

Metamorphic Testing of Cross-Language Sentiment Analysis

Boyang

September 28, 2018

Contents

1	Introduction	1
2	Test Data	1
3	Normalization	2
4	Assessing Machine translation tool quality	2
4.1	Method	3
4.1.1	Result	5
5	Assessing Sentiment analysis tool quality	7
5.1	Chinese Sentiment analysis quality	7
6	Method for better compound mode for sentiment analysis tool and machine translation tool	9

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 combining 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

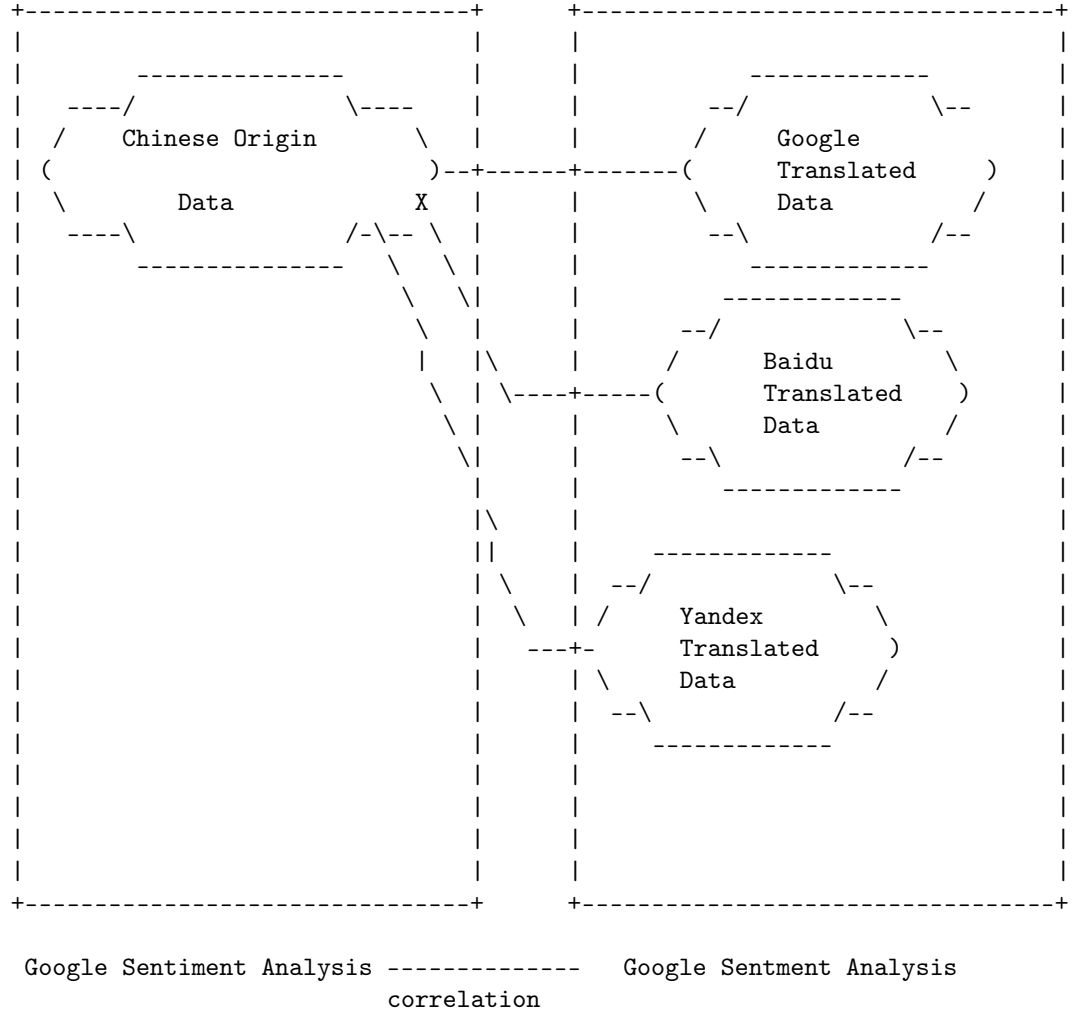
Total have 46180 movies reviews.

