتورولا', ۰/۰۱۳۶۶۳۳۳۶۲۲۸۴۰۹۲۲۵)
 ('ایسوس', ۰/۰۱۰۰۸۰۸۷۶۱۱۹۷۲۸۲۲۳)
 ('بلک بری', ۰/۰۰۸۶۳۱۷۱۹۵۴۷۷۲۵۸۳۸)
 ('لنوو', ۰/۰۰۸۲۲۲۵۳۲۷۹۴۶۰۷۴۲۸)
 ('اوپو', ۰/۰۰۶۰۳۵۰۲۴۴۹۸۴۳۸۸۹۳)
 ('وان پلاس', ۰/۰۰۵۸۶۹۰۲۳۹۷۰۸۱۳۱۴۴)
 ('مایکروسافت', ۰/۰۰۱۳۳۶۴۷۹۷۸۸۵۳۲۸۳۳۷)
 ('گوگل', ۰/۰۰۱۰۵۸۴۹۱۹۹۲۵۳۴۴۶۹۷)
 ('جی ال ایکس', ۰/۰۰۰۸۳۶۹۲۸۳۵۱۶۰۶۲۲۹۴)
 ('سونی اریکسون', ۰/۰۰۰۶۵۹۸۳۹۱۶۴۱۷۰۷۳۳۴)
 ('میزو', ۰/۰۰۰۴۴۶۷۶۶۱۰۰۷۲۵۷۰۸۵)
 ('آمازون', e-05۹/۱۶۴۴۳۲۸۳۳۱۰۸۰۴)
 ('آی میت', e-05۶/۸۷۳۳۲۴۶۲۰۲۴۸۸۱۳)
 ('زد تی ای', ۰/۰)
 ('آلکاتل', ۰/۰)
 ('ویوو', ۰/۰۰۰۴۷۱۳۱۳۶۸۸۶۴۹۶۱۶۱)

رویکرد سوم: پیاده‌سازی با استفاده از رویکرد ترکیبی

پیاده‌سازی در فایل sentiment/hybrid.py موجود است.

روش پیشنهادی: در این قسمت از هر دو رویکرد SVM و مبتنی بر لغت‌نامه استفاده می‌کنیم. بدین صورت که ابتدا SVM توسط داده دارای برچسب (دیجیکالا) آموزش دیده می‌شود. سپس دقت به ازای داده تست, هم برای روش SVM و هم مبتنی بر لغت‌نامه به‌دست می‌آید. تا به اینجا کار قطبیت نظرات هم با svm و هم با لغت نامه پیش بینی شده است. سپس براساس confidence measure برای svm و تعداد کلمات مثبت یا منفی موجود در نظر برای روش مبتنی بر لغت نامه, تصمیم گیری می‌شود که قطبیت نهایی در روش ترکیبی کدام یک از قطبیت ها باشد.

خروجی به شرح زیر است:

hybrid approach accuracy: 0.8293577981651377

svm accuracy: 0.8165137614678899

lexicon based accuracy: 0.7027522935779816

تعیین محبوب‌ترین برند موبایل:

برای این منظور از داده خزیده شده از سایت mobile.ir استفاده شد. خروجی به شرح زیر است:

('سامسونگ', ۰/۰۸۳۹۱۰۷۲۳۸۰۳۸۴۷۱۶)
 ('شیائومی', ۰/۰۶۶۶۴۹۵۰۳۹۳۳۸۸۱۸۶)

