Cross-language Sentiment Analysis for Social Media

Boyang Yan
School of Computing and Information Technology
University of Wollongong
Wollongong, Australia
by932@uowmail.edu.au

Xiaoxia Pu
School of Computing and Information Technology
University of Wollongong
Wollongong, Australia
xp816@uowmail.edu.au

Xudong Zhang
School of Computing and Information Technology
University of Wollongong
Wollongong, Australia
xz944@uowmail.edu.au

Helene Tran
School of Computing and Information Technology
University of Wollongong
Wollongong, Australia
ht185@uowmail.edu.au

Abstract—This document is a model and instructions for LaTeX. This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. *CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.

Index Terms—sentiment analysis, machine translation

I. Introduction

Social Media is becoming more and more popular. There are lots of text comments on different discussion topics every day. It would be impossible to analysise the huge amount of data generated manually. These topics are discussed by speakers of different languages, from different cultural backgrounds, further complicating any analysis. Most people enounter language and culture barriers during cross-language communication. In this paper, the use of machine translation and sentiment analysis tools to solve this problem of analysing cross-cultural and cross-language data is explored and discussed. Sentiment analysis is a part of text data mining. The aim of sentiment analysis is to determine the attitude of speakers or writers with respect to some topic or the overall contextual polarity or emotional reaction to a text document. It is usually equated with opinion mining, which involves the use of natural language processing and machine learning to ascertain the possibility of positive or negative opinions [1]. Sentiment analysis is useful for analyzing a huge amount of data relating to personal opinions. It can be used in an e-business context. For example, business managers can analysise customers' attitudes, as to whether they like or dislike their product or service. Also, government can use sentiment analysis to analyze citizen perspectives. In a word, sentiment analysis is coming into widespread use. As Dr. Haiyun mentions, English language sentiment analysis research has undergone major developments in recent years [2]. However, less research has been undertaken in other languages, such as Chinese. Today, a lot of English language sentiment analysis theories have been developed. Also, lots of Machine Translation tools are

available, such as Google translation, Bing translation and Yandex translation. Manchine Translation uses computational linguistic programs and natural language processing theory [3]. However, nobody working in a combination of these two fields of research has undertaken non-English sentiment analysis. Therefore, the research described and discussed in this document aimed to create a testing model to find the best-combination of English sentiment analysis tools and machine translation tools to obtain reliable sentiment analysis results from non-English texts, for non-English speakers who do not have such sentiment analysis tools to analyze their own language. In the end, everyone to be able to understand different language speakers' attitudes (positive or negative), emotions and opinions.

II. Literature Review

This research consists of three components. Firstly, measuring machine translation service quality. Secondly, testing sentiment analysis service quality. Thirdly, finding the best compound mode for machine translation service and sentiment analysis service. As a result, this section will fouced doing literature review on machine translation, Testing methodology and sentiment analysis.

A. Testing in Machine translation

There are two research articles about testing modeling of machine translation(MT).

a) Around Trip Translation method: As Somers argue that they establish an around trip translation (RTT) method for detect the quality of machine translation [4]. For example, testing English to Chinese translation tool. Firstly, using English to Chinese translation tool translate test data to Chinese. Secondly, translation Chinese translated data back to English. Finally, compare the similarity for two English data set. They also metion two similarity metrics BLEU and F-score, to judge the translations. The limitation of RTT modle are it cannot tell which one is good MT from all bad MT tools as well

as cannot find which sentences are easier for translation and which sentences are harder for translation.

b) A monte carlo method for machine translation services: Another article is about using third-party language to test the quality of machine translation [5]. For example, if testing English to Chinese translation tool. Firstly, random choice an Intermediate third-party language. Secondly, translation English test data to the third-party language, after translation third-party language to Chinese. These two steps are only finished one path. Another path is a translation in English directly to Chinese. In the end, the comparison the two paths of results similarity. In this article, the main finding is Google Translate is the best machine translation. In addition, there is a tendency to be produced better results in European languages, which use ANOVA Statistics method and Pairwise t tests getting this conclusion. In my experiment, I also got Google Translator is the best machine translation compare with Yandex and Baidu.

B. Testing methodology

Accounting to Sethi said that there are two categorized testing techniques, which are Static Testing and Dynamic Testing. The Dynamic Testing are divided into three categories, which are Functional Testing, Structural testing and Non-Functional Testing [**?**]. In my research project, I will focus on Functional Testing on this project.

