

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修正

田中 貴志[†] 中島 伸介[†] 田中 克己[†]

[†] 京都大学大学院情報学研究科社会情報学専攻 〒606-8501 京都市左京区吉田本町

E-mail: †{tksh,nakajima,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし 従来の検索では、ユーザの検索意図が不明確であることやデータに関する知識が不十分であることにより、うまくいかないことがある。現実社会では、複数の人が集まって対話することで、お互いの知識を補い、検索意図を明確化するなど、この問題を解決することが出来る。本研究ではこの考えを元にして、ユーザ同士で対話を行いながら検索をし、この対話から意図を抽出して、システムにフィードバックする手法を提案する。このときに、ユーザの対話状況で用いる検索手法を変えることも検討する。これにより、効率のよい情報空間の探索の実現を目指す。

キーワード 情報探索, 協調型検索

Dynamic Query Modification from User Dialog in Relevance Feedback

Takashi TANAKA[†], Shinsuke NAKAJIMA[†], and Katsumi TANAKA[†]

[†] Department of Social Informatics, Graduate School of Informatics, Kyoto University

Yoshida Honmachi, Sakyo, Kyoto 606-8501, Japan

E-mail: †{tksh,nakajima,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract It is difficult for users to gain information they want in an existing system when they don't grasp exactly what they want. In the real world, communication with other people helps them know their demand. Users can obtain information for their purposes more easily after they understand precisely what they really need through an interaction with others. In this research, we propose a information retrieval system based on the idea above. Our system enables users to talk to others when they're seeking information. Our system can automatically create queries based on user's demand which it extract from their dialog. We take particular note of change of query generation method according to dialog conditions. We think that our system enables to explore information space efficiently.

Key words information exploration, collaboration search

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。Google [10] 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。しかし、このようなキーワード型の検索エンジンを利用するためには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する必要がある。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解を得るのは困難である。

この問題を解決するために、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ふ。

このような問題が表れる例として現実世界における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、店員が介入することにより商品を探す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。また、店員側にとっても出来るだけ客に興味のある商品を見せ、購入意欲を高める必要がある。そのためには、客からの何らかのフィードバックを得る方式が必要となる。

我々はこの問題を現実世界からインターネット上に置き換えて、考慮してきた。その結果、我々は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 [3]。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求めるユーザが



図1 システム概念図

ループとそのユーザグループに情報を呈示するエージェントから成る。ユーザグループは呈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、ユーザの呈示情報に対する評価や意図を抽出し、検索質問を生成する。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。

本手法では情報検索において以下の三点を目標とし、実現を目指している。

- 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する。
- 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、その明確化を助ける。また、検索過程における検索意図の移り変わりに対応する。

我々は今まで上記の手法を実現するためにベクトル空間モデルにおける適合フィードバックを利用して、対話から得られる呈示事例への全ての評価を一つの検索質問に反映させる方式をとっていた[3]。しかし、この方式ではすぐに解が類似したものばかりになってしまい、探索が手詰まりになってしまうことが多かった。これは検索意図の明確化をし、その移り変わりに対応するには不十分である。そこで、本研究では適合フィードバックを基にした手法で検索対象空間を検索するための「従属型検索」と探索が手詰まりになったときに一部の評価のみを用いて、解空間を検索する「独立型検索」を用意し、対話状況によって使い分けることによって、検索対象空間を効率よく探索する。

以下、2章では関連研究を、3章では本手法の基本的な考えを、4章から7章ではシステムの構成と各部の詳細を、8章ではプロトタイプシステムを、9章ではまとめについて述べる。

2. 関連研究

本研究における検索方式は適合フィードバック[2]を利用している。これは、一回の検索で最終的な結果を得るのではなく、検索を繰り返し行い、結果をユーザの要求に近づけていく方法である。これは Rocchio により、提唱された方法であり、前回の検索質問 Q_0 から Q_1 を生成する式は以下のように表される。

$$Q_1 = Q_0 + \frac{1}{n_1} \sum_{i=1}^{n_1} R_i - \frac{1}{n_2} \sum_{i=1}^{n_2} R_i$$

ここで、 R はユーザが適合と判断した事例（正例）の集合、 S は不適合と判断した事例（負例）の集合である。 n_1, n_2 はそれぞれ R, S に含まれる事例の数である。