('هوای', ۰/۰۴۶۱۳۱۰۶۳۲۴۵۶۱۴۶۹)
 ('آنر', ۰/۰۳۳۲۵۸۹۸۴۴۴۴۰۵۱۸۵)
 ('سونی', ۰/۰۲۱۵۲۰۱۶۳۷۳۷۰۰۳۸۰۴)
 ('نوکیا', ۰/۰۱۹۸۰۲۱۴۶۲۸۵۴۴۸۶۱۶)
 ('اچ تی سی', ۰/۰۱۱۰۶۵۷۶۵۰۶۲۸۳۹۰۱۳)
 ('اپل', ۰/۰۱۰۷۰۲۵۴۵۸۹۳۵۹۵۳)
 ('موتورولا', ۰/۰۱۰۶۴۳۶۸۸۱۲۸۸۶۰۰۵۵)
 ('لنوو', ۰/۰۰۶۷۶۵۶۷۰۱۴۴۹۶۹۸۳۱)
 ('ایسوس', ۰/۰۰۶۳۴۲۱۴۳۴۲۲۸۹۸۲۲۴)
 ('بلک بری', ۰/۰۰۵۷۹۱۵۲۳۰۹۸۸۲۱۸۷۶)
 ('اوپو', ۰/۰۰۴۱۱۲۶۲۹۴۳۶۳۰۶۲۸۸)
 ('وان پلاس', ۰/۰۰۳۹۸۸۶۴۳۱۵۲۸۱۶۱۹۷)
 ('ال جی', ۰/۰۰۳۳۱۰۶۰۹۰۹۰۶۸۸۵۵۱۲)
 ('مایکروسافت', ۰/۰۰۰۸۵۹۱۶۵۵۷۸۳۴۷۳۰۹۱)
 ('ویوو', ۰/۰۰۰۷۳۳۱۵۴۶۲۶۸۳۱۹۳۱۸)
 ('گوگل', ۰/۰۰۰۵۱۵۴۹۹۳۴۷۰۱۵۷۴۹۷)
 ('سونی اریکسون', ۰/۰۰۰۵۰۲۰۵۱۵۳۷۹۵۹۲۶۷۱)
 ('جی ال ایکس', ۰/۰۰۰۳۵۱۳۰۳۲۵۸۷۰۰۵۹۱)
 ('زد تی ای', ۰/۰۰۰۲۷۴۹۳۲۹۸۵۰۱۶۱۵۲۳)
 ('آمازون', ۰۵۹/۱۶۴۴۳۲۸۳۳۱۰۸۰۴-e-05)
 ('آلکاتل', ۰۲۴۸۸۱۳/۸۷۳۳۲۴۶۲۰۲۴۸۸۱۳-e-05)
 ('میزو', ۰/۰)
 ('آی میت', ۰۲۴۸۸۱۳-۸۷۳۳۲۴۶۲۰۲۴۸۸۱۳-e-05)

مقایسه و ارزیابی نهایی:

به منظور ارزیابی نهایی لیست محبوبترین برند های به دست آمده در پروژه، از رتبه بندی های اعلام شده توسط ۳ مرجع مختلف استفاده می کنیم:

مرجع اول: سایت gs.statcounter.com

لیست رتبه بندی این سایت: (در فایل [dataset/gs.statcounter_ranking.txt](#) نیز موجود است).

سامسونگ

هوای

اپل

ال جی

سونی

اچ تی سی

لنوو

شیائومی

نوکیا

مرجع دوم: سایت دیجیکالا:

لیست رتبه بندی این سایت: (در فایل `dataset/digikala_ranking.txt` نیز موجود است).

سامسونگ

نوکیا

هواوی

اپل

آنر

ایسوس

جی ال ایکس

جیمو

میزو

مرجع سوم: گزارش فصلی کافه‌بازار, مربوط به بهار ۹۸

لیست رتبه بندی این سایت (تنها برای گوشی‌های اندرویدی): (در فایل `dataset/cafebazar_ranking.txt` نیز موجود است).

سامسونگ

هواوی

ال جی

سونی

لنوو

اچ تی سی

شیائومی

ایسوس

جی ال ایکس

برای تعیین همبستگی میان لیست‌ها از معیار `spearman` استفاده شد. پیاده‌سازی در ادرس `/sentiment/correlation.py` موجود است. مقایسه لیست محبوب‌ترین برند تعیین شده توسط الگوریتم‌های ما و رتبه‌بندی‌های موجود در ۳ مرجع فوق در جدول ۱ آمده است:

	دجیکالا	کافهبازار	gs.statcounter.com
SVM	۰/۶۷۶	۰/۸۴۰	۰/۹۱۷
Lexicon based	۰/۷۲۳	۰/۸۷	۰/۹۳۳
Hybrid	۰/۶۳	۰/۸۱۸	۰/۹۰۷

جدول ۱ همبستگی اسپیرمن میان سه الگوریتم پیاده‌سازی شده و سه مرجع دیجیکالا, کافهبازار و سایت gs.statcounter.com

