© 2013 FIZ Karlsruhe

io-port 06098512

Bollegala, Danushka; Okazaki, Naoaki; Ishizuka, Mitsuru

A preference learning approach to sentence ordering for multi-document summarization.

Inf. Sci. 217, 78-95 (2012).

Summary: Ordering information is a difficult but an important task for applications generating natural-language texts such as multi-document summarization, question answering, and concept-to-text generation. In multi-document summarization, information is selected from a set of source documents. Therefore, the optimal ordering of those selected pieces of information to create a coherent summary is not obvious. Improper ordering of information in a summary can both confuse the reader and deteriorate the readability of the summary. Therefore, it is vital to properly order the information in multi-document summarization. We model the problem of sentence ordering in multi-document summarization as a one of learning the optimal combination of preference experts that determine the ordering between two given sentences. To capture the preference of a sentence against another sentence, we define five preference experts: *chronology, probabilistic, topical-closeness, precedence,* and *succession*. We use summaries ordered by human annotators as training data to learn the optimal combination of the different preference experts. Finally, the learnt combination is applied to order sentences extracted in a multi-document summarization system. The proposed sentence ordering algorithm considers pairwise comparisons between sentences to determine a total ordering, using a greedy search algorithm, thereby avoiding the combinatorial time complexity typically associated with total ordering tasks. This enables us to efficiently order sentences in longer summaries, thereby rendering the proposed approach useable in real-world text summarization systems. We evaluate the sentence orderings produced by the proposed method and numerous other baselines using both semi-automatic evaluation measures as well as performing a subjective evaluation.

Keywords: sentence ordering; multi-document summarization; natural language processing
doi:10.1016/j.ins.2012.06.015