

相対的質問とその質問処理方式

Relative Queries and Their Query Processing

中島 伸介[◆] 木下 真一[♥]
 小山 聡[◆] 角谷 和俊[◆]
 田中 克己[◆]

Shinsuke NAKAJIMA Shinichi KINOSHITA
 Satoshi OYAMA Kazutoshi SUMIYA
 Katsumi TANAKA

従来までの情報検索の多くは、キーワード検索や属性条件による検索、コンテンツベース検索などのように、絶対的条件に基づく問合せにより行われている。この場合の問題点としては、検索しようとするデータについて、何らかの絶対的条件を事前に把握している必要がある。一方、人と人とのコミュニケーションにおいては、絶対的な問合せのみならず、相対的な問合せも行われている。したがって、本研究では人間としては自然ともいえる相対的評価を用いた、相対的情報検索方法を提案する。そこで本論文では、効果的な情報検索を行うための高度なノウハウを要求することなく、サンプルデータ集合の中から解候補を選択するという直感的な操作により情報検索を実現する相対的検索質問処理方式について検討する。

Most conventional information retrieval systems retrieve data based on absolute conditions, such as keywords or metadata. However, we cannot say that users must know the metadata even if they have images of the data that they want. On the other hand, we often ask other person about information we want to know using relative conditions. We believe that it is natural for us to express a favorite data using relative expressions. Therefore, we propose relative retrieval queries based on user's relative expression. Furthermore, this paper describes their query processing.

1. はじめに

近年のWWWの発達により、人々がアクセス可能な情報が急増している。この状況に伴い、情報検索の重要性も高まってきているといえる。従来までの検索技術の多くは、ユーザが得ようとする情報の絶対的条件を指定することによって検索するものが多い。Web情報に対する検索エンジンでのキーワード検索、データベースに対する属性条件による検索、画像データベース等に対するコンテンツベース検索など

がこれに当てはまる。これらの方法は、データを検索しようとする領域に対して、ユーザが数値やキーワード等の絶対的条件を指定することにより、この条件に合致したデータを検索する。

一方、人と人とのコミュニケーションにおいては、絶対的な問合せのみならず、相対的な問合せも行われている。例えば、「今見た黄色い花の写真の中では、この写真が最も気に入ったのですが、ここにある赤い花の写真の中で、私の好みに最も合致するのはどの写真でしょうか。」というように問合せることがこれに当てはまる。このような例示を用いた相対的な質問により、自分の要求を表現することは、人がしばしば振舞う行動の1つである。

相対的な質問による情報検索の利点を以下に示す。

- ・ 検索対象の絶対的条件を把握していなくても、相対的要求を提示することにより問合せが可能
- ・ 例示による情報検索であるため、比較的容易に検索質問を提示することが可能
- ・ 他人の判断による相対的条件の再利用が可能

以上のとおり、本研究では相対的質問による情報検索[1]の概念を提案し、この検索手法の実装に向けた検討を行う。

2. 基本概念

本論文で提案する相対的な情報検索とは、相対的質問に対して、何らかの質問処理を行うことを示す。以下に相対的質問および、その質問処理の基本概念について述べる。

2.1 相対的質問

本研究では、相対的質問を以下のように定義する。

“あるサンプルデータ集合の中からユーザが最も良いと思われるデータを選択した時の、サンプルデータ集合に対する選択されたデータの相対関係”を相対的質問とする。

つまり、比較検討されながら、ユーザに選択されなかったデータと、選択されたデータとの差異に、データ検索におけるユーザの要求が含まれていると考え、これをそのまま質問とするものである。そもそもユーザが比較検討できるデータ量には限りがあり、ユーザが選択するデータは、サンプルデータ集合の中ではユーザの要求に最も近かったというものでしかない。したがって、ユーザが選択するデータそのものをユーザの要求と考えるのではなく、どんなサンプル集合の中で、データが選択されたのか、ということに注目し、これをそのまま質問とする。

ここで従来の質問について考えてみると、画像データベースに対するコンテンツベース検索では、ある画像データを質問とすることでデータベースへの問合せとし、そのデータとの類似度の高いデータ群をユーザに提示する。ただし、この時のジレンマとして、始めに選択したデータがユーザの要求に完全に合致しているのであれば、これ以上データを検索する必要はなく、選択したデータが完全にはユーザの要求に合致していないのであれば、わざわざ合致しないデータに類似するものを質問とすることになる。

