

お好みの長さで筋のおった要約を作ります ～修辞構造木の最適刈り込みによる文書要約～

どんな研究

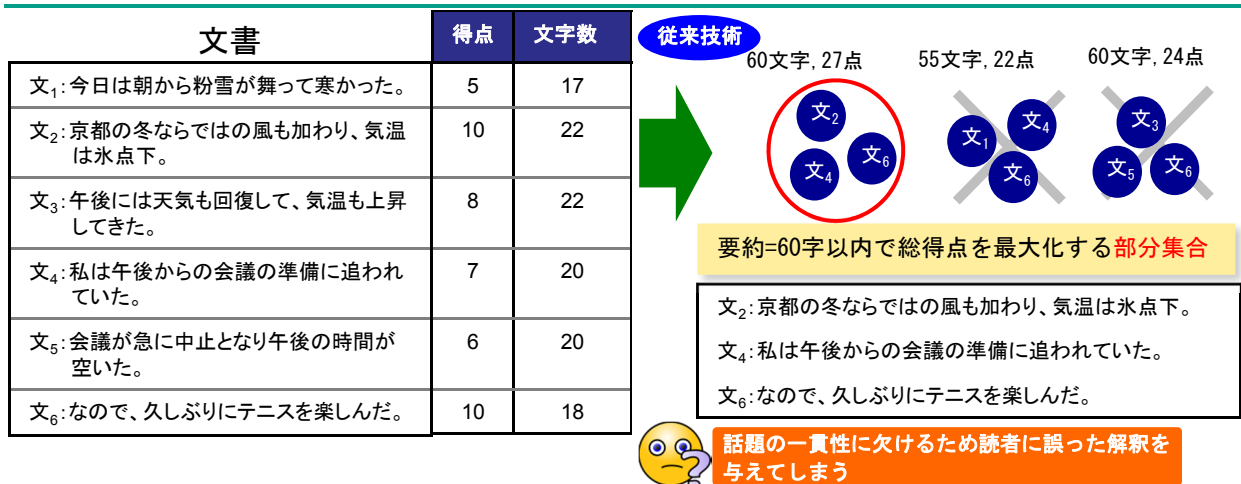
従来の自動要約法は、文書を文の集合としてとらえ、その部分集合を要約としていました。しかし、作成された要約に一貫性が欠けてしまい読みにくいという問題がありました。本展示では、文書が本来持つ**話題構造**を利用して**一貫性のある要約**を生成する技術を紹介します。

どこが凄い

文書を文間の依存関係、文の重要度を反映した木として表現し、根付き部分木を抽出することで一貫性のある要約を生成する手法を考案しました。これは、**ツリーナップサック問題**という計算困難な問題として定式化されるので、効率的に解くためのアルゴリズムも考案しました。

目指す未来

一貫性のある要約を生成することによって、文書を素早く、正確に理解できるようになります。文章を読む手間を減らすことができれば、より時間を有効活用できるようになります。将来的には文短縮を用いてより少ない文字数でも一貫性を保った要約を生成できるようにします。

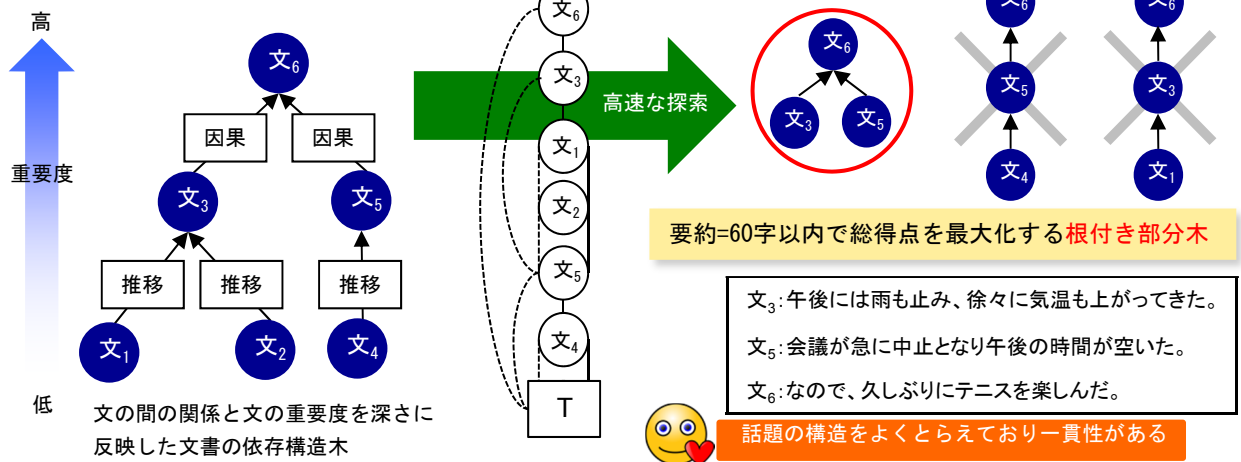


提案技術 【本展示のポイント】

依存構造に基づく修辞構造木 部分木集合の簡潔な表現

【本展示のポイント】

60文字, 24点 58文字, 23点 57文字, 23点



関連文献

- [1] T. Hirao, Y. Yoshida, M. Nishino, N. Yasuda, M. Nagata, "Single-document summarization as a tree knapsack problem," in *Proc. EMNLP-2013*, pp. 1515-1520, 2013.
- [2] 安田宜仁, "制約をともなう動的計画法のZDDを用いた解法," ERATO 湊離散構造処理系プロジェクトセミナー, 2013.
- [3] 吉田康久, 鈴木潤, 平尾努, 永田昌明, "係り受け木に基づく談話構造の提案," 言語処理学会第20回年次大会, pp. 468-471, 2014.

連絡先

平尾 努 (Tsutomu Hirao) 協創情報研究部 言語知能研究グループ
E-mail: hirao.tsutomu{at}lab.ntt.co.jp ({at}の部分をも@に置き換えてください)