

興味ベクトル間の距離を関連性とみなした情報の整理

Organization of Information Regarding Distance of Interest Vectors as Relevance

江坂 直紀†
Naoki Esaka

樋口 賢治‡
Kenji Higuchi

原田 史子‡
Fumiko Harada

島川 博光‡
Hiromitsu Shimakawa

1. はじめに

情報化社会が進む昨今、情報はいたるところに溢れている。情報を収集することは容易になってきているが、収集しているだけでは情報は意味をもたない。情報はユーザに活用されて初めて意味をもつ。しかし、情報の種類は多岐にわたるため、ユーザは自らもつ興味にどの程度関連しているかがわからないと、その情報が役立つ度合いを把握できない。役立つ度合いが不明な情報が活用されることはないで、結局これらの情報は収集されただけに終る。よって、情報が活用されるには、ユーザの興味と各情報の関連の度合いを示すことが重要である。

本研究では、情報とユーザがもつ興味との関連性を考慮した情報整理を行うことで、情報の活用の支援を目指す。そのために、ユーザの収集した情報からユーザの興味を抽出し、距離を関連性とみなす手法により、抽出した興味を基に情報を自動で整理するモデルを提案する。

2. 情報整理の必要・重要性

2.1 溢れる情報

現在、我々は大量の情報の中で生活している。街には広告やポスターが溢れ、電車では吊革広告を必ず目にする。このように我々は情報に常に接することで、情報を収集する機会が増加している。しかし、収集した情報が自身にどれほど役立つのかは、きちんと整理しなければ把握困難である。また、情報は整理されていなければ、一時的に利用されるものでしかないが、実際にはユーザは情報を再び利用したいと考えるケースが多い。

このような問題点は、情報の活用を妨げる要因になっている。そこで、情報の活用を支援するために、情報の有益性を定量的に示し、再利用を考慮した情報の検索が可能なる情報整理の枠組みが必要であると考えられる。

2.2 興味による情報整理

情報の活用を支援する整理モデルとして、情報をユーザの興味を基に整理することが有効であると考えられる。人はあるときは勉学、あるときはサークル活動といったように、複数の興味をもっており、それを切り替えながら生活しているととらえることができるからである。ユーザの興味ごとに必要とする情報は異なることから、興味と情報との関連性を定量的に示すことで、情報の活用を支援できると考えられる。

また、人の興味は時間とともに変化するものである。興味の変化にも対応できる情報整理を行うことで、「あのとき興味があったあの情報」といった、ユーザにとって使いやすい情報の検索が可能であると考えられる。

3. 距離を関連性とみなした情報整理

3.1 興味ベクトルの抽出

本研究では、人の興味や情報の対応する興味はスポーツ・健康・学習といったジャンルの値の強弱から構成できるものとしている。これらをベクトルとして表現したものを、「興味ベクトル」と定義する[1]。

人の興味ベクトルはユーザが収集した情報の興味ベクトルから抽出する。ユーザが一定の情報を収集したのち、収集した情報の興味ベクトルに対し、主成分分析を行う。それにより、情報の興味ベクトルを二次元平面上で表し、可視化する。つぎに図 1(a)に示すように、ユーザは情報を収集した意図に合わせてフリーハンドでクラスタ分けを行う。これにより、大まかなユーザの興味が見れる。その後、図 1(a)の G_1 , G_2 , G_3 , G_4 のように、システムはクラスタ内の情報の興味ベクトルの重心を同定する。この重心をユーザの興味ベクトルとして抽出する[2]。

3.2 距離による関連性の表現

システムは抽出したユーザの興味ベクトルを基に情報を自動整理する。ただし、情報はひとつの興味にのみ関連するとは限らず、ユーザは情報がどの興味にどの程度関連するかを比較したい場合がある。そのため、本研究ではひとつ、またはふたつの興味に着目した、二種類の情報整理手法を提案する。

ひとつの興味に着目して情報を整理する場合は、情報と興味の関連性を互いのユークリッド距離により測定する。システムは、興味により近い座標にある情報を関連性が強いととらえて整理する。興味 G_1 を基に情報 I_1 を整理する例を図 1(a)に示す。システムはユークリッド距離 d_1 を求め、それを関連性とみなして情報を整理する。

ふたつの興味を基に情報を整理する場合は、整理の基準となる興味に関連が弱いであろう情報はユーザに示す必要がない。よって、整理の対象となる情報は図 1(b)に示すように、それぞれのクラスタの接線 4 本により囲まれ

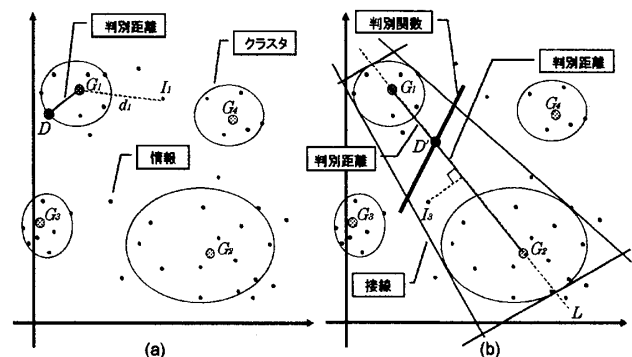


図 1: 二次元平面による情報と興味の表現

†立命館大学大学院 理工学研究科

‡立命館大学 情報理工学部

た四角形に含まれる情報に限定する。これにより、情報をどちらかの興味にある程度関連があるものみに絞ることができる。情報の座標から直線 L に垂線を下ろし、その交点からそれぞれの興味へのユークリッド距離を、情報と興味との関連性としてみなす。

