Analysis and Evaluation of Terminology Translation Consistency in Scientific and Technical Literature

Author(s)

Baosheng Yin

Human-Comput. Intell. Res. Center, Shenyang Aerosp. Univ., Shenyang, China

Xiaodong Yue ; Dongfeng Cai ; Guiping Zhang

In large-scale scientific and technical literature translation in which many people are involved, inconsistency in the translation of the same terminology is inevitable. Firstly, this paper carried out a comprehensive analysis to terminology translation inconsistency, finding that most are translations with same meaning but different indications, which influences the readability of the whole article.

Then, we put forward a semantic similarity-based calculation method to identify this category of terminology in consistency, selected translation with high frequency of network using Internet search engines, and carried out unionization.

Finally, evaluate the literature improvement (post-editing) with two indexes of precision and consistency of terminology translation. In the experimental analysis, 100000-word patent document (English-Chinese human translation) is selected. The consistency index of original translation is 0.494, the consistency index of the processed translation is 0.763. The experimental result indicated that the method effectively improves the terminology translation consistency on the premise of correctly replacing terminology, thus significantly improves the readability of translation.

1. Ture F, Oard D W, Resnik P. Encouraging consistent translation choices[C]//Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2012: 417-426.

2. Itagaki M, Aikawa T, He X. Validation of the consistency of automatic terminology translation: U.S. Patent Application 11/897,197[P]. 2007-8-29.

3. Liu, S., Li, C. H., Li, M., &Zhou, M.Learning Translation Consensus with Structured Label Propagation. The Association for Computational Linguistics (ACL),Jeju Island, Korea, July 2012.

4. Xiao T, Zhu J, Yao S, et al. Document-level consistency verification in machine translation[C]//Machine Translation Summit. 2011, 13: 131-138.

5. Xiaodong Yue Baosheng Yin and Dongfeng Cai. Terminology Translation Consistency Check of Sci-tech Document The 6th Youth Conference of Computational Linguistics, 2012: 276-281.

6. Bao Liu, Guiping Zhang and Dongfeng Cai. Technical term automatic extraction research based on statistics and rules [J]. Computer Engineering and Applications, 2008, 44(23): 147-150.

7. Och F J, Ney H. Giza++: Training of statistical translation models[J]. 2000.

8. Brown P F, Pietra V J D, Pietra S A D, et al. The mathematics of statistical machine translation: Parameter estimation[J]. Computational linguistics, 1993, 19(2): 263-311.

9. Och F J, Ney H. A systematic comparison of various statistical alignment models[J]. Computational linguistics, 2003, 29(1): 19-51. [CrossRef]

10. Koehn P, Och F J, Marcu D. Statistical phrase-based translation[C]// Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1. Association for Computational Linguistics, 2003: 48-54.

11. Qun Liu, Sujian Li. Word Similarity Computing Based on How-net .J . International Journal of Computational Linguistics &Chinese Language Processing, 2002, 7(2): 59-76.

12. Tian Xia. Study on Chinese Words Semantic Similarity Computation. J. Computer Engineering, 2007, 33(6): 191-194.

13. Agirre E, Martinez D. Exploring automatic word sense disambiguation with decision lists and the Web[C]//Proceedings of the COLING-2000 Workshop on Semantic Annotation and Intelligent Content. Association for Computational Linguistics, 2000: 11-19.