Contents

1   Each colom represent mean

Column	Mean	remark
F (6)	ranking	
G (7)	Chinese origin data	
H (8)	Google translated data	
I (9)	Baidu Chinese sentiment analysis positive probability	base on Chinese origin data
J (10)	Baidu Chinese sentiment analysis confidence	base on Chinese origin data
K (11)	Baidu Chinese sentiment analysis Negative probability	base on Chinese origin data
L (12)	Baidu Chinese sentiment analysis the category	base on Chinese origin data
M (13)	Baidu English sentiment analysis positive probability	base on Google translated data
N (14)	Baidu English sentiment analysis confidence	base on Google translated data
O (15)	Baidu English sentiment analysis Negative probability	base on Google translated data
P (16)	Baidu English sentiment analysis the category	base on Google translated data
Q (17)	Google Chinese sentiment analysis score	base on Chinese origin data
R (18)	Google Chinese sentiment analysis manitude	base on Chinese origin data
S (19)	Google English sentiment analysis score	base on Google translated data
T (20)	Google English sentiment analysis manitude	base on Google translated data
U (21)	Yandex translated data	base on Chinese origin data
V (22)	Google English sentiment analysis score	base on Yandex translated data
W (23)	Google English sentiment analysis manitude	base on Yandex translated data
X (24)	Baidu translated data	
Y (25)	Google English sentiment analysis score	base on Baidu translated data
Z (26)	Google English sentiment analysis manitude	base on Baidu translated data

Continued from previous page

Column	Mean	remark
AA (27)	Baidu English sentiment analysis postitive probability	base on baidu translated data
AB (28)	Baidu English sentiment analysis confidence	base on baidu translated data
AC (29)	Baidu English sentiment analysis Negative probability	base on baidu translated data
AD (30)	Baidu English sentiment analysis the category	base on baidu translated data
AE (31)	Baidu postitive probability change to Google score standard	base on origin data Baidu sentiment analysis
AF (32)	Baidu postitive probability tranform to Google score standard	base on column M
AG (33)	Baidu English sentiment analysis postitive probability	base on Yandex translated data
AH (34)	Baidu English sentiment analysis confidence	base on Yandex translated data
AI (35)	Baidu English sentiment analysis Negative probability	base on Yandex translated data
AJ (36)	Baidu English sentiment analysis the category	base on Yandex translated data

1.1 note

- Baidu sentiment analysis Category Note

  - 2 mean to belong to the positive category, 1 mean to belong to the neutral category, and 0 mean to belong to the Negative category

- Google sentiment analysis Note

  - The score of a document's sentiment indicates the overall emotion of a document. The magnitude of a document's sentiment indicates how much emotional content is present within the document, and this value is often proportional to the length of the document.

  - A document with a neutral score (around 0.0) may indicate a low-emotion document, or may indicate mixed emotions, with both high positive and negative values which cancel each out. Generally, you can use magnitude values to disambiguate these cases, as truly neutral documents will have a low magnitude value, while mixed documents will have higher magnitude values.

  - "Clearly positive" and "clearly negative" sentiment varies for different use cases and customers. You might find differing results for your specific scenario. We recommend that you define a threshold that works for you, and then adjust the threshold after testing and verifying the results. For example, you may define a threshold of any score over 0.25 as clearly positive, and then modify the score threshold to 0.15 after reviewing your data and results and finding that scores from 0.15-0.25 should be considered positive as well.

