

ユーザの利用履歴を利用した 動的なインデックス半自動生成手法の提案

三石 大 佐々木 淳 船生 豊
岩手県立大学ソフトウェア情報学部

動画像や音声など、その内容を表すインデックスを明示的に作成することが難しいマルチメディアデータベース等の検索のために、半自動かつ動的にインデックスを生成するための手法について提案する。本稿で提案する手法は、利用者のデータベースの利用が、その利用者の関心や興味によって偏りがあることを利用するものである。これは、予め内容毎にある程度分類されているデータベースへの個々の利用者のアクセス履歴をもとに、それぞれの分野に対する利用者の興味の方向を示すベクトル空間(ユーザモデル)を推測し、同時に、個々のデータがどのような利用者からアクセスされたかの履歴をもとに、このユーザモデルからそのデータの内容を示すインデックスとしてのベクトル空間(タイトルモデル)を推測するものである。これを繰り返すことにより、データベースの利用時に各データの内容の方向性を示す値を動的に更新し、適切なインデックス付けが可能となると同時に、そのデータが持つ潜在的な意味情報の抽出が可能となると予想され、従来のシステムでは利用者が気付かなかったデータを発見することが可能となる。本稿では、このモデルに基づく漸化式を定義するとともに、シミュレーションによりその有効性を示す。

A Proposal of Dynamic and Semi-automatic Indexing Algorithm with User's History

Takashi MITSUISHI Jun SASAKI Yutaka FUNYU
Faculty of Software & Information Science, Iwate Prefectural University

We propose a semi-automatic indexing algorithm with emotional keywords such as genre names for multi-media database(e.g. movie files, audio files) according to user's sensitivity by using user's access histories for database. At first, we simply categorize data, and create a vector space of each user's interest(user model) from the history of which categories the data, the user had accessed, belong to. Next, we create vector space of each data(title model) from the history of which users the data had been accessed from. By continuing the above method, we could create suitable indexes, which show emotional content of data. In this paper, we define the recurrence formula based on the proposed algorithm. We also show the effectiveness of the algorithm by simulation result.

1 はじめに

近年のコンピュータ技術の発達に伴い、音声データベースや動画像データベースなどのマルチメディアデータベースが構築され、これらの技術を応用した VoD やオンラインミュージックショップなどが現実的なものとなりつつある。このようなマルチメディアデータベースを効果的に利用するためには、目的のデータを効率良く検索できることが重要であ

り、そのための適切なインデックスを作成することが必要である。

しかしながら、データの内容そのものやその方向性による検索の実現のためには、そのための適切なインデックスを作成することが必要であるが、個々のデータの内容が複雑なマルチメディアデータベースの場合、データの周波数解析のような分析による内容の推測は困難であり、また、データベースの構

築者が個々のデータの内容を把握し、これを全てのデータに対し行うことは難しい。

そこで我々は、映画や音楽などのマルチメディアデータに対し、利用者の関心や興味に応じてデータの利用に偏りがあることに着目し、その利用履歴を利用することで、内容を表すインデックスを半自動かつ動的に作成するための手法を提案する。これは、あらかじめだまかに分類されたデータベースに対する利用者の利用履歴に基づき、その利用傾向から内容の方向性を推測するものである。

本稿では、我々の提案する手法のモデルについて述べ、このモデルに基づき、インデックス作成のための漸化式を示す。次に、この利用者のデータ利用のシミュレーションを行い、我々の提案するモデルおよびそのモデルを示す漸化式によるインデックス作成の有効性について議論する。

2 利用履歴に基づくインデックス半自動生成

ここでは、マルチメディアデータなどのような、その内容を表すインデックスを作成することが難しいデータに対するインデックスの半自動生成手法について提案する。

2.1 内容の方向性による検索

データベース上のデータの効果的な検索を行うためには、そのデータを特徴付ける適切なインデックスの作成が必要とされる。

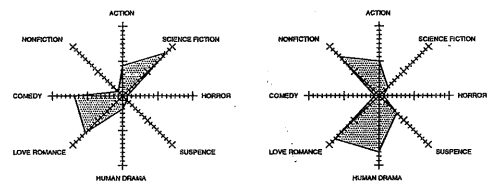
動画像や音声のようなマルチメディアデータの場合、そのデータに付随的な予め判っている情報、例えば映画のデータベースであれば、各作品のタイトル、製作者、出演者などの情報をインデックスとして利用する事は容易である。