また、対話文を利用して検索を行うものには Riding the Net [1] がある。Riding the Net のシステムの流れを説明する。まず2人のユーザがスクリーンのついたシステムの前にすわり、互いに普通に話して対話を行う。このとき同時に、システムの音声認識エンジンが二人の対話を拾って、そこからキーワードを抽出する。抽出されたキーワードに対応する画像を web 上から探し、システムへとダウンロードする。このダウンロードした画像がユーザの前のスクリーンに生成された3D空間上を次々と通り過ぎていく。ユーザはこの画像を見ながら、対話を行い、もし興味ある画像があったら、スクリーン上のその画像にタッチすることで、その画像をとめじっくり観察することができる。またその画像をひいてきた URL を調べることもできる。

Riding the Net ではが発想、対話支援を目的としているのに対し、本研究は検索を目的としている点で異なっている。本研究の手法はユーザが何らかの情報検索要求を持っている状況での使用を想定している。

3. 基本的な考え

3.1 本手法の利用の流れ

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事例をユーザグループに呈示する。

(2) ユーザはこの事例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事例に対するユーザの評価を抽出する。

(4) ユーザが新たな事例を取得するために、事例の「再呈示」命令をシステムに送る。

(5) 抽出された評価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する。

(6) 新しく生成された検索質問を DB に発行し、新たな呈示事例を決定する。

(7) 得られた呈示事例をユーザに対し、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。

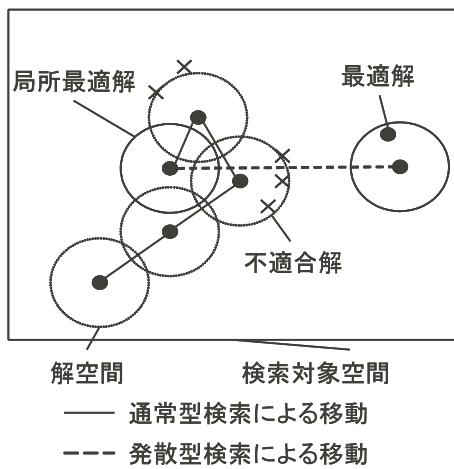


図2 従属型検索と独立型検索による探索の概念図

3.2 本手法の特徴

本研究ではユーザの意図の明確化とその移り変わりに対応するために、質問生成において以下の二点を実現する。

- 従属型検索と独立型検索を用いし、検索対象空間において、行き詰まることを防ぐ。
- 上記検索手法の切り替えをユーザの対話状況等から読み取り、スムーズな情報探索を行う。

3.3 従属型検索と独立型検索

適合フィードバックを用いて情報検索を行う場合、検索対象の特徴をあらわすベクトル空間において、最適解に近づくように解空間を徐々に移動させる。この移動は最適解への収束を目指したものである。しかし、この収束の結果は、局所最適解である可能性がある。つまり、本来ユーザが求めている最適解とは別の解に収束している可能性がある。このとき、ユーザにとって局所最適解が好ましい結果でなかった場合、好ましい最適解の存在があるにもかかわらず、探索を中止する可能性が高くなる。

これを改善するために本研究では、局所的な最適解から抜け出すための検索手法を新しく用意する。つまり、この手法とはあらかじめ使用されていた検索手法によって、収束しつつある解を別の観点から見た検索手法を用いることで解空間の意味的位置を大きく変更するものである。この検索を独立型検索、元の検索手法を従属型検索と呼ぶことにする(図2)。

本研究では従属型検索と独立型検索における質問生成を以下のように定式化する。

- 従属型検索 $Q_{t+1} = Q_t + f(E_t)$
- 独立型検索 $Q_{t+1} = g(E_{0\sim t})$

ここで、 Q_t は第 t 回目の検索質問であり、 E_t は t 回目の事例呈示中に得られたユーザの評価集合、 $E_{0\sim t}$ は検索開始時から t 回目の事例呈示中までに得られた評価集合である。ここで、本研究で扱う評価の種類は評価対象となる事例、評価した範囲およびそれに対する肯定もしくは否定により表1のようにまとめられている。関数 f および g の詳細については6章で述べる。

