

ハイパーメディアシステムにおけるデータ管理機構の開発

7G-2

中川晃一 宗像浩一 西内信実 前川隆昭 嶋 憲司

三菱電機(株)産業システム研究所

1. はじめに

公共システムや産業プラントにおける図面、文書、写真、地図などの設備情報を管理するために、ハイパーメディアシステムを開発している[1]。このような設備情報では一般にデータ量が多く、ナビゲーションだけに頼ってデータを検索するのは困難な場合がある[2]。しかしデータの種類とリンク構造には、ある程度の規則性があることが期待できる。この点に着目して、ある時点で参照中のデータに類似したデータを自動的に検索する手法を提案する。まず各データをその型と、前後にリンクしたデータの型とによって分類する。そしてナビゲーションの履歴を手がかりにして、参照中のデータと同じ分類に属するデータを検索して提示する。

本稿ではデータの分類手法と類似データの検索アルゴリズムを述べる。

2. データの分類

ハイパーメディアにおけるリンク構造のノードとなるデータの集合をDとする。Dの元は全て型を持つものとし、データdの型をT(d)で表す。型としては、例えばイメージ型、テキスト型、音声型等のメディアの種類が考えられる。dへの直接のリンク元データの集合をD⁻={d⁻₁, d⁻₂, ..., d⁻_m}とし、dからの直接のリンク先データの集合をD⁺={d⁺₁, d⁺₂, ..., d⁺_n}とする。するとdと、D⁻、D⁺の各要素となるデータの接続関係を表す有向グラフは、一般に図1のようになる。ここで、この有向グラフの各ノードを、データ自身ではなく、その型に置き換えると、図2のようになる。この有向グラフをdの隣接型グラフと呼び、G(d)で表す。

データが別でも隣接型グラフが等しくなることがある。そこでこの等価性に基づいてデータを分類する。すなわち、あるハイパーメディアのリンク構造について、全てのデータの隣接型グラフを調べた結果 p 種類の隣接型グラフG₁, ..., G_pがあるとすると、Dは

$$D_i = \{d \mid d \in D, G(d) = G_i\} \quad (i=1, \dots, p) \quad (1)$$

の直和に分割される。このときD_iをクラスと呼び、データdの属するクラスをD(d)で表す。

A Data Retrieval Mechanism for Hypermedia Systems.
Koichi Nakagawa, Koichi, Munakata, Nobumi Nishiuchi,
Takaaki Maekawa, Kenji Sima,
Industrial Electronics & Systems Lab., Mitsubishi
Electric Corp.

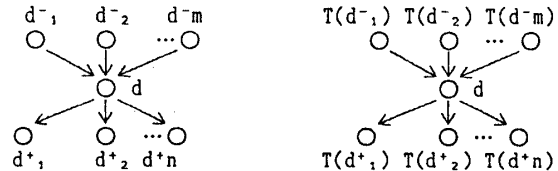


図1 データdの隣接関係

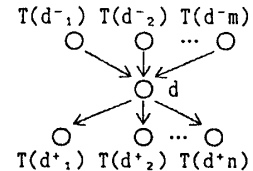


図2 データdの隣接型グラフG(d)

3. 類似データの検索

本手法で検索する、参照中のデータと類似したデータについて図3を用いて説明する。データd₁を起点としてハイパーメディアのリンクを順にたどり、d₂, d₃, ..., d_{N-1}を経由して現在参照中のデータd_Nに到ったとする。但し、この間リンクをたどらずにノード間を移動することはないものとする。このとき、データ列H={d₁, d₂, ..., d_N}を履歴と呼ぶ。ここでHのk番目の要素 H(k)=d_k (1 ≤ k ≤ N)の直接のリンク先データの中に、D(d'_{k+1})=D(d_{k+1}) (d'_{k+1} ≠ d_{k+1})となるデータd'_{k+1}が存在すると仮定する。このとき、d'_{k+1}からのリンクを順にたどって得られる N-k個のデータのクラスの列 {D(d'_{k+1}), D(d'_{k+2}), ..., D(d'_{k+N})}が、列 {D(d_{k+1}), D(d_{k+2}), ..., D(d_N)}と一致するとき、d'_{k+N}をd'_{k+1}に関するd_Nの類似データと定義する。

さて、