Sentence Compression for the LSA-based Summarizer

Josef Steinberger [*]
jstein@kiv.zcu.cz

Karel Jezek[*]
jezek_ka@kiv.zcu.cz

Abstract: We present a simple sentence compression approach for our summarizer based on latent semantic analysis (LSA). The summarization method assesses each sentence by an LSA score. The compression algorithm removes unimportant clauses from a full sentence. Firstly, a sentence is divided into clauses by Charniak parser, then compression candidates are generated and finally, the best candidate is selected to represent the sentence. The candidates gain an importance score which is directly proportional to its LSA score and indirectly to its length. We evaluated the approach in two ways. By intrinsic evaluation we found that the compressions produced by our algorithm are better than baseline ones but still worse than what humans can make. Then we compared the resulting summaries with human abstracts by a standard n-gram based ROUGE measure.

Keywords: Test Summarization, Sentence Compression, Latent Semantic Analysis

---

[*] Department of Computer Science and Engineering, University of West Bohemia, Univerzitni 22, 306 14 Plzen