独立型検索、従属型検索による検索は本研究の応用例として想定しているショッピングにおいてもあてはまる。つまり、店

員が客の意図を全て汲んで商品を紹介していく動作が本研究の従属型検索に対応する。また、客の意図を全て汲んだものが客を満足させなかった、またはそのようなものがなかった場合は客の気に入りそうな代替品を薦める。これが独立型検索に対応する。

このように従属型検索と独立型検索を使い分けて、システムを実現している研究としては他に、Nakajimaらの研究がある[4]。彼らは新しい検索質問を生成する際に、現在の検索質問を基にする生成法を収束型、基にしない生成法を発散型とし、探索を行っている。

3.4 対話からの検索手法切り替え

二つの手法のうち、どちらかを利用するのかを適切に判断することは容易ではない。本研究では対話から二つの手法のうちどちらを使用するとよいのかを推定し、検索を行うことを提案する。基本的には従属型検索で検索を行う。独立型検索を使用すべき場合と考えるのは以下のとおりである。

- 呈示された事例全体に対する否定があった場合
ユーザにとって、よいものが見出せない解空間に収束している可能性があり、独立型検索を行う。
- 呈示されている事例とは関係のない話題が多い場合
ユーザが呈示事例群に興味なくなってきたと考え、独立型検索により、新しい事例を呈示する。
- 検索の結果の上位が変わらない状況が続いた場合
局所最適解に陥っていると考え、独立型検索を行う。

上記三つの概念を元に通常型が発散型を使用するか決める。詳細は6章で述べる。

4. システムの概要

システムのサーバ部分における各部の構成と処理の流れを図3に示す。全体は対話文処理部、検索手法決定部、検索処理部の3つの処理部と呈示事例を保持しているデータベースから成る。

4.1 データベース作成

本手法ではまず、検索対象となる事例をデータベースに保存しておく必要がある。情報の種別、形態は問わないが、事例を表す説明文が付随していることを要求する。この説明文を形態素解析にかけて、全事例から名詞と未知語(形態素解析器の辞書にない単語)を抜き出し、これをキーワードとする特徴ベクトルで各々の事例を表現する。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。また、形態素解析には茶筌[8]を使用している。

4.2 処理部の概要

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。対話文処理部はこの入力を形態素解析にかけて、対話の特徴ベクトルを生成し、これに加えて呈示事例に対する評価等を抽出していく。ユーザから事例の再呈示命令が出されたとき、検索手法判定部の処理を行う。ここでは対話から得られた評価をもとに従属型または独立型どちらの検索手法を用いるかを判定する。次に検索処理部において、決められた検索手法を用いて検索対象DBを検索し、該当上位の事例をユーザに

表 1 呈事例に対する評価の種類

評価事例	個別 (local)		複合 (complex)			
			呈事例全体 (global)		比較 (comparison)	
評価範囲	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)

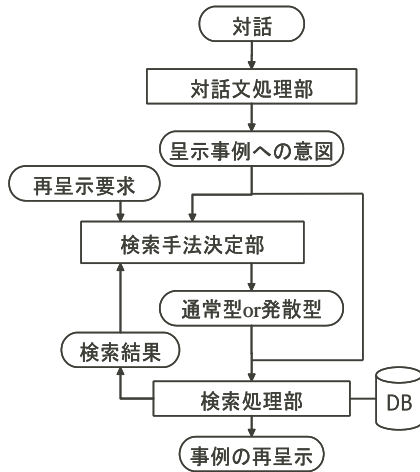


図 3 システム構成と流れ

再呈示する。

5. 対話文処理部

対話文処理部ではユーザ間でなされた対話をあとの処理部で使用できる形に加工する。ユーザの発言から取得する情報は以下のとおりである。

- (1) 発言特徴ベクトル
- (2) 呈事例に対する評価

(1) の発言特徴ベクトルとは、発言を形態素解析し、出現する名詞と未知語を次元とし、その出現頻度を重みとした特徴ベクトルである。

(2) の事例に対する評価は次の 3 点により、12 種類に分類している。

- 評価対象となる事例の範囲 (個別, 呈事例全体, 比較)
- 事例内の注目要素の範囲 (事例の要素全体への評価, 部分への評価)
- 評価内容 (肯定, 否定)