2 Baidu sentiment analysis

2.1 Baidu Chinese sentiment analysis

2.1.1 Baidu Chinese sentiment analysis positive probability histogram
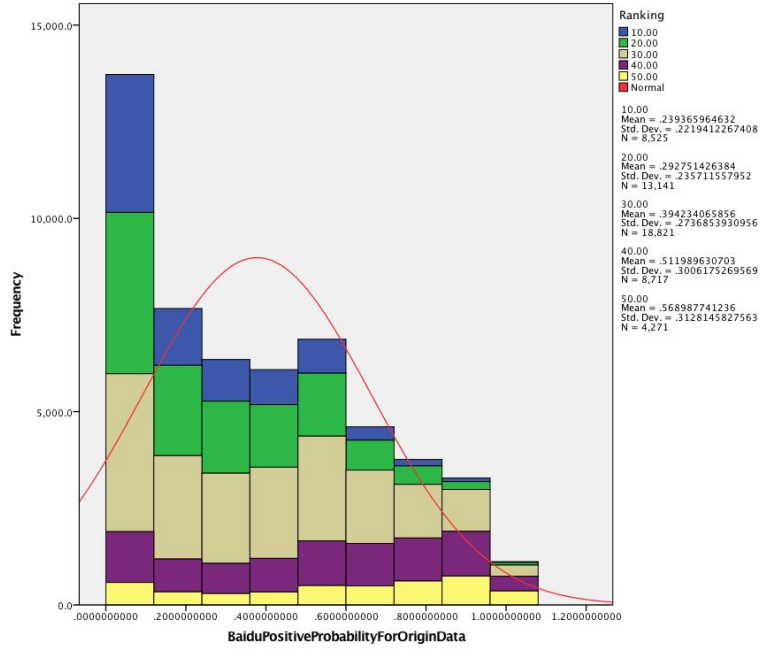
2.1.2 Baidu Chinese sentiment analysis postitive probability compare with different ranking(origin data)

Ranking	Mean	Valid N	std.deviation	Total N	Minimum	Maximum
Ranking 10	0.239365965000	8525	0.2219412270000	8572	0.000106	1.000000
Ranking 20	0.292751426000	13141	0.2357115580000	13226	0.000162	1.000000
Ranking 30	0.394234	18821	0.273685	18974	0.000214	1.000000
Ranking 40	0.511990	8717	0.300618	8790	0.001050	1.000000
Ranking 50	0.568988	4271	0.312815	4307	0.000536	1.000000

Estimated Marginal Means of BaiduPositiveProbabilityForOriginData
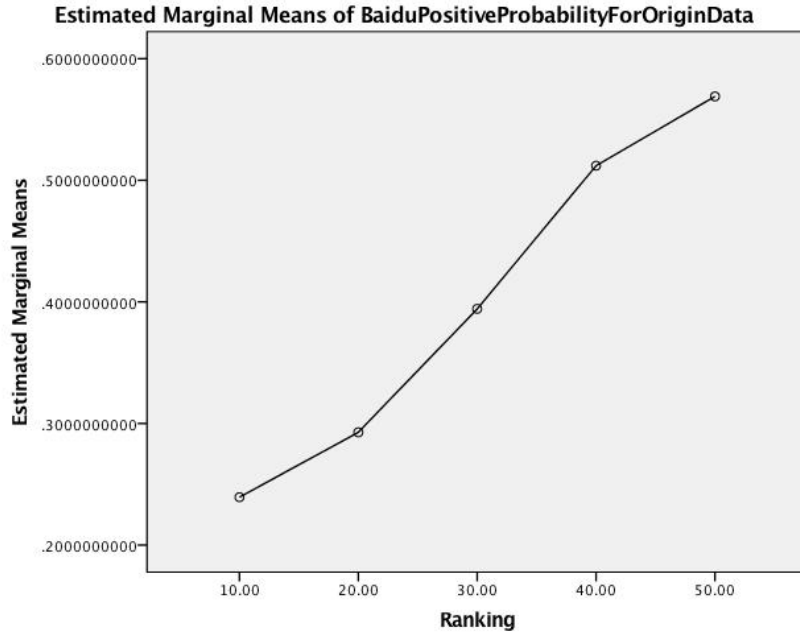
- Baidu Chinese sentiment analysis positive probability values are valid.

