

格構造解析による 自動要約生成文の冗長性の除去

橋本 哲弥¹, 三輪 誠², 鶴岡 慶雅¹, 近山 隆¹
1 東京大学 2 マンチェスター大学

背景

• 一般的な自動要約のプロセス

人間は、日常生活の中で日本語や英語といった自然言語をごく当たり前のように使っている。しかし、コンピュータにそれと同じようなことをさせるのは容易ではない。これまで、世界中で半世紀以上にわたって研究を積み重ねてきた。それにもかかわらず、言葉の「意味」をコンピュータ上でどのように表現すべきかといった基本的な問題さえまだ手探りの状態である。



重要文抽出

人間は、日常生活の中で日本語や英語といった自然言語をごく当たり前のように使っている。それにもかかわらず、言葉の「意味」をコンピュータ上でどのように表現すべきかといった基本的な問題さえまだ手探りの状態である。



文圧縮

人間は、日常生活の中で自然言語をごく当たり前のように使っている。それにもかかわらず、言葉の「意味」をコンピュータ上でどのように表現すべきかといった問題さえまだ手探りの状態である。

→ 文圧縮に焦点をあてる

概要

既存手法

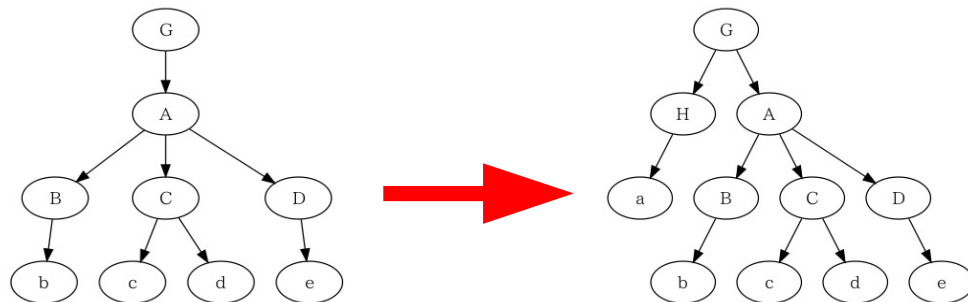
- ・構文木から、学習に基づいた削除規則に従って Noisy-Channel Modelによって句を取り除く

$$P(s/t) \propto P(t/s) * P(s)$$

s: 生成される要約文の構文木

t: 原文の構文木

P(s): sの文としての正しさ



P(s|t)が最大且つ決められた
閾値を超えているs

原文 t

文法的な評価しか行っていないため、文法的に正しいが意味的におかしい文章を生成してしまうことがある。

意味的な評価を考慮した文圧縮を行いたい

概要

提案手法

原文-要約文データから構文木の削除規則を学習する際に、それぞれの句の持つ深層格も学習する



$$P(s/t) \propto P(t/s) * P(s)$$

$P(t/s)$ の計算の際に句の深層格を考慮

それぞれの句が意味的にどのような働きを持つかを、省略可能か不可能かの判断に利用する