この内容をまとめたものが図 1 である。ただし、現段階では比較の評価抽出は考慮しておらず、呈事例全体への評価も検索質問生成には利用していない。

5.1 評価の抽出法

本節では対話からの評価の抽出法について述べる。対話からの評価抽出にあたって以下の単語に注目する。

事例指定語 ユーザが話題にしている呈事例を表す語。事例につけられた名称および表示された位置等がこれに含まれる。これには、「どれも」など呈事例全てを対象とする語も含む。
要素指定語 ユーザが話題にしている事例の部分を表す語。特

徴ベクトル空間の基底に対応している語がこれにあてはまる。
対象肯定語 「いい」や「良い」などの対象を肯定する語。
対象否定語 「悪い」や「だめ」などの対象を否定する語。
否定語 「ない」などの前の単語を否定する語。
同意語 「うん」や「そうですね」など相手への同意を表す語。
非同意語 「そうかな」など相手への非同意を表す語。
逆接語 「けれど」や「でも」など前文からの逆接を表す語。
疑問語 「ですか」や「？」など疑問を表す語。
比較語 「より」など比較表現を表す時に使用する語。
デリミタ 「。」や「」など文の区切りを表す語。

次にこれらの単語を利用して評価を決定していく。処理はデリミタで区切られた区間ごとに以下のルールを与えて行う。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事例を対象とした文章であるとする。このとき、複数の事例指定語が含まれていれば、その事例指定語間に上記の単語 (2) ~ (7) 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、複数の事例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、前文での対象事例を現在の文でも対象事例とする。

(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、対象事例の全体に対する評価を述べた文章とする。

(5) 文中に対象肯定語および対象否定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、肯定および否定の評価をつける。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) 文頭に非同義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。

5.2 評価の統合

本手法ではユーザグループに対して、同じ呈事例群を呈示する。そのため、ユーザごとに得ら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

検索手法決定では従属型検索を使用するか独立型検索を使用するかを決定を行う。この判断は独立型検索選択指数 H の値により行う。 H は 2.2 節で述べた概念をそれぞれ反映した三つの指数、場の否定指数 P 、話題乖離指数 T 、解連続出現指数 C によって以下のように表される。

$$H = \alpha P + \beta T + (1 - \alpha - \beta) C$$

α, β はどの指数を重要視するかを決めるヒューリスティックな変数である。 P, T, C については後で詳細を述べる。この H がある値 θ 以上の場合は独立型検索を行う。それ以外は従属型検索を行う。

• 場の否定指数

概要 本指数はユーザの呈示されている事例群に対する否定の度合いを表す。呈事例群に対して否定の度合いが大きいと、ユーザが適合した事例を見出せない解空間に収束している可能性があり、探索を中止してしまう可能性が高くなると考える。
定義 画面上に呈示される事例数を n とする。前回の呈示以降に抽出された評価のうち、LTP, LTN, GTP, GTN の数をそれぞれ $N_{LTP}, N_{LTN}, N_{GTP}, N_{GTN}$ とする。場の否定指数 P を次式で定義する。

$$P = \begin{cases} \frac{n - N_{LTP}}{n} & (N_{GTN} > 0) \\ \frac{N_{LTN}}{n} & (N_{GTN} = 0 \wedge N_{LTP} = 0 \wedge N_{GTP} = 0) \\ 0 & ((N_{LTP} > 0 \wedge N_{GTN} = 0) \vee N_{GTP} > 0) \end{cases}$$

• 話題乖離指数

概要 本指数はユーザグループにおいてなされている対話と呈示されている事例の説明文の内容の乖離の度合いを表す。本指数が大きい場合は、ユーザが現在の呈示事例への興味が薄れていると考えられ、独立型検索を行う。

定義 前回の呈示が行われてから、今回の再呈示要求までの発言を $U_i (i = 1, \dots, n)$ とし、その発言特徴ベクトルを K_i とする。さらに、呈示されていた事例 $O_j (j = 1, \dots, l)$ の特徴ベクトル F_j から以下のように F を生成する。

$$F = (w_1, w_2, \dots, w_m)^t \quad F_j = (w_{j1}, w_{j2}, \dots, w_{jm})^t$$

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

ここで、 $K_i (i = 1, \dots, n), F$ 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

