

インデックス半自動生成のための ユーザの利用履歴を利用した内容推測手法の検討

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
〒 020-0193 岩手県滝沢村滝沢字巣子 152-52
TEL: 019-694-2570 / FAX: 019-694-2571
{takashi, jsasaki, funyu}@soft.iwate-pu.ac.jp

あらし 動画像や音声など、その内容を示すインデックスを明示的に作成することが難しいデータベースの検索のために、利用者の興味に応じてデータの利用に偏りがあることを利用し、インデックスを半自動的に生成する手法を提案し、その検討を行う。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利用履歴をもとに、利用者の興味の方角を示すベクトル空間(ユーザモデル)を推測し、さらに、このユーザモデルをもとに、データの特徴としてのベクトル空間(タイトルモデル)を推測し、インデックスを生成するものである。本稿では、この特徴推測のために5種類の方式を定義し、シミュレーションによる評価を行った。

キーワード マルチメディアデータベース、データマイニング、インデックス半自動生成、感性検索、感性情報

A Study on Presumption of Characteristics of Data with User's History for Semi-automatic Indexing

Takashi MITSUISHI Jun SASAKI Yutaka FUNYU

Faculty of Software & Information Science, Iwate Prefectural University
Sugo 152-52, Takizawa, Iwate 020-0193, Japan
TEL: 019-694-2570 / FAX: 019-694-2571
{takashi, jsasaki, funyu}@soft.iwate-pu.ac.jp

Abstract We proposed a semi-automatic indexing methods 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. Next, we presume a vector space of each user's interest(user model) from the history of which data he had accessed, and presume a vector space of each data(title model) from the history of which users the data had been accessed from. In this paper, we define five sorts of formulas based on the proposed methods, and evaluate the effectiveness of these formulas by simulation result.

key words multi-media database, data mining, semi-automatic indexing, kansei retrieval, emotional information

1 はじめに

近年、コンピュータ技術やネットワーク技術の発達に伴い、VoD やオンラインミュージックショップなど、音声データベースや動画画像データベースなどのマルチメディアデータベースを利用したサービス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アプリケーションに限らず、データベースの効果的な利用のためには、目的のデータを効率良く検索できることが重要である。そのためには、データの内容そのものやその方向性等、その特徴を適切に示すインデックスを作成することが必要である。

しかし、データの内容が複雑で変化に富むようなマルチメディアデータベースの場合、個々のデータを個別に分析し、その特徴を示すインデックスを作成することは容易ではない。そこで我々は、映画や音楽などのマルチメディアデータベースに対し、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、個々のデータの特徴を推測し、インデックスを半自動的に生成するための手法を提案する。

本稿では、まず、マルチメディアデータベースの検索のためのインデックス生成と既存の問題について述べる。次に、我々の提案する手法のモデル化を行い、これに基づき、その実現手法について検討を行う。

2 インデックス生成のためのデータ内容の特徴推測

本章では、マルチメディアデータベースにおけるインデックス生成のための内容推測の必要性、および既存の手法によるインデックス生成の問題について述べる。

2.1 データの内容を示すインデックス

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、そのデータの特徴付け、かつ利用者が直観的に推測可能な、適切なインデックスの作成が必要である。映画や音楽などのマルチメディアデータの場合にはこのようなインデックスとして、(1) 各作品のタイトル、製作者、出演者などの付随的な情報、および (2) データの内容そのもの、内容の方向性、内容から受ける印象を表す抽象的な感性語句を用いることが出来る。

(1) のようなデータに付随的な予め判っている情報をインデックスとして利用する事は比較的容易であるといえる。しかし、その内容が複雑で変化に富む映画や音楽

などのデータの場合、(2) のような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生成するためには、データベースの構築者が個々のデータを詳しく吟味し、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、十分に理解することが必要であり、これを全てのデータに対し行うことは容易ではない。本研究では、このような内容の方向性に基づく特徴を示すインデックスの生成を対象とする。

