

ユーザ間対話に基づくオンライン質問生成

田中 貴志[†] 赤星 祐平[†] 中島 伸介[†]
角谷 和俊[†] 田中 克己[†]

ユーザ間の対話に基づきシステムがリアルタイムに質問を生成し、提示する方式について述べる。我々は、これまでにユーザ間の対話から意図を抽出し、話題の対象物を提示する方式について検討を行っている。本稿では、ユーザの検索意図により、オブジェクト型検索とプロパティ型検索を判別し、それぞれに応じて質問を生成する方式を提案する。また、本方式に基づく、システムの実現方法に関して検討を行う。

Online Query Generation from User Dialog

TAKASHI TANAKA,[†] YUHEI AKAHOSHI,[†] SHINSUKE NAKAJIMA,[†]
KAZUTOSHI SUMIYA,[†] and KATSUMI TANAKA[†]

We describe an online query generation and realtime presentation method from user dialog. We have discussed an extraction of users' intention from their dialog and presentation of topic objects for discussion to encourage their conversation. In this paper, we propose two types of queries: object-type query and property-type query and focus on the adjudication of the types. We also describe the implementation issues of the method and discuss the validity.

1. はじめに

多くの情報の中から自分にとって望ましい情報を取り出す情報検索は、インターネットの急激な普及により、我々の生活において非常に身近なものとなった。Google¹⁾などのweb検索エンジンは、現在多くの人に利用されている。これらの検索システムを作成する上で、使用時におけるユーザの負担を抑えつつ、得られる検索結果をよりよいものにすることが重要である。

Googleに代表されるキーワード型の検索エンジンを利用する際、ユーザが自分の検索要求を正確に把握し、その要求を満たすキーワードを選択しなければならない点である。これらの情報が判明していない場合、一度の検索でユーザが望む検索結果を得るのは困難である。

この問題を解決する手段として、適合フィードバック²⁾やシステムとの対話に基づく検索などがある。ただ、これらの手法はシステムとのインタラクションを何度も繰り返す必要がある。そのため、ユーザには検索を行っているという意識が強くなる。これはユーザ

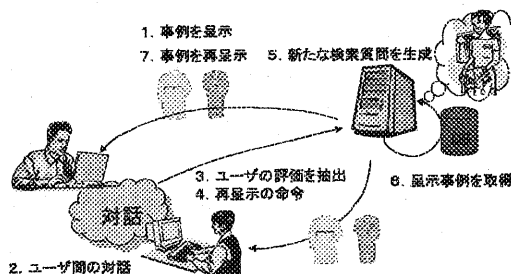


図1 システム概念図

に対して検索への強い意欲を要求し、ユーザの負担を大きくしてしまう。逆にインタラクションを減らすと、ユーザからの十分なフィードバックが得づらくなってしまう。

我々はこの問題を解決するために、対話する相手をシステムではなく現実の人間とする方式を提案した⁴⁾。本方式の概念図を図1に示す。この検索方式は同一の情報を探るユーザグループとそのユーザグループに情報を提示するエージェントから成る。ユーザグループは提示された情報をもとにユーザ同士で対話を行う。エージェントは対話を監視し、ユーザの提示情報に対する評価(ex. 適合フィードバックにおける正例や負

[†] 京都大学大学院 情報学研究科 社会情報学専攻
Division of Social Informatics,
Graduate School of Informatics, Kyoto University

例)等の検索意図を抽出し、リアルタイムで検索質問を生成する(オンライン質問生成)。本手法において、ユーザ側からのシステムとのインタラクションは情報の再呈示の要求だけであり、ユーザの意図のフィードバックはユーザ同士の対話から自動的に抽出される。