次に、ユーザからのフィードバックにより次回の質問を修正する適合フィードバックについて考える。適合フィードバックでは、ユーザはチェックしたサンプルデータの中から、気に入ったデータを選択し、システムは選択されたデータの特徴に基づいて次回の質問修正を行う。この時の質問修正は、ユーザが選択されたデータを気に入っているという前提で行われているが、もしこのデータがユーザの要求を完全に満たしているのであれば、これ以上データ検索を繰り返す必要

[◆] 正会員 京都大学大学院情報学研究所博士後期課程
nakajima@dl.kuis.kyoto-u.ac.jp

[♥] 学生会員 神戸大学大学院自然科学研究科博士前期課程
kinosita@db.cs.scitec.kobe-u.ac.jp

[◆] 正会員 京都大学大学院情報学研究所
oyama.sumiya.tanaka@dl.kuis.kyoto-u.ac.jp

はない。この時のデータ選択は、ユーザがチェックした中では、ユーザの要求に最も近かったということであり、むしろサンプルデータに対する選択されたデータの相対関係にこそ意味がある。

この相対的質問の適用範囲としては、Web情報検索や画像検索、属性条件を持ったデータベースへの問合せも可能である。著者らがこれまでに行ってきたコンテキストブックマーク[2][3][4]による問合せも、相対的質問をWeb情報検索への問合せに利用したものである。このコンテキストブックマークは、ブックマークしたWebページだけに注目するのではなく、どのような状況(コンテキスト)でブックマークをしたのかということブックマークのメタデータとして残すものである。

2.2 質問処理

2.1にて述べたように、相対的質問を決定するために、ユーザが提示すべき情報は、サンプルデータ集合Sと、そのサンプル集合の中から選択したデータxのみである。本研究では、相対的質問に対する質問処理を、扱う各データから特徴ベクトルを抽出して行う。図1に特徴ベクトル空間上での相対的質問のイメージを示す。ただし、Cは、サンプルデータの平均特徴ベクトル、Bはブックマークしたページの特徴ベクトルを意味する。

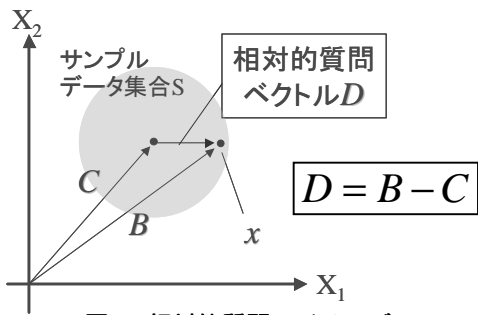


図1 相対的質問のイメージ
Fig.1 Relative Query

選択サンプルの特徴ベクトルBからサンプルデータ集合の平均特徴ベクトルCを差し引いたもの、すなわちB - Cを相対的質問ベクトルDとして保存する。

さて、この相対的質問に対する質問処理方式として、著者らのこれまでの研究において差異増幅処理[5]について提案している。本研究では、新たに相対的処理を提案する。

以下にそれぞれの処理方法について説明する。

1) 差異増幅処理

この処理では、選択されたデータxは、ユーザの要求を完全に満たすものではないという立場から、あえて選択されたデータxとデータ集合Sの平均特徴ベクトルCとの差異ベクトルである相対的質問ベクトルDを増幅することにより、ユーザの検索意図を推定して質問処理を行うものである。前述のコンテキストブックマークによるWeb情報検索では、この処理方法にて行う。(図2参照)

2) 相対的処理

相対的処理とは、ユーザのデータ選択による決定される相対的質問を任意のターゲットデータ集合へ問合せを可能にする質問処理である。すなわち、1章で述べた“黄色い花の画像集合の中から、気に入った画像を選択し、この選択意図に基づいて、赤い花の画像集合の中から気に入った画像を検

