Impact of Translation on Sentiment Analysis: A Case-Study on Telugu Reviews

by

Rama Rohit Reddy Gangula, Radhika Mamidi

in

19th International Conference on Computational Linguistics and Intelligent Text Processing
(*CICLing-2018*)

Hanoi, Vietnam

Impact of Translation on Sentiment Analysis: A Case-Study on Telugu Reviews

Gangula Rama Rohit Reddy* and Radhika Mamidi

Language Technologies Research Center(LTRC)
Kohli Center On Intelligent Systems (KCIS)
International Institute of Information Technology Hyderabad (IIIT-H) - 500032
ramarohitreddy.g@research.iiit.ac.in
radhika.mamidi@iiit.ac.in
https://www.iiit.ac.in

Abstract. Sentiment analysis research has predominantly been on English texts. There exists many sentiment resources for English but very less exist for other languages. To improve sentiment analysis in a low resource language, sentiment labeled corpora are translated from English into the focus language and use them as additional resources for sentiment analysis research in the focus language [3]. But when text is translated from one language into another, sentiment is preserved to varying degrees. In this paper, we use product and book reviews in English as stand-in for source language text and determine loss in sentiment and sentiment predictability when they are translated into Telugu (a low resource South Asian language), manually and automatically. For this purpose, we use manually and automatically determined sentiment labels of the English text as a benchmark. We show that sentiment analysis of Telugu manual translations of English text produces competitive results w.r.t English sentiment analysis. We discover that even though machine translation significantly reduces the human ability to recover sentiment, automatic sentiment systems are still able to capture sentiment information from the translations in certain cases. In the process, we created a Telugu-English parallel corpus that is independently annotated for sentiment using a 5-value scale by Telugu and English speakers. We also created a Telugu lexicon annotated at both sentiment and emphasis level.

Keywords: Sentiment analysis, Machine Learning, Machine Translation, Reviews, Support Vector Machine, Parallel Corpus, Lexicon.

1 Introduction

The term sentiment analysis is most commonly used to refer to the goal of determining the polarity of a piece of text. Automated sentiment analysis of text,

---

* Please note that the LNCS Editorial assumes that all authors have used the western naming convention, with given names preceding surnames. This determines the structure of the names in the running heads and the author index.

especially reviews has many applications in commerce and product development. In the past two decades, a vast majority of research has been on English texts [10–12]. Furthermore, many sentiment resources essential for automated sentiment analysis such as sentiment labeled corpora exist mainly in English. There is a growing need to analyze texts from other languages such as Telugu, but with very few resources this cannot be done effectively. Thus for the automated sentiment analysis of low resource languages, we decided to translate the English resources to the focus-language and use them for automated sentiment analysis of text in that language. To our knowledge, this is the first resource containing English reviews and their translations into Telugu (both manually and automatically produced) each manually labeled for sentiment using 5-value scale. We also created first Telugu lexicon annotated at both sentiment and emphasis level though sentiment level lexicon was available [9]. The lexicon consists of 6000 words. We use English product and book reviews as specific instances. We use Telugu and English sentiment analysis systems as well as a English to Telugu translation system. We outline the advantages and disadvantages of manual and machine translation methods and conduct quantitative and qualitative experiments to determine the impact of translation on sentiment. As benchmarks we use manually and automatically determined 5-value scale sentiment labels of the English product and book reviews. The results will determine the best suited methods.

Through our experiments on the parallel corpus, we show that sentiment analysis of Telugu manual translations of English texts produces competitive results w.r.t English sentiment analysis. We also show that translations (both manual and automatic) introduces marked changes in sentiment carried by the text: Highly positive and highly negative texts can often be translated into the texts that are just positive and negative. Positive and negative texts can often be translated into texts that are neutral. Highly negative text can be translated to Highly positive text. We also find that in some cases, certain attributes of automatically translated text that mislead humans with regards to the true sentiment of the source text, do not seem to affect the automatic sentiment analysis system.

2 Related Work

2.1 Sentiment Analysis in English

Sentiment analysis systems have been applied to many different kinds of texts including customer reviews [11, 10, 12], newspaper headlines [4], blogs [18], novels [5],emails [15]. Often these systems have to cater to the specific needs of the text such as formality versus informality, length of utterances, etc.

2.2 Sentiment Analysis in Telugu

Sentiment analysis of Telugu social media texts has several challenges. Telugu is an agglutinative Dravidian language spoken widely in India. It is morphologically complex language. Very little work is done on sentiment analysis in Telugu.

Sentiment analysis systems have been applied to different kinds of Telugu texts including Song Lyrics [1], News [16, 17].