本研究では上記の探索の過程において、オブジェクト検索とプロパティ検索、発散と収束から成るフェーズを導入する。前者はユーザが必要としている情報の種別である。オブジェクトとはユーザが検索対象として認識するモノであり、プロパティはそれを説明、表現する情報である。後者は探索におけるユーザの解への位置を表す。発散であれば解には遠く、荒い粒度で様々な情報を呈示する。収束であれば、解に近く、ユーザの検索意図を全て満たすような情報を呈示する。これらの組み合わせによる四種類のフェーズにあわせた検索を行うことによって、ユーザの検索意図に沿った効率のよい情報探索の実現を目指す。

以上の手法により、情報探索において以下の利点をもたらせると考える。

- 検索を行っているというユーザの意識を軽減し、逆にその過程を楽しめるものにできる。
- ユーザの意図の発露を促し、ユーザの意図を十分に反映した検索を可能にする。
- ユーザが自身の検索意図を明確に把握できていない場合、その明確化を促せる。

本手法の応用例としてはウィンドウショッピングが挙げられる。ウィンドウショッピングでは客はあらかじめ自分の欲する商品を特定しているわけではなく、様々な商品を見ながら気に入る商品を探す。このとき、客は商品を探すことだけが目的ではなく、商品を探することを意識させられることを嫌うことがある。しかし、ただ闇雲に商品を眺めるだけでは、客にとって興味のない商品も見ることになり、有益ではない。本手法を利用することにより、検索を意識せずに、その検索意図をフィードバックすることが可能となり、客に適した商品を見せることが可能になると考えられる。

2. 本研究のアプローチ

2.1 対話から呈示までの流れ

本手法はインターネット上においてクライアント-サーバモデルで実現することを想定している。情報を探索するユーザグループ(クライアント側)とそのユーザグループに情報を呈示するサーバから全体のシステムは構成される。ユーザはそれぞれ自身の端末を持ち、他のユーザおよびサーバとの通信を行い、情報の探索を行う。このときの利用の流れを図1を用いて

説明する。

- (1) サーバ側が検索対象の中からいくつかのオブジェクトをユーザに呈示する。
- (2) ユーザはこのオブジェクトをもとに対話を行う。現段階ではユーザ同士の対話はテキストチャットとしている。
- (3) ユーザ間の対話の裏で、システムは対話から呈示オブジェクトに対するユーザの評価等の検索意図を抽出する。
- (4) ユーザが新たなオブジェクトを取得するために、オブジェクトの「再呈示」命令をシステムに送る。
- (5) 抽出された検索意図と直前に使用された検索質問をもとに、新たな検索質問を生成、発行する。新たな呈示オブジェクトを決定する。
- (6) 得られた呈示オブジェクトをユーザに対し、再呈示する。(ステップ2へ戻る)

以上の動作をユーザが満足するか、もしくは検索を諦めるまで繰り返す。以下に対話例を示す。場面は携帯電話を見ている所である。

A「なんだかどれもいまいちだね」

B「全くだよ」

—ユーザが興味を持つプロパティを検索

A「100万画素だってさ」

B「最近のカメラ機能はすごいね」

—該当する携帯を検索

A「最近は動画まで取れるんだね」

B「もっとないかなあ」

—呈示された携帯に類似したものを検索

A「このAA-01って携帯がいいなあ」

B「そうだね」

—AA-01のプロパティを検索

A「この携帯にしよう」

2.2 オブジェクトとプロパティ

検索においてユーザが求める情報は大きく二つに分けられる。ユーザが一つのモノとして認識する情報の名称(以降、オブジェクトと呼ぶ)とそのオブジェクトを説明する情報の総体(属性または評判等、以降、プロパティと呼ぶ)である。例えば、車を検索するユーザにとって、「プリウス」(正確にはプリウスが指すモノ)はオブジェクトであり、「プリウス」の持つ値段やハイブリッドカーという特徴、経験者の運転の感想等がプロパティとなる。

このとき、赤星らの研究¹⁾ではプロパティに対する要求を用いてオブジェクトを検索する方式をオブジェ

クト検索, オブジェクトを特定してそのプロパティを求める検索をプロパティ検索と呼び, 文章中出现する品詞の比率からどちらの検索を使用するか決定している.