しかしながら、利用目的に応じてデータを効果的に検索するためには、データの内容そのものや、その内容の方向性を示すインデックスを作成することが必要と考えられる。本研究では、後者の内容の方向性を示すインデックスの生成を対象とする。例えば簡単なものとして、SFやアクション、ラブロマンスといった各作品のジャンル、更に複雑なものとしては、「格好いい」、「感動的な」などの抽象的な

感性語句がこれにあたる。

一般に、映画や音楽のようなデータの場合、このような内容の方向性はジャンルと呼ばれ、多くのデータは予め複数のジャンルに大別されていることが多く、このようなジャンル名を検索のためのインデックスとして利用することは可能である。しかし、実際にはデータの方向性は必ずしも単一の方角ではなく、実際には、個々の作品は複数のジャンルの要素を複合的に合わせ持つことが多い。

例えば、通常SFに分類される「バック・トゥ・ザ・フューチャー」は、ラブロマンスやコメディ的な要素を少なからず持っているといえる。また同様に、一般にラブロマンスに分類される「タイタニック」は、史実に基づく事件の中に、架空のラブストーリーを折り交ぜた人間ドラマを描いたものである(図1)¹。



(1) バック・トゥ・ザ・フューチャー (2) タイタニック

図 1: 内容の方向の多様性

このように、個々のデータの内容は多様な方向性を持ち、この方向性を示す適切なインデックスを作成することができれば、「SFかつ、ラブロマンスやコメディ的な要素を持つもの」といった検索により「バック・トゥ・ザ・フューチャー」を発見することができ、より効果的な検索を行うことができる。ことに加え、データの詳しい内容を知らない利用者が、新たなデータを発見することも可能となる。例えば、「バック・トゥ・ザ・フューチャー」の詳しい内容を知らない利用者が、「ラブロマンス」をキーワードに検索を行い、この作品を発見し、新たな見知を得ることができる。

さらに、このようなジャンル名によるデータの内容の方向性を示すことができれば、そのジャンル名に対し利用者が抱くイメージとの対応関係を導入する

¹ 詳しい分析によるものではない。

ことで、抽象的な感性語句による検索もある程度行えると予想される。

2.2 既存の分類による問題

しかしながら、一般的に行われる単純なジャンル分けでは、どれか1つのジャンルに分類されてしまっていることが多く、必ずしもその内容を性格に示していることにはならない。そのため、既存の分類によるジャンル名をインデックスとした場合、適切な検索を行うことが難しい。

例えば、そのデータ群の内容に2種類の方向性G1、G2がある場合、個々のデータの方向性に基づく分布は図2-(1)のように示すことができる。しかし、単純なジャンル分けによる分類では、図2-(2)のように分類されることとなる。その結果、データt1は、G1、G2両方の方向性を多分に持つが、G1に分類され、G2の方向性を持つという情報を示すことができない。またデータt2は、G1、G2の方向性とも少ないが、G1の方向性を多少ながらも持つという情報が失われるばかりではなく、G2の方向性を多く持つわけではないにも関わらずG2に分類されることになる。

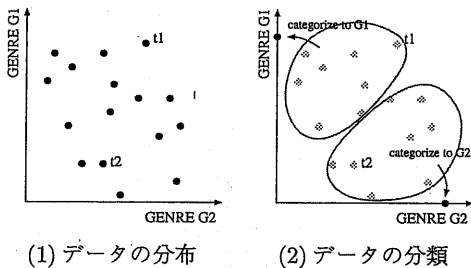


図2: データの方向性とその分類

そのため、既存の一般的な単純な分類により付けられたジャンル名だけでは、必ずしもその内容を性格に示していることにはならず、結果的にこのジャンル名をインデックスとした場合、適切な検索を行うことが難しいと言える。