a) Metamorphic Testing: The proposed of this research project is based on Metamorphic Testing. Metamorphic Testing is for testing function correctness. There is a research article written by Zhou in 2016. This article is well-explained for what is Metamorphic Testing. As Zhou said, Metamorphic testing (MT) is a property-based software testing method developed for automated test case generation and automated result verification, based on the effects of some expected properties of the target program [**?**]. These properties, recognized as metamorphic relations (MRs), serve as essential relations among the inputs and outcomes of multiple executions of the target program. For instance, test calculator function correctness will be shown if input 1 + 1, and the result will be 2. In this example people can easily make the judgment, correct or not. However, people will not easy to make this judgment quickly if input sin (3.7), and the calculator gives a output. In a generally acknowledged, Sin (3.7) = sin (3.7 + 360) is correct. In metamorphic testing, the name of 3.7 is the test case. The name of 3.7 + 360 is the follow-up test case. Metamorphic Relation is the relationship between two input test cases as well as the two outputs. Metamorphic testing is based on Metamorphic Relation. The two outputs need an existing mathematical relation. In this example, the relation is "=". The advantage of Metamorphic testing (MT) method, which can automate result verification and test case generation. The disadvantage is that cannot detect memory leak or some others insensitivity failure situation. However, Metamorphic Testing is appropriate for testing translation tools and sentiment analysis tools. Effectiveness of Metamorphic Relations I have read an article which was written by Zhou

in 2013. The main purpose of reading this article is to find the best one among the three testing modelings. This article is based on white-box testing, which have source code, as well as the most important conclusion is if the Metamorphic Relations can get bigger distance (dissimilarity) that will have more chance to detect failures. In other words, MRs with very different initial and follow-up execution are more likely to detect failures than those with similar initial and follow-up executions. The concept of "difference" are defined in namely coverage Manhattan distance (CMD), frequency Manhattan distance (FMD), and frequency Hamming distance (FHD) in regard to adaptive random testing (ART), where CMD metric on the basis of branch coverage execution profiles performs the best fault-detection effectiveness. I think the advantage of this article is suitable for finding the most effectiveness of Metamorphic Relations in White-box. However, this article is not suitable for Black - Box Testing. The reason is Black Box Testing have not source code available, so it cannot calculation the program's distance. In this research project, my testing modeling have not source code available. So, I have found another research article (Henard2016)talking about the difference between black-box testing and white-box testing.

White-box VS Black-box Henard (2016) have done some research for difference between white box testing and black - box testing in 2016. They have two finding is useful in my research black-box testing and white-box testing performance just have a little difference (at most 4

The three different machine translation testing modeling, we can call the three different Metamorphic Relations. So, I have found a research article(Cao,2013) about the effectiveness of metamorphic relations.

Also, when I finished reading all of the details, I have found Two-way ANOVA is also suitable for part 3 of my research (finding the best compound mode for machine translation service and sentiment analysis service). I think that the disadvantages of those three-testing modeling will be involved some of the noise. It is clear the Round-Trip will be involved translation back path noise. Using the third language will be involved in third language translation path noise. My modeling will be involved sentiment analysis tools' noise. However, my modeling does not involve any noise from translation tool. I think my modeling is better than the other two testing modeling. I am total from another aspect to test translation tool.

III. Method

IV. 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 %

V. IQR, Mean, Median, Q1, Q3, lower extreme, upper extreme, mean slope, median slope

A. Baidu Chinese Sentiment analysis Positive Probability Base On (Chinese Origin Data)

r0.5 A wrapped table going nicely inside the text.

Ranking	IQR	Mean	Median	Q1	Q3	lowerExtreme	upperExtreme
10	0.3426533	0.2404781	0.1789910	0.0489987	0.3916520	-0.4649812	0.9056319
20	0.3817485	0.2949725	0.2489580	0.0878135	0.4695620	-0.4848092	1.0421847
30	0.4504405	0.3980354	0.3808120	0.1516095	0.6020500	-0.5240512	1.2777107
40	0.5255830	0.5123376	0.5385910	0.2461685	0.7717515	-0.5422060	1.5601260
50	0.5368080	0.5712128	0.6188395	0.3101650	0.8469730	-0.4950470	1.6521850

Mean Slope	Median Slope
0.008788344	0.0116933

VI. Questionnaire

References

[1] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current state of text sentiment analysis from opinion to emotion mining," *ACM Computing Surveys (CSUR)*, vol. 50, no. 2, p. 25, 2017.

[2] H. Peng, E. Cambria, and A. Hussain, "A review of sentiment analysis research in chinese language," *Cognitive Computation*, vol. 9, no. 4, pp. 423–435, 2017.

[3] "machine translation (1950s) linguistics." *Dictionary of Theories*, p. 323, 2002. [Online]. Available: http://ezproxy.uow.edu.au/login?url=https://search-ebscohost-com.ezproxy.uow.edu.au/login.aspx?direct=true&db=f6h&AN=40423164&site=eds-live

[4] H. Somers, "Round-trip translation: What is it good for?" in *Proceedings of the Australasian Language Technology Workshop 2005*, 2005, pp. 127–133.

[5] D. Pesu, Z. Q. Zhou, J. Zhen, and D. Towey, "A monte carlo method for metamorphic testing of machine translation services," in *Proceedings of the 3rd International Workshop on Metamorphic Testing*. ACM, 2018, pp. 38–45.