

対話的文書読解支援

河野 匡彦 寺田 実¹

東京大学大学院情報理工学系研究科知能機械情報学専攻

1 はじめに

近年 WWW や電子メールなど，ネットワークを通して電子的文書情報に触れる機会が飛躍的に増大した．この膨大な文書の中から求める情報を含む文書を探すことは Google[3] などのサーチエンジンによって精度良く，素早く行なえるようになってきたが，そうやって選び出した文書を実際に読んで理解する部分はいまだに人力に任せている．

そこで本論文では「ユーザーがその文書に何を求めているか」を，自然言語処理を用いて文書の表示にインタラクティブに反映することで文書の理解を助ける，対話的文書読解支援 (Interactive Reading Helper, IRH) を提案する．

2 対話的文書読解支援

対話的文書読解支援はユーザーに「良い要約」を提供するのではなく、「その文書がどういう性格のものかを容易に把握でき，ユーザーが自分の読みやすいように文書の表示を変更できるシステム」を提供することで，ユーザーの読解を支援する．これを以下のように実現した (図1)．

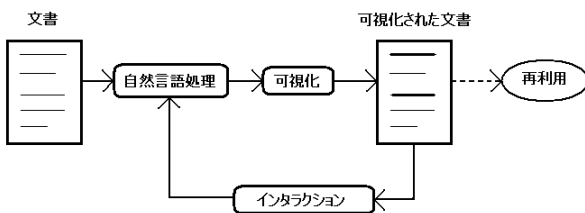


図 1: 対話的文書読解支援

- 自然言語処理技術を利用して単語や文の重要度を計算する．
- 重要度を反映した文書の可視化を行なう．
 - キーワードリストを提示する．

- 重要度の高い文を強調 (もしくは低い文章を目立たないように) 表示する．

* 文を削らないため，詳しく知りたい箇所については自然に周辺の文を読むことができる．

* 斜め読みが容易になる．

- ユーザーが指定した単語をハイライトするなどして，これも強調する．

- ユーザーが指定した単語を重要度計算に反映する．

- 以上の処理を繰り返して対話的に文書の読解を支援する．

こうしてシステムとのインタラクションを繰り返すことで，ユーザーは自分自身にとって読みやすい文書の表示を得ることができる．

そしてこのようなインタラクションの過程そのものが文書理解を促す．なぜならユーザーは読み進みながら自分にとって重要な文，単語を意欲的に拾い上げていくようになり，表示を自分にとってより良いものにすることで更にそれが容易になっていくというポジティブなフィードバックを実現できる．ノートを取りながら文書を読むように，対話的文書読解支援はユーザーが重要な単語や文に注意を向けることを促し，文脈を掴み内容を理解することを支援する．

3 実装

3.1 利用した自然言語処理

文の重要度は以下のように計算する．

1. 名詞の重要度 = $\text{単語出現頻度}(tf) * \log(1 / \text{文書出現頻度}(df))$

2. その他の品詞の重要度 = 0

¹Masahiko Kono, Minoru Terada. university of Tokyo, Graduate School of Information Science and Technology, Mechano-Informatics.

3. ユーザが指定した単語(クエリー)の重要度 = 元の重要度 * 定数(現在は 10)
4. クエリーを含む文に現れた単語の重要度 = 元の重要度 * 定数(現在は 10) * クエリーを含む文に現れた回数 / クエリーを含む文の数
5. 文の重要度 = 名詞と動詞の重要度の和 / 名詞と動詞の単語数

文章要約の研究 [2] において, そもそも人が文章を要約する時に重要度計算 + 重要文抽出という過程を経ているのかという疑問がある [1]. 本システムでは, 対話的処理によって要約生成過程を文章理解の過程と一致させることでこの問題を解消している.

3.2 対話的処理の実現

使いやすさ, 汎用性を考えプロキシサーバ (DeleGate[4]) のフィルタとして実装した. WWW ブラウザから使用できるようになっている. また, これによりユーザとシステムのインタラクションは URL (アドレス + 指定単語) のみで表現されるため, 他の用途 (文章要約など) に容易に利用可能になっている.

このシステムは現在以下のページで使用可能である.

<http://www.sp.ice.ac.jp/~mk/research/>

4 対話的文書読解支援の使用例

本論文の第 2 章「対話的文書読解支援」に対して対話的文書読解支援を行った場合のキーワードリストは例えば以下ようになる. ユーザが「対話」「読解」をキーワードとして選択すると, 選択はしていないがキーワードと関連の深い「文脈」「要約」と言った単語の重要度が上昇し, 表示もそれを反映したものに变化する.

重要度	初期状態	キーワード選択後
1	ユーザ	読解
2	読解	単語
3	箇所	ユーザ
4	フィードバック	文脈
5	可視	要約

5 評価

アンケートを中心とした評価実験を 1 月中に行い, デモの席で発表する予定である.

6 課題と展望

頻出単語は他の単語とともに出現する機会が多い (相関度が高い) ため, 初期重要度を抑えても重要度が上がりやすい. したがってインタラクションが進むにつれてユーザの意図と異なる表示になってしまう. これについては対話的文書読解支援に向けた重要度計算方法を考える必要があると思われる.

7 まとめ

本研究では文章要約と文書の可視化の問題と限界を検討し, ユーザとは独立に要約を生成するシステムではなく, ユーザの要求をインタラクティブに文書の表示に反映することで文書を読みやすくし, 理解を促す対話的文書読解支援を提案した.

ユーザが文書を読みやすくしようとするインタラクションそのものが文書理解を促し, 得られた結果が本人だけでなく, 後でその文書を読む他のユーザーや, 要約システムやサーチエンジンなどにとっても有用なものであることも述べた.

参考文献

- [1] 望主 雅子, 荻野 紫穂, 太田 公子, 井佐原 均. 重要文と要約の差異に基づく要約手法の調査. 情報処理学会研究報告「自然言語処理」Vol.135, No.13, 2000.
- [2] 佐藤 理史, 奥村 学. 電脳文章要約術. 情報処理 Vol.40, No.2, pp.157-161, 1999.
- [3] Google, <http://www.google.com>.
- [4] Delegate, <http://www.delegate.org/delegate/>.