我々は本手法をオンラインショッピングへ応用しているが, この場合におけるオブジェクト検索とは商品を検索することであり, プロパティ検索とはその商品の特徴や評判を検索することにあたる.

2.3 発散と収束

我々は対話を基にして検索をしていく過程を発散と収束の二種類のフェーズに分けた⁴⁾. 発散のフェーズとは自身の要求する情報が不明確, 要求を持っていない, または要求する情報がデータベースに存在しなかったユーザに対して, ユーザが興味を持つと類推される情報を荒い粒度で検索し, ユーザの要求の明確化, 新たな要求の喚起を促すフェーズである. 収束のフェーズとは自身の要求が明確化しているユーザに対して, ユーザの検索意図を全て満たすような情報を検索し, ユーザの要求を満たす情報を呈示するフェーズである.

我々は発散と収束のフェーズにおける目的を満たすように, 独立型検索, 従属型検索を以下のように定式化した⁴⁾.

- 独立型検索 $Q_{t+1} = f(I_s) (I_s \subset I_{0 \sim t})$
- 従属型検索 $Q_{t+1} = Q_t + g(I_t)$

ここで, Q_t は一連の対話における第 t 回目の検索質問であり, I_t は t 回目の検索質問の発行以降の対話から現在までに得られたユーザの検索意図の集合, $I_{0 \sim t}$ は対話開始時から現在までに得られた意図の集合である. 本研究で扱う対話から得られる検索意図の種類は後述する.

独立型検索では前回の検索質問は利用せずに, 対話開始時から現在までに得られた検索意図の一部を使用して検索質問を生成する. ユーザの検索意図を緩和することにより, 検索結果間の類似度を低下させ, 様々な結果を呈示する. 発散フェーズで使用する.

従属型検索では前回の検索質問とその検索結果がある状況での対話から得られた全ての意図を利用して検索質問を生成する. 収束フェーズで使用する.

2.4 四種類の検索手法

本研究ではオブジェクト検索とプロパティ検索に対して, 前述の発散および収束の概念をそれぞれ与えて, 合計四種類の検索手法を用意する. 検索質問の生成において, ユーザの検索意図および前回の検索質問が重要となる. 以下に検索手法ごとにその目的と概要を説明する. O はオブジェクトの集合, P はプロパティの集合, I は検索意図の集合を表す. 検索意図の種類は

5.1 節で述べる.

● 独立型オブジェクト検索

過去に抽出された検索意図を利用してユーザが興味を持つと類推されるオブジェクトを検索する. ユーザの検索要求がある程度形成されており, 現在のオブジェクトに否定的または興味が無い場合に使用される.

– 検索質問 $Q_{t+1} = f(I_s) (I_s \subset I_{0 \sim t})$

– 取得結果 $R_{t+1} \subset O$

● 従属型オブジェクト検索

前回の検索質問とそれ以降に得られたユーザの検索意図を利用して, 検索質問をユーザの検索要求に沿うように修正し, 検索する. ユーザの検索要求が明確であり, 現在の呈示オブジェクトがその要求に沿ったものである場合に使用される.

– 検索質問 $Q_{t+1} = Q_t + g(I_t)$

– 取得結果 $R_{t+1} \subset O$

● 発散型プロパティ検索

ユーザが興味を持っていない, 話題にしていないプロパティを検索する. ユーザがオブジェクトに対する知識が不足しており, どのようなプロパティを持っているかわからない場合に使用される.

– 検索質問 $Q'_{t+1} = f'(I_{0 \sim t})$

– 取得結果 $R_{t+1} \subset P$

● 収束型プロパティ検索

ユーザが興味を持っている, 話題にしているプロパティを検索する. 話題にしているプロパティの詳細な情報, 関連情報を得たい場合に使用される

– 検索質問 $Q'_{t+1} = Q'_t + g'(I_t)$

– 取得結果 $R_{t+1} \subset P$

状況に応じてこれら四種類の検索手法を使い分けることにより, 情報の効率的な探索を可能にする.