2.3 Multilingual Sentiment Analysis

Mihalcea et al. used English resources to automatically generate Romanian subjectivity lexicon using an English-Romanian dictionary [13]. The generated lexicon is then used to classify Romanian text. Wan translated Chinese customer reviews to English using a machine translation system [19]. The translated reviews are then classified with a rule-based system that relies on English lexicons. Mohammad et al. studied the affect of translation on sentiment using Arabic social media posts[14]. Balahur and Turchi conducted a study to assess the performance of statistical sentiment analysis techniques on machine-translated texts [2]. The authors stated that sentiment analysis can be performed on automatically translated texts without a substantial loss in accuracy. On contrary to this, our study shows that it may not be true for all languages and there may be a loss of sentiment while translating in some languages.

3 Experimental Setup

To study the impact of translation on sentiment analysis, we propose the following experimental setup: (Here H is human, M is machine, A is annotation, T is translation)

- Identify an English review dataset. Lets refer it as En.
- Manually translate the chosen dataset into Telugu. Lets refer it as Te(HT).
- Automatically translate the chosen dataset into Telugu. Lets refer it as Te(MT).
- Manually annotate English review dataset for sentiment. Lets refer it as En(HA).
- Manually annotate all translated (both manually and automatically) Telugu datasets. Lets refer them as Te(HT.HA) and Te (MT.HA)
- Run an English sentiment analysis system on English review dataset and generate automatically annotated English review dataset. Lets refer it as En(MA)
- Run a Telugu sentiment analysis system on both the Telugu datasets and generate automatically annotated Telugu datasets. Lets refer them as Te(HT.MA) and Te (MT.MA).

Figure 1 depicts the experiment setup. Once various sentiment labeled datasets are created, we compare them to draw inferences. For example, comparing the labels for En(HA) and Te(HT.HA) will show how different the sentiment labels tend to be when the text is manually translated from English to Telugu. The results will also show how feasible it is to first translate English text into Telugu automatically and then use automatic sentiment analysis (En(HA) vs. Te(MT.MA)).
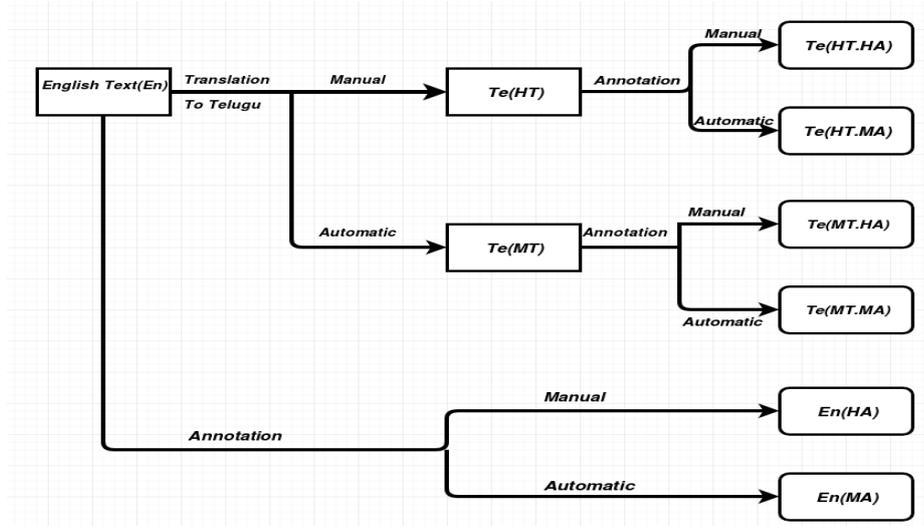
Fig. 1. Experimental setup to determine the impact of translation on sentiment. We compare the sentiment labels of En(HA), a manually annotated English dataset with other datasets shown on the right side of the figure.

DATA: For the experiment we created an English-Telugu parallel corpus which consists of reviews with around 3500 entity level sentences.

The paper is structured as follows. In Section 4, we discuss the task of generating translations manually and automatically. In Section 5, we describe the procedure of annotating Telugu and English texts. In Section 6, we discuss the Telugu sentiment analysis. In Section 7, we discuss English sentiment analysis, In Section 8, we present the observations on sentiment after translation and In Section 9, we present our conclusions.

4 Generating Telugu Translation

The manual translation was carried out in the following way: First the data was normalized and all the spelling mistakes were corrected. The numbers were retained as in Roman script. The translators were instructed to be faithful to the original text as much as possible and retain the same sentiment value. The translators first understood the exact meaning and emotion of the review sentences. Later they generated the review in Telugu such that it fits Telugu grammar and syntax and also carries the same emotion and meaning as the sentence. Though this process is accurate, it is time consuming.