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

3 Normalization

$$v' = (v - \min) / (\max - \min) * (\text{newmax} - \text{newmin}) + \text{newmin}$$

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

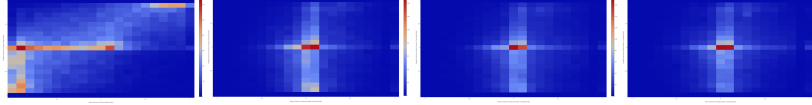
Google Score for Google translated data		Google Score for Yandex translated data		Google Score Baidu translated data	
Google Score for origin data	0.512 (Pearson Correlations) p-value: 0.0	0.506 (Pearson Correlations) p-value: 0.0	0.490 (Pearson Correlations) p-value: 0.0	0.490 (Pearson Correlations) p-value: 0.0	0.490 (Pearson Correlations) p-value: 0.0
Google Score for origin data	0.381 (Kendall Correlations) p-value: 0.0	0.375 (Kendall Correlations) p-value: 0.0	0.363 (Kendall Correlations) p-value: 0.0	0.363 (Kendall Correlations) p-value: 0.0	0.363 (Kendall Correlations) p-value: 0.0
Google Score for origin data	0.504 (Spearman Correlations) p-value: 0.0	0.497 (Spearman Correlations) p-value: 0.0	0.482 (Spearman Correlations) p-value: 0.0	0.482 (Spearman Correlations) p-value: 0.0	0.482 (Spearman Correlations) p-value: 0.0
Google Score for origin data	0.512 (Point Biserial) p-value: 0.0	0.506 (Point Biserial) p-value: 0.0	0.490 (Point Biserial) p-value: 0.0	0.490 (Point Biserial) p-value: 0.0	0.490 (Point Biserial) p-value: 0.0

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

- 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	
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 (Google standard) for Yandex Translated data	
ranking	0.1638 (pearson correlation) p-value: 0.0	0.1645 (pearson correlation) p-value: 0.0	0
ranking	0.1687 (spearman correlation) p-value: 0.0	0.1682 (spearman correlation) p-value: 0.0	0
ranking	0.1638 (point biserial correlation) p-value: 0.0	0.1645 (point biserial correlation) p-value: 0.0	0
ranking	0.1455 (kendall correlation) p-value: 0.0	0.1447 (kendall correlation) p-value: 0.0	0

1. draw heatmap

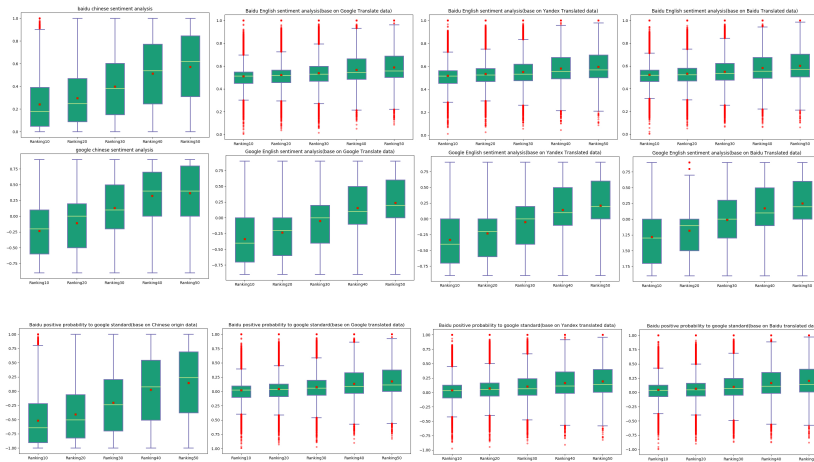
- (a) Divide the sentiment analysis scores between $[-1, 1]$ into 20 regions, 0, 1, 2, 3 ... 18, 19. Each interval is 0.1