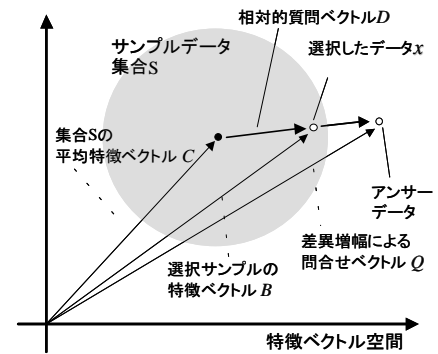


図2 差異増幅処理のイメージ
Fig.2 Difference Amplification Method

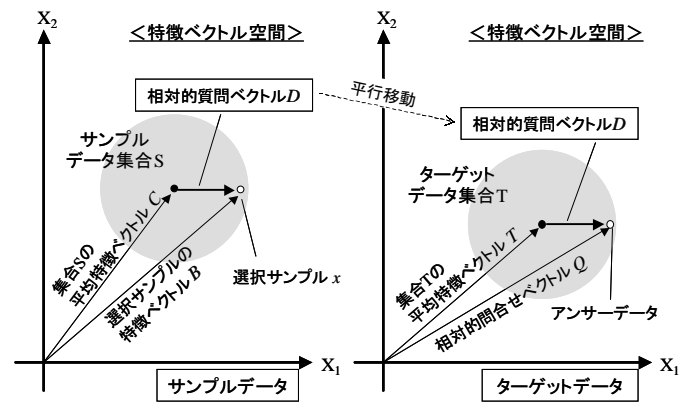


図3 相対的処理のイメージ
Fig.3 Relative Querying Method

索する”という問合せを実現することができる。

この処理では、この相対的質問によって情報検索しようとするターゲットデータ集合の平均特徴ベクトルTに、相対的質問ベクトルDを加えることによって、問合せベクトルQを得る(図3参照)。したがって、相対的検索質問における問合せベクトルQの算出式の基本形は $Q = T + D$ と表現できる。つまり、相対的検索質問における問合せベクトルの算出では、サンプルデータにおけるユーザの相対的評価からのみ、問合せベクトルが生成されるのではなく、ターゲットデータが与えられてはじめて、問合せベクトルを算出することができる。

ここで、相対的問合せ質問 Q_r は、相対的質問に加え、ターゲット集合を指定することで生成することができる。すなわち、選択されたサンプルx、このxの周辺データであるサンプルデータ集合S、相対的問合せを適用するターゲットデータ集合T、の3つのパラメータから成る。ただし、サンプルxはサンプルデータ集合Sの要素である。 $Q_r = (x, S, T)$

あるデータベースDBに対して、この相対的質問 Q_r により問合せた際の答え $Q_r(DB)$ は、xとSの間の相対的な関連を示す評価関数を、 $relative(x, S)$ とすると、以下のようになる。

$$Ans = Q_r(DB) = \{y \mid y \in T, relative(x, S) \approx relative(y, T)\}$$

ここで、 $relative(x, S) \approx relative(y, T)$ の解釈について説明する(図4参照)。

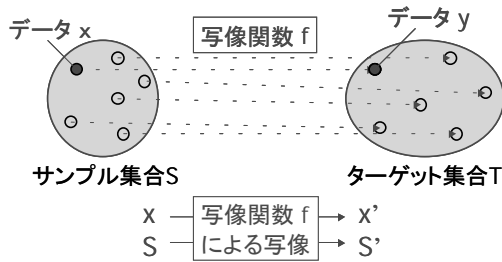


図4 相対的処理の解釈

Fig.4 Interpretation of Relative Querying

サンプル集合Sの各データを、ターゲット集合Tの各データに写像する関数 f が存在するものとし、データ x の写像が x' 、サンプル集合Sの各データを写像したデータ集合を S' とする。この時、集合Sの全ての x において $relative(x, S)$ と $relative(x', S')$ の評価関数の差異が最小となる写像関数 f を求めることができれば、相対的問合せに対する出力データ y は、 $y = f(x)$ とすることができる。

しかしながら、上記の条件を示す写像関数 f を求めることは容易でない。したがって、本研究では集合Tのデータのうち、集合Sにおけるデータ x への相対的特徴ベクトルとの類似度が最も高い特徴ベクトルを持つ、集合T内のデータ y をこの問合せの解とする(図5参照)。

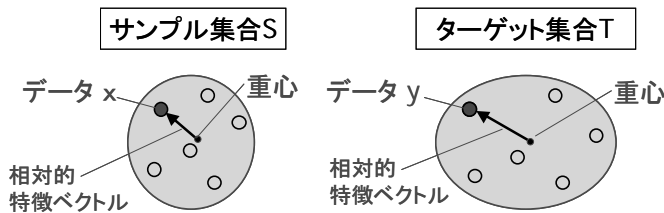


図5 相対的問合せの近似処理

Fig.5 Approximation of Relative Query

3. 相対的問合せに対する相対的処理手順

前章にて相対的情報検索の基本概念について述べた。ここでは本研究で新たに提案する相対的問合せの相対処理の具体的な方法について述べる。以下に処理手順を示す。

- (1) ユーザによる相対的問合せ質問 Qr の提示
- (2) Qr に基づいた相対的問合せベクトル D の抽出
- (3) 相対的問合せベクトル Q の算出
- (4) ターゲットデータ集合T内のデータの検索