また模様などの静止画像や楽譜データなどに対し、予めどのようなパターンがどのような印象をうけるかを分析し、各データの周波数成分の解析により、その中に含まれるパターンやその繰り返しを

分析することにより、その方向性を推測する手法がある。

しかし、動画像データや録音されたなど、1つのデータ量が多く、その内容が複雑で変化に富むようなデータの場合、その繰り返しパターンの解析だけでは内容の方向性を分析することは難しい。

抽象的なキーワードによる効果的な検索の実現のために、各データの内容の方向性を示す適切なインデックスを作成するには、データベースの構築者が個々の作品を吟味、理解し、さらには、そのデータが潜在的に持つような方向性に至るまで内容を分析し、適切なインデックスを作成する必要があるが、これを全てのデータに対し行うことは容易ではない。

2.3 利用履歴を利用したインデックス半自動生成手法の提案

そこで我々は、データベースに対する利用者の利用履歴から、個々のデータの内容の方向性を推測し、インデックスを半自動生成するための手法を提案する。

映画や音楽のようなデータの場合、そのデータの内容そのものを詳しくは知らなくともある程度その方向性を予め知っている利用者は多く、その関心や興味の方向性に応じてデータの利用に偏りが生じることが予想される。そこでデータベースの利用者の利用履歴を観察することにより、その利用傾向から、データが潜在的に持つ方向性を推測することができると思われる。

例えば、先の図2におけるデータt1は実際の内容としてはG2の方向性を多分に持っているため、このことを知っており、かつG2に関心の高い利用者から多くの利用があることが予想される。すなわちt1は、ジャンルG1に分類されているものの、G2に分類されているデータを多く利用する利用者からも多く利用される傾向が予想される。したがって、t1を利用した利用者の利用履歴から、他のデータへの利用傾向を分析することにより、単純なジャンル分けによりG1に分類されたt1が、G2の方向性を持つことを推測できると考えられる。

このように利用者の利用履歴を観察することにより、利用者の興味の方向性を推測し、この利用者の興味の方向性から、各データが潜在的に持つ内容の

方向性を推測することが可能であると考えられ、この推測をもとに、各データの内容の方向を示すインデックスを作成することができる。データベースの利用時に、これを逐次行うことで、予め各データの方向性に基づき単純に分類されたデータベースの利用履歴から、半自動かつ動的にインデックスを生成することができる。

2.4 インデックスの半自動生成モデル

まず、データの集合 $\{m|m \in M\}$ および利用者の集合 $\{n|n \in N\}$ が存在すると仮定し、ある時刻 t におけるデータ m の内容の方向性、および利用者 n の興味の方角性をそれぞれ各成分が 0 から 1 の大きさを持つベクトル空間により表現し、これをタイトルモデル $\vec{T}_m(t)$ 、ユーザモデル $\vec{U}_n(t)$ と定義する。

例えば仮に、ここではタイトルモデルおよびユーザモデルのそれぞれが G1 から G8 までの 8 次元のベクトル空間により表現できるものとし、G1 に分類されたデータ m のタイトルモデルの初期値、すなわち時刻 0 における $\vec{T}_m(t)$ は、

$$\vec{T}_m(0) = (1, 0, 0, 0, 0, 0, 0, 0)$$

のように表現することができる。また例えば、時刻 t_1 において推測されているユーザモデル $\vec{U}_n(t_1)$ を

$$\vec{U}_n(t_1) = (0.9, 0.4, 0.3, 0.3, 0.2, 0.5, 0.8, 0.1)$$

のように表現することができる。

ただし、一般的なジャンル分けによる分類では、比較的同様な内容、近い内容を示すようなジャンルが存在する場合があるが、ここでは全てのジャンルが直行しているものと仮定する。

