

複数の意味構造を持つユーザモデルを用いた問い合わせ理解\*

3C-6

加納 康男 木下 茂行 高橋 友一†  
ATR 通信システム研究所‡

1 まえがき

文献検索システムにおける問い合わせ文にはあいまい性を含むことが多い。我々は、これまでに文献検索システムにおける問い合わせ文を理解する方法を提案してきた。問い合わせの理解を検索意図に相当するキーワード集合を抽出することであると考え、キーワードの検索効率を指標として定量的な評価を行ってきた [1]。また得られた文献の検索効率を指標としてシステム全体の評価も行っている [2]。

我々は問い合わせを理解するためのアプローチとしてユーザモデルを導入し、ユーザモデルの表現方法、構築方法、利用方法を提案した。ユーザモデルは利用者ごとの検索意図を表す概念構造であり、利用者独自の言葉を理解するとともに隠れた検索要求を推論するのに用いる。

本稿では、評価実験を通して明らかになったユーザモデルの問題点およびその解決方法について述べる。

2 現在の検索システムにおけるユーザモデル

ユーザモデルは、名詞(句)で表された概念ノードと概念ノード間の関係を表すリンクを持つ意味ネットワークで表現される。

概念ノードは、検索キーワードに対応するキーワードノードと利用者の独自語に対応する独自語ノードの2種類から構成される。

リンクは、次のように定義されている。

((ノード  $N_p$ ) (ノード  $N_q$ ) (関係の種類  $R$ ) (意図の重み))  
関係の種類  $R$  は、問い合わせ文の係り受け構造や利用者との対話から決定され、階層、同義、関連、限定の4つの種類がある。意図の重みは、'初期状態'、'意図する'、'意図しない'の3段階ある。

ユーザモデルは、問い合わせ文の構造および利用者との対話によって、キーワードノード間を初期状態のリンクで結んだ初期シソーラスの構造を修正することによって構築される。

本検索システムは、ユーザモデルを用いて、問い合わせ文に対応したユーザモデル上の構造から'意図する'重みのリンクをトレースして検索意図に相当するキーワードを抽出するものである。

3 現在の検索システムの評価と問題点

検索システムの評価

検索システムの性能を、キーワードの検索効率(適合率・再現率)および検索意図に相当するキーワード集合の抽出に要した対話回数を指標として評価した。評価実験では被験者に3つの異なる検索意図の問い合わせ文  $Q_1$ 、 $Q_2$ 、

\*Query Understanding with User Models Including Multiple Requirements' Structures

†Yasuo KANOU Shigeyuki KINOSHITA Tomoichi TAKAHASHI

‡ATR Communication Systems Research Laboratories

表1. 従来のユーザモデルを用いた問い合わせ文理解における対話回数推移の例

対話の目的	1回目			2回目			3回目		
	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>
Construction	9	13	1	0	1	0	0	0	0
Editing	0	0	3	0	1	1	0	1	3
Confirmation	1	1	3	3	2	3	1	1	1
Total	10	14	7	3	4	4	1	2	4

表2. 複数の検索意図を表現するユーザモデルを用いた問い合わせ文理解における対話回数推移の例

対話の目的	1回目			2回目			3回目		
	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>	Q <sub>1</sub>	Q <sub>2</sub>	Q <sub>3</sub>
Construction	9	13	2	0	1	0	0	0	0
Selection	0	0	2	1	1	0	0	0	0
Confirmation	1	1	3	3	2	3	1	1	1
Total	10	14	7	4	4	3	1	1	1

Construction: 新しい意味構造の構築

Editing: 他の検索意図の影響の修正

Confirmation: 最終確認と意味構造の精練化

Selection: 候補キーワードの選択

$Q_3$ を作成してもらい、これらを順にシステムに入力して対話を行いキーワードを抽出した。この評価実験は3回繰り返した。このとき同一の検索意図であれば2、3回目の問い合わせ文は変えてもよいとした。問い合わせ文を入力してからキーワードが得られるまでの対話回数の推移の例を表1に示す。

評価実験の結果、1回目の対話でユーザモデルが構築され、このとき対話が進むにつれてキーワードの適合率・再現率が上昇することがわかった。2回目以降、構築されたユーザモデルを利用することによって対話回数は減少し、対話の直後でもキーワードの検索効率が高いことがわかった。

検索システムの問題点

評価実験において、一つの検索意図の意味構造と他の検索意図の意味構造の一部が重なる場合に、干渉が起こるため、意味構造を修正するための対話が生じる場合があった(意味構造は、ある検索意図に相当した概念を含むユーザモデル上の部分をさす)。即ち、このような場合は次の3つの問題が生じている。

(1) 過剰な推論

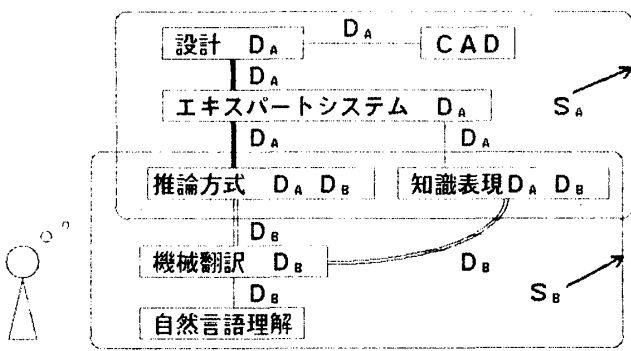
他の検索意図の意味構造にあるキーワードまで余分に推論して出力する。

(2) リンクの削除による影響

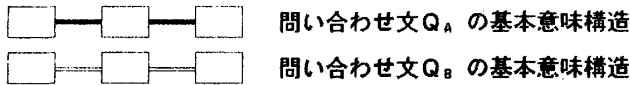
ある検索意図の意味構造のリンクの重みが、他の検索意図での対話で'意図しない'に修正されると、前者の検索意図での再検索時に本来必要なキーワードが推論されない。