3. システムの構成

2章で述べたアプローチに従ってシステム的设计を行う. このシステム的设计法はひとつの例であり, 唯一のものではない.

システムのサーバ部分における構成と処理の流れを図2に示す. 全体はオブジェクトを保持しているデータベースと対話文処理部, 検索手法決定部, 検索処理部の3つの処理部から成る. システム使用時において, 本システムはユーザの対話を全体の入力とする.

対話文処理部はユーザ興味, ユーザ評価からなる検索意図を抽出する. この検索意図は以降の処理部で使用されると同時に, オブジェクトの属性評価生成にも使用される.

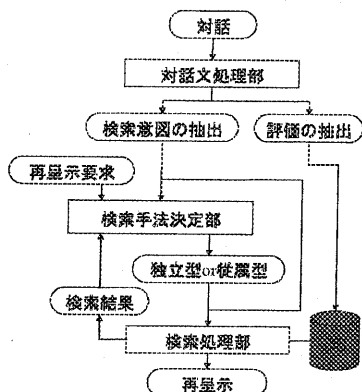


図2 システム構成と流れ

ユーザからオブジェクトの再表示命令が出されたとき、検索手法判定部の処理を行う。当処理部では2.4節で述べられた四種類の検索手法のうち使用する手法を決定する。このとき、オブジェクト検索とプロパティ検索の判定には対話中の一般性の度合い¹⁾を分散と収束の判定には表示事例のユーザへの訴求度⁴⁾を主に使用している。

検索処理部では決定された検索手法を用いて検索対象DBを検索し、該当上位のオブジェクトまたはプロパティをユーザに再表示する。

4. データベースの生成

4.1 データ構造

本研究では検索対象となるオブジェクトをあらかじめデータベースに蓄積しておく。オブジェクト及びそのプロパティがどのような形態で蓄積されているかを説明する。

データはオブジェクト単位でデータベースに保存され、各オブジェクトに対して、プロパティが付随している。プロパティにはユーザへのオブジェクト表示時に使用されるhtml形式の表示用情報と検索時に使用される属性と評価がある。属性はそのオブジェクトを構成する要素を表しており、名詞で表現される。各属性にはオブジェクト内での重要度を表す重みを保持している。評価は属性もしくはオブジェクトそのものに付与されるものであり、形容詞、もしくは形容動詞で表現される。異なる評価が単一の属性に付与されることも認めている。データの一例を図3に示す。

4.2 データ生成方法

次にデータ構造に則ったデータの生成法を述べる。まず、web上からオブジェクトを紹介しているページ

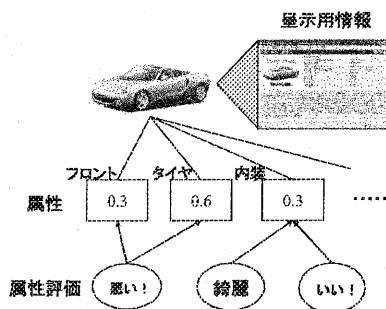


図3 データ構造例

を探し、そのページを表示用情報とする。次に表示用情報としたページ上のテキストを形態素解析にかける。このとき全オブジェクトから得られた名詞と未知語(形態素解析器の辞書にない単語)を抜き出し、これを属性とする。重みは説明文におけるキーワードの出現回数をその説明文の持つ最大の出現回数で割って正規化したものとする。形態素解析には茶筌⁹⁾を使用している。

次に属性評価の生成を行う。評価の生成は表示用情報からは行わず、ユーザの対話から生成していく。5.1節で紹介するユーザ評価のうち評価表現のタイプが形容表現となっているものから属性評価の生成を行う。つまり、評価するオブジェクトおよび評価する範囲で属性評価を付与するオブジェクトおよび属性を特定し、形容表現に当たる評価が付与する。例えば、評価するオブジェクトA、評価する範囲「デザイン」に対して、「綺麗だ」という形容表現があれば、オブジェクトAの属性「デザイン」に対する属性評価「綺麗だ」を付与する。