2.1.3 Baidu Chinese sentiment analysis postitive probability tranform to Google Score standard compare with different ranking (origin data)

Ranking	Mean	Valid N	std.deviation	Total N	Minimum	Maximum	Variance
Ranking 10	-0.598875	8525	0.557595		-0.999894	1.000000	0.310912
Ranking 20	-0.488772	13141	0.617021		-0.999838	1.000000	0.380715
Ranking 30	-0.236524	18821	0.728420		-0.999786	1.000000	0.530596
Ranking 40	0.054493	8717	0.773410		-0.998950	1.000000	0.598164
Ranking 50	0.188983	4271	0.774245		-0.999464	1.000000	0.599456
Total	-0.274854	53475	0.733884		-0.999894	1.000000	0.538586

5

Estimated Marginal Means of
BaiduPositiveProbababilityToGoogleScoreStandardForOriginData
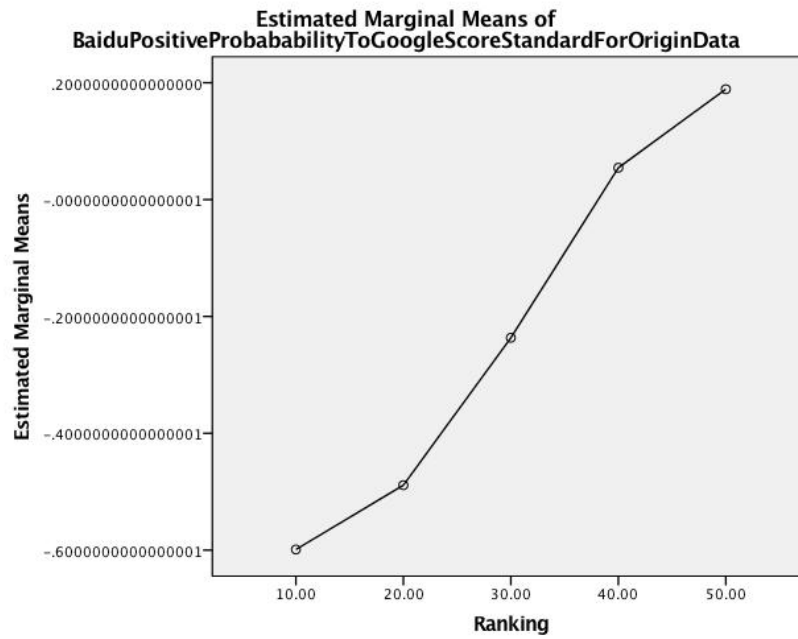
2.1.4 Baidu Chinese sentiment analysis category value compare with different ranking (origin data)

Estimated Marginal Means of BaiduCategoryForOriginData

- Baidu Chinese sentiment analysis category values are valid.

2.1.5   Chinese sentiment analysis Error Rate

Ranking	Error Rate
Ranking 10	0.0054829678
Ranking 20	0.0064267352
Ranking 30	0.0080636661
Ranking 40	0.0083048919
Ranking 50	0.0083584862

- Total Error Rate: 0.0073140396

2.1.6   Baidu Chinese sentiment analysis Summary

- Baidu Chinese sentiment analysis positive probability values are valid.

- Baidu Chinese sentiment analysis category values are valid.

2.2   Baidu English sentiment analysis

2.2.1   Baidu English sentiment analysis postitive probability compare with different ranking (based on Google translated data)

Ranking	Mean	Valid N	Std.deviation	Total N	Minimum	Maximum	Variance
Ranking 10	0.517526	7968	0.134711		0.005045	1.000000	0.018147
Ranking 20	0.531020	12225	0.141214		0.037275	1.000000	0.019941
Ranking 30	0.540824	17457	0.137174		0.014443	1.000000	0.018817
Ranking 40	0.567782	8163	0.144971		0.051860	1.000000	0.021016
Ranking 50	0.589054	4006	0.150737		0.086614	1.000000	0.022722