2.2 内容の多様性と分類

一般に、映画や音楽等の場合、各データの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容易である。しかし、実際にはデータの方向性は必ずしも単一ではなく、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。

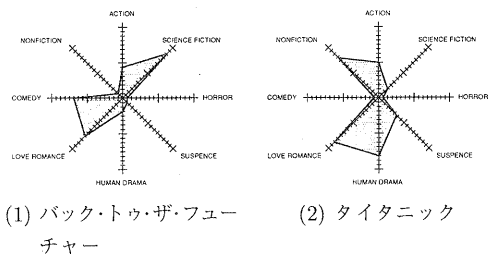


図 1: 内容の方向の多様性

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

¹ 詳しい分析によるものではない。

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。

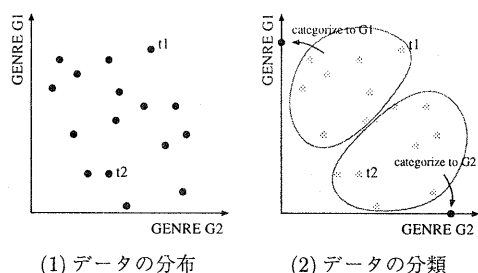


図 2: データの方向性の分布と分類

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。

このように、個々のデータの内容は多様な方向性を持つが、この方向性に基づく特徴を適切に示す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、新たな知見を得ることができる。

さらに、ジャンル名等による内容の特徴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印象との対応関係を導入することで、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。

2.3 既存の内容推測手法

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][7][9]。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。

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 [8]。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応することはできない。

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 [4][6]。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 [5] 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。

3 データベースの利用履歴を利用したデータ内容の推測

本章では、利用者のデータベースの利用履歴を利用してデータの内容の方向性に基づく特徴の推測手法を提案し、その実現のための式を検討する。

3.1 利用履歴を利用したインデックス半自動生成の提案

我々は、既存のインデックス生成手法による問題解決のために、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してきた [10][3]。

これは、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、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考えるものである。

例えば、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の特徴(ユーザモデル)を推測し、このユーザモデルから、各データが潜在的に持つ内容の方向性の特徴(タイトルモデル)を推測することが可能であると考えられる。この推測結果をもとに、個々のデータについて各方向毎に重み付けを行うことで、その特徴を示すインデックスを半自動的に生成することが可能となる。

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。

3.2 利用履歴を利用した内容推測手法の検討

利用履歴を利用したデータの特徴推測のために3種類5方式の特徴推測手法を提案し、利用者のデータベース利用シミュレーションによる実験をおこない、その有効性確認を行う。そのため先ず、ユーザモデルとタイトルモデルのベクトル表現、および実験方法とその評価方法を定義する。

3.2.1 ユーザモデルとタイトルモデルのベクトル表現

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、データ m の推測される特徴、および利用者 n の推測される興味の方向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 $\vec{T}_m(c_m)$ 、ユーザモデル $\vec{U}_n(c_n)$ と定義する。ただし c_m および c_n は、それぞれデータ m が利用者から利用された回数、および利用者 n がなんらかのデータを利用した回数を示す。

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 $c_m = 0$ におけるタイトルモデル $\vec{T}_m(0)$ 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、利用回数 c_n において推測されているユーザモデル $\vec{U}_n(c_n)$ を

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

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。

3.2.2 実験方法

ここでは、データを0から $M-1$ 番までの M 個、利用者0から $N-1$ の N 人、データの内容および利用者興味の方向性として G 種類があるとする。

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 $\vec{P}\vec{T}_m$ 、利用者の興味の方向 $\vec{P}\vec{U}_n$ 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向 $\vec{P}\vec{U}_n$ と各データ内容の潜在的な方向 $\vec{P}\vec{T}_m$ の適合度の大きさの割合に応じて、ランダムにデータを選択し、利用するものとする。利用者の興味の方向 $\vec{P}\vec{U}_n$ とデータの内容が潜在的な方向 $\vec{P}\vec{T}_m$ との適合度は、 $\vec{P}\vec{U}_n$ と $\vec{P}\vec{T}_m$ のなす角 θ の余弦から、この p 乗の値を利用し、次式により定義する。