We also used Google translate Api[1] for the task of automatic translation of review sentences. It translates considerably good because it uses Neural Machine translation and has huge data of different languages [20].

---

[1] https://www.npmjs.com/package/google-translate-api

5 Creating sentiment labeled data in English and Telugu

Manual sentiment annotations were performed for three datasets as follows:

1. Original English sentences were annotated by native English speakers.
2. Manual Telugu translations were annotated by native Telugu speakers.
3. Automatic Telugu translations were annotated by native Telugu speaker.

Each sentence was annotated by 5 annotators and the sentiment label marked by majority was chosen. The annotators annotated each sentence using a 5-value scale, distinguishing between highly negative, negative, neutral, positive and highly positive. 5-value scale annotation was chosen so as to determine the loss in sentiment accurately because by using 3-value scale we may not capture the loss of emphasis after translation in the text. Figure 2 shows the class distribution of sentiment labels in various datasets. We can observe different distribution of sentiment labels in each dataset. We can say that there is loss in sentiment after translation from the change in the distribution. For each post, we determine the count of the most frequent annotation divided by the total number of annotations. This score is averaged for all posts to determine the inter-annotator agreement shown in last column of Figure 2.

Data	Highly Negative	Negative	Neutral	Positive	Highly Positive	Agreement
En(HA)	24.3	11.05	16.7	11.08	36.87	70.2
Te(HT.HA)	25.48	9.1	16.6	9.57	39.25	65.4
Te(MT.HA)	25.88	7.07	24.76	7.28	35.01	57.2

Fig. 2. Class distribution (in percentage) of the sentiment annotated datasets.

6 Telugu Sentiment Analysis

Using Telugu SentiWordNet [9][2], we created a sentiment lexicon which consisted of around 6000 words. Words which are synonyms and antonyms are grouped

---

[2] http://amitavadas.com/sentiwordnet.php

together and are marked as positive, negative, neutral along with the emphasis or level of sentiment conveyed by each individual word on a scale of 4 where 1-no emphasis, 2-little emphasis, 3-emphasis, 4- high emphasis. Then we built Telugu sentiment analysis system by using the following method:

- Assign score for each word in sentence:
    1. If a word is neutral, then assign a score of 3 irrespective of emphasis.
    2. If a word is positive, then assign a score of 4 if it is of no emphasis or little emphasis and score of 5 if there is emphasis or high emphasis.
    3. If a word is negative, then assign a score of 2 if it is of no emphasis or little emphasis and score of 1 if there is emphasis or high emphasis.
    4. If a word is not present in lexicon, treat it as a neutral word.
- score of the word is changed if a negation is present before the word.
- Sum the score of words in a sentence and average them to get score of a sentence.
- We use this score along with number of negated contexts, word and character n-grams as features and train a linear-kernel Support Vector Machine classifier [6] on the available training data and label each sentence using a 5-value scale namely highly positive, positive, neutral, negative, highly negative.

7 English Sentiment Analysis

First the data is normalized and a TF-IDF vector is generated for each sentence. A linear-kernel Support Vector Machine [6] classifier is trained on the available training data. This is used to predict the labels of the sentences in our dataset. An accuracy of 67.5% is achieved.

8 Sentiment After Translation

Using the methods described in the above sections, we have generated all the manually and automatically labeled datasets mentioned in Experimental setup. Figure 3 shows the class distribution in datasets that have been automatically labeled with sentiment. These percentages can be compared with those in Figure 2 which shows true sentiment distribution in the English review dataset.

Observe that the automatic system has difficulty in assigning positive and negative classes to posts. This is probably because of the small percentage (about 10%) of positive and negative sentences in the training data. Also notice that the system predominantly guesses highly positive, which is also a reflection of the distribution in the training data. The strong bias to highly positives is increased in the Telugu translations.

Main Result: Table 1 shows how similar labels are across various pairs of datasets. Column 1 shows the data pairs and Column 2 shows the percentage of instances where sentiment labels match. Row a. shows the match percentage of 67.5% which shows the accuracy of English automatic sentiment analysis system. Row b. shows the difference in labels when text is manually translated

Data	Highly Negative	Negative	Neutral	Positive	Highly Positive
En(MA)	25.7	6.83	14.4	5.8	47.27
Te(HT.MA)	25.6	3.23	12.74	5.46	52.97
Te(MT.MA)	22.2	3.62	12.75	6.03	55.4

Fig. 3. Class distribution (in percentage) resulting from automatic sentiment analysis.