$K'_i (i = 1, \dots, n), F'$ とする。ここで $K'_i (i = 1, \dots, n)$ の和を K' とする。

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

このとき、話題乖離指数 T を次式で定義する。

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

• 解連続出現指数

概要 本指数は現在の呈示事例群と直前の呈示事例群との一致の度合いを表す。本指数が大きい場合はユーザが再呈示要求を行っているにもかかわらず、新しい事例を呈示できない状況であるので、独立型検索を行う。

定義 再呈示要求を行う直前の呈示事例集合を O_n 、一つ前の再呈示要求を行う直前の呈示事例集合を O_{n-1} とする。解連続出現指数 C は次式で定義する。

$$C = \frac{\text{num}(O_{n-1} \cap O_n)}{\text{num}(O_n)}$$

7. 検索処理部

検索処理部では対話から得られた評価と選択された検索手法を用いて検索を行う。従属型検索と独立型検索の説明をする。

7.1 従属型検索

従属型検索では LTP, LPP, LTN, LPN を用いて検索質問を生成、検索をする。従属型検索は得られた評価を一度にすべて適用して検索質問を生成するのではなく、

- 1 ある商品に対する全体への評価 (LTP, LTN) からの検索質問生成
- 2 ある商品に対する部分への評価 (LPP, LPN) からの検索質問生成

を順々に適用して、再帰的に検索質問を生成していく。

これがどういうことかを説明するために、まず以下のように定義を行う。

定義

呈示された事例の数を m とし、その一つを O_i と表す。LTP と評価された事例の集合を $SLTP$ とし、LTN と評価された事例の集合を $SLTN$ とする。

また、 O_i の要素で LPP と評価された要素の集合を $SLPP_i$ とし、LP と評価された要素の集合を $SLPN_i$ 、何の評価もつけられなかった属性の集合を SPI_i とする。このとき、 i を 1 から m まで動かしたとき、 $SLPP_i, SLPN_i$ に含まれる要素が等しいものを一つの組とし、これを要素とする集合を SPA とする。この集合の要素は $SPA_j (j = 0, 1, \dots, l, 0 \leq l \leq 2m)$ とする。

このような定義の関係を図 4 にわかりやすく示す。この図において、 $m = 4, n = 9$ である。 SPA は次のように表される。

$$SPA = \{SPA_1, SPA_2, SPA_3, SPA_4\}$$

$$= \{\{SLPP_1, SLPP_2, SLPN_3\}, \{SLPN_1\}, \{SLPN_2, SLPP_3\}, \{SLPP_4\}\}$$

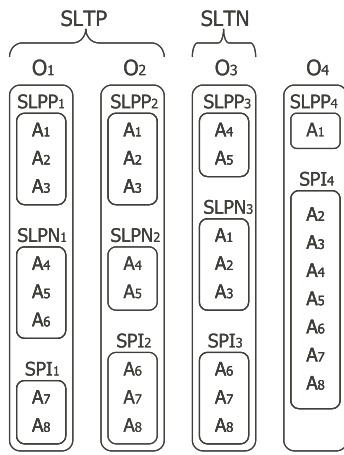


図 4 評価の付与例

$$= \{\{A_1, A_2, A_3\}, \{A_4, A_5, A_6\}, \{A_4, A_5\}, \{A_1\}\}$$

したがって, (*) を言い換えると

1' SLTP, SLTN からの検索質問生成

2' SPA_i からの検索質問生成

ということである。手順 1' は 1 ステップであり, 手順 2' は l ステップ行われる。1 ステップごとに検索質問 Q とユーザの評価から新しい検索質問 Q' を生成し, 次のステップでは Q' を Q と読み替え, さらに新しい検索質問を生成していく。このようにして, すべてのステップを消化したときに生成された検索質問を検索に使用する。

6.1.1 節で手順 1', 6.1.2 節で手順 2' による検索質問ベクトル生成法の具体的な説明を行う。

7.1.1 全体への評価からの生成

手順 1' による新しい検索質問ベクトル Q' の生成法を述べる。

全体の評価からの生成は Rocchio の式を利用する。つまり, 新しい検索質問ベクトル Q' は以下の式であらわされる。

$$Q' = Q + \alpha \sum_{i=1}^{m_1} \frac{R_i}{m_1} - \beta \sum_{i=1}^{m_2} \frac{N_i}{m_2}$$