5. 対話文処理部

5.1 ユーザ興味とユーザ評価

対話文処理部ではユーザ同士の対話から各ユーザの検索意図を抽出する。検索意図はユーザ興味とユーザ評価の二種類に大別している。

● ユーザ興味

ユーザ興味とはユーザが関心を示す事柄であり、特徴ベクトルで表現する。この特徴ベクトルは、発言を形態素解析し、出現する名詞と未知語、形容(動)詞を次元とし、その出現頻度を重みとした特徴ベクトルである。

t 回目の再表示要求時におけるユーザ興味特徴ベクトル I_t は以下のように定義される。前回の表示が行われてから、今回の再表示要求までの発言を

$U_i(i=1,\dots,n)$ とし、その特徴ベクトルを K_i とする。ここで、 $K_i(i=1,\dots,n)$ の次元が等しくなるように、それぞれのベクトルを拡張する。このとき新規に拡張された次元の重みは0とし、新たに生成されたベクトルをそれぞれ $K'_i(i=1,\dots,n)$ とする。このとき $K'_i(i=1,\dots,n)$ の和を I_i とする。

ユーザ興味は独立型オブジェクト検索、従属（独立）型プロパティ検索で使用される。

● ユーザ評価

ユーザ評価は呈示されたオブジェクトに対するユーザの評価である。ユーザ評価は以下の観点で分類している。この内容をまとめたものが表1である。

- － 評価対象オブジェクト（個別、呈示オブジェクト全体、比較）
- － 評価する範囲（オブジェクトそのものへの評価、属性への評価）
- － 評価内容（肯定、否定、形容表現）

ユーザ評価は従属型オブジェクト検索、およびデータ生成時の属性評価を生成する際に使用される。

5.2 ユーザ評価の抽出法

本節では対話からのユーザ評価の抽出法について述べる。

ユーザ評価の抽出にはあらかじめ決められたキーワードを使用する。使用されるキーワードの種類は以下の通りである。

事例指定語 ユーザが話題にしている呈示オブジェクトを表す語。オブジェクトにつけられた名称等がこれにあたる。これには、「どれも」など呈示オブジェクト全てを対象とする語も含む。