Data Pair	Match %
a. En(HA)-En(MA)	67.5
b. En(HA)-Te(HT.HA)	71.2
c. En(HA)-Te(HT.MA)	64.6
d. En(HA)-Te(MT.HA)	50.8
e. En(HA)-Te(MT.MA)	54.1
f. Te(HT.HA)-Te(MT.HA)	51.2
g. Te(HT.HA)-Te(HT.MA)	57.8
h. Te(MT.HA)-Te(MT.MA)	47.2

Table 1. Match percentage between pairs of sentiment labeled datasets.

from English to Telugu and also annotated manually. There is only 71.2% match which shows that translation does affect sentiment. Row c. shows that Telugu automatic sentiment analysis system performance is less compared to manual annotation. Rows d. and e. shows that instances of sentiment label match are very less when translated automatically as the task of automatic translation is fairly difficult. Row f. shows that manual and automatic translation lead to only about 51% match in manually annotated sentiment labels with each other. Row g. shows result of 57.8% which is close to human agreement on manually translated data (65%). Row h. shows accuracy of the Telugu automatic sentiment analysis system on the automatically translated text (assuming the Telugu sentiment labels(Te(MT.HA)) as gold).

We manually examined several sentences to understand why humans incorrectly annotate a sentence automatically translated. Most of the cases were due to bad translation where sentiment word has either disappeared or it's emphasis is changed or the sentiment of word is altered. In some cases, translations were

affected by typos on English side. Figure 4 shows some examples. In some cases the automatic sentiment analysis system annotates correctly (where manual annotations of translation may fail) because it learns an appropriate model even for mistranslations.

1. Bad auto. translation: wrong translation of Sentence		
Post	I would not suggest you this TV	Highly Negative
Auto.Trans	నేను ఈ టీవిని సూచించాలనుకుంటున్నాను Nneu ee TV ni suchinchalanukuntunna	Highly positive
Manl. Trans	నేను మీకు ఈ టీవిని సూచించను Nenu meeku ee TV ni suchinchanu	Highly negative
2. Bad auto. translation: wrong translation , passive sentences are hard to translate		
Post	Main thing this tv is picture Quality which every one want, best picture quality And solid sound which are those you wont get in this TV	Highly Negative
Auto.Trans	ప్రధానంగా ఈ టీవీ చిత్రం నాణ్యత ప్రతి ఒక్కటి టీవిలో మంచి చిత్ర నాణ్యతను మరియు ఘన ధ్వనిని కోరుతుంది	Neutral
Manl. Trans	ప్రధాన విషయం ఈ TV పిక్చర్ నాణ్యత మరియు మంచి సౌండ్ ఈ TV లో మీరు రెండు పొందలేరు	Highly Negative
3. Bad auto. translation: Sarcasm is Hard to translate		
Post	This phone can also be used to iron clothes	Highly Negative
Auto.Trans	ఈ ఫోన్ ఇనుము బట్టలు చాలా ఉపయోగకరంగా ఉంటుంది Ee phone Inumu battalu chala upayogakaranga untundi	Highly positive
Manl.Trans	ఈ ఫోన్ ని ఇస్త్రి చేయడానికి కూడా వాడొచ్చు Ee phone ni istri cheyadani kuda vadochu	Highly negative

Fig. 4. Class distribution (in percentage) resulting from automatic sentiment analysis.

9 Conclusions

We present a set of experiments to systematically study the impact of translation of Telugu sentiment analysis. Our experiments show that the impact of translation at fine-grain level is high i.e level of sentiment or emotion is largely lost in automatic translation where it is fairly retained if translated manually and also Telugu automated sentiment analysis system works much better on manual translation of data than on automatic translations. This is because the task of automatic translation is fairly difficult.

References

1. Abburi, H., Akkireddy, E.S.A., Gangashetti, S., Mamidi, R.: Multimodal sentiment analysis of telugu songs. In: Proceedings of the 4th Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2016) co-located with 25th International Joint Conference on Artificial Intelligence (IJCAI 2016), New York City, USA, July 10, 2016. pp. 48–52 (2016), http://ceur-ws.org/Vol-1619/paper8.pdf

2. Balahur, A., Turchi, M.: Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. Computer Speech & Language 28(1), 56–75 (2014), `https://doi.org/10.1016/j.csl.2013.03.004`

