Abstract

Evaluation plays a vital role in checking the quality of MT output. It is done either manually or automatically. Manual evaluation is very time consuming and subjective, hence use of automatic metrics is done most of the times. This paper evaluates the translation quality of different MT Engines for Hindi-English (Hindi data is provided as input and English is obtained as output) using various automatic metrics like BLEU, METEOR etc. Further the comparison automatic evaluation results with Human ranking have also been given.
Assessing the Quality of MT Systems for Hindi to English Translation

- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu, "BLEU: A Method for Automatic Evaluation of Machine Translation", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, pp. 311-318, July 2002.

- Ankush Gupta, Sriram Venkatapathy and Rajeev Sangal, "METEOR-Hindi: Automatic MT Evaluation Metric for Hindi as a Target Language", Proceedings of ICON-2010: 8th International conference on Natural language processing, Macmillan Publishers, India.

- George Doddington, "Automatic Evaluation of Machine Translation Quality using N-gram Cooccurrence Statistics", Proceedings of 2nd HumanLanguage Technologies Conference (HLT-02), SanDiego, CA, pp. 128-132, 2002.

- Banerjee, S. and Lavie, A. "METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments", Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43rd Annual Meeting of the Association of Computational Linguistics (ACL-2005), Ann Arbor, Michigan, June 2005.

- Niladri Chatterjee, Anish Johnson, Madhav Krishna, "Some Improvements over the BLEU Metric for Measuring Translation Quality for Hindi", International Conference on Computing: Theory and Applications (ICICTA 2007), Kolkata, India, March 2007.

- Koehn, P. and Monz, "Manual and automatic evaluation of machine translation between European languages", Proceedings of the Workshop on Statistical Machine Translation, New York City, pp. 102-121, June 2006.

- Ananthakrishnan R, Pushpak Bhattacharyya, M Sasikumar, Ritesh M Shah, "Some Issues in Automatic Evaluation of English-Hindi MT: More Blues for BLEU", Proceedings of 5th International conference on Natural Language Processing, Macmillan Publishers, India.

- Aaron L. F. HAN Derek F. WONG Lidia S. CHAO, "LEPOR: A Robust Evaluation Metric for Machine Translation with Augmented Factors", Proceedings of COLING 2012: Posters, Mumbai, pp. 441-450, December 2012.

- Aaron Li-Feng Han, Derek F. Wong, Lidia S. Chao, Liangye He Yi Lu, Junwen Xing, and Xiaodong Zeng, "Language-independent Model for Machine Translation Evaluation with Reinforced Factors", Proceedings of the XIV Machine Translation Summit, Nice, France, pp. 215-222, September 2-6, 2013.

- Alon Lavie, Kenji Sagae, and Shyamsundar Jayaraman. "The Significance of Recall in Automatic Metrics for MT Evaluation", Proceedings of the 6th Conference of the Association for Machine Translation in the Americas (AMTA-2004), Washington, DC, pp. 134–143, September 2004.

- Machine Translation Evaluation: Human Evaluators Meet Automated Metrics, Workshop at the LREC 2002 Conference, 27 May 2002, Las Palmas, Canary Islands, Spain.

- Xingyi Song and Trevor Cohn and Lucia Specia, "BLEU deconstructed: Designing a Better MT Evaluation Metric", Proceedings of the 14th International conference on Intelligent text processing and computational linguistics (CICLING), March 2013.

- Nisheeth Joshi, Hemant Darbari, Iti Mathur. Human and Automatic Evaluation of English-Hindi Machine Translation Systems, In Proceedings of International Conference on Soft Computing and Artificial Intelligence, Organized by AIRCC, India. Published in Advances in Intelligent and Soft Computing Series, Vol. 166, pp. 423-432. 2012.

- EuroMatrix, "Survey of Machine Translation Evaluation", Available: www.euromatrix.net/deliverables/Euromatrix_D1.3_Revised.pdf
Assessing the Quality of MT Systems for Hindi to English Translation

- Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics. Available: http://www.nist.gov/speech/tests/mt
- Terrence Szymanski, "Recognizing Textual Entailment with a Modified BLEU Algorithm".

Index Terms

Computer Science
Artificial Intelligence

Keywords

Automatic MT Evaluation BLEU METEOR NIST Morphologically Rich Languages.