ここで, R_i は LTP と評価された事例 $O_i (\in SLTP)$ に対する特徴ベクトルであり, N_i は LTN と評価された事例 $O_i (\in SLTN)$ に対する特徴ベクトルである。 m_1, m_2 はそれぞれ適合事例数, 不適合事例数である。 β, γ はそれぞれヒューリスティックな変数である。

7.1.2 部分への評価からの生成

次に手順 2' による新しい検索質問ベクトル Q' の生成法を述べる。

- 基本的な考え方

SPA_i に対し, 肯定, 否定の評価がつけられたとき, これは次の 2 通りの意味をもつ。

(1) SPA_i 内の要素の重要度にだけに注目して, 判断を下した。つまり, SPA_i の要素のみを次元としたベクトル空間に着目し, 判断を行った。

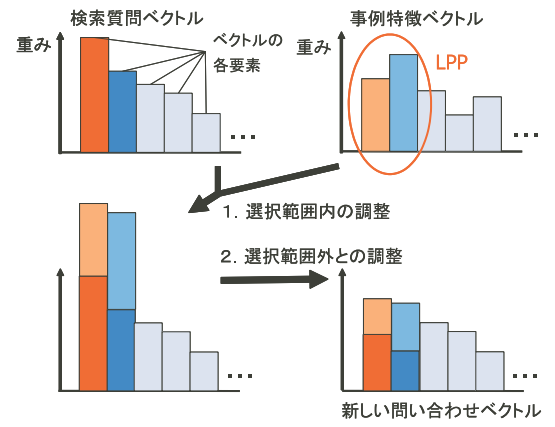


図 5 LPP 評価からの検索質問生成

(2) SPA_i 内の要素の重要度と, SPA_i 外の要素の重要度と比較して, 判断を行った。

この二つの観点を満たすように新しい検索質問ベクトル Q' を生成する。まず, 検索質問ベクトル Q と SPA_i を肯定した事例, SPA_i を否定した事例から一時的な検索質問ベクトル Q_t を生成する (1 を満たす)。その後, Q_t 全体のノルムと Q_t を SPA_i の要素からなるベクトル空間に射影したベクトルのノルム Q^{SPA_i} の比が Q_t 全体のノルムと Q_t を SPA_i に射影したベクトルのノルムの比に近づくように, Q^{SPA_i} を調節する (2 を満たす)。こうして Q' を生成する。

図 5 に部分肯定による検索質問生成の模式図を示す。

- 定義

Q' の計算法を具体的に述べる。

SPA_s に含まれている要素の集合を $A = \{A_1, \dots, A_k\}$ とし, 含まれない要素の集合を $B = \{B_1, \dots, B_l\}$ と表す ($k+l=n$) とする。このとき, O_i ($O_i \in$ 呈示事例のうち評価をつけられた部分への範囲が SPA_s と等しいものの集合 $1 \leq i \leq 2m$) における A_j ($1 \leq j \leq k$) の値を f_{ij}^a とし, B_j ($1 \leq j \leq l$) の値を f_{ij}^b とする。したがって, このときの O_i の特徴ベクトル F_i は以下のように表される。

$$F_i = \underbrace{(f_{i1}^a, f_{i2}^a, \dots, f_{ik}^a)}_{A \text{ に含まれる}} \underbrace{(f_{i1}^b, f_{i2}^b, \dots, f_{il}^b)}_{B \text{ に含まれる}}$$

同様に, 前回の検索質問ベクトルを Q は, B_j の値を q_{bj} , C_j の値を q_{cj} とし, 以下のように表される。

$$Q = (q_1^a, q_2^a, \dots, q_k^a, q_1^b, q_2^b, \dots, q_l^b)$$

また, f_{ij}^a のうち肯定されたものを r_{ij} , 否定されたものを n_{ij} とする。このとき, 新しい検索質問ベクトル Q_t を次のように定義する。

$$Q_t = \underbrace{(q_1^a + g(1), (q_2^a + g(2)) \dots, (q_k^a + g(k)))}_X \underbrace{(q_1^b, q_2^b, \dots, q_l^b)}_Y$$

$$g(j) = \alpha \sum_{i=1}^{m_1} \frac{r_{ij}}{m_1} - \beta \sum_{i=1}^{m_2} \frac{n_{ij}}{m_2}$$