Estimated Marginal Means of
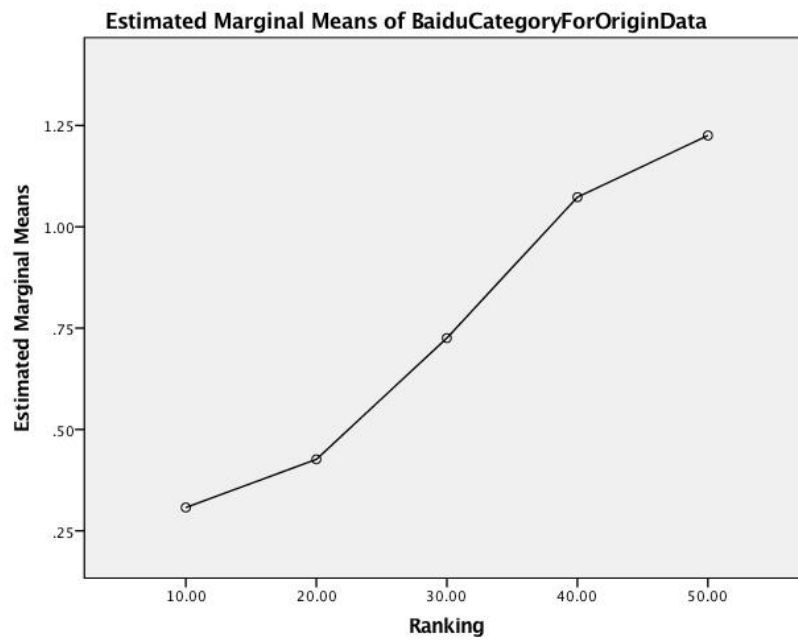BaiduPositiveProbabilityForGoogleTranslatedData
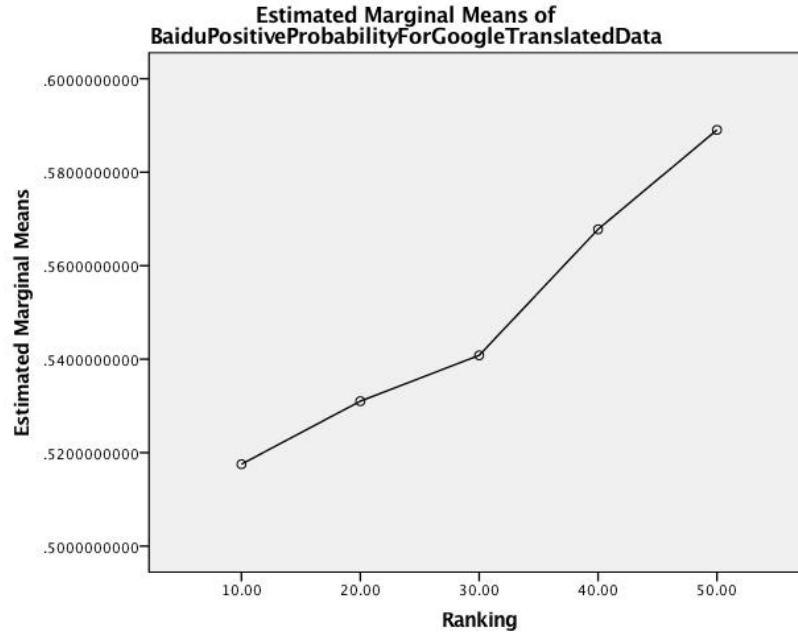
- Baidu English sentiment analysis positive probability values are valid.

2.2.2 Baidu English sentiment analysis positive probability tranform to Google Score Standard (based on Google translated data)

Ranking	Mean	Valid N	Std.deviation	Total N	Minimum	Maximum	Variance
Ranking 10	0.112029	7968	0.586700		-0.994955	1.000000	0.344216
Ranking 20	0.150325	12225	0.587147		-0.962725	1.000000	0.344742
Ranking 30	0.193858	17457	0.577416		-0.985557	1.000000	0.333410
Ranking 40	0.276835	8163	0.564339		-0.948140	1.000000	0.318479
Ranking 50	0.352409	4006	0.537822		-0.913386	1.000000	0.289253

Estimated Marginal Means of
BaiduPositiveProbababilityToGoogleStandardForGoogleTranslated

