Sentiment Analysis in English-Hindi Translation: A Review

Shubhnandan S. Jamwal

Department of Computer Science and IT, University of Jammu

Abstract: The objective of the Machine translation systems is to enhance the knowledgeable society of Indians without any language barrier. The users of the Indian language use number of tones for representing number of sentiments on social media platforms the individuals are also sharing information, opinions, comments and suggestions which results in proliferation of user generated large volume of text data available for interpretation. In this paper, we have analyzed very common translation technology of Google and it is observed that most of the sentences after translation by the Google engine have got very poor score of 2 or even 1 out of 5.

Introduction

India is a multilingual country and but all the Indians are not polyglots. There are eighteen constitutional languages and ten prominent scripts for the Indian languages. Majority of the Indians do not understand, read or write English, therefore implementing an efficient language translator especially reflecting the sentiments which are represented in different texts, is needed to remove the digital divide of the language.

As reported by the different resources that from 2016, a research team at Google announced GNMT that is Google Neural Machine Translation system and after this the translation on Google began using neural machine translation instead of statistical methods. Initially the translation of the language was preferred by using SMT and now also statistical machine translation remains as an important option of research for the researchers of the NLP. When we come to the research of NLP on the Indian languages, it has been observed that Indian languages are very rich in morphology and the SMT with richer inflectional morphologies pose greater challenge for translation. The conventional SMT approaches tend to perform poorly when either source or target language has rich morphology.

During the past decade the work on Sentiment Analysis has got momentum in contrast to code-mixed Indian language text. However, the non-availability of linguistic and lexical vocabulary and some other resources for the NLP tasks like annotated resources, the task of Sentiment Analysis of Indian Languages becomes somehow difficult. Sentiments can be classified into a negative or positive group but most of the research works in SA is conducted for English. On the other side, the research work on Indian languages has not studied well and it is therefore necessary to perform SA for Indian languages as well.

Review of Literature

T. P. Nagarhalli, V. Vaze and N. K. Rana [1] showed that Machine Translation is especially necessitated in the Indian perspective because more than 50% of the data generated online is in English which is known by only 12%. It is observed that many systems has been proposed in the Indian perspective including rule-based, example-based, statistical based and a hybrid of these machine translation techniques. But, recent study has shown that Neural Machine Translation provides better results. In recent times Google and Facebook have also developed

Neural Machine Translation system. These systems are one-fits-all kind of systems which do not take into consideration the complexities in a language, like Indian languages and proposed a broad framework for implementing Neural Machine Translation for Indian-English languages.

B. N. V. N. Raju and M. S. V. S. B. Raju [2] carried out the research on SMT for English and Telugu parallel corpus which have been used for training the system. They also observed that the translation mainly depends on quality and quantity of corpus. K. M. Chaman Kumar and others [3] presented a survey of several research works of different machine translation approaches and also compared these papers using several parameters such as the tools & technologies, datasets, advantages, disadvantages, and accuracy. They also proposed a machine translation model based on Statistical Machine Translation that will help us to translate from the Indian regional language Konkani to English and vice versa. J. Nair and others [4] proposed English to Hindi machine translation system design based on declension rules and also discussed the different approaches of Machine Translation.

G. I. Ahmad and others [5] has also worked on Sentiment Analysis (SA) of code-mixed text that provides useful information in the field of politics, marketing, business, health, sports etc. In their research, they attempted to provide a detailed summary of Sentiment Analysis of Indian languages with a special focus on code mixed Indian Languages.

N. Hadiya and N. Nanavati [6] discussed various available lexicon resources that are used for sentiment analysis in some Indian languages and presented the theoretical parametric evaluation of the studied techniques. They also further discussed challenges, which were identified during SA in Indian Languages. B. Gupta et al. [7] studied transfer learning and implemented it for Sentiment Analysis of Tweets by using the knowledge of Yelp reviews and observed that transfer learning approach is faster than the conventional machine learning approach and give comparable accuracy at much smaller dataset. Y. Sharma and et al., [8] extracted sentiments by using vector representation of the words using unsupervised technique. Singh, Avinash and others [9] have also carried out research on very low resourced languages like Dogri and used statistical machine translation system to train translation models for Dogri English language pair.

