Vector space model using semantic relatedness

Dmitry Bondarchuk
Ural State University of Railway Transport, Yekaterinburg, Russia

Computers understand very little of the meaning of human language. This profoundly limits our ability to give instructions to computers, the ability of computers to explain their actions to us, and the ability of computers to analyse and to process text. Vector space models (VSM) [1] are used to overcome these limitations. However, classic VSM cannot identify semantic information [2], which results is a significantly lower expert recognition. To solve this problem, we propose a new model based on semantic relatedness (similarity) estimation. Measuring the semantic relatedness of words is a fundamental problem in natural language processing and has many useful applications, including textual entailment, word sense disambiguation, information retrieval and automatic thesaurus discovery. Experimental results indicates that the proposed model outperforms the classic VSM. All experiments are done on several linguistic resources such as dictionaries, corpuses or free encyclopedias etc.

References

[1] G. Salton, A vector space model automatic indexing. *Communications of the ACM.* **18(11)** (1975) 613-620.

[2] J. Mitchell, M. Lapata, Vector-based models of semantic composition. *In Proceedings of ACL.* (2008) 236-244.