

知的ヘルプシステムのための意味を考慮した テキストマッチング手法の改良と評価

柿間 俊高[†] 村松 幸樹[‡] 杉本 徹[‡]

芝浦工業大学 大学院工学研究科 電気電子情報工学専攻[†]

芝浦工業大学 工学部 情報工学科[‡]

1. はじめに

現在、多種多様なソフトウェアが開発され、ユーザに利用されている。ソフトウェアには使い方の説明をするためにヘルプ機能が備わっているが、ユーザはヘルプ文書の検索や理解に多くの時間を費やしている。これはヘルプ文書特有の言葉遣いが原因と考えられる。そこで自然言語文の意味理解と対話要素を取り入れ、ユーザとヘルプテキストの間に生じる具体性のズレといったギャップを埋め、ユーザの知識に合わせて言い換えたヘルプテキストを提供するシステムが提案されている[1]。しかしこのシステムは、意味理解に用いる言語資源の構築に多くの手間がかかるという課題がある。そこで我々はより容易に構築可能な辞書を用いて、同様の知的ヘルプシステムを構築することを目指しており、システムにおける入力文とヘルプテキストの意味を考慮したマッチング手法を提案している[2]。

本研究では、さらに修飾語や動詞の種類に応じた処理を追加することにより、マッチング手法を改良し妥当性の評価を行う。

2. 知的ヘルプシステム

2.1 システムの概要

知的ヘルプシステムの処理の流れを図1に示す。

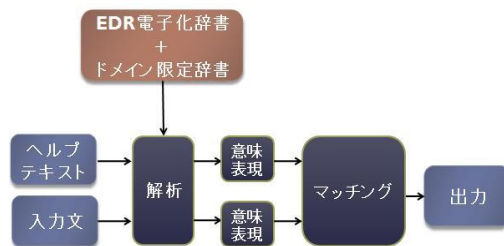


図1 システムの処理の流れ

ユーザが入力した文とヘルプテキストに対して電子化辞書を用いた自然言語解析を行い、解析結果

Improvement and Evaluation of a Semantic Text Matching Method for an Intelligent Help System

†Toshitaka Kakima

Graduate School of Engineering, Shibaura Institute of Technology

‡Koki Muramatsu, Toru Sugimoto

Department of Information Science and Engineering, Shibaura Institute of Technology

として意味表現を生成する。そして入力文とヘルプテキストから生成された意味表現のマッチングを行い、その結果としてユーザの入力に最も関連していると思われるヘルプテキストのタイトルの一覧を出力する。

本システムの特徴は EDR 電子化辞書[3]とドメイン限定辞書[4]という 2 つの電子化辞書を併用することにより、ヘルプ文書特有の言葉の解析およびマッチングを行う点である。

2.2 テキスト解析

マッチングに用いる意味表現を生成するために、本研究で開発中の自然言語解析ライブラリ SEM を用いて入力文とヘルプテキストの解析を行う。SEM が行う解析処理は、形態素解析、係り受け解析、電子化辞書を用いた単語解析と意味解析からなる。単語解析では、単語の概念的意味候補の設定を行い、意味解析では概念候補のフィルタリングと深層格の付与を行う。出力される解析結果は形態素情報や係り受け情報に加え、木構造で表した概念的意味構造を含む。

例として「図形をクリックして文書に挿入する」という文の解析結果を図2に示す。

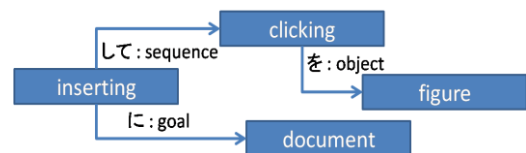


図2 解析結果の例

3. マッチング

本研究では単語解析で付与された単語の概念と、意味解析で付与された深層格の情報を利用して意味と文の構造を考慮したマッチングを行う。この際に、概念同士の類似度は、概念体系上の距離に基づいて定める。マッチングのアルゴリズムについて図3を例に説明する。ここでは、入力文を「吹き出しの背景の色を青に変えたい」、ヘルプテキストを「図形または WordArt の塗りつぶしを追加または変更する」としている。

1. 入力文側とヘルプ側の、木構造における根となる概念を比較し類似度を計算する。根から伸びる枝の深層格が「or」、「and」の場合、それらも比較対象とし、類似度が高い組合せ