要素指定語 話題にされているオブジェクトの属性を表す語。特徴ベクトル空間の基底に対応している語がこれにあたる。

対象肯定語 「良い」などの対象を肯定する語。

対象否定語 「悪い」などの対象を否定する語。

形容語 「安い」や「綺麗」などの形容詞または形容動詞語幹。

否定語 「ない」などの前の単語を否定する語。

同意語 「うん」など相手への同意を表す語。

非同意語 「そうかな」などの相手への非同意を表す語。

逆接語 「けれど」などの前文からの逆接を表す語。

疑問語 「ですか」などの疑問を表す語。

比較語 「より」などの比較表現を表す時に使用する語。

デリミタ 「。」などの文の区切りを表す語。

ユーザ評価の抽出は以上のキーワードを用いて、デリミタで区切られた区間ごとに抽出ルールを適用して行う。以下に比較を除いた評価の抽出ルールを示す。

- (1) 文中に疑問語が存在しないか調べる。もし、存在していれば、その語が文の末尾でなければ、その後の直後で文を区切り、残った文は次の文として調べる。また、疑問語のあった文の評価は直接検索質問には反映しない。
- (2) 文中に逆接語が存在しないか調べる。もし、存在していれば、(文頭でなければ)その直前で文を区切る。
- (3) 文中に事例指定語が含まれているか調べる。含まれていれば、その文章はその事例指定語が指し示すオブジェクトを対象とした文章であるとする。このとき、複数のオブジェクト指定語が含まれていれば、その事例指定語間に上記のキーワードが含まれていないか調べる。含まれていなければ、その文は事例指定語がさす、複数のオブジェクトを対象とした文とし、含まれていれば、二つ目の事例指定語の直前で文を区切る。また、事例指定語が含まれていなければ、前文での対象オブジェクトを現在の文でも対象オブジェクトとする。
- (4) 文中に要素指定語が含まれているか調べる。含まれていれば、その文章はその要素指定語が指し示す要素を対象とした文章であるとする（複数存在すれば、複数を対象としているとする）。含まれていなければ、対象オブジェクトの全体に対する評価を述べた文章とする。
- (5) 文中に対象肯定語、対象否定語および形容語が含まれていないか調べる。含まれていれば、対象に対して、それぞれ、肯定、否定および形容詞からなる評価をつける。ただし、直後に否定語が存在する場合は、それぞれつける評価を逆にする。また、対象肯定語および対象否定語が含まれておらず、逆接語が含まれている場合は前文の評価の逆を現在の文の対象につける。
- (6) 文頭に同意語が存在している場合、前文の評価を現在の文の評価として抽出する。
- (7) 文頭に非同意語が存在している場合、直近の他ユーザの評価を否定したものを現在の文の評価として抽出する。

5.3 評価の統合

本手法ではユーザグループに対して、同じ呈示オブジェクト群を呈示する。そのため、ユーザごとに得ら

表 1 ユーザ評価の種類

評価対象オブジェクト	個別		呈示オブジェクト全体 (global)		比較 (comparison)	
	部分的 (partial)	全体的 (total)	部分的 (partial)	全体的 (total)	部分的 (partial)	全体的 (total)
肯定 (positive)	LPP(ex. この車のデザインはよい)	LTP(この車はよい)	GPP(どの車もデザインがよい)	GTP(どの車もよい)	CPP(この車よりあの車のデザインがよい)	CTP(この車よりあの車がよい)
否定 (negative)	LPN(この車のデザインは悪い)	LTN(この車は悪い)	GPN(どの車もデザインが悪い)	GTN(どの車も悪い)	CPN(この車よりあの車のデザインが悪い)	CTN(この車よりあの車が悪い)
形容 (動) 詞 (adjective)	LPA(この車のデザインは綺麗だ)	LTA(この車は綺麗だ)	GPA(どの車もデザインが綺麗だ)	GTA(どの車も綺麗だ)	CPA(この車よりあの車のデザインが綺麗だ)	CTA(この車よりあの車綺麗だ)

れた評価を一つのものに統合して検索手法の決定、検索処理を行っている。

相反する評価が抽出された場合の統合は、あらかじめユーザに与えた評価尊重度によって、優先する評価を決定して行う。

6. 検索手法決定部

検索手法決定部では検索手法の決定を行う。

まず、再呈示要求時点での訴求度 S 、および潜在的訴求度 PS を以下のように定義する。

• 訴求度

呈示されていたオブジェクト $O_j (j = 1, \dots, l)$ の属性特徴ベクトル F_j から以下のように F を生成する。

$$F = (w_1, w_2, \dots, w_m)^t \quad F_j = (w_{j1}, \dots, w_{jm})^t$$

$$w_s = g\left(\sum_{k=1}^l w_{ks}\right) \quad g(x) = \begin{cases} 1 & x > \theta \\ 0 & 0 \leq x \leq \theta \end{cases}$$

ここで、ユーザ興味特徴ベクトル I_i 、 F の次元が等しくなるように、それぞれのベクトルを拡張する。このとき新規に拡張された次元の重みは 0 とし、新たに生成されたベクトルをそれぞれ $I'_i (i = 1, \dots, n)$ 、 F' とする。

$$I'_i = (\alpha_1, \alpha_2, \dots, \alpha_t)^t \quad F' = (\beta_1, \beta_2, \dots, \beta_t)^t$$