この式では X の部分にだけ着目したとき、確かに部分集合 A に関して Q_{k-1} から F_i に近づいている。しかし、ベクトル全体について、着目したとき、 X における要素が大きくなりすぎてしまい、適切であるとはいえない。だからといって、 Y の部分の値を変更すると、部分集合 A への評価であることに反してしまう。したがって、 X の部分に γ をかけることにより、全体としてのバランスも適切にしてやる必要がある。よって、 Q_t は次のように Q' へと書き直される。

$$Q' = (\gamma(q_1^a + g(1)), \gamma(q_2^a + g(2)), \dots, \\ \gamma(q_k^a + g(k)), q_1^b, q_2^b, \dots, q_l^b)$$

ここで、 γ は次の式を満たすように取る。ここで、 F_i は O_i のうち部分肯定をつけられた事例の特徴ベクトルであり、 m_3 はその個数である。

$$\frac{\|Q'^1\|}{\|Q'\|} = \delta \frac{\|Q^1\|}{\|Q\|} + (1 - \delta) \sum_i^{m_3} \frac{\|F_i^1\|}{\|F_i\|}$$

$$Q'^1 = (\gamma(q_1^a + g(1)), \gamma(q_2^a + g(2)), \dots,$$

$$\underbrace{\gamma(q_k^a + g(k)), 0, \dots, 0}_{i \text{ 個}}$$

$$Q^1 = (q_1^a, q_2^a, \dots, q_k^a, \underbrace{0, \dots, 0}_{i \text{ 個}})$$

$$F_i^1 = (f_{i1}^a, f_{i2}^a, \dots, f_{ik}^a, \underbrace{0, \dots, 0}_{i \text{ 個}})$$

このように Q' を計算することにより、 Q' のノルムと Q' を A に関して射影してできたベクトル Q'^1 のノルムの比が、 Q のそれと F_i のその平均の間（言い換えると、 Q' と Q'^1 の余弦の値が Q と Q^1 の余弦と F_i と F_i^1 の余弦の平均の間）にくる。ここで、 δ の値が大きいくほど Q_k における比に近づき、小さいほど F_i における比に近づく。

7.2 独立型検索

独立型検索は直近に得られた LTP, LPP を使用して質問生成および検索を行う。このとき使用する評価の種類によって、以下のように特徴ベクトル空間の次元をリダクションする。

- LTP である場合

評価の対象となっている事例の特徴ベクトルの重みの大きさ上位 m 個を選択し、選択された m 次元からなるベクトル空間を作成する。

- LPP である場合

評価の対象となっている事例の肯定された特徴（要素）を次元とするベクトル空間を作成する。

このようにして、生成されたベクトル空間において、類似度を計算する。このときの検索質問は評価対象となった事例の特徴ベクトルとし、このときの類似尺度は特徴ベクトルのユークリッド距離に基づいて算出する。この結果、解の上位 n 個において、未知の事例が含まれる割合がある値 θ 以上ならば、その解をユーザに呈示する。 θ 以下ならば、次に最近につけられた LTP, LPP に対して同様の操作を行う。

上記手順では結果的に質問に使用する評価は一つだが、これ

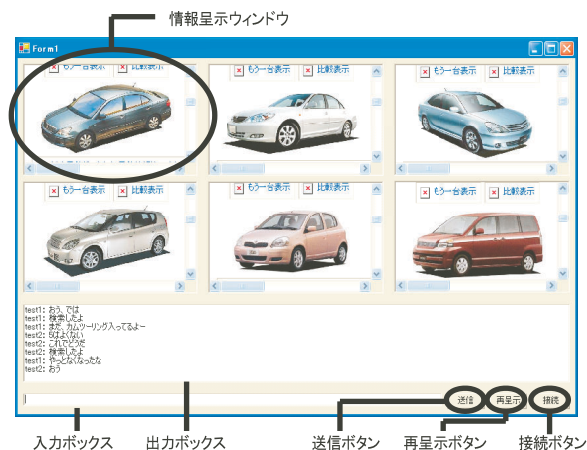


図 6 プロトタイプ画面

に比較および呈示事例全体に対してつけられた評価を組み合わせたことを検討する。

8. プロトタイプシステム

図 6 に本手法を使用したプロトタイプシステムのクライアント側の画面を示す。このプロトタイプでは対象事例を車としている。車情報の呈示用 html および特徴ベクトル生成用の説明文は Goo-net [9] から取得した。