を根の類似度とする。図3の場合は、入力側の単語「変える」とヘルプ側の単語「変更」、「追加」の概念を比較し類似度を計算する。

2. 1で対象となった概念から伸びる枝の深層格をチェックする。そして、深層格が等しい節点の概念同士を比較し類似度を計算する。図3では、深層格が「object」である入力文側の「色」とヘルプ側の「塗りつぶし」を比較する。この際、比較対象の節点から伸びる枝の深層格が「or」、「and」、「modifier」の場合、それらも比較対象とし、類似度が最も高い組合せを選ぶ。図3では、「色」と「WordArt」、「色」と「図形」、「塗りつぶし」と「背景」、「塗りつぶし」と「吹き出し」を比較する。
3. 2の処理を係り元の文節がなくなるまで繰り返す。そして、算出された最大類似度を足し合わせ、適用したマッチング回数で割ったものを全体の類似度とする。

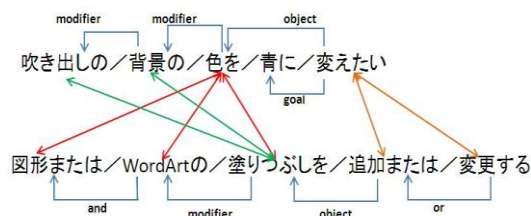


図3 マッチングの例

4. 実験と考察

マッチング手法の有効性を確かめるために入力文のサンプルを与えて複数のヘルプテキストの中から適合するものを検索する実験を行った。検索の対象とするヘルプテキストは、Microsoft Office Word 2007 付属のヘルプテキストで「表」「図やグラフを使った作業」というカテゴリーに属す 89 ページの中から抽出したタイトル 248 文とした。表1は検索対象となるヘルプテキストの例である。

表1 検索対象となるテキストの例

| |
|------------------|
| 表を挿入または作成する |
| 表の書式を設定する |
| 罫線を追加または削除する |
| ワードアートを追加する |
| クリップのコレクションを削除する |
| 透かしや背景を変更する |

入力文のサンプルに関しては理研対話コーパス[5]から収集した。理研対話コーパスは、ワープロ使用に関するユーザとインストラクターとの質疑応答の実験で得た対話データをまとめたコーパスである。この音声コーパスを基にした書き起こしコーパスから収集した 74 文と、新たに被験者実験を行い収集した 41 文の計 115 文とした。表2に入力文とマッチすべきヘルプテキストの例を示す

表2 入力文とヘルプテキストの例

| 入力文 | ヘルプテキスト |
|----------------------|---------------------|
| WordArt を入れたいな | ワードアートを追加または削除する |
| 表のデータを消す | 表の内容を消去する |
| セルの内容をコピーして行全体に貼り付ける | 表の書式を設定する |
| 写真を傾けたい | 図、図形、またはワードアートを回転する |

ドメイン限定辞書を解析に用いているため、ヘルプ特有の言葉を含む場合でも、正解のヘルプテキストとのマッチングを行うことができた。また、深層格が複数付与されたため、文の構造が異なってもマッチングがうまく行われたケースが多かった。しかし、ユーザが用いた言葉と正解のヘルプテキストに含まれる言葉が異なったために、本来は正解となるヘルプテキストとの類似度が低くなるケースがあった。

例えば入力文「写真を傾けたい」と正解のヘルプテキスト「図、図形、またはワードアートを回転する」の2文を比較する際、ユーザが用いた「写真」という言葉とヘルプテキストの「図」という言葉が比較されたが、これらは概念体系上遠い位置にあるため、類似度が低くなった。

5. まとめ

本研究では、[2]で提案したユーザの入力文とヘルプテキストの意味を考慮したマッチング手法の改良を行った。また、ヘルプテキストの検索実験を行うことにより提案手法の妥当性を評価した。

今後は、ユーザの意図を判別しより適切なヘルプテキストを返すために、対話を組み込んだ応答手法を検討していきたい。また、解析に用いているドメイン限定辞書の拡張も行っていく。

参考文献

- [1] 岩下志乃 他: マニュアルテキストを用いた個人化ヘルプシステム, 第 18 回人工知能学会全国大会, 2004.
- [2] 柿間俊高 他: 知的ヘルプシステムにおける入力文とヘルプテキストの意味を考慮したマッチングの評価, 第 90 回人工知能学会知識ベースシステム研究会, 2010
- [3] 日本電子化研究所: EDR 電子化辞書第二版, 2001.
- [4] 荒木亮 他: 知的ヘルプシステムのためのドメイン限定辞書構築手法の提案, 第 72 回情報処理学会全国大会, 2010.
- [5] 理研ワープロ操作対話音声コーパス: 音声資源コンソーシアム,
<http://research.nii.ac.jp/src/index.htm>