このとき、訴求度 S を次式で定義する。

$$S = \frac{K'^t F'}{\sum_{k=1}^t \alpha_k}$$

• 潜在的訴求度

全てのオブジェクト $O_j (j = 1, \dots, n)$ の属性特徴ベクトル F_j から、訴求度の場合と同様に F_a を生成する。

$$F_a = (w_1, w_2, \dots, w_m)^t \quad F_j = (w_{j1}, \dots, w_{jm})^t$$

$$w_s = g\left(\sum_{k=1}^n w_{ks}\right)$$

ここで、ユーザ興味特徴ベクトル I_i 、 F_a の次元が等しくなるように、それぞれのベクトルを拡張する。

このとき新規に拡張された次元の重みは 0 とし、新たに生成されたベクトルをそれぞれ $I'_i (i = 1, \dots, n)$ 、 F_a' とする。このとき、潜在的訴求度 PS を次式で定義する。

$$PS = \frac{K'^t F_a'}{\sum_{k=1}^t \alpha_k}$$

訴求度とは対話内容と呈示オブジェクトの属性の類似度であり、この値が高いほど呈示内容がユーザの間で話題としてのぼっており、呈示オブジェクトのユーザへの訴求度が高いと考えられる。潜在的訴求度は対話内容と検索対象となる全てのオブジェクトの属性の類似度であり、この値が高いほど、探索中に訴求度の高いオブジェクトを得られると考えられる。

このとき、 S 、 PS 、閾値 α 、 β を用いて、以下のように場合分けを行う。

- (1) $S > \alpha$
- (2) $S \leq \alpha \wedge PS > \beta$
- (3) $S \leq \alpha \wedge PS \leq \beta$

(1) の場合は現在の呈示オブジェクト、プロパティがユーザの検索要求に合っていると考えて、検索手法として、従属型オブジェクト検索、もしくは収束型プロパティ検索を選択する。

(2) の場合は現在の呈示オブジェクトではユーザは不満であるが、話題内容は検索対象に適しており、興味のあるオブジェクトもしくはプロパティを得られる可能性は高い。このときは独立型オブジェクト検索、もしくは収束型プロパティ検索を選択する。

(3) の場合はユーザが持つ認識では検索対象に対して興味を抱かない可能性が高く、ユーザの興味を喚起する必要がある。したがって、発散型プロパティ検索を選択する。

ここで、(3) の場合は一意に検索手法が決定するが、(1)(2) の場合は決まらない。(1)(2) の場合はさらに、一般性度合い G を計算する。 G は以下のようにして計算する。

$$G = \frac{\sum_{i=1}^n g(x_i)}{n}$$

ここで x_i はユーザ興味特徴ベクトルの次元となっている単語であり、 $g(x_i)$ は x_i の一般性の度合いである。この値は x_i が形容詞の場合に高く、地名や固有名詞の場合に最も低くなる。このとき、閾値 γ を用いて、 $G > \gamma$ の場合は従属型オブジェクト検索、独立型オブジェクト検索を選択し、 $G \leq \gamma$ の場合は収束型プロパティ検索を行う。

以上により、検索手法が決定する。

7. 検索処理部

検索処理部では対話から得られた検索意図と選択された検索手法を用いて検索を行う。四種類の検索法の説明を述べる。

7.1 独立型オブジェクト検索

ユーザ興味特徴ベクトルの要素のうち形容(動)詞である要素のみを持つベクトル空間にユーザ興味特徴ベクトルを射影して、新たに形容特徴ベクトル A を生成する。同時に、オブジェクト $O_i (i=1, \dots, n)$ が持つ属性に対してつけられた属性評価を次元とするベクトル空間を生成し、各属性評価の出現頻度を重みとした属性評価特徴ベクトル V_i を生成する。ここで、 A, V_i の次元が等しくなるように、それぞれのベクトルを拡張する。このときオブジェクト O_i のスコアを $A \cdot V_i$ で定義する。全てのオブジェクトに対してスコアを計算し、値の高いものから順に呈示オブジェクトとする。