Analysis of Google NMT

Google translate is the technology used by most of the professionals working on the languages. Google Translate's uses NMT system that uses a large artificial neural network capable of deep learning. It has been observed by many researchers that, quality of translation is improved by adding number of examples in any of the NMT systems. Initially in 2016, Google used translation engine for converting English to eight other languages i.e., French, Portuguese, Chinese, German, Spanish, Japanese, Korean and Turkish and later in 2017, Russian, Hindi and Vietnamese were also added. In 2021 other Indian languages were also added for the translations which are using the same GNMT.

Analysis of English-Hindi translation

In this paper, we have selected some famous quotes that are reflecting some sentiments. The sentences selected for the translation are reflected in the table 1 and these sentences are translated by the Google translator as reflected in table 1. The sentences translated are manually checked by the linguist. The score to these sentences is allotted by the expert manually between 1 to 5 as 1 being the poor translation and 5 being as best translation.

Table 1: English-Hindi Translation.

S.No	English	Translated by Google	Score Allotted	Manual Translation
1.	I'm taking your birthday as an opportunity to thank you for all that you are to me. I will never be able to forget the support that you have given me. Happy Birthday, my sugar lumps! May your birthday be as sweet as you are.	मैं आपके जन्मदिन को एक अवसर के रूप में ले रहा हूं कि आप सभी के लिए धन्यवाद। आपने मुझे जो समर्थन दिया है, मैं उसे कभी नहीं भूल पाऊंगा! जन्मदिन मुबारक हो, मेरी चीनी गांठ! हो सकता है आपका जन्मदिन आपके लिए उतना ही मीठा हो।	3	आप मेरे लिए सब कुछ हैं । आपको धन्यवाद देने के लिए आपके जन्म दिन को मैं एक अवसर के रूप में ले रहा हूँ । आपने जो समर्थन मुझे दिया मैं उसे कभी भूल नहीं पाऊँगा । जन्म दिनकी शुभकामनाएँ । तुम्हारा जन्म दिन उतना ही आनंद प्रधान करने वाला हो, जितने मधुर तुम स्वयंम हो ।
2.	Yippee your birthday is finally here. It is time to party! It is time to sing and dance, and it is time for birthday bumps! May you have a delicious and partylicious birthday!	Yippee आपका जन्मदिन अंत में यहाँ है। यह पार्टी का समय है! यह गाने और नृत्य करने का समय है, और यह जन्मदिन के धक्कों का समय है! हो सकता है कि आपका जन्मदिन बहुत स्वादिष्ट और आनंदमय हो!	3	अहा ! अंततः तुम्हारा जन्म दिन आ ही गया है । यह समय प्रीतिभोज, गायन और नृतय का है । यह समय अति उत्साह से जन्म दिन में सम्मिलित होने वालों के लिए है । तुम्हारा जन्म दिन आनन्ददायक और उत्सव जैसा हो।
3.	May your smile get brighter with each passing year.	आपकी मुस्कुराहट हर बीतते साल के साथ शानदार हो सकती है।	2	हर साल बीतने के साथ तुम्हारी मुस्कराहट अधिक-से-अधिक खिलती जाए ।
4.	On your birthday, I just want to remind you that you are the best person I have ever met in my whole life.	आपके जन्मदिन पर, मैं बस आपको याद दिलाना चाहता हूं कि आप मेरे जीवन में कभी मिले सबसे अच्छे व्यक्ति हैं।	2	तुम्हारे जन्म दिन पर बस मैं याद दिलाना चाहता हूँ कि मैं आज तक जितने भी व्यक्तियों से मिला हूँ, उनमें से तुम श्रेष्ठ हो ।
5.	Don't bother what anybody else has to say, you have changed my life and I will never forget that.	किसी और को क्या कहना है परेशान मत करो, तुमने मेरी जिंदगी बदल दी है और मैं इसे कभी नहीं भूलूंगा।	3	किसी दूसरे का मत क्या है, इसकी परवाह मत करो । मैं यह नहीं भूल सकता कि तुमने मेरा जीवन बदला दिया है ।
6.	Happy Birthday to you, my dearest friend.	आपको मेरा सबसे प्रिय दोस्त, जन्मदिन मुबारक हो।	2	मेरे प्रिय मित्र तुम्हें जन्म दिन की शुभकामनाएँ ।
7.	I was so wrong, and I promise I won't let it	मैं बहुत गलत था, और मैं वादा करता हूं कि मैंने इसे दोबारा नहीं	3	मुझसे गलती हो गई । मैं वादा करता हूँ कि ऐसी गलती पुन: नहीं