3.1 相対的問合せ Qr の提示

ユーザにより提示される相対的問合せ質問 Qr は、前章で述べたとおり、選択されたサンプル x 、その x の周辺データであるサンプルデータ集合S、相対的問合せを適用するターゲットデータ集合T、の3つのパラメータから成る。

$$Qr = (x, S, T)$$

3.2 相対的問合せ Qr に基づいた相対的問合せベクトル D の抽出

相対的問合せベクトルは、サンプル集合の中における、ユーザが選択したデータの相対的関係を示すものであり、この相対的問合せベクトルからユーザの選択意図が推定できると考える。ここで、サンプルデータ集合Sの中でユーザに選択されたデータを x を選択した場合を考える(図6参照)。また、データ x の特徴ベクトルを B とする。

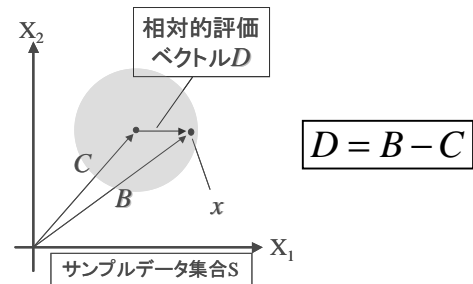


図6 相対的問合せベクトルの算出

Fig.6 Calculating Relative Vector

この時、ユーザは x 自体に興味があった訳ではなく、集合S内の他のサンプルとの違いに注目して x を選択したものと考えられる。したがって、相対的問合せの相対処理においては、集合Sのサンプルの平均特徴ベクトル C を算出し、 B と C の差をとることで、相対的問合せベクトル D を算出する。この相対的問合せベクトル D が集合S内のサンプルに対するユーザの相対的問合せであり、 x に対する相対的評価であると考えることができる。

3.3 相対的問合せベクトル Q の算出

相対的問合せベクトル Q の算出方法について述べる(図7参照)。相対的問合せベクトル Q は、相対的問合せベクトル D とターゲットデータ集合の平均特徴ベクトル T の和からなる。つまり、サンプルデータ集合のデータにおけるユーザの検索意図を、ターゲットデータ集合に重ね合わせることで、サンプルデータ集合Sから x を選択した際のユーザの検索意図をターゲットデータからのデータ検索に利用しようとするものである。

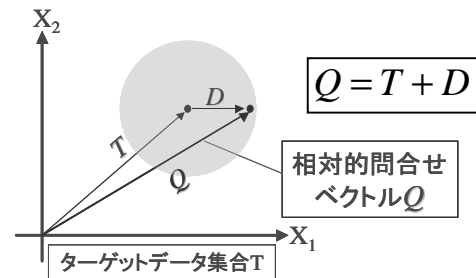


図7 相対的問合せベクトルの算出

Fig.7 Calculating Relative Vector

ただし、ベクトル要素のばらつきが、集合Sと集合Tの間で異なる場合は、ベクトル要素毎の標準偏差を基に相対的問合せベクトルの値を補正することを考える(図8左参照)。ベクトル D, T, Q の i 番目の要素を、 d_i, t_i, q_i とし、集合SとTの i 番目のベクトル次元の標準偏差を s_i, u_i とすると、 q_i は図8に示す算出式に基づいて補正される。

3.4 ターゲットデータ集合T内のデータの検索

相対的問合せベクトル D とターゲットデータ集合の平均特徴ベクトル T から得られる相対的問合せベクトル Q により、ターゲットデータ集合T内のデータの検索を行う。検索方法としては単純に、問合せベクトル Q とターゲットデータ集合内の各データとの類似度計算を行い、最も類似度が高いものをアンサーデータとしてユーザへ提示する。

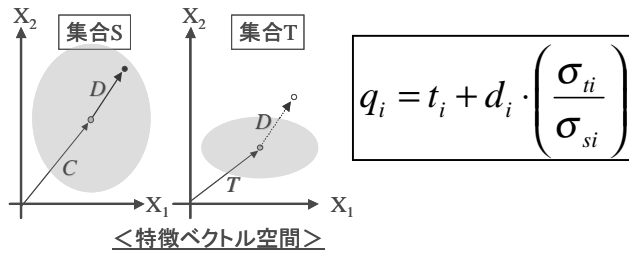


図8 問合せベクトルの各要素のばらつきを補正
Fig.8 Modifying an Element of Query Vector

4. 相対的検索質問の簡易実験

本章では、画像データを用いて行った、相対的検索質問の簡易実験とその結果について述べる。

サンプルデータ集合Sとして“黄色い花”の画像データ集合、ターゲットデータ集合として“赤い花”の画像データ集合を用いた。図9に示すように、黄色い花集合の中から中ぶりの花が中央に写っている写真を選択し、この黄色い花集合の中での選択したデータの相対的評価を基に、同じような赤い花を、ターゲット集合である赤い花集合から検索した。