7.2 従属型オブジェクト検索

ユーザ評価を使用して、Rocchio の式を拡張させた手法でオブジェクトを検索する。詳細は⁴⁾に記述されている。

7.3 発散型プロパティ検索

プロパティ検索では全プロパティ中の属性とその属性につけられた属性評価のペアの中からユーザに呈示するものを検索する。

属性 a_i 、属性評価 v_j からなるペアのスコア S を属性 a_i に対応する属性特徴ベクトルの次元の重みを x_i 、属性評価 v_j に対応する属性評価特徴ベクトルの次元の重みを y_j 、両者に対応するユーザ興味特徴ベクトルの次元をそれぞれ w_i, w_j としたときに以下のように計算する。

$$S(a_i, v_j) = x_i * (1 - w_i) + y_j * (1 - w_j)$$

過去に呈示されたことのない全てのペアに対してスコアを計算し、値の高いものから順に呈示プロパティとする。

7.4 収束型プロパティ検索

発散型プロパティ検索と同様に変数において、属性 a_i 、属性評価 v_j からなるペアのスコア S を以下のよ

うに検索する。

$$S(a_i, v_j) = w_i + w_j$$

過去に呈示されたことのない全てのペアに対してスコアを計算し、値の高いものから順に呈示プロパティとする。

7.5 オブジェクト及びプロパティの呈示

● プロパティの呈示

プロパティ検索において検索された属性、属性評価のペアを同時に呈示する。このとき、属性、属性評価を助詞で連結し、短文を生成して表示を行う。

● オブジェクトの呈示

オブジェクトの呈示は呈示用情報を呈示することにより行う。また、オブジェクト呈示中にそのオブジェクトに関する属性、属性評価のペアが存在する場合はそのペアも呈示する。

8. おわりに

本研究では、検索への意識を軽減しつつ、ユーザの評価、意図を十分に反映した検索を行うために、複数のユーザの対話を利用した検索方式を提案した。さらに、その対話過程において、発散と収束、オブジェクト検索型とプロパティ検索型という軸を導入し、それらに従って検索方法を使い分けることを提案した。今後はシステムの実装と手法の評価を行っていく。

謝 辞

本研究の一部は、平成 15 年度科研費特定領域研究(2)「Web の意味構造に基づく新しい Web 検索サービス方式に関する研究」(課題番号: 15017249, 代表: 田中克己) による。

参 考 文 献

- 1) 赤星祐平, 小山聡, 角谷和俊, 田中克己携帯端末による電子メール交換に基づく Web 検索日本データベース学会 Letters, Vol.2, No.1, pp.111-114, (2003)
- 2) Christa Sommerer, Laurent Mignonneau and Roberto Lopez-Gulliver: Riding the Net, ACM SIGGRAPH 2001, Conference Abstracts and Applications, (2001).
- 3) J. Rocchio. Relevance feedback in information retrieval. In The SMART Retrieval System: Experiments in Automatic Document Processing, pages 313-323. Prentice-Hall Inc., (1971).
- 4) 田中 貴志, 中島 伸介, 田中 克己適合フィードバックにおける複数ユーザの対話からの動的質問生成第 14 回データ工学ワークショップ (DEWS2003), (2003).

- 5) Shinsuke Nakajima, Shinichi Kinoshita and Katsumi Tanaka: "Context-Dependent Information Exploration", Eleventh International World Wide Web Conference (Poster session) (2002).
- 6) 木下真一, 中島伸介, 田中克己: 差異増幅型適合フィードバックと相対的質問評価に基づく画像検索システム, Proc. of DBWeb2002, 情報処理学会シンポジウムシリーズ Vol.2001, No.19, pp.121-128 (2002)
- 7) 徳永健伸: 情報検索と言語処理, 東京大学出版会, pp 151-171, (1999).
- 8) 長尾真: 自然言語処理, 岩波書店 (1996)
- 9) 茶筌 HP: <http://chasen.aist-nara.ac.jp/>
- 10) Goo-net: <http://goo-net.com/>
- 11) Google: <http://www.google.com/>