2.2.3 Baidu English sentiment analysis category values compare with different ranking (based on Google translated data)
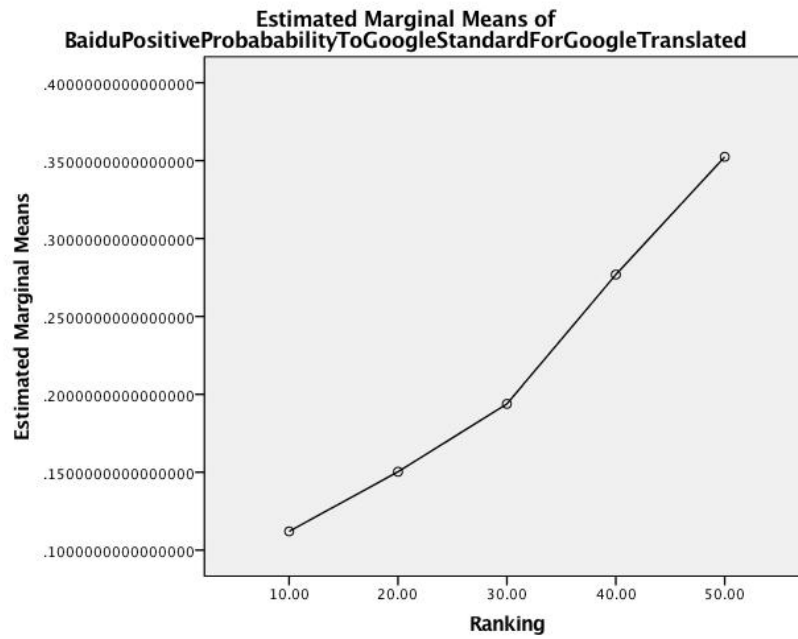

Estimated Marginal Means of BaiduCategoryForGoogleTranslatedData
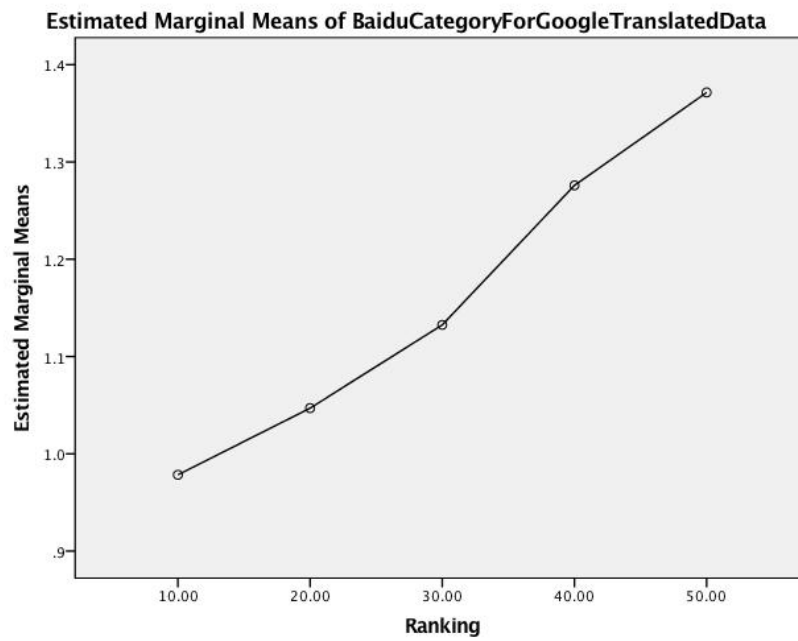
- Baidu English sentiment analysis category values are valid.