$$\cos^p \theta = \left\{ \frac{\vec{P}\vec{T}_m \cdot \vec{P}\vec{U}_n}{|\vec{P}\vec{T}_m| |\vec{P}\vec{U}_n|} \right\}^p$$

p の値が大きいく程、 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。

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 $\vec{P}\vec{T}_m$ 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再現度として評価する。このとき、再現度としては、適合度と同様に $\vec{T}_m$ と $\vec{P}\vec{T}_m$ のなす角 θ の余弦から、次式により定義する。

$$\cos \theta = \frac{\vec{T}_m(c_m) \cdot \vec{P}\vec{T}_m}{|\vec{T}_m(c_m)| |\vec{P}\vec{T}_m|}$$

3.2.3 タイトルモデルの初期値を利用した推測

先ず、単純に分類されたデータのタイトルモデルの初期値と利用者の利用履歴から、逐次、ユーザモデルとタイトルモデルを推測する手法について検討する。

これは、データベースの利用履歴を全て記録しておき、ある時点での利用者のユーザモデルをその利用者が利用した全データのタイトルモデルの初期値から推測し、また、タイトルモデルをそのデータを利用した各利用者全てのその時点におけるユーザモデルから推測するものである。

すなわち、利用者 n がデータ m を利用したとして、このときデータ m はそれまでに c_m 回の利用があったとし、また利用者 n は c_n 回の利用を行っていたとした場合、両モデルの推測方法として平均を用いると、ユーザモデル $\vec{U}_n(c_n+1)$ およびタイトルモデル $\vec{T}_m(c_m+1)$ はそれぞれ次式により求めることが出来る。

$$\left. \begin{aligned} \vec{U}_n(c_n+1) &= \frac{\sum_{m(n, c_n+1)} \vec{T}_m(0)}{c_n+1} \\ &= \frac{\vec{U}_n(c_n) + \vec{T}_m(0)}{c_n+1} \\ \vec{T}_m(c_m+1) &= \frac{\sum_{n(m, c_m+1)} \vec{U}_n(c_n)}{c_m+1} \end{aligned} \right\} \quad (1)$$

ただし、 $n(m, c_m+1)$ はデータ m を 0 回から c_m+1 回まで利用した利用者の集合、また $\{m(n, c_n+1)\}$ は利用者 n が 0 回から c_n+1 回まで利用したデータの集合とする。

$M = 100, N = 100, G = 2, C = 1000$ として、 $p = 1 \sim 5$ についてシミュレーションを行い、(1) 式に基づきタイトルモデルを求めた場合、利用の偏りに応じて 0.916 から 0.980 程度の平均再現度を得ることができた (図 3)。

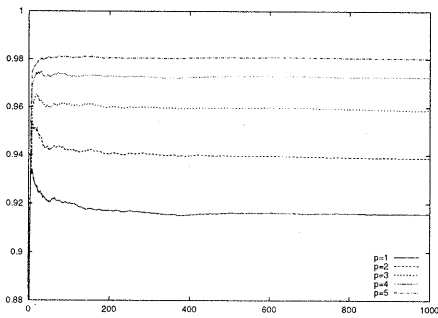


図 3: (1) 式によるタイトルモデルの再現度の平均

しかしながらこの方式では、(1) ユーザモデルの推測のために、各利用者が利用した全てのデータのタイトルモデルの初期値を利用するため、実際の利用において利用者の好みが変わった場合など、利用傾向に変化が生じた場合の影響を反映しにくい、(2) 利用履歴を全て記録

しておき、タイトルモデルの推測においてそのデータの利用者全てのユーザモデルをこの利用履歴から調べ毎回計算し直す必要があり、データ数や利用者数、利用回数が増えた場合のスケーラビリティが低い、(3) 利用の偏りの大きさが再現度与える影響が大きく、偏りが小さい場合に高い再現度を得ることができない、といった問題がある。