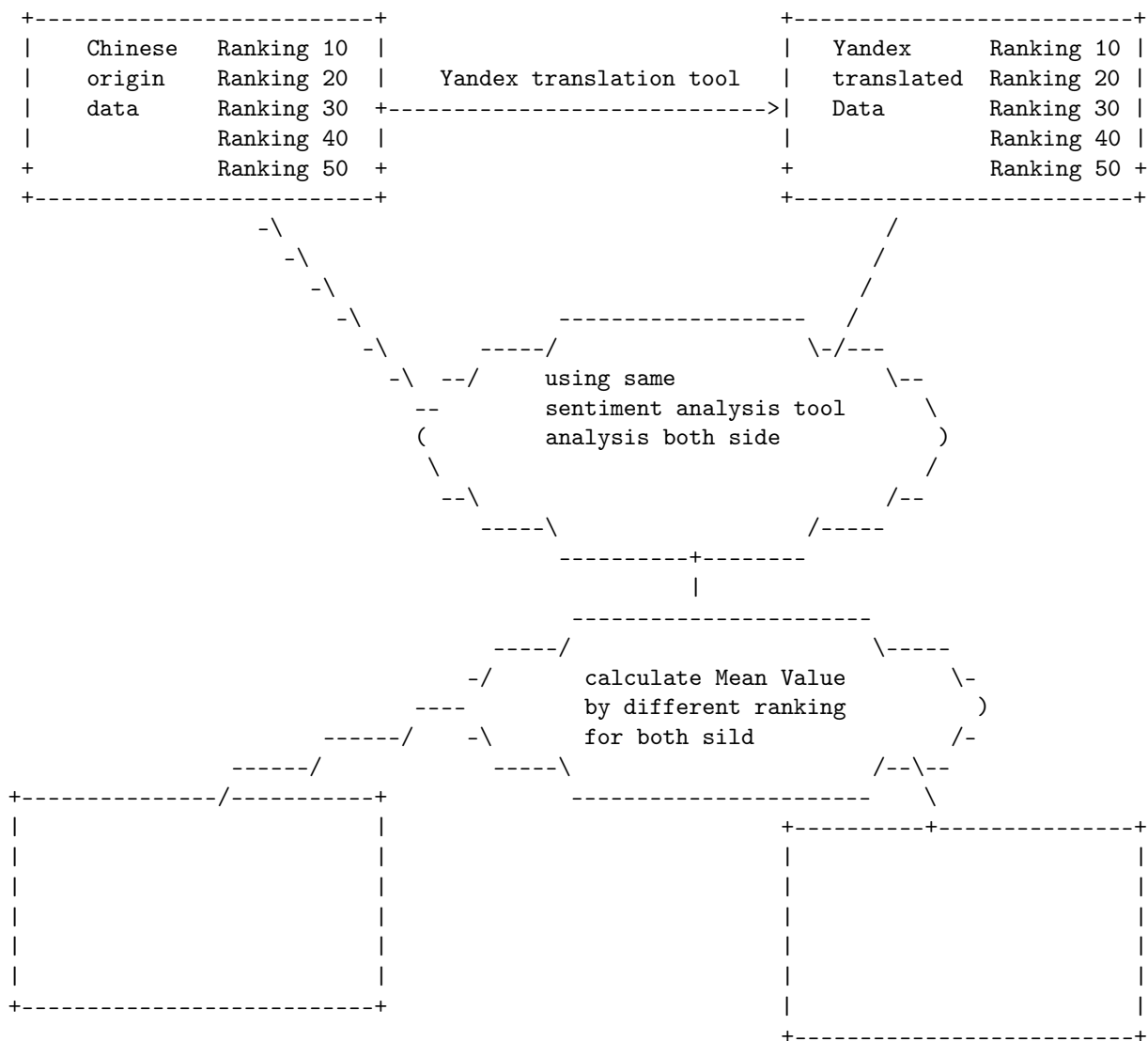
and then calculate the correlation and draw the heatmaps between the user rating (i.e., 10, 20, ..., 50) and sentiment analysis scores (for heatmap, use higher resolutions by dividing the region $[-1,1]$ into 20 subregions to give a 20x5 heatmap).

5 Assessing Sentiment analysis tool quality

5.1 Chinese Sentiment analysis quality

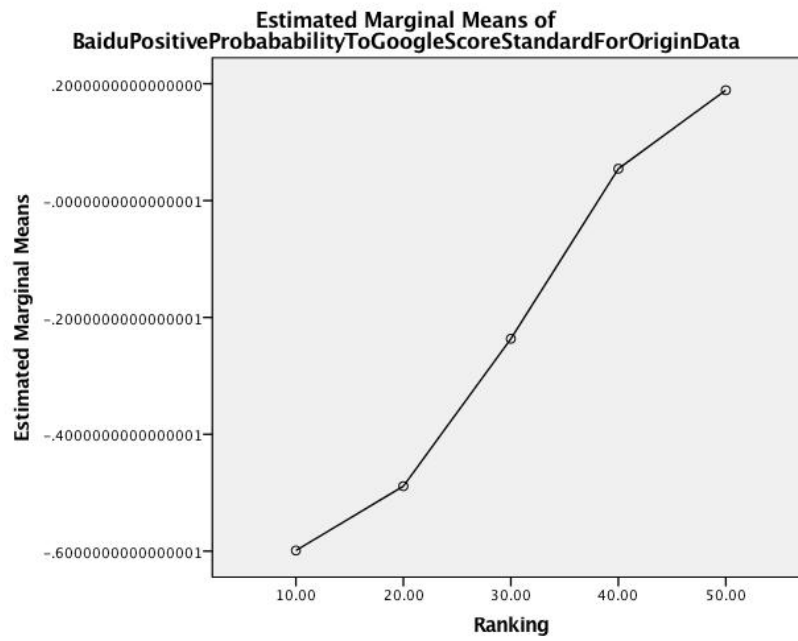
	Baidu Chinese sentiment analysis tool's score (Google Standard)	Google Chinese sentiment analysis tool's score (Google Standard)
ranking	0.3552 (pearson Correlation) p-value: 0.0	0.3897 (pearson correlation) p-value: 0.0
ranking	0.3419 (spearman correlation) p-value: 0.0	0.3982 (spearman correlation)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





slope

slope



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