なお、興味間の関連性、情報同士の類似性は、それぞれのユークリッド距離で測定する。このように測定した関連性により、システムは情報を整理する。整理結果は後述の可視化ツールにより、定量的かつ時系列でユーザに表示する。

3.3 興味の判別

情報と興味の距離だけでは、互いの関連性を十分に表現できない場合がある。たとえば、情報が複数の興味から同じ距離にあるからといって、同様の関連性を示すとは限らない。なぜなら、興味ごとにどの範囲まで情報が必要とするのが異なるためである。つまり、興味ごとに距離に対する関連性の重みが異なるのである。そこで本手法では、ここまではこの興味に関連する可能性が高いという距離を求め、それを示すことで、関連性の重みを表現する。その距離を判別距離とよぶことにする。

判別距離は、ひとつの興味に着目した場合と、ふたつの興味に着目した場合で求める方法が異なる。ひとつの興味に着目した場合は、図1(a)のように、重心からクラスタまでの最大の距離にある点 D と重心との距離を判別距離とする。ふたつの興味に着目した場合は、判別分析を用いる。着目したふたつの興味のクラスタ内に存在する情報の座標を説明変数とし、どちらのクラスタに含まれるかを群として判別分析を行い、線形判別関数を求める。その後、図1(b)のように、判別関数と直線 L との交点 D' と、それぞれの重心との距離を判別距離とする。

情報が判別距離よりも重心に近いかどうかを測ることで、その情報と興味との関連性の有無が距離のみの場合と比べて詳細化される。つまり、判別距離とは情報と着目した興味との間の関連の有無を判別する閾値である。閾値を示すことで、同じ距離でも関連性としては異なることを示すことが可能になる。また、システムは判別距離を任意のタイミングで更新する。それにより、ユーザの整理における情報の判別基準を変化させることができる。本手法ではこの判別基準の変化により、時間とともに変化する興味に対応した情報整理を実現する。

3.4 可視化ツール

本研究では図2のようなツールにより、システムが情報整理した結果をユーザに示すことを想定している。(a)はひとつの興味に着目した場合を示し、(b)はふたつの興味に着目した場合を示す。横軸は着目した興味と時間軸を、縦軸はその興味からの距離を表す。(b)ではもうひとつの着目した興味を横軸に平行な直線として追加し、横軸と直線の距離が興味間の関連性を示している。また、 D 、 D' はそれぞれの判別距離を縦軸に示したものである。四角は情報を表しており、興味の軸との距離の遠近において、判別距離を閾値とした関連性の強弱を表現する。なお、情報同士の類似性は、ユーザが類似性を調べたい情報を指定したさいにシステムが計算を行う。そして、類似性の高い情報を関連性と同様、定量的に示す。

このツールにより、ユーザに自身の興味と収集した情報の関連性を定量的に把握させ、情報の活用を支援する。

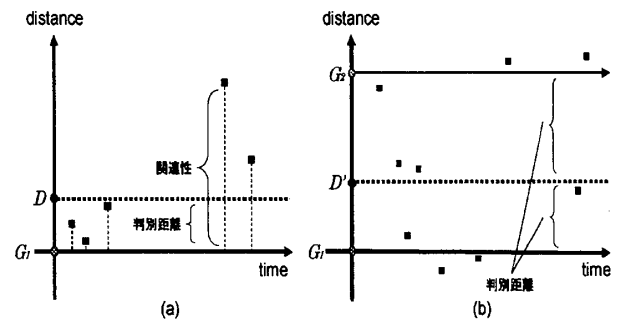


図2：可視化ツール

4. 既存研究

本システムの有用性を検証するために、既存研究との比較を行う。情報を整理し、その活用を支援する研究として、大澤の俺デスク[3]と近藤の Retrospector[4]が挙げられる。

俺デスクは、ユーザの情報に対する着目度と情報間の関連度を基に情報を整理し、情報の検索を支援する。検索結果として関連度が高いものから順に示すが、ユーザは関連性を定量的に把握しづらい。本研究では、距離を関連性として示すことで定量的な把握を可能にする。しかし、本研究は情報に対する着目度は考慮しておらず、ユーザの情報に対するアクセス履歴は利用できない。

Retrospector は、キーワードやサムネイルなどから過去の活動や情報を検索できるように情報を整理する。ユーザの記憶に残っているキーワードなどから情報を検索するので、あのとときのあの興味といった検索はできない。

どちらの研究も、興味といったユーザと情報の関連は考慮していないため、情報の有益性などは判断できない。本研究は情報をユーザの興味に対する関連性を考慮して整理するため、ユーザに有益な情報を定量的に把握することができる。

5. おわりに

本稿では、情報とユーザがもつ興味との関連性を距離により測定することで、情報の活用を支援する情報整理手法を提案した。

今後は、本システムの有用性を数値的に検証するために、システムの実装、実験を行っていく。

参考文献

- [1] 江坂 直紀, 高田 秀志, 島川 博光, Pen-less Recorder による情報収集に基づく興味ベクトルの把握, 組込みシステムシンポジウム 2006, pp.132-135, Oct. 2006.
- [2] 樋口 賢治, 江坂 直紀, 原田 史子, 島川 博光, 多次元ベクトル表現を用いた興味のクラスタリング, FIT2007, Sep. 2007.
- [3] 大澤亮, 高汐一紀, 徳田英幸, 俺デスク: ユーザ操作履歴に基づく情報想起支援ツール, 情報処理学会 プログラミングシンポジウム, Vol.1 (1), pp.1-1, Jan. 2006.
- [4] 近藤 秀樹, 三宅 芳雄, 計算機上での活動履歴を利用する記憶の拡張システムの評価, 第 13 回インタラクティブシステムとソフトウェアに関するワークショップ, pp.131-132, Dec. 2005.