2.3 Baidu Chinese sentiment analysis positive probability tranform to Google Score standard Method

```python
if (inputValue < 0.5):
    result = (1 - inputValue) * -1
elif (inputValue > 0.5):
    result = inputValue
else:
    result = 0
```

3 Google sentiment analysis

3.1 Google Chinese sentiment analysis

3.1.1 Google Chinese sentiment analysis scores compare with different ranking (origin data)

Ranking	Mean	Valid N	std.deviation	Total N	Minimum	Maximum
Ranking 10	-0.238742	8567	0.445384	8572	-0.900000	0.900000
Ranking 20	-0.118380	13210	0.448064	13226	-0.900000	0.900000
Ranking 30	0.117291	18940	0.462095	18974	-0.900000	0.900000
Ranking 40	0.315915	8778	0.458128	8790	-0.900000	0.900000
Ranking 50	0.361626	4305	0.441309	4307	-0.900000	0.900000

Estimated Marginal Means of GoogleScoreForOriginData
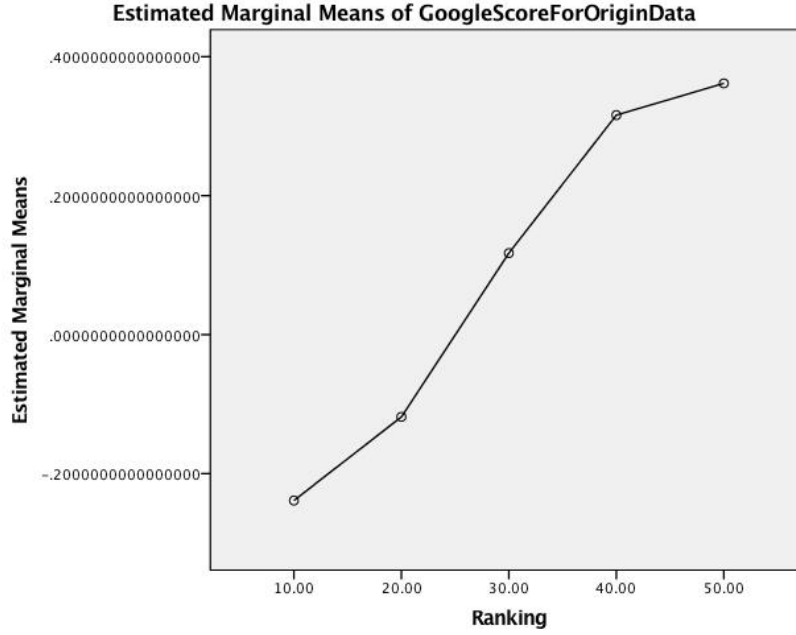
- Google Chinese sentiment analysis score values are valid.

3.1.2 Google Chinese sentiment analysis Error Rate

Ranking	Error Rate
Ranking 10	0.0005832944
Ranking 20	0.0012097384
Ranking 30	0.0017919258
Ranking 40	0.0013651877
Ranking 50	0.0004643603

- Total Error Rate: 0.0012808851

3.2 Google English sentiment analysis

3.2.1 Google English sentiment analysis score compare with different ranking (based on Google translated data)

Ranking	Mean	Valid N	Std.deviation	Total N	Minimum	Maximum	Variance
Ranking 10	-0.338431	8566	0.430581		-0.900000	0.900000	0.185400
Ranking 20	-0.244312	13204	0.437549		-0.900000	0.900000	0.191449
Ranking 30	-0.057978	18940	0.447353		-0.900000	0.900000	0.200125
Ranking 40	0.147830	8777	0.455342		-0.900000	0.900000	0.207336
Ranking 50	0.225000	4304	0.453471		-0.900000	0.900000	0.205636