(3) 意味構造の固定化

ある検索意図において、一つの概念に接続される意



Q<sub>A</sub> 設計のためのエキスパートシステムの推論方式について  
 Q<sub>B</sub> 機械翻訳における知識表現と推論方式について



S<sub>A</sub>、D<sub>A</sub> ; 問い合わせ文Q<sub>A</sub>の意味構造、意味構造識別子  
 S<sub>B</sub>、D<sub>B</sub> ; 問い合わせ文Q<sub>B</sub>の意味構造、意味構造識別子  
 図1. 複数の意味構造を持つユーザーモデルの例

意味構造が一度構築されると、他の検索意図に適応した意味構造がシステム主導で構築できない。

意味構造が複数存在することによる上記の問題は、CAIで用いられているオーバーレイモデル(利用者の知識構造をシステムの知識構造のサブセットとする)では存在せず、検索システムのような意図を理解するシステムのユーザーモデルに特有である。

4 複数の検索意図を表現するユーザーモデル

本手法は、検索意図ごとの意味構造を区別することによって上記問題点を解決しようとするものである。

問い合わせ理解は、問い合わせ文を検索意図に相当する意味構造にマッピングすることと考える。

ユーザーモデルの表現方法

ユーザーモデル上に検索履歴に応じた複数の意味構造が存在した場合に、これらを識別するために意味構造識別子を導入する。

全てのリンクは、次のように定義される。

((ノード N<sub>p</sub>) (ノード N<sub>q</sub>) (関係の種類 R) (意図の重み) (意味構造識別子 D<sub>i</sub>))

問い合わせ文に対応するユーザーモデル上の構造(基本意味構造)に含まれるノードに対しても意味構造識別子 D<sub>i</sub>を付与する。

複数の検索意図が入力された時のユーザーモデルの例を図1に示す。図1において問い合わせ文Q<sub>A</sub>が入力された時の意味構造をS<sub>A</sub>、問い合わせ文Q<sub>B</sub>が入力された時の意味構造をS<sub>B</sub>で示す。このようにユーザーモデルU<sub>k</sub>(k; 利用者)は各意味構造識別子D<sub>i</sub>が付与された意味構造S<sub>i</sub>の和集合で表現される。

$$U_k = \bigcup_{i=1}^n S_i$$

複数の検索意図の識別

問い合わせ文の意味構造の識別には、その基本意味構造内のリンクの既存の意味構造識別子を用いる。問い合わせ文が既存の意味構造と同一の検索意図で異なる表現

であったときに、既存の意味構造を利用するために、基本意味構造内のノードの意味構造識別子を用いる。

問い合わせ文Q<sub>n</sub>の基本意味構造内のリンクの既存の意味構造識別子D<sub>i</sub>が一意に決定できるとき、D<sub>i</sub>を問い合わせ文Q<sub>n</sub>の意味構造識別子とする。

基本意味構造内のリンクに既存の意味構造識別子がないかもしくは複数の意味構造識別子D<sub>i</sub>があるときは、問い合わせ文Q<sub>n</sub>の意味構造識別子として新しい意味構造識別子D<sub>new</sub>を与える。そして基本意味構造内のノードまたはリンクの意味構造識別子D<sub>i</sub>と既存の意味構造を用いて、次のように新しい意味構造を構築する。

1. 意味構造識別子D<sub>i</sub>が付与された意味構造S<sub>i</sub>に含まれるキーワードを利用者が意図しているかを問い合わせる。
2. 意図していると答えたキーワードに対するリンクに現在の意味構造識別子D<sub>new</sub>を与える。

5 本手法の評価

上記の手法によって、3章の問題点は次のようにして解決される。

- (1) 現在の問い合わせ文に対応する意味構造識別子が付与されたリンクのみを推論に利用することによって、他の意味構造のキーワードの混入を防ぐ。
- (2) キーワードが不要とする指示に対しては、その時の問い合わせに対する意味構造識別子が付与されたリンクの重みのみを修正することによって、他の意味構造に影響を及ぼさない。
- (3) ある概念に対し他の意味構造識別子の‘意図する’リンクで接続されたキーワードが全て不要とされた場合、新たな意味構造識別子を持つ意味構造を再構築する。

前回の評価実験と同一の問い合わせ文に対して前章の手法を施した場合の対話回数 of 推移を表2に示す。表2より、他の検索意図による影響の修正のための対話が、候補キーワードの選択のための対話になり、その対話数は減少していることがわかる。

本手法の対話は、現在の検索意図とは関係のないキーワードが必要なキーワードに紛れて示されることがなく、検索意図ごとのキーワードの集合が示されるので、利用者によりなじみやすいものになっている。

6 むすび

知的文献検索システムにおいて、複数の検索意図による干渉の問題点を明確化した。その解決策として複数の検索意図を表現する意味構造の表現方法、構築方法、利用方法を提案しその評価について述べた。

今後はこのような機構をもった文献検索システム全体に対して、キーワードおよび文献の検索効率(適合率・再現率)による評価を行う。

参考文献

- [1] 木下, 加納他, “知的文献検索システムにおける問合せ理解とその評価”, 情報処理学会, 知識工学と人工知能研究会, No.62, 62-2
- [2] 木下, 加納他, “文献データベースの知的インタフェースとその性能評価”, 情報処理学会, データベース・システム研究会, No.72