次に、この定義から、我々の提案するタイトルモデルおよびユーザモデルの更新手法について定義する。

ある時刻 t に利用者 n が利用したデータの集合 $\{m(n, t)\}$ を

$$\{m(n, t)\} = \{m|\text{accessed } m \in M \text{ by } n \text{ at } t\}$$

とすると、ある時間 (t_1, t_2) 内に利用者 n が利用したデータのタイトルモデルの集合 $\{\vec{T}(n, t_1, t_2)\}$ は

$$\{\vec{T}(n, t_1, t_2)\} =$$

$$\{\vec{T}_m(\tau)|m \in \{m(n, t), \tau \in (t_1, t_2)\}\}$$

と表わされ、その平均 $aver(\{\vec{T}(n, t_1, t_2)\})$ は、

$$\begin{aligned} & aver(\{\vec{T}(n, t_1, t_2)\}) \\ &= \frac{\sum_{\tau \in (t_1, t_2)} \sum_{m \in \{m(n, \tau)\}} \vec{T}_m(\tau)}{\#\{\vec{T}(n, t_1, t_2)\}} \end{aligned}$$

と求めることができる。

ここで、時刻 t_1 における利用者 n のユーザモデル $\vec{U}_n(t_1)$ および時間 $[t_1, t_2)$ にその利用者が利用したデータのタイトルモデルの平均 $aver(\{\vec{T}(n, t_1, t_2)\})$ がそれぞれ図 3、図 4 のようになっていたとする。

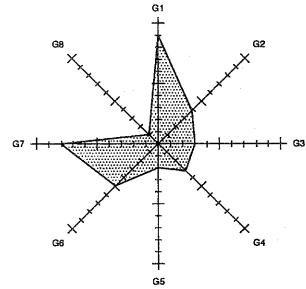


図 3: t_1 における n のユーザモデル

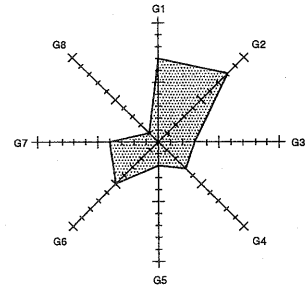


図 4: (t_1, t_2) に利用したタイトルモデルの平均

これを見ると、 $\vec{U}_n(t_1)$ と $aver(\{\vec{T}(n, t_1, t_2)\})$ との間では、G1 に関しては若干の違いがあり、また G2 に関しては、推測された興味の高さより利用データの平均値の方が大幅に高く、G6 に関しては、逆に平均値の方が大幅に低く、その他はほぼ同じであることが判る。

そこで、この結果に基づき両者の差を縮める方向で、時刻 t_2 の後におけるユーザモデル $\bar{U}_n(t_2 + \Delta t)$ を図5のように推測することができる。

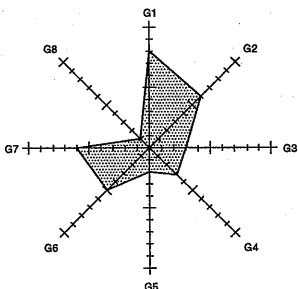


図5: t_2 の後におけるユーザモデルの推測

すなわち、時刻 t_2 において推測される利用者 n のユーザモデル $\bar{U}_n(t_2)$ を時刻 t_1 におけるユーザモデル、および時間 $(t_1, t_2]$ 内にその利用者が利用したデータのタイトルモデルから、次式により定義することができる。

$$\bar{U}_n(t_2 + \Delta t) = \bar{U}_n(t_1) - \alpha_U \left\{ \bar{U}_n(t_1) - \text{aver}(\{\bar{T}(n, t_1, t_2)\}) \right\} \quad (1)$$

