

対話型ヘルプシステムにおける項目の絞り込み手法

江口 潤, 佐川 雄二, 田中 敏光
(名城大学大学院理工学研究科)

1. まえがき

ソフトウェアが多機能で便利になるにつれて、そのヘルプシステムが抱える情報量も多くなっていく。そのため、検索にもある程度の知識やコツが必要となり、本当に知りたい項目をすばやく見つけ出すのは困難である。

一般に、ユーザは求めているものを一度の質問で完全に表現する事は困難であり、対話という形で、さらに、どのような情報が必要かをシステムが提示し、誘導できる事が必要である。さらに、質問の表面的な意味だけではなく、より深い意味や、ユーザが何を知りたいのか推測できるような機能を検索システムに取り込めば、より早くと確に知りたい項目にたどり着けるであろう。

本研究では対話式検索システムを作成し、ユーザと対話する事により候補を絞り込んでゆく上で、より少ないやり取りで最適な項目に導くための絞り込み手法について述べる。

2. 対話ヘルプシステム

2.1 基本的動作

システムの使用するヘルプは、図1のような木構造の形で表現されていると仮定する。葉のノードに、ヘルプとして表示する具体的な説明テキストが与えられており、これを「ヘルプ項目」と呼ぶ。中間のノードにはその子孫にあたるヘルプ項目に共通する項目や対象などを表す句が与えられている。

システムはユーザの発話からキーワードを抽出し、3.1節で述べる方法で尤度計算した結果、最も高いヘルプ項目を表示する。ヘルプ項目が1つに限定できない場合、候補を絞り込むための質問をユーザに対して行う。例えばユーザの入力したキーワードから、ノード「文書の書式を設定する」の下位項目の幾つかとノード「図やグラフを使って作業する」の下位項目の幾つかが候補となる場合には、「<文書の書式を設定する>に
関係ありますか」などと質問し、Yesの場合には該当ノードから始まる部分木のみからさらに絞り込んでいく。

図1 ヘルプの木構造[1]

2.2 システムとユーザの概念構造のずれ

前節で述べた方法は、システムに用意されたヘルプに大きく依存する。したがって例えば、ヘルプの木構造とユーザの概念構造に違いがあると、間違った項目へ誘導されてしまう場合がある。

例えば、ユーザが知りたい項目が、「図表目次」であるとする。システムが上位項目である「文書の書式を設定する」までたどり着いて例えば「<文書の書式を設定する>に
関係ありますか」と質問した場合、ユーザが、それが自分の知りたいことと関係があると思わなければ、答えは当然「ノー」であり、結果誤誘導となってユーザは知りたい項目にたどり着けない。

また、システムが「図やグラフを使って作業する」までたどり着いたと仮定する。ユーザが、関係があると判断し「はい」と返答しても、その下位の項目には存在していないため、これも誤誘導となってしまふ。

3. システムの概要

ユーザと木構造との概念構造のずれを発見するには、対話をシステム主導だけで行うのではなく、ユーザからも積極的に情報を追加してもらう必要がある。このため、本システムでは、キーワードの追加入力を許すこととする。図2に本システムの概要を示す。

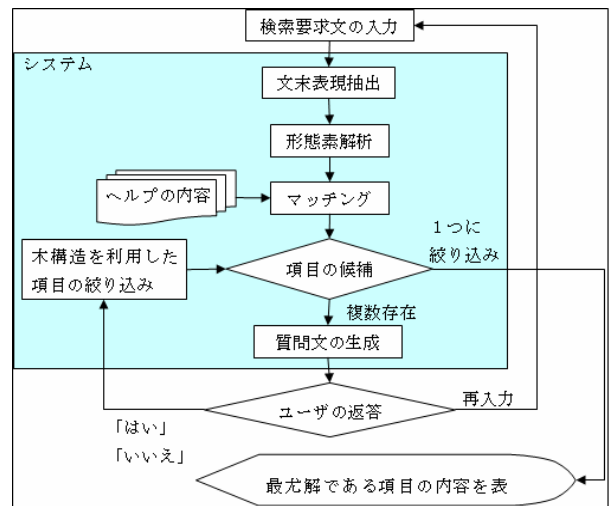


図2 システムの概要

追加入力されたキーワードで項目の尤度を算出し、元の尤度との組み合わせにより尤度を再計算する手法を考案する。

3.1 マッチング

ユーザから入力されたキーワードでヘルプの内容を全文検索し、式①にて各項目の尤度を求める。

$$(\text{葉ノードの尤度}) = \sum_i \left(f_i \cdot \left(\log \left(\frac{N}{d_i} \right) + 1 \right) \right) \cdots \textcircled{1}$$