Estimated Marginal Means of GoogleScoreForGoogleTranslatedData
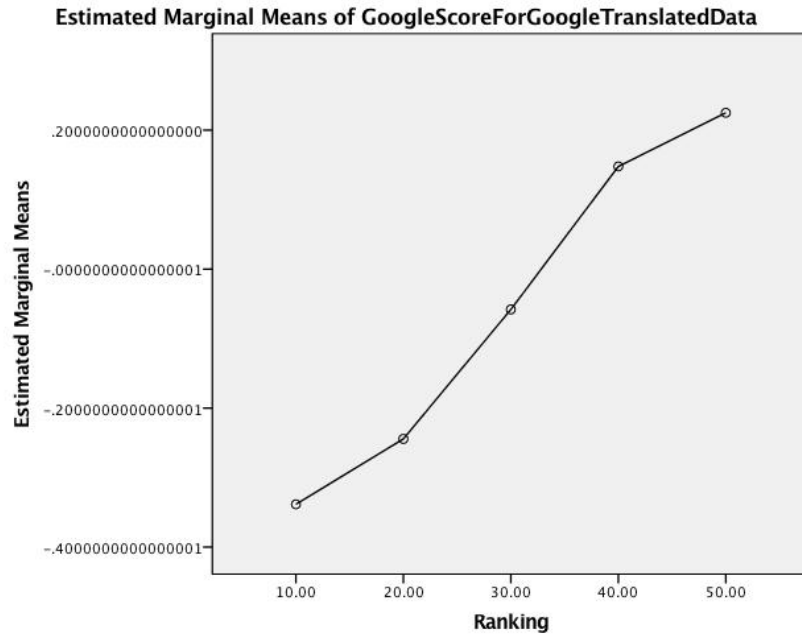
- Google English sentiment analysis score values are valid based on Google translated data.

3.2.2 Google English sentiment analysis score compare with different ranking (base on Yandex translated data)

Ranking	Mean	Valid N	Std.deviation	Total N	Minimum	Maximum	Variance
Ranking 10	-0.337873	8568	0.416416		-0.900000	0.900000	0.173403
Ranking 20	-0.233371	13221.000000	0.422133		-0.900000	0.900000	0.178196
Ranking 30	-0.055703	18972.000000	0.429758		-0.900000	0.900000	0.184692
Ranking 40	0.138917	8788.000000	0.447876		-0.900000	0.900000	0.200593
Ranking 50	0.208268	4306.000000	0.449598		-0.900000	0.900000	0.202138

- Google English sentiment analysis score values are valid based on Yandex translated data.

3.2.3 Google English sentiment analysis score compare with different ranking (base on Baidu translated data)

Ranking	Mean	Valid N	Std.deviation	Total N	Minimum	Maximum	Variance
Ranking 10	-0.284984	8491.000000	0.416185		-0.900000	0.900000	0.173210
Ranking 20	-0.192064	13092.000000	0.417855		-0.900000	0.900000	0.174603
Ranking 30	-0.017125	18820.000000	0.429167		-0.900000	0.900000	0.184185
Ranking 40	0.167667	8734.000000	0.432601		-0.900000	0.900000	0.187144
Ranking 50	0.244657	4286.000000	0.430004		-0.900000	0.900000	0.184904

- Google English sentiment analysis score values are valid based on Baidu translated data.

3.2.4 Correlations Between Origin data, Google Translated data, Yandex Translated and Baidu Translated data (each element)

		GoogleScoreForGoogleTranslatedData	GoogleScoreForOriginData	GoogleScoreForYandexTranslatedData	GoogleScoreForBaiduTranslatedData
GoogleScoreForGoogleTranslatedData	Pearson Correlation	1	.524[**]	.731[**]	.705[**]
	Sig. (2-tailed)		.000	.000	.000
	Sum of Squares and Cross-products	12396.029	6721.968	8776.759	8283.063
	Covariance	.230	.125	.163	.155
	N	53791	53791	53777	53416
GoogleScoreForOriginData	Pearson Correlation	.524[**]	1	.516[**]	.498[**]
	Sig. (2-tailed)	.000		.000	.000
	Sum of Squares and Cross-products	6721.968	13287.617	6418.213	6057.717
	Covariance	.125	.247	.119	.113
	N	53791	53800	53786	53423
GoogleScoreForYandexTranslatedData	Pearson Correlation	.731[**]	.516[**]	1	.690[**]
	Sig. (2-tailed)	.000	.000		.000
	Sum of Squares and Cross-products	8776.759	6418.213	11641.908	7847.940
	Covariance	.163	.119	.216	.147
	N	53777	53786	53855	53409
GoogleScoreForBaiduTranslatedData	Pearson Correlation	.705[**]	.498[**]	.690[**]	1
	Sig. (2-tailed)	.000	.000	.000	
	Sum of Squares and Cross-products	8283.063	6057.717	7847.940	11188.339
	Covariance	.155	.113	.147	.209
	N	53416	53423	53409	53423