خروجی‌ها برای روش ترکیبی :

```
***** calculate spearman correlation *****
['سامسونگ' و 'هواوی' و 'اپل' و 'ال جی' و 'سونی' و 'لنوو' و 'اچ تی سی' و 'شیائومی' و 'ایسوس' و 'جی ال ایکس' و 'آئر' و 'نوکیا' و 'اپل' و 'موتورولا' و 'بلک بری' و
'اوپو' و 'وان پلاس' و 'مایکروسافت' و 'گوگل' و 'سونی اریکسون' و 'زد تی ای' و 'آلکاتل' و 'آی میت' و 'آمازون' و 'ویوو' و 'میزو']
['سامسونگ' و 'شیائومی' و 'هواوی' و 'آئر' و 'سونی' و 'نوکیا' و 'اپل' و 'اچ تی سی' و 'موتورولا' و 'لنوو' و 'ایسوس' و 'بلک بری' و 'اوپو' و 'وان پلاس' و 'ال
جی' و 'مایکروسافت' و 'گوگل' و 'سونی اریکسون' و 'زد تی ای' و 'آلکاتل' و 'جی ال ایکس' و 'آی میت' و 'آمازون' و 'ویوو' و 'میزو']
list1 ranks: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]
list2 ranks: [1, 3, 15, 5, 10, 8, 2, 11, 21, 4, 6, 7, 9, 12, 13, 14, 16, 17, 18, 19, 20, 22, 23, 24, 25]
cafebazar correlation: 0.8184615384615385
*****
['سامسونگ' و 'هواوی' و 'اپل' و 'ال جی' و 'سونی' و 'لنوو' و 'اچ تی سی' و 'شیائومی' و 'نوکیا' و 'آئر' و 'موتورولا' و 'ایسوس' و 'بلک بری' و 'اوپو' و 'وان
پلاس' و 'مایکروسافت' و 'گوگل' و 'سونی اریکسون' و 'زد تی ای' و 'آلکاتل' و 'جی ال ایکس' و 'آی میت' و 'آمازون' و 'ویوو' و 'میزو']
['سامسونگ' و 'شیائومی' و 'هواوی' و 'آئر' و 'سونی' و 'نوکیا' و 'اپل' و 'اچ تی سی' و 'موتورولا' و 'لنوو' و 'ایسوس' و 'بلک بری' و 'اوپو' و 'وان پلاس' و 'ال
جی' و 'مایکروسافت' و 'گوگل' و 'سونی اریکسون' و 'زد تی ای' و 'آلکاتل' و 'جی ال ایکس' و 'آی میت' و 'آمازون' و 'ویوو' و 'میزو']
list1 ranks: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]
list2 ranks: [1, 3, 7, 15, 5, 8, 10, 2, 6, 4, 9, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]
gs statcounter correlation: 0.9076923076923077
*****
['سامسونگ' و 'نوکیا' و 'هواوی' و 'اپل' و 'آئر' و 'ایسوس' و 'جی ال ایکس' و 'جیمو' و 'میزو' و 'شیائومی' و 'سونی' و 'اچ تی سی' و 'موتورولا' و 'لنوو' و 'بلک
بری' و 'اوپو' و 'وان پلاس' و 'ال جی' و 'مایکروسافت' و 'گوگل' و 'سونی اریکسون' و 'زد تی ای' و 'آلکاتل' و 'آی میت' و 'آمازون' و 'ویوو']
['سامسونگ' و 'شیائومی' و 'هواوی' و 'آئر' و 'سونی' و 'نوکیا' و 'اپل' و 'اچ تی سی' و 'موتورولا' و 'لنوو' و 'ایسوس' و 'بلک بری' و 'اوپو' و 'وان پلاس' و 'ال
جی' و 'مایکروسافت' و 'گوگل' و 'سونی اریکسون' و 'زد تی ای' و 'آلکاتل' و 'جی ال ایکس' و 'آی میت' و 'آمازون' و 'ویوو' و 'جیمو']
list1 ranks: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26]
list2 ranks: [1, 6, 3, 7, 4, 11, 21, 26, 25, 2, 5, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24]
digikala correlation: 0.6348717948717948
end :)
```