ここで、 α_U は t_1 におけるユーザモデルと $(t_1, t_2]$ において利用したタイトルモデルの平均との差を縮める割合を決定する比例係数である。

同様に、時刻 $t_2 + \Delta t$ におけるデータ m のタイトルモデル $\bar{T}_m(t_2 + \Delta t)$ を時刻 t_1 におけるタイトルモデル、および時間 $(t_1, t_2]$ 内にそのデータを利用した利用者のユーザモデルから、次式により推測することができる。

$$\bar{T}_m(t_2 + \Delta t) = \bar{T}_m(t_1) - \alpha_T \left\{ \bar{T}_m(t_1) - \text{aver}(\{\bar{U}(m, t_1, t_2)\}) \right\} \quad (2)$$

ここで、 $(t_1, t_2]$ を十分小さくとり $t_1 = t_2 = t$ とし、その間にある利用者 n があるデータ m を利用したとすると、 $(t_1, t_2]$ においてどのユーザモデル、タイトルモデルとも変化はないので、(1)、(2)式はそれぞれ

$$\begin{aligned} \bar{U}_n(t + \Delta t) &= \bar{U}_n(t) - \alpha_U \{ \bar{U}_n(t) - \bar{T}_m(t) \} \\ \bar{T}_m(t + \Delta t) &= \bar{T}_m(t) - \alpha_T \{ \bar{T}_m(t) - \bar{U}_n(t) \} \end{aligned}$$

となる。時刻を t 、利用者 n が t までに利用したデータの個数 c_{U_n} 、およびデータ m を t までに利用した利用者数 c_{T_m} として $U_n(t)$ および $T_m(t)$ を書き換えると、それぞれ

$$\left. \begin{aligned} \bar{U}_n(c_{U_n} + 1) &= \\ \bar{U}_n(c_{U_n}) - \alpha_U \{ \bar{U}_n(c_{U_n}) - \bar{T}_m(c_{T_m}) \} & \\ \bar{T}_m(c_{T_m} + 1) &= \\ \bar{T}_m(c_{T_m}) - \alpha_T \{ \bar{T}_m(c_{T_m}) - \bar{U}_n(c_{U_n}) \} & \end{aligned} \right\} \quad (3)$$

とすることができる。これは、データの1回の利用毎に、利用者のユーザモデルと利用されたデータのタイトルモデルの差に基づき、それぞれを更新することを意味する。

また、 $t_1 = 0$ 、 $\alpha_U = \alpha_T = 1$ とすると、(1)、(2)式はそれぞれ

$$\begin{aligned} \bar{U}_n(t_2 + \Delta t) &= \text{aver}(\{\bar{T}(n, 0, t_2)\}) \\ \bar{T}_m(t_2 + \Delta t) &= \text{aver}(\{\bar{U}(m, 0, t_2)\}) \end{aligned}$$

となる。ここで、1回に1人の利用者から1つのデータが利用されると仮定し、(3)と同様、利用者 n が t までに利用したデータの個数 c_{U_n} 、およびデータ m を t までに利用した利用者数 c_{T_m} を用いて書き換えると、

$$\left. \begin{aligned} \bar{U}_n(c_{U_n} + 1) &= \\ \frac{c_{U_n} \times \bar{U}_n(c_{U_n}) + \bar{T}_m(c_{T_m})}{c_{U_n} + 1} & \\ \bar{T}_m(c_{T_m} + 1) &= \\ \frac{c_{T_m} \times \bar{T}_m(c_{T_m}) + \bar{U}_n(c_{U_n})}{c_{T_m} + 1} & \end{aligned} \right\} \quad (4)$$

を得る。これは、データの1回の利用毎に、利用者 n がそれまでに利用したデータのタイトルモデルの平均、およびデータ m をそれまで利用した利用者のユーザモデルの平均に基づき、それぞれユーザモデルおよびタイトルモデルを更新することを意味する。

3 実験

本研究で提案する、ユーザモデルおよびタイトルモデルの更新方法の有効性確認のために、データベースの利用シミュレーションを行い、(3)式および(4)式を用いたタイトルモデルの更新を評価する。

