Abstract

This paper presents a novel effect of Pronoun Resolution on measurement of document similarity. In this paper we have studied the effect of pronoun resolution within the framework of the Vector Space Model and Probabilistic Latent Semantic Analysis. For this purpose we have developed a Benchmark Corpus consisting of documents whose similarity scores have been given by human beings. We measured the inter-document similarity on these documents using VSM and PLSA. We then performed pronoun resolution on these documents and again calculated the similarity using both methods. Next, the correlation coefficient of the scores was taken with those of the human generated scores. The correlation coefficients clearly demonstrated substantial and consistent improvements of the similarity score after pronoun resolution.

Reference


[11] University of Birmingham, School of computer science http://www.cs.bham.ac.uk/~7Eaxk/ML_PLSA.ppt


Index Terms

Computer Science

Natural Language
Effect of Pronoun Resolution on Document Similarity

Key words

Document Similarity

Pronoun Resolution

Information Retrieval

Statistical Algorithm