**. Correlation is significant at the 0.01 level (2-tailed).

- assumption Google English sentiment analysis tool and Google Chinese sentiment analysis tool are same

    - Google translation sentence quality > Yandex translation sentence quality > baidu translation sentence quality

    - analysis same langeuage corrlations always bigger than cross langeuage corrlations

3.2.5 Correlations between origin data Mean, Google translated data Mean, Yandex translated Mean and baidu translated data Mean

Correlations		GoogleEnglishSentimentScoreForYandexTranslatedMean	GoogleEnglishSentimentScoreForBaiduTranslatedMean	GoogleChineseMeanValue	GoogleEnglishSentimentScoreForGoogleTranslatedMean
GoogleEnglishSentimentScoreForYandexTranslatedMean	Pearson Correlation	1	1.000**	.997**	1.000**
	Sig. (1–tailed)		.000	.000	.000
	Sum of Squares and Cross–products	.219	.212	.245	.227
	Covariance	.055	.053	.061	.057
	N	5	5	5	5
GoogleEnglishSentimentScoreForBaiduTranslatedMean	Pearson Correlation	1.000**	1	.996**	1.000**
	Sig. (1–tailed)	.000		.000	.000
	Sum of Squares and Cross–products	.212	.205	.237	.220
	Covariance	.053	.051	.059	.055
	N	5	5	5	5
GoogleChineseMeanValue	Pearson Correlation	.997**	.996**	1	.996**
	Sig. (1–tailed)	.000	.000		.000
	Sum of Squares and Cross–products	.245	.237	.277	.254
	Covariance	.061	.059	.069	.064
	N	5	5	5	5
GoogleEnglishSentimentScoreForGoogleTranslatedMean	Pearson Correlation	1.000**	1.000**	.996**	1
	Sig. (1–tailed)	.000	.000	.000	
	Sum of Squares and Cross–products	.227	.220	.254	.236
	Covariance	.057	.055	.064	.059
	N	5	5	5	5

**. Correlation is significant at the 0.01 level (1–tailed).

- translation sentence tools' quality have NOT significant impact sentiment analysis results because same data use different translation tool to analysis and all three results have highest correlations between each other.

- I guess translation key word quality more importance compare with sentence translation quality

- Using sentiment analysis results compare different translation tools' quality are NOT reliable.

4 Baidu sentiment analysis VS Google sentiment analysis

4.1 Baidu Chinese sentiment analysis VS Google Chinese sentiment analysis

4.1.1 Mean Value Correlation

- Pearson Correlation 0.991

- sig. 0.001

- N 5

- Conclusion Baidu Chinese sentiment analysis and Google Chinese sentiment analysis have higher liner relationship.

4.1.2 Error Rate

- Baidu Chinese sentiment analysis Total Error Rate = 0.0073140396

- Google Chinese sentiment analysis Total Error Rate = 0.0012808851

- conclusion

  - Baidu sentiment analysis error rate high than Google sentiment analysis error rate

4.1.3 Tendency

- chinese sentiment analysis results given by both Baidu and Google are valid because when the ranking group ID increases from 11 to 50, the sentiment analysis score also strictly increases accordingly.

4.2 Baidu English sentiment analysis VS Google English sentiment analysis

4.2.1 Mean Value Correlation (based on Google translation)

- Pearson Correlation 0.978

- sig. 0.004

- N 5

- Conclusion Baidu English sentiment analysis and Google English sentiment analysis have higher liner relationship.

4.2.2 Tendency

- English sentiment analysis results given by both Baidu and Google are valid because when the ranking group ID increases from 11 to 50, the sentiment analysis score also strictly increases accordingly.