3.1 実験内容

ここでは、データを0から $M-1$ 番までの M 個、利用者を0から $N-1$ の N 人、データの内容および利用者興味の方角性として G 種類があるとす。タイトルモデル \vec{T}_m およびユーザモデル \vec{U}_n とは別に、あらかじめデータの内容が潜在的に持つ方向 $\vec{P}T_m$ 、利用者の興味の方角 $\vec{P}U_n$ を各要素毎に0から1の実数としてランダムに設定し、タイトルモデルとしては、この潜在的な方向の中で最も大きな要素に相当するものを1とし、他の値を0とする。またユーザモデルの初期値は、各要素とも0とする。

利用者0から $M-1$ までが順番に、その利用者の興味の方角 $\vec{P}U_n$ と各データ内容の潜在的な方向 $\vec{P}T_m$ の適合率の大きさの割合に応じて、ランダムにデータを選択し、利用するものとし、これを1セットとして C 回繰り返す。その結果、データの内容の潜在的な方向 $\vec{P}T_m$ に応じたタイトルモデル \vec{T}_m が形成できるかどうかを評価する。

このとき、タイトルモデルの再現率としては、 \vec{T}_m と $\vec{P}T_m$ のなす角 θ の余弦

$$\cos\theta = \frac{\vec{T}_m \cdot \vec{P}T_m}{|\vec{T}_m| |\vec{P}T_m|}$$

とする。

また利用者の興味の方角 $\vec{P}U_n$ とデータの内容が潜在的な方向 $\vec{P}T_m$ との適合率は、再現率と同様に $\vec{P}U_n$ と $\vec{P}T_m$ のなす角 θ の余弦を用い、この p 乗の値

$$\cos^p\theta = \left\{ \frac{\vec{P}T_m \cdot \vec{P}U_n}{|\vec{P}T_m| |\vec{P}U_n|} \right\}^p$$

を利用する。 p の値が大きい程、 θ に応じて $\cos^p\theta$ の大きさに差が出るため、データ選択時の偏りが大きいことになる。

3.2 結果

まず、データの個数 $M = 100$ 、利用者の人数 $N = 100$ 、データの内容および利用者の興味の方角性 $G = 2$ とし、またデータ選択時の偏りを決定する値を $p = 1$ として100回繰り返す、(3)式、すなわち、データを利用した利用者のユーザモデルと、そのデータのタイトルモデルの値の差に基づきそれぞれを更新した結果、および(4)式、すなわち、利用

したデータのタイトルモデルおよび利用した利用者のユーザモデルの平均に基づき、ユーザモデルおよびタイトルモデルを更新した結果、利用回数に対する利用者0のユーザモデル \vec{U}_0 の再現率の変化を図6に示す。また同様に、データ0のタイトルモデル \vec{T}_0 の再現率の変化を図7に示す。ただし、ここでは $\alpha_U = \alpha_T = 0.01$ とした。

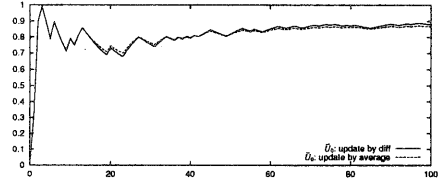


図6: ユーザモデル \vec{U}_0 の再現率の変化

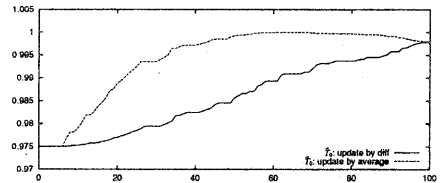


図7: タイトルモデル \vec{T}_0 の再現率の変化

次に、 p の値を1から5まで変化させ、1000回繰り返した場合のデータ選択の偏りに応じたタイトルモデルの再現率の違いを確認する。図8および図9にそれぞれ(3)式および(4)式に基づき更新を行ったタイトルモデルの再現率の平均の変化を示す。