f_i : キーワード i の文書に含まれる数

N : 項目 (葉ノード) の総数

d_i : キーワード i が含まれる項目 (葉ノード) の数

最終的には正規化されたものが各項目の尤度となる。

3.2 木構造を利用した項目の絞り込み

ユーザからの「はい」または「いいえ」の入力に対して、図2のように木構造をたどっていく。

システムが赤丸で囲まれたノードで質問文を生成したと仮定する。

ユーザが「はい」と返答した場合は、そのノードの子孫ノード以外を検索対象から除く。逆にユーザが「いいえ」と返答した場合は、

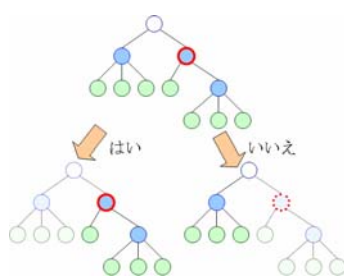


図3 項目の絞り込み

そのノードを含め子孫ノードを検索対象から除く。これを繰り返す、目的項目を絞り込む。

3.3 質問の決定方法

より少ないやり取りで最適な項目に誘導するために、質問コスト[2]を定義して使用した。それに伴い質問文選出方法も考案した。

質問コストは効率的な質問ほど値が小さくなるように質問関数を定義する。質問コストの値は、項目の尤度や検索による候補数などから求める。質問コストが最小となるノードが質問に使われる。計算方法は以下の3種類である。

(1) 手法1 [2]

中間ノードの尤度は子孫ノードの合計とする
式②にて質問コストを算出する

$$(\text{質問コスト}) = |(\text{項目の尤度}) - 0.5| \cdots \textcircled{2}$$

この手法では、Microsoft Word のヘルプのように項目が多数ある場合は、ルートノードに近いノードの項目名が質問文に使われてしまう。そのため、システムからの質問文がより抽象的となってしまう。よって、さらに2つの手法を考案した。

(2) 手法2

中間ノードの尤度は子孫ノードの合計値とする。
ただし、子孫ノードに尤度が0のノードが含まれる場合はそのノードの尤度は0とする
式③にて質問コストを算出する

$$(\text{質問コスト}) = 1.0 - (\text{ノードの尤度}) \cdots \textcircled{3}$$

(3) 手法3

中間ノードの尤度は子ノードの平均値とする
全体の尤度の平均値に近い項目ほど質問に使われるように、式④にて質問コストを算出する

(質問コスト)

$$= \left| (\text{ノードの尤度}) - \frac{1.0}{(\text{尤度が0でない葉ノードの数})} \right| \cdots \textcircled{4}$$

3.4 ユーザによるキーワード再入力

ユーザはシステムからの質問に対して、「はい」「いいえ」ではなく、別のキーワードを含む発話を行うことができる。これは、システムの質問がユーザの意図とずれているとユーザが感じた時に起こりうる。本システムでは、ユーザからキーワードの追加入力があった場合、2.2節で述べた概念構造のずれが生じていると判定して、以下の対応を行う。

追加入力されたキーワードで項目の尤度を算出し、元の尤度との組み合わせにより新たな尤度を求める手法を2つ考案する。

(1) 手法1

元の尤度と新たに算出された尤度を足し合わせる

(2) 手法2

尤度が下がった葉ノードは尤度を0とする。逆に尤度が上がった項目は高くなった尤度を採用する

4. 動作例

以下に質問の決定方法手法2において検索した結果を載せる。

- 1:システム : 質問をどうぞ...
- 2:ユーザ : グラフの作成
- 3:システム : グラフを作成するについてですか?
- 4:ユーザ : はい
- 5:システム : 該当項目:グラフを作成する

5. まとめ

システムからの質問に使う3種類の質問生成方法と、2種類のユーザから追加入力されたキーワードの処理方法を用い、それらを切り替えて対話的に検索できるヘルプシステムを作成した。それぞれの組み合わせによって検索処理を比べ、ユーザとシステムとの対話回数を調査する。

対話回数が少ないほど、システムとして優秀であると考える。

参考文献

- [1] Microsoft Office 2003, Microsoft Office Word ヘルプ。
- [2] 伊藤亮介, 駒谷和範, 河原達也: 電化製品のマニュアルの知識と構造を利用した音声対話ヘルプシステム, 情報処理学会研究報告, Vol.2001-SLP-37, No.68, pp.1-6, (2001) .