システムを利用するときはまずはじめに、接続ボタンを押し、サーバに接続する。サーバに接続すると、事例（車）が呈示される。この呈示事例について、他のユーザとの対話（チャット）を行う。この対話により、検索質問が生成される。たとえば、
A:「左上のやつがいいよ。だって、燃費がいいみたいだし。」
B:「うん。そうだね。でも、値段がね。」
という会話がなされれば、A の一文目から左上の商品 (a とする) に対する LTP を抽出し、二文目から a の「燃費」に対する部分肯定 (LPP) を抽出する。次に B の一文目および二文目から a の「燃費」に対する部分肯定 (LPP) を抽出し、三文目から a の「値段」に対する LPN を抽出する。その後、再呈示ボタンをおしたときに、抽出された評価を用いて、検索質問を生成する。これをデータベースに発行し、新たな結果を得て、それをユーザに呈示する。

8.1 評価方針

現在、プロトタイプシステムの評価は行っていない。今後、プロトタイプシステムの評価を行う際の方針を述べる。

本研究は以下の 3 つの観点から評価を行う予定である。

- 対話からの評価抽出における評価

対話からの評価抽出時における評価の尺度としては、抽出した評価の正答率とユーザの対話に表れた評価の数により行う。後者は特に一人で検索を行う際に得られるフィードバックとの比較を行う。

- 従属型検索と独立型検索の選択の精度

手動で選択した場合と比較して、本手法で選択した場合の精度を調べる。また、精度を最大にする三つの指数の関係を調べる。

- 検索システムとしての評価検索システムの評価としては、ユーザが探索前に適合すると知らなかった事例を発見した割合、

新規率を重要視する。また、探索終了までにかかった時間は重要ではないと考えている。

9. 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、検索対象空間を効率よく探索するために、対話の状況やそこから得られる意図によって、従属型検索と独立型検索を使い分けることを提案した。今回のプロトタイプシステムでは車を対象事例としたが、今後は旅行の目的地決め等の協調作業に応用も考慮していく。また、課題としては比較および呈示事例全体に対する意図の利用法の検討および手法の評価が挙げられる。

謝 辞

本研究の一部は、平成 14 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: 14019048, 代表: 田中克己), および、平成 14 年度 NEC 共同研究「クロスメディア情報流通システムにおける情報メディアの活性化の研究」(代表: 田中克己) による。

文 献

- [1] Christa SOMMERER, Laurent MIGNONNEAU and Roberto LOPEZ-GULLIVER: RIDING THE NET, ACM SIGGRAPH 2001, Conference Abstracts and Applications,(2001).
- [2] J. Rocchio. Relevance feedback in information retrieval. In The SMART Retrieval System: Experiments in Automatic Document Processing, pages 313-323. Prentice-Hall Inc.,(1971).
- [3] 田中 貴志, 小山 聡, 角谷 和俊, 田中 克己:ユーザ間の対話からのユーザプロフィール抽出と動的情報探索, 情報処理学会研究報告, Vol.2002, No.67, 2002-DBS-128-6, pp.39-46,(2002).
- [4] Shinsuke Nakajima, Shinichi Kinoshita and Katsumi Tanaka: " Context-Dependent Information Exploration ", Eleventh International World Wide Web Conference (Poster session) (2002).
- [5] 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 DB-Web2002, 情報処理学会シンポジウムシリーズ Vol.2001 ,No.19, pp.121-128 (2002)
- [6] 徳永健伸:情報検索と言語処理, 東京大学出版会, pp 151-171,(1999).
- [7] 長尾真:自然言語処理, 岩波書店 (1996)
- [8] 茶筌 HP:<http://chasen.aist-nara.ac.jp/>
- [9] Goo-net:<http://goo-net.com/>
- [10] Google:<http://www.google.com/>