3.3 評価

以上の結果をみると、200回程度の利用の繰り返すでは、(3)式に基づき更新する方式、および(4)式に基づき更新する方式でも、ユーザモデル、タイトルモデルとも予め設定した潜在的な値をある程度再現できていることが判る。タイトルモデルの平均を取った結果でも、もともとの再現率の平均が0.88程度であったものが、 p の値に応じて違いはあるものの両方式とも利用を繰り返すに伴い再現率が上がり、200回程度の利用で0.98前後まで上がったことが確認できる。

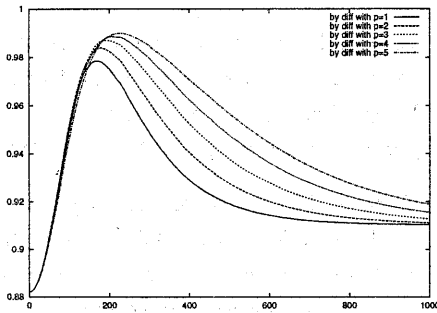


図 8: (3) 式によるタイトルモデルの再現率

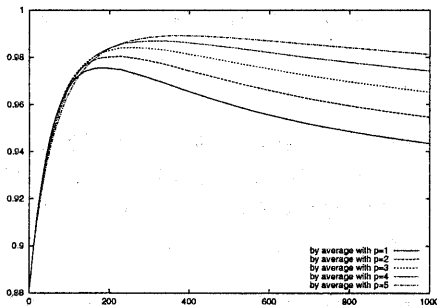


図 9: (4) 式によるタイトルモデルの再現率

しかし、利用の繰り返しが多い場合、その後どちらの手法とも再現率は下がり始め、特に、(3)式に基づく手法では、それが顕著であり、0.91程度にまで下がり、収束していることが判る。(4)式に基づく手法については、この図からは判断できないが、繰り返し回数を増やして別に行った実験で、同様の値に収束していることを確認している。

すなわち、(3)式、(4)式の両手法とも、ある程度の繰り返し回数内では、その利用回数に応じて比較的高いタイトルモデルの再現率を得ることができ、本研究で提案する手法によりデータの内容の方向性を示すインデックスを動的かつ半自動で生成できると言える。

しかしながら、データの利用回数がある程度以上増えると再現率は下がり、そのままでは内容の方向性を示すものとはならなくなってしまうことも確認された。

4 関連研究

画像や音楽などのデータを分析し、その内容の方向性を推測し、検索を実現する研究は多い [1][2][6][8]。しかしながら、これらは平面デザインなどの静止画像や、MIDIデータを対象としたものでありデータ1つあたりのデータ量も少なく、その個々のデータの分析が比較的容易であり、かつ、データの持つ繰り返しパターンなどの周波数成分と利用者が受ける印象との対応が予め研究されている、もしくは得やすいといった傾向がある。サンプリングによる音声データ、さらには動画像データ等の場合、データ1つあたりのデータ量が非常に多くその処理が容易ではない上、周波数成分の分析だけでは内容の方向性を必ずしも得ることができないようなデータに対し、これらの手法を適用することは難しい。

また、文章に対する検索において、各単語のベクトル空間により表現される意味付けを行い、この意味に基づき検索を行う手法がある [7]。しかしながら、これにより各文章の方向性を特徴付けることは可能であるが、これを動画像データや音声データそのものに適用することはできない。

これに対し我々の研究は、データそのものを分析するのではなく、データの利用者の利用傾向を分析するものである。これは、データそのものの内容を正確に把握することはできないが、そのデータのおおよその方向性を推測することが可能である。

WEB等の利用において個人毎の利用履歴から各利用者の興味の方向を推測し、この利用者の興味に基づき効率的な検索や積極的な情報提供を実現する研究がある [3][5]。しかしながら、これは提供されるデータの内容が予め何らかの方法で適切に分類されている必要があり、さらに、データに潜在的に含まれる方向を抽出することは難しい。利用履歴からデータ分類手法に関する研究として、Webの利用履歴からHTML文書の相関関係を推測するもの [4]があるが、相関関係のみでは、検索のためのインデックスを作成することは困難である。

