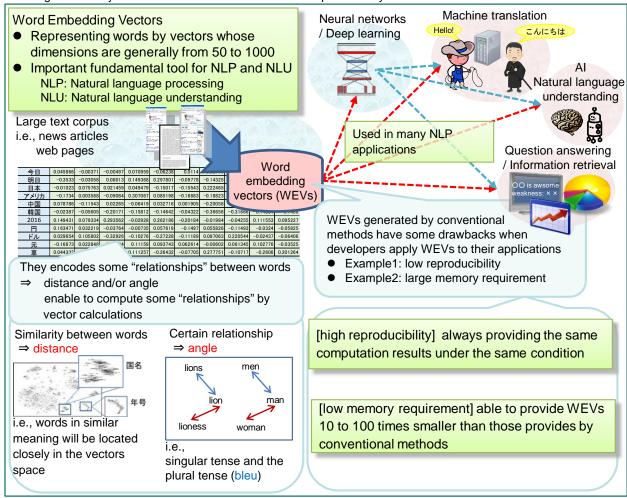
Learning 'neat' semantic representations of words

- Neural word embeddings with high usability -

Abstract

A set of word embedding vectors (WEVs) is a "computer-friendly" dictionary that encodes semantic relationships between words into a vector space. It also enables computers to analogically compute semantic relationships among words. Therefore, WEVs have become an important fundamental resource for many natural language processing (NLP) and understanding (NLU) systems, i.e., machine translation, question answering, and communication with robots. We introduce a method for obtaining computer-friendly WEVs. For example, our method can reduce the memory requirement to store WEVs into memory approximately100 times smaller than those required by conventional methods. As a result, it becomes possible to apply WEVs into applications worked in low-resource devices, such as smartphones. Moreover, our method can also provide high reproducible WEVs whereas WEVs generated by the conventional methods lack such reproducibility.



References

 J. Suzuki, M. Nagata, "Right-truncatable neural word embeddings," in Proc. of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), pp. 1145-1151, 2016.
J. Suzuki, M. Nagata, "Learning compact neural word embeddings by parameter space sharing," in Proc. of the 25th International Joint Conference on Artificial Intelligence (IJCAI), pp. 2046-2052, 2016.

Contact

Jun Suzuki Linguistic Intelligence Research Group, Innovative Communication Laboratory Email: suzuki.jun(at)lab.ntt.co.jp