Abstract

Machine Translation (MT) is exigent because it involves several thorny subtasks such as intrinsic language ambiguities, linguistic complexities and diversities between source and target language. Usually MT depends upon rules that provide linguistic information. At present, the corpus based MT approaches are used that include techniques like Example Based MT (EBMT) and Statistical MT (SMT). In addition to others, both of these corpus based techniques have different frameworks in the contemporary data-driven paradigm. SMT systems generate outputs using probabilities, whereas EBMT systems translate input text by matching examples from large amount of training data. Urdu MT is in its infancy with very limited availability of required data and computational resources. In this paper, we analyzed and evaluated the main MT techniques using qualitative as well as quantitative approaches. Strengths and weaknesses of each technique have been brought to light through special focus and discussion on examples from Urdu language MT literature. We evaluated the automated machine translated outputs using Bilingual Evaluation Understudy (BLEU). The EBMT approach produced the highest accuracy of 84.21% whereas the accuracy of the online SMT system is 62.68%. We found that BLUE scores of machine translated long Urdu sentences are low in comparison with long sentences. Similarly source text containing low frequency words affect the quality of Urdu machine translation negatively. Experiments and findings section of this paper explicate our
reported results in detail. The paper concludes with proposal of future directions for research in Urdu machine translation.

References

- B. Jawaid, "Statistical Machine Translation between Languages with Significant Word Order Difference", University of Malta & Charles University in Prague, 2010
- S. D. Samantaray, "Example Based Machine Translation approach for Indian Language", International Center for Chemical and Biological Sciences, 2004, 1-10
- N. Karamat, "Verb Transfer for English to Urdu Machine Translation (Using Lexical Functional Grammar (LFG))", National University of Computer & Emerging Sciences, Lahore, Pakistan, 2006

Index Terms

Computer Science          Artificial Intelligence

Keywords

Machine Translation Comparison Rule Based Machine Translation Statistical Machine Translation
Example Based Machine Translation
Bilingual Evaluation Understudy
Urdu to English Machine Translation.