تصویر خروجی همبستگی اسپیرمن در روش ترکیبی

خروجی‌ها برای روش مبتنی بر لغت‌نامه:

```
***** calculate spearman correlation *****
['سامسونگ' و 'هواوی' و 'ال جی' و 'سونی' و 'لنوو' و 'اچ تی سی' و 'شیائومی' و 'ایسوس' و 'جی ال ایکس' و 'آنر' و 'نوکیا' و 'اپل' و 'موتورولا' و 'بلک‌بری' و
'اوپو' و 'وان پلاس' و 'مایکروسافت' و 'گوگل' و 'سونی اریکسون' و 'میزو' و 'آمازون' و 'آی میت' و 'زد تی ای' و 'آلکاتل' و 'ویوو']
['سامسونگ' و 'شیائومی' و 'هواوی' و 'آنر' و 'نوکیا' و 'سونی' و 'اپل' و 'اچ تی سی' و 'ال جی' و 'موتورولا' و 'ایسوس' و 'بلک‌بری' و 'لنوو' و 'اوپو' و 'وان
پلاس' و 'مایکروسافت' و 'گوگل' و 'جی ال ایکس' و 'سونی اریکسون' و 'میزو' و 'آمازون' و 'آی میت' و 'زد تی ای' و 'آلکاتل' و 'ویوو']
list1 ranks: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]
list2 ranks: [1, 3, 9, 6, 13, 8, 2, 11, 18, 4, 5, 7, 10, 12, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25]
cafebazar correlation: 0.87
*****
['سامسونگ' و 'هواوی' و 'اپل' و 'ال جی' و 'سونی' و 'اچ تی سی' و 'لنوو' و 'شیائومی' و 'نوکیا' و 'آنر' و 'موتورولا' و 'ایسوس' و 'بلک‌بری' و 'اوپو' و 'وان
پلاس' و 'مایکروسافت' و 'گوگل' و 'جی ال ایکس' و 'سونی اریکسون' و 'میزو' و 'آمازون' و 'آی میت' و 'زد تی ای' و 'آلکاتل' و 'ویوو']
['سامسونگ' و 'شیائومی' و 'هواوی' و 'آنر' و 'نوکیا' و 'سونی' و 'اپل' و 'اچ تی سی' و 'ال جی' و 'موتورولا' و 'ایسوس' و 'بلک‌بری' و 'لنوو' و 'اوپو' و 'وان
پلاس' و 'مایکروسافت' و 'گوگل' و 'جی ال ایکس' و 'سونی اریکسون' و 'میزو' و 'آمازون' و 'آی میت' و 'زد تی ای' و 'آلکاتل' و 'ویوو']
list1 ranks: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]
list2 ranks: [1, 3, 7, 9, 6, 8, 13, 2, 5, 4, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]
gs statcounter correlation: 0.933076923076923
*****
['سامسونگ' و 'نوکیا' و 'هواوی' و 'اپل' و 'آنر' و 'ایسوس' و 'جی ال ایکس' و 'جیمو' و 'میزو' و 'شیائومی' و 'سونی' و 'اچ تی سی' و 'ال جی' و 'موتورولا' و 'بلک
بری' و 'لنوو' و 'اوپو' و 'وان پلاس' و 'مایکروسافت' و 'گوگل' و 'سونی اریکسون' و 'آمازون' و 'آی میت' و 'زد تی ای' و 'آلکاتل' و 'ویوو']
['سامسونگ' و 'شیائومی' و 'هواوی' و 'آنر' و 'نوکیا' و 'سونی' و 'اپل' و 'اچ تی سی' و 'ال جی' و 'موتورولا' و 'ایسوس' و 'بلک‌بری' و 'لنوو' و 'اوپو' و 'وان
پلاس' و 'مایکروسافت' و 'گوگل' و 'جی ال ایکس' و 'سونی اریکسون' و 'میزو' و 'آمازون' و 'آی میت' و 'زد تی ای' و 'آلکاتل' و 'جیمو']
list1 ranks: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26]
list2 ranks: [1, 5, 3, 7, 4, 11, 18, 26, 20, 2, 6, 8, 9, 10, 12, 13, 14, 15, 16, 17, 19, 21, 22, 23, 24, 25]
```

