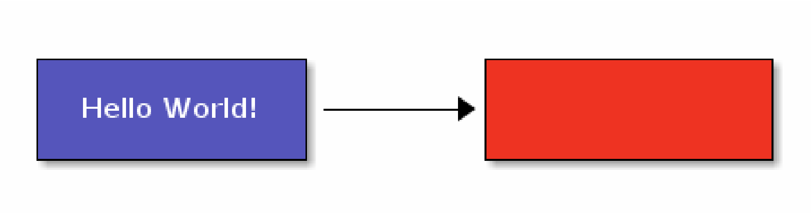
Metamorphic Testing of Cross-Language Sentiment Analysis

Boyang

October 26, 2018

Contents



1 Introduction

The purpose of this research is assessing the quality of translation tool and quality of sentiment analysis tool. Finally, we will achieve a method, finding which translation tool conbining with which sentiment analysis tool together, for getting better sentiment analysis result. Currently, most of sentiment analysis tool only support English. We want to find a method for let non-English people using English Sentiment Analysis tool with Machine translation tool, analysis their non-English text. In the lit review I will include Metamorphic Testing Method and Machine translated. Currently, I have not found

2 Test Data

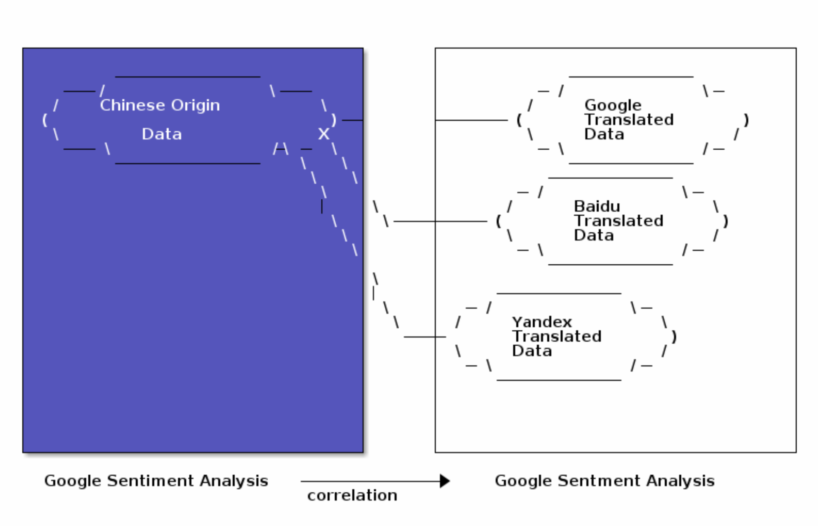
Total have 46180 movies reviews.

Ranking	Number of Test Data	Percentage
Ranking 10	7353	15.92 %
Ramking 20	11209	24.27 %
Ranking 30	16223	35.13 %
Ranking 40	7663	16.59 %
Ranking 50	3732	8.08 %

3 Normalization

v = (v-min)/(max-min) * (newmax-newmin) + newmin

4 Assessing Machine translation tool quality



4.1 Method

1. Compare correlation coefficient between Chinese sentiment analysis results and English sentiment analysis results by each

(a) Using Google, Baidu, Yandex translation tools, translated original Chinese data to English data

(b) Using same sentiment analysis tool analysis original chinese dataset and translated dataset

(c) Calculate correlation coefficient between Chinese sentiment analysis results and English sentiment analysis results

(d) Compare correlation coefficient values. if value is bigger than others, we can say this translation tool, which use in original dataset to English dataset, can achieve better results than others.

4.1.1 Result

- Base on Google sentiment analysis tool

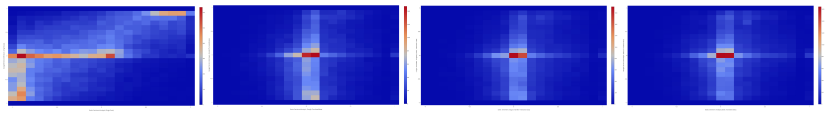
	Google Score for Google translated data	Google Score for Yandex trans
Gooogle Score for origin data	0.512 (Pearson Correlations) p-value: 0.0	0.506 (Pearson Correlations) p
Google Score for origin data	0.381 (Kendall Correlations) p-value: 0.0	0.375 (Kendall Correlations) p
Google Score for origin data	0.504 (Spearman Correlations) p-value: 0.0	0.497 (Spearman Correlations)
Gooogle Score for origin data	0.512 (Point Biserial) p-value: 0.0	0.506 (Point Biserial) p-value:

- Google translation tool quality > Yandex translation tool quality > Baidu translation tool quality

- Base on Baidu sentiment analysis tool

	Baidu Positive Probability for Google translated data	Baidu P
Baidu Positive Probability for origin data	0.288 (Pearson Correlations) p-value: 0.0	0.280 (P
Baidu Positive Probability for origin data	0.188 (Kendall Correlations) p-value: 0.0	0.174 (K
Baidu Positive Probability for origin data	0.271 (Spearman Correlations) p-value: 0.0	0.249 (S
Baidu Positive Probability for origin data	0.288 (Point Biserial) p-value: 0.0	0.280 (P

- Google translation tool quality > Yandex translation tool quality > Baidu translation tool quality



1. Compare correlation coefficient between sentiment analysis results and user rating

   (a) Divide the sentiment analysis scores between [-1,1] into 5 regions, which are 10, 20, 30, 40, 50. Each interval is 0.4.

regions	scope
10	[-1, -0.6]
20	(-0.6, -0.2]
30	(-0.2, 0.2)
40	[0.2, 0.6)
50	[0.6, 1]

1. calculate correlation coefficient between regions and user rating

- Base on Google sentiment analysis tool

	Google score for Google Translated data	Google score for Yandex Translated data	G
ranking	0.3639 (pearson Correlation) p-value: 0.0	0.3621 (pearson correlation) p-value: 0.0	0
ranking	0.3645 (spearman correlation) p-value: 0.0	0.3623 (spearman correlation) p-value: 0.0	0
ranking	0.3689 (point biserial correlation) p-value: 0.0	0.3621 (point biserial correlation) p-value: 0.0	0
ranking	0.3009 (kendall correlation) p-value: 0.0	0.2999 (kendall correlation) p-value: 0.0	0

- Base on Baidu sentiment analysis tool

	Baidu sentiment score (Google standard) for Google Translated data	Baidu sentiment score (Go
ranking	0.1638 (pearson correlation) p-value: 0.0	0.1645 (pearson correlation
ranking	0.1687 (spearman correlation) p-value: 0.0	0.1682 (spearman correlati
ranking	0.1638 (point biserial correlation) p-value: 0.0	0.1645 (point biserial corre
ranking	0.1455 (kendall correlation) p-value: 0.0	0.1447 (kendall correlation

1. draw heatmap

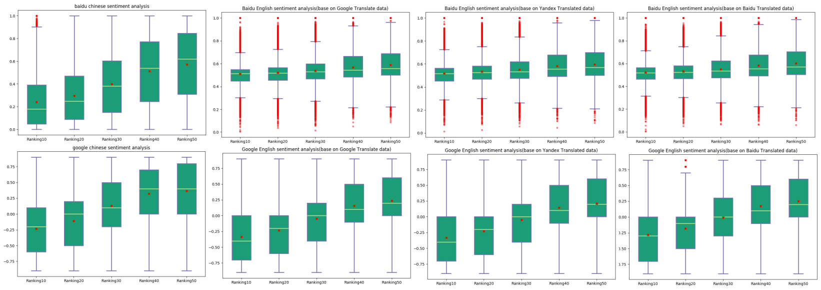
   (a) Divide the sentiment analysis scores between [-1, 1] into 20 regions, 0, 1, 2, 3 . . . 18, 19. Each interval is 0.1

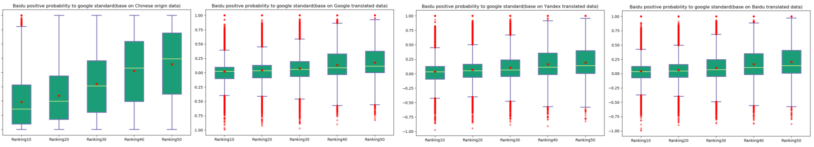
and then calculate the correlation and draw the heatmaps between the user rating (i.e., 10, 20, , 50) and sentiment analysis scores (for heatmap, use higher resolutions by dividing the region [-1,1] into 20 subregions to give a 20x5 heatmap).