我々の提案する手法では、利用者の興味の方向を推測するとともに、この興味の方向からデータの内容の方向を推測することにより、データの潜在的な方向性に対しても推測することが可能であり、これを利用して検索のためのインデックスを作成することが可能である。

5 まとめ

本研究では、マルチメディアデータ等、データの内容を表す明示的なインデックスの作成が困難なデータに対する検索のために、その方向性を示すインデックスを動的かつ半自動的に作成するための手法について提案した。

本手法は、映画や音楽等のデータの場合、利用者が予めある程度そのデータの内容の方向性を知っており、かつその利用者の興味の方向に応じて、データの利用に偏りがあることに着目し、データベースの利用履歴から利用者の興味の方向(ユーザモデル)、およびデータの内容の方向(タイトルモデル)を推測し、これを逐次更新するものであり、このモデルに基づき、ユーザモデルおよびタイトルモデル更新のための2種類の漸化式を定義した。

我々は、本提案モデルおよびそのモデルに基づく漸化式の有効性確認のために、データベース利用シミュレーションを行った。その結果、どちらの漸化式においても、各データの持つ潜在的な内容の方向性のある程度再現できることが確認された。しかしながら、利用を繰り返すうちにタイトルモデルおよびユーザモデルの値が平均化されてしまい、特に利用の偏りが小さい場合、顕著であることが確認された。今後、利用を繰り返しても値が平均化されない手法を検討する必要があると言える。

さらに、今回の実験では、利用者の興味とデータの内容の方向の適合率として、各ベクトル間の余弦を用いた。しかしながらこの場合、その方向は一致するもののその大きさは考慮されていないことになり、必ずしも興味と内容が一致しているとは言えない。これは、実験結果の再現率の評価方法についても同じことが言え、適合率、再現率の計算方法についても検討の必要があると言える。

今後、これらの検討を行うと同時に、新たなデータ利用者が加わった場合や方向性の数が3以上の多数の場合におけるユーザモデルおよびタイトルモデルの変化についても実験を行い、その上で、具体的なデータベースを構築し利用実験を行うことにより、実際の評価を行う予定である。

参考文献

- [1] Fukuda, M., Sugita, K. and Shibata, Y.: Perceptual Retrieving Method for Distributed Design Image Database System, *Trans. IPS Japan*, Vol. 39, No. 2, pp. 158-169 (1998).
- [2] Ishihara, S., Ishihara, K. and Nagamachi, M.: Analysis of Individual Differences in Kansei Evaluation Data Based on Cluster Analysis, *KANSEI Engineering International*, Vol. 1, No. 1, pp. 49-58 (1999).
- [3] Okada, R., Lee, E.-S., Kinoshita, T. and Shiratori, N.: A Method for Personalized Web Searching with Hierarchical Document Clustering, *Trans. IPS Japan*, Vol. 39, No. 4, pp. 867-877 (1998).
- [4] 風間一洋, 佐藤進也, 清水奨, 神林隆: WWWのユーザ操作履歴によるHTML文書の相関関係の解析, *情報処理学会論文誌*, Vol. 40, No. 5, pp. 2450-2459 (1999).
- [5] 桑田喜隆, 谷津正志, 小泉宣夫: ユーザモデルに基づく技術支援情報の自動配信サービス, *情報処理学会論文誌*, Vol. 40, No. 11, pp. 3896-3905 (1999).
- [6] 原田将治, 伊藤幸宏, 中谷広正: 感性語句を含む自然言語文による画像検索のための形状特徴空間の構築, *情報処理学会論文誌*, Vol. 40, No. 5, pp. 2356-2366 (1999).
- [7] 吉田尚史, 清木康, 北川高嗣: 意味的連想検索機能を持つメディア情報検索システムの実現方式, *情報処理学会論文誌*, Vol. 39, No. 4, pp. 911-922 (1998).
- [8] 佐藤聡, 菊地幸平, 北上始: 音楽データを対象としたイメージ検索のための感情価の自動生成, *情報処理学会研究会報告 99-DBS-118*, pp. 57-64 (1999).