3.2.4 ユーザモデルとタイトルモデル間の相互フィードバックによる推測

次に、タイトルモデルとユーザモデル間の相互フィードバックによる推測手法 (図 4) について検討する。

この手法は、ある時点において推測されているユーザモデルと、その利用者が、その後に利用したデータのタイトルモデルから、次の時点におけるユーザモデルを推測し、また同様に、ある時点において推測されているタイトルモデルと、その後、そのデータを利用した利用者のユーザモデルから、次の時点におけるタイトルモデルを推測するものである。このように、ユーザモデルとタイトルモデルを相互に反映させることにより、それぞれのモデルを動的に推測することが可能であると考えられる。

このとき、フィードバックの方法としては、1 回のデータの利用毎に、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、これを縮める方向でそれぞれを更新する方式と、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 をそれまでに利用した利用者のユーザモデルの平均に基づき、それぞれ更新する方式を考えることができる。

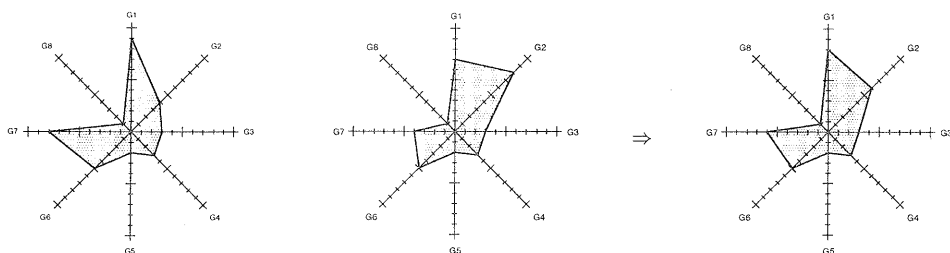
差に基づきそれぞれを更新する方式は、次式により定義することが出来る。

$$\left. \begin{aligned} \vec{U}_n(c_n+1) &= \\ &\vec{U}_n(c_n) - \alpha_U \{ \vec{U}_n(c_n) - \vec{T}_m(c_m) \} \\ \vec{T}_m(c_m+1) &= \\ &\vec{T}_m(c_m) - \alpha_T \{ \vec{T}_m(c_m) - \vec{U}_n(c_n) \} \end{aligned} \right\} \quad (2)$$

ここで、 α_U, α_T は c_n, c_m におけるユーザモデルおよびタイトルモデルと c_n+1, c_m+1 におけるユーザモデルおよびタイトルモデルとの差を縮める割合を決定する比例係数である。また、平均に基づきそれぞれ更新する方式は、次式により定義することが出来る。

$$\left. \begin{aligned} \vec{U}_n(c_n+1) &= \frac{c_n \times \vec{U}_n(c_n) + \vec{T}_m(c_m)}{c_n+1} \\ \vec{T}_m(c_m+1) &= \frac{c_m \times \vec{T}_m(c_m) + \vec{U}_n(c_n)}{c_m+1} \end{aligned} \right\} \quad (3)$$

これらの方式では、データの利用があった際に、その時点でユーザモデルとタイトルモデルのみから次の



(1) c_m におけるデータ m のタイトルモデル (2) c_m+1 に利用した利用者 n のユーザモデル (3) c_m+1 におけるデータ m のタイトルモデル