3. Balahur, A., Turchi, M., Steinberger, R., Ortega, J.M.P., Jacquet, G., Küçük, D., Zavarella, V., Ghali, A.E.: Resource creation and evaluation for multilingual sentiment analysis in social media texts. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014. pp. 4265–4269 (2014), `http://www.lrec-conf.org/proceedings/lrec2014/summaries/965.html`

4. Bellegarda, J.R.: Data-driven analysis of emotion in text using latent affective folding and embedding. Computational Intelligence 29(3), 506–526 (2013), `https://doi.org/10.1111/j.1467-8640.2012.00457.x`

5. Boucouvalas, A.C.: Real time text-to-emotion engine for expressive internet communications. In: Proceedings of International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP-2002) (2002)

6. Chang, C., Lin, C.: LIBSVM: A library for support vector machines. ACM TIST 2(3), 27:1–27:27 (2011), `http://doi.acm.org/10.1145/1961189.1961199`

7. Das, A., Bandyopadhyay, S.: Sentiwordnet for indian languages. In: In the 8th Workshop on Asian Language Resources (ALR), COLING 2010, August, Beijing, China. pp. 56–63 (2010), `http://aclweb.org/anthology//W/W10/W10-3208.pdf`

8. Das, A., Bandyopadhyay, S.: Dr sentiment knows everything! In: The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA - System Demonstrations. pp. 50–55 (2011), `http://www.aclweb.org/anthology/P11-4009`

9. Das, A., Gambäck, B.: Sentimantics: Conceptual spaces for lexical sentiment polarity representation with contextuality. In: Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis, WASSA@ACL 2012, July 12, 2012, Jeju Island, Republic of Korea. pp. 38–46 (2012), `http://aclweb.org/anthology/W/W12/W12-3707.pdf`

10. Hu, M., Liu, B.: Mining opinion features in customer reviews. In: McGuinness and Ferguson [12], pp. 755–760, `http://www.aaai.org/Library/AAAI/2004/aaai04-119.php`

11. Liu, B.: Sentiment Analysis - Mining Opinions, Sentiments, and Emotions. Cambridge University Press (2015), `http://www.cambridge.org/us/academic/subjects/computer-science/knowledge-management-databases-and-data-mining/sentiment-analysis-mining-opinions-sentiments-and-emotions`

12. McGuinness, D.L., Ferguson, G. (eds.): Proceedings of the Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications of Artificial Intelligence, July 25-29, 2004, San Jose, California, USA. AAAI Press / The MIT Press (2004)

13. Mihalcea, R., Banea, C., Wiebe, J.: Learning multilingual subjective language via cross-lingual projections. In: ACL 2007, Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, June 23-30, 2007, Prague, Czech Republic (2007), `http://aclweb.org/anthology/P07-1123`

14. Mohammad, S.M., Salameh, M., Kiritchenko, S.: How translation alters sentiment. J. Artif. Intell. Res. 55, 95–130 (2016), `https://doi.org/10.1613/jair.4787`

15. Mohammad, S.M., Yang, T.: Tracking sentiment in mail: How genders differ on emotional axes. CoRR abs/1309.6347 (2013), `http://arxiv.org/abs/1309.6347`

16. Mukku, S.S., Choudhary, N., Mamidi, R.: Enhanced sentiment classification of telugu text using ML techniques. In: Proceedings of the 4th Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2016) co-located with 25th International Joint Conference on Artificial Intelligence (IJCAI 2016), New York City, USA, July 10, 2016. pp. 29–34 (2016), `http://ceur-ws.org/Vol-1619/paper5.pdf`

17. Naidu, R., Bharti, S.K., Babu, K.S., Mohapatra, R.K.: Text summarization with automatic keyword extraction in telugu e-newspapers. In: Smart Computing and Informatics, pp. 555–564. Springer (2018)

18. Neviarouskaya, A., Prendinger, H., Ishizuka, M.: Affect analysis model: novel rule-based approach to affect sensing from text. Natural Language Engineering 17(1), 95–135 (2011), `https://doi.org/10.1017/S1351324910000239`

19. Wan, X.: Using bilingual knowledge and ensemble techniques for unsupervised chinese sentiment analysis. In: 2008 Conference on Empirical Methods in Natural Language Processing, EMNLP 2008, Proceedings of the Conference, 25-27 October 2008, Honolulu, Hawaii, USA, A meeting of SIGDAT, a Special Interest Group of the ACL. pp. 553–561 (2008), `http://www.aclweb.org/anthology/D08-1058`

20. Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., Dean, J.: Google's neural machine translation system: Bridging the gap between human and machine translation. CoRR abs/1609.08144 (2016), `http://arxiv.org/abs/1609.08144`