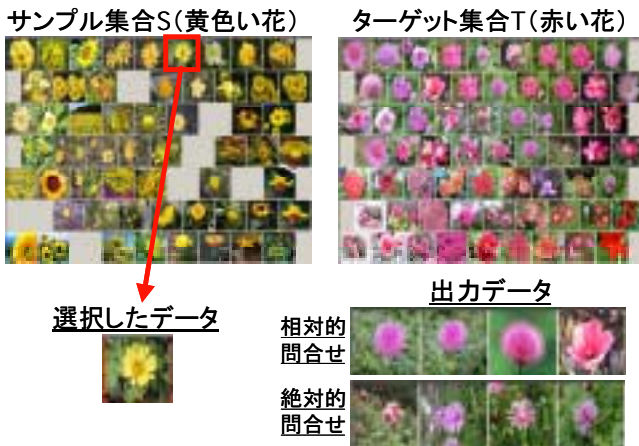


図9 相対的検索質問の実験結果
Fig.9 Result of a Relative Query Test

出力データのうち、上段が相対的問合せによる結果であり、下段が絶対的問合せによる結果である。左側から検索結果上位のものを並べた。この絶対的問合せは、選択したデータとの類似度を計算し、赤い花データのうち類似度の高いデータを解とした。相対的問合せによる結果では、中ぶりの赤い花が中央に位置している画像データが出力されているが、絶対的問合せの結果では、選択したデータに対して適切とはいえない結果が出力されている。当然のことながら、今後十分な評価および検討を行う必要があるが、相対的検索質問の有効性を示すことができたと考えている。

5. おわりに

本論文のまとめを以下に示す。

- ・ 相対的質問を基にした情報検索方法の必要性を示すとともに、その基本概念について検討した。
- ・ 相対的質問に対する質問処理方式として、相対的処理を提案した。また、この処理方式に基づいた簡易実験を行い、この方式の有効性について議論した。

今後は、本格的な実装に向けた課題を整理し、システムの実装および検証を行うつもりである。

[謝辞]

本研究の一部は、文部科学省科学研究費特定領域研究(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)による。ここに記して謝意を表します。

[文献]

- [1] 中島伸介, 小山 聡, 角谷和俊, 田中克己: 相対的検索質問とその質問処理方式, DBWS2002 情報処理学会研究報告 2002-DBS-128 pp.403-410(2002) .
- [2] Shinsuke Nakajima, et al: "Context-Dependent Information Exploration". Eleventh International World Wide Web Conference (Poster session) (2002).
- [3] 中島伸介, 黒田慎介, 田中克己: 閲覧履歴を反映したコンテキスト依存型Webブックマーク, 情報処理学会論文誌: データベース, Vol.43, No.SIG 5(TOD14), pp.23-36 (2002) .
- [4] Shinsuke Nakajima, et al: "Context-Dependent Web Bookmarks and Their Usage as Queries". Proc. of the 3rd International Conference on Web Information Systems Engineering. WISE2002 (2002) (to appear).
- [5] 木下真一, 中島伸介, 田中克己: 差異増幅機能を有する適合フィードバック検索, DBWS2001 情報処理学会研究報告 2001-DBS-125 pp.241-248(2001) .

中島 伸介 Shinsuke NAKAJIMA

京都大学大学院情報学研究科博士後期課程在学中。1997 神戸大学大学院自然科学研究科機械工学専攻博士前期課程修了。日本データベース学会会員, 情報処理学会, 人工知能学会, 環境システム計測制御学会各学生会員。

木下 真一 Shinichi KINOSHITA

神戸大学大学院自然科学研究科博士前期課程在学中。2001 神戸大学工学部情報知能工学科卒業。日本データベース学会学生会員。

小山 聡 Satoshi OYAMA

京都大学大学院情報学研究科助手。2002 京都大学大学院情報学研究科博士後期課程修了。博士(情報学)。情報検索, データマイニングなどの研究に従事。日本データベース学会, 電子情報通信学会, 人工知能学会, ACM, AAAI 各会員。

角谷 和俊 Kazutoshi SUMIYA

京都大学大学院情報学研究科助教授。1998 神戸大学大学院自然科学研究科博士後期課程修了, 博士(工学)。情報処理学会, 日本データベース学会, ACM, IEEE Computer Society, 映像情報メディア学会各会員。

田中 克己 Katsumi TANAKA

京都大学大学院情報学研究科教授。1976 京都大学大学院修士課程修了。工学博士。主にデータベースの研究に従事。情報処理学会, 日本データベース学会, 人工知能学会, ACM, IEEE Computer Society 等各会員。