図 4: タイトルモデルとユーザモデルの相互フィードバックによる推測

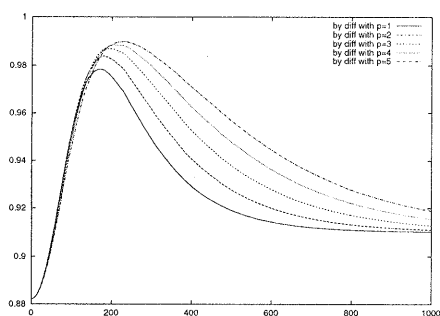


図 5: (2) 式によるタイトルモデルの再現度の平均

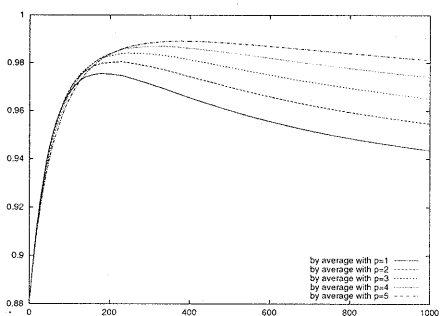


図 6: (3) 式によるタイトルモデルの再現度の平均

時点での推測を行うため、利用履歴を記録しておく必要がなく、また計算も容易となり、スケーラビリティが高いと言える。また、その時点で推測されているユーザモデルとタイトルモデルを利用することにより、利用者の利用傾向の動的な変化を反映しやすいといった利点がある。

ある。

しかしながら、3.2.3 節と同様の実験を行った結果、ある程度利用回数では、(2) 式および (3) 式に基づく方式の両者ともあまり偏りの大きさによらず 0.975 から 0.990 程度と比較的高い再現度を得ることができたが、その後利用回数を重ねると、どちらも再現度が下がり 0.910 程度に収束してしまうことが確認された (図 5, 図 6)。これらは、相互にフィードバックを繰り返すことにより、ユーザモデルおよびタイトルモデルの各成分が平均化してしまっているためと考えられる。ただし、(3) 式に基づく方式では、再現度の低下は比較的緩やかであり、長い繰り返し期間の間、高い再現度を保つことが出来ている。

3.2.5 相互フィードバックを行わない推測

そこで次に、(2) 式もしくは (3) 式と (1) 式を組み合わせることにより、相互フィードバックを行わない手法について検討を行った。

この方法は、ユーザモデルの推測方法として (1) 式を用い、また、タイトルモデルの推測方法としては (2) 式もしくは (3) 式を用いるものであり、両方式は次のように定義することが出来る。

$$\left. \begin{aligned} \bar{U}_n(c_n+1) &= \frac{\bar{U}_n(c_n) + \bar{T}_m(0)}{c_n+1} \\ \bar{T}_m(c_m+1) &= \frac{\bar{T}_m(c_m) - \alpha_T \{ \bar{T}_m(c_m) - \bar{U}_n(c_n) \}}{c_m+1} \end{aligned} \right\} \quad (4)$$

$$\left. \begin{aligned} \bar{U}_n(c_n+1) &= \frac{\bar{U}_n(c_n) + \bar{T}_m(0)}{c_n+1} \\ \bar{T}_m(c_m+1) &= \frac{c_m \times \bar{T}_m(c_m) + \bar{U}_n(c_n)}{c_m+1} \end{aligned} \right\} \quad (5)$$

これらの方式では、タイトルモデルの初期値からユーザモデルの推測を行うため、利用者の利用傾向の変化を

反映しにくいといった問題があるが、利用履歴を残す必要がなく計算も容易であり、スケーラビリティは高いと言える。

同じく実験を行い、タイトルモデルの再現度をもとめた結果、(4)式による方式では、利用の偏りが小さい場合には(2)式と同様に各成分が平均化することにより再現度の低下を招くが、偏りが大きい場合には再現度が高いまま安定することが判った(図7)。また(5)式による方式では、(1)式によるものとはほぼ同様であるが、若干高い推測結果を得ることが出来た。