تصویر ۲ خروجی همبستگی اسپیرمن در روش مبتنی بر لغت‌نامه

خروجی‌ها برای روش SVM:

```
***** calculate spearman correlation *****
['سامسونگ' و 'هواوی' و 'ال جی' و 'سونی' و 'لنوو' و 'اچ تی سی' و 'شیائومی' و 'ایسوس' و 'جی ال ایکس' و 'آنر' و 'نوکیا' و 'اپل' و 'موتورولا' و 'بلک‌بری' و
'اوپو' و 'وان پلاس' و 'سونی اریکسون' و 'مایکروسافت' و 'ویوو' و 'گوگل' و 'میزو' و 'زد تی ای' و 'آلکاتل' و 'آی میت' و 'آمازون']
['سامسونگ' و 'شیائومی' و 'هواوی' و 'آنر' و 'سونی' و 'نوکیا' و 'اچ تی سی' و 'اپل' و 'موتورولا' و 'بلک‌بری' و 'لنوو' و 'ایسوس' و 'ال جی' و 'اوپو' و 'وان
پلاس' و 'سونی اریکسون' و 'مایکروسافت' و 'ویوو' و 'گوگل' و 'جی ال ایکس' و 'میزو' و 'زد تی ای' و 'آلکاتل' و 'آی میت' و 'آمازون']
list1 ranks: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]
list2 ranks: [1, 3, 13, 5, 11, 7, 2, 12, 20, 4, 6, 8, 9, 10, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25]
cafebazar correlation: 0.8400000000000001
*****
['سامسونگ' و 'هواوی' و 'اپل' و 'ال جی' و 'سونی' و 'اچ تی سی' و 'لنوو' و 'شیائومی' و 'نوکیا' و 'آنر' و 'موتورولا' و 'بلک‌بری' و 'ایسوس' و 'اوپو' و 'وان
پلاس' و 'سونی اریکسون' و 'مایکروسافت' و 'ویوو' و 'گوگل' و 'جی ال ایکس' و 'میزو' و 'زد تی ای' و 'آلکاتل' و 'آی میت' و 'آمازون']
['سامسونگ' و 'شیائومی' و 'هواوی' و 'آنر' و 'سونی' و 'نوکیا' و 'اچ تی سی' و 'اپل' و 'موتورولا' و 'بلک‌بری' و 'لنوو' و 'ایسوس' و 'ال جی' و 'اوپو' و 'وان
پلاس' و 'سونی اریکسون' و 'مایکروسافت' و 'ویوو' و 'گوگل' و 'جی ال ایکس' و 'میزو' و 'زد تی ای' و 'آلکاتل' و 'آی میت' و 'آمازون']
list1 ranks: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]
list2 ranks: [1, 3, 8, 13, 5, 7, 11, 2, 6, 4, 9, 10, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]
gs statcounter correlation: 0.9176923076923077
*****
['سامسونگ' و 'نوکیا' و 'هواوی' و 'اپل' و 'آنر' و 'ایسوس' و 'جی ال ایکس' و 'جیمو' و 'میزو' و 'شیائومی' و 'سونی' و 'اچ تی سی' و 'موتورولا' و 'بلک‌بری' و
'لنوو' و 'ال جی' و 'اوپو' و 'وان پلاس' و 'سونی اریکسون' و 'مایکروسافت' و 'ویوو' و 'گوگل' و 'زد تی ای' و 'آلکاتل' و 'آی میت' و 'آمازون']
['سامسونگ' و 'شیائومی' و 'هواوی' و 'آنر' و 'سونی' و 'نوکیا' و 'اچ تی سی' و 'اپل' و 'موتورولا' و 'بلک‌بری' و 'لنوو' و 'ایسوس' و 'ال جی' و 'اوپو' و 'وان
پلاس' و 'سونی اریکسون' و 'مایکروسافت' و 'ویوو' و 'گوگل' و 'جی ال ایکس' و 'میزو' و 'زد تی ای' و 'آلکاتل' و 'آی میت' و 'آمازون']
list1 ranks: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26]
list2 ranks: [1, 6, 3, 8, 4, 12, 20, 26, 21, 2, 5, 7, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 22, 23, 24, 25]
digikala correlation: 0.6765811965811965
end :)
```

تصویر ۱ خروجی همبستگی اسپیرمن در روش مائسین بردار پشتیبان