5 Assessing Sentiment analysis tool quality

5.1 Chinese Sentiment analysis quality

	Baidu Chinese sentiment analysis tool's score (Google Standard)	Google Chinese sentiment anal
ranking	0.3552 (pearson Correlation) p-value: 0.0	0.3897 (pearson correlation) p-
ranking	0.3419 (spearman correlation) p-value: 0.0	0.3982 (spearman correlation)
ranking	0.3552 (point biserial) p-value: 0.0	0.3897 (point biserial correlatio
ranking	0.2842 (kendall correlation) p-value: 0.0	0.3281 (kendall correlation) p-





```
 +--------------------------+                              +--------------------------+
 |    Chinese   Ranking 10  |                              |   Yandex      Ranking 10 |
 |    origin    Ranking 20  |    Yandex translation tool   |   translated  Ranking 20 |
 |    data      Ranking 30  +----------------------------->|   Data        Ranking 30 |
 |              Ranking 40  |                              |               Ranking 40 |
 +              Ranking 50  +                              +               Ranking 50 +
 +--------------------------+                              +--------------------------+
                   -\                                              /
                     -\                                           /
                       -\                                        /
                         -\               -------------------   /
                           -\       -----/                   \-/---
                             -\  --/     using same                \--
                               --        sentiment analysis tool      \
                               (         analysis both side            )
                                \                                     /
                                 --\                               /--
                                    -----\                   /-----
                                          ----------+--------
                                                    |
                                           -----------------------
                                     -----/                       \-----
                                   -/       calculate Mean Value        \-
                               ----         by different ranking          )
                        ------/    -\       for both sild               /-
                 ------/             -----\                       /--\--
+---------------/-----------+              -----------------------    \
|                          |
|                          |
|                          |
|                          |
|                          |
|                          |
+--------------------------+
```

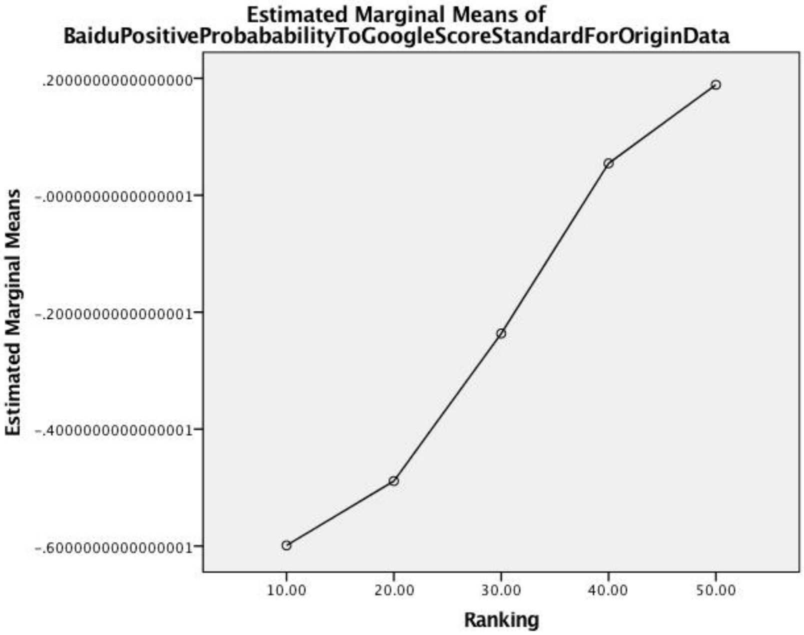
```
+----------+--------------+
|                         |
|                         |
|                         |
|                         |
|                         |
|                         |
+-------------------------+
```

```
 linear regression
```

```
liner regression
```

```
slope
```

```
slope
```



5.2 Google Chinese sentiment analysis boxplot

language=Python,label= ,caption= ,captionpos=b,numbers=none import numpy as np from openpyxl import load$_{w}orkbookimportmatplotlibasmpl$

agg backend is used to create plot as a .png file mpl.use('agg')

import matplotlib.pyplot as plt

def drawBoxPlots(title, data$_t o_p lot, ax)$:

Create the boxplot bp = ax.boxplot(data$_t o_p lot, patch_a rtist = True, showmeans = True)changeoutlinecolor, fillcolorandlinewidthoftheboxesforboxinbp['boxes'] : changeoutlinecolorbox.set(color =' 7570b3', linewidth = 2)changefillcolorbox.set(facecolor =' 1b9e77')$

change color and linewidth of the whiskers for whisker in bp['whiskers']: whisker.set(color='7570b3', linewidth=2)

change color and linewidth of the caps for cap in bp['caps']: cap.set(color='7570b3', linewidth=2)

change color and linewidth of the medians for median in bp['medians']: median.set(color='b2df8a', linewidth=2)

change the style of fliers and their fill for flier in bp['fliers']: flier.set(marker='o', markerfacecolor='red', markersize=5, markeredgewidth=0.0, alpha=0.5)

for mean in bp['means']: mean.set(marker = 's', markerfacecolor='red')

Custom x-axis labels ax.set$_x ticklabels([' Ranking10',' Ranking20',' Ranking30',' Ranking40',' Ranking50']$

Create data "' np.random.seed(10) ranking10 = np.random.normal(100, 10, 200) ranking20 = np.random.normal(80, 30, 200) ranking30 = np.random.normal(90, 20, 200) ranking40 = np.random.normal(70, 25, 200) ranking50 = np.random.normal(70, 25, 200) "' ranking10 = np.array([]) ranking20 = np.array([]) ranking30 = np.array([]) ranking40 = np.array([]) ranking50 = np.array([]) wb = load$_w orkbook(filename =' good.xlsx', read_o nly = True)ws = wb['Sheet1']$

for row in range(1, 46181): for row in range(1, 10): ranking = ws.cell(row=row, column=20).value) value = ws.cell(row=row, column=17) if ranking == 10: ranking10 = np.append(ranking10 , value) elif ranking == 20: ranking20 = np.append (ranking20, value) elif ranking == 30: ranking30 = np.append (ranking30, value) elif ranking == 40: ranking40 = np.append (ranking40, value) elif ranking == 50: ranking50 = np.append (ranking50, value) "' ranking20.append([0]) ranking30.append([1]) ranking40.append([3]) ranking50.append([4]) "' combine these different collections into a list data$_t o_p lot = [ranking10, ranking20, ranking30, ranking40, ran$

fig, axes = plt.subplots(nrows=2, ncols=4, figsize=(9, 4)) Create a figure instance fig = plt.figure(1, figsize=(9, 6)) Create an axes instance ax = fig.add$_s ubplot(111)addpatch_a rtist = Trueoptiontoax.boxplot()togetfillcolor$

drawBoxPlots("google chinese sentiment analysis", data$_t o_p lot, ax)Savethefigurefig.savefig("googleChine$ $tight')$

File "<stdin>", line 1, in <module> File "/tmp/babel-MzHCZL/python-Me3jED", line 64 ranking = ws.cell(row=row, column=20).value) ^ SyntaxError: invalid syntax]] File "<stdin>", line 1, in <module> File "/tmp/babel-MzHCZL/python-foVeeD", line 64 ranking = ws.cell(row=row, column=3).value) ^ SyntaxError: invalid syntax]] [[`Python` 3.6.4 (default, Jan 5 2018, 02:35:40) [GCC 7.2.1 20171224] on linux Type "help", "copyright", "credits" or "license" for more information. Traceback (most recent call last): File "<stdin>", line 1, in <module> File "/tmp/babel-MzHCZL/python-kKjo78", line 64 ranking = ws.cell(row=row, column=3).value) ^ SyntaxError: invalid syntax python.el: native completion setup loaded]] File "<stdin>", line 1, in <module> File

"/tmp/babel-Xbwqve/python-4683rz", line 63, in <module> print (ws.cell(row=row, column=7).value) File "/usr/lib/python3.6/site-packages/openpyxl/worksheet/worksheet.py", line 307, in cell raise ValueError("Row or column values must be at least 1") ValueError: Row or column values must be at least 1]] [[Python 3.6.4 (default, Jan 5 2018, 02:35:40) [GCC 7.2.1 20171224] on linux Type "help", "copyright", "credits" or "license" for more information. Traceback (most recent call last): File "<stdin>", line 1, in <module> File "/tmp/babel-Xbwqve/python-3h66Xh", line 63, in <module> print (ws.cell(row=row, column=7).value) File "/usr/lib/python3.6/site-packages/openpyxl/worksheet/worksheet.py", line 306, in cell if row < 1 or column < 1: TypeError: '<' not supported between instances of 'tuple' and 'int' python.el: native completion setup loaded]]

chinese origin data	Google translated data	baidu tran

Linear regression slope

6 Method for better compound mode for sentiment analysis tool and machine translation tool