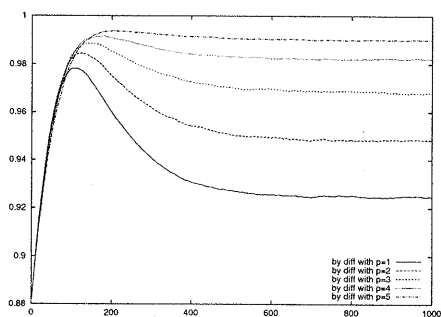


図 7: (4) 式によるタイトルモデルの再現度の平均

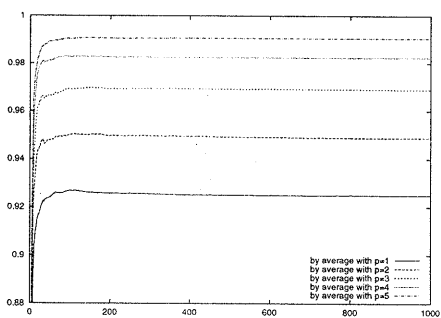


図 8: (5) 式によるタイトルモデルの再現度の平均

3.2.6 結論

(1)式に基づく実験結果から、単純に分類されたデータベースの利用履歴から、インデックス自動生成のための個々のデータの特徴推測が可能であることが判る。しかしこの方法では、利用者の利用傾向の動的な変化に対応できない、スケーラビリティが低い、データ利用の偏りの大きさが再現度に大きく影響するといった問題がある。

(2)式および(3)式に基づく実験結果から、ユーザモデルとタイトルモデル間で相互にフィードバックを行い、ある時点で推測されている各モデルから次の時点での各モデルを推測することで、スケーラビリティや利用の偏りの大きさによる影響の問題を解決し、利用回数がある程度少ない場合には比較的高い再現度を得ることができるが、利用回数を重ねると、相互にフィードバックを行うことにより値が平均化してしまうという問題があることが判る。

また、(4)式および(5)式に基づく実験により、フィードバックを行わない方式では値の平均化をある程度押さえることが出来、また(1)に比較し高い推測結果を得ることが出来るが、利用の偏りの大きさによる影響が出てしまうことが判る。

4 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、その特徴を示すインデックスの半自動的生成を提案し、そのための手法について検討を行った。

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を知っており、かつその利用者の興味や関心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や関心の方向(ユーザモデル)を推測し、またこのユーザモデルから、データの内容の特徴(タイトルモデル)を推測するものである。

我々は本提案に基づき、まず、(1)式に基づきタイトルモデルの推測実験を行った。その結果、単純に分類されたデータベースの利用履歴から、個々のデータのインデックス自動生成のための特徴の推測が可能であることが判った。

しかしこの方法では、利用者の利用傾向の動的な変化に対応できない、スケーラビリティが低い、データ利用の偏りの大きさが再現度に大きく影響する、といった問題があるため、(2)式、もしくは(3)式により、ある時点で推測されている各モデルから次の時点での各モデルを推測する方法として、ユーザモデルとタイトルモデル間で相互にフィードバックを行う方式、および、これと(1)式を組合せた(4)式、もしくは(5)式によるフィードバックを行わない方式を提案し、実験を行った。

その結果、フィードバックを行った場合には偏りの大きさによらず比較的高い再現度を得ることができるが、利用回数を重ねると値が平均化してしまう、フィードバックを行わない場合には値の平均化をある程度押さえる

ることが出来るが、(1) 式と同様に利用の偏りの大きさによる影響が出てしまうことが判った。

今後、利用の偏りによらず高い再現度を得られ、かつ利用回数を重ねた場合でも値が平均化されない手法を検討し、また、利用者の興味の変化などによる利用傾向の変化や、新たなデータや利用者が加わった場合など、動的な変化を想定した実験を行うと同時に、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] Mitsuishi, T., Sasaki, J. and Funyu, Y.: A Proposal of Semi-automatic Indexing Algorithm for Multi-media Database with Users' Sensibility, *Proc. of the 2000 Spring Conference of KOSES & International Sensibility Ergonomics Symposium*, pp. 120-125 (2000).
- [4] 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).
- [5] 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).
- [6] 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).
- [7] 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).
- [8] 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).
- [9] 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).
- [10] 三石大, 佐々木淳, 船生豊: 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, 情報処理学会研究報告 2000-DBS-121, Vol. 2000, No. 44, pp. 53-60 (2000).