	happen again.	होने दिया।		करूँगा ।
8.	I am lucky to have a teacher as wonderful as you are.	मैं भाग्यशाली हूं कि एक शिक्षक के रूप में आप अद्भुत हैं।	3	आप जैसा अद्भुत शिक्षक पा कर स्वयं को मैं भाग्यशाली समझता हूँ ।
9.	Wishing you a meaningful and happy Teachers Day ahead!	आगे आपको एक सार्थक और खुश शिक्षक दिवस की शुभकामनाएं !	2	मेरी ओर से आने वाले अर्थपूर्ण शिक्षक दिवस की शुभकामनाएँ ।
10.	Heaven is indeed very lucky to have you there unlike us.	स्वर्ग वास्तव में बहुत खुशकिस्मत है कि आप हमारे विपरीत हैं।	3	स्वर्ग वास्तव में आपके वहाँ होने से बहुत भाग्यशाली है । हम आप जैसे भाग्यशाली नहीं हैं ।
11.	We are nothing without having you around.	हम आपके आस-पास होने के बिना कुछ भी नहीं हैं।	1	अगर आप हमारे सहायक न हों, तो हम कुछ भी नहीं हैं ।

Results and Discussions:

After the translation of the sentences by the Google, the scores were allotted and also the exact translation of the sentences was also done which is reflected in the last column. The score in the range of 1 to 5 has been allotted and it is observed that no sentence out of the eleven sentences has got the 5 score. Most of the sentences have got very poor score of 2 or even 1. It is observed that the sentences in which any type of sentiments are involved the translation becomes poor.

References:

[1] T. P. Nagarhalli, V. Vaze and N. K. Rana, "A Novel Framework for Neural Machine Translation of Indian-English Languages," 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2020, pp. 676-682, doi: 10.1109/ICICT48043.2020.9112513.

[2] B. N. V. N. Raju and M. S. V. S. B. Raju, "Statistical Machine Translation System for Indian Languages," 2016 IEEE 6th International Conference on Advanced Computing (IACC), Bhimavaram, India, 2016, pp. 174-177, doi: 10.1109/IACC.2016.41.

[3] K. M. Chaman Kumar, S. Aswale, P. Shetgaonkar, V. Pawar, D. Kale and S. Kamat, "A Survey of Machine Translation Approaches for Konkani to English," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, pp. 1-6, doi: 10.1109/ic-ETITE47903.2020.110.

[4] J. Nair, K. A. Krishnan and R. Deetha, "An efficient English to Hindi machine translation system using hybrid mechanism," 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 2016, pp. 2109-2113, doi: 10.1109/ICACCI.2016.7732363.

[5] G. I. Ahmad, J. Singla and N. Nikita, "Review on Sentiment Analysis of Indian Languages with a Special Focus on Code Mixed Indian Languages," 2019 International Conference on Automation, Computational and Technology Management (ICACTM), London, UK, 2019, pp. 352-356, doi: 10.1109/ICACTM.2019.8776796.

[6] N. Hadiya and N. Nanavati, "Indic SentiReview: Natural Language Processing based Sentiment Analysis on major Indian Languages," 2019 3rd International Conference on Computing

Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 322-327, doi: 10.1109/ICCMC.2019.8819786.

[7] B. Gupta et al., "Cross domain sentiment analysis using transfer learning," 2017 IEEE International Conference on Industrial and Information Systems (ICIIS), Peradeniya, 2017, pp. 1-5, doi: 10.1109/ICIINFS.2017.8300363.

[8] Y. Sharma, G. Agrawal, P. Jain and T. Kumar, "Vector representation of words for sentiment analysis using GloVe," 2017 International Conference on Intelligent Communication and Computational Techniques (ICCT), Jaipur, India, 2017, pp. 279-284, doi: 10.1109/INTELCCT.2017.8324059.

[9] Singh, Avinash & Kour, Asmeet & Jamwal, Shubhnandan. (2016). English-Dogri Translation System using MOSES. Circulation in Computer Science. 1. 45-49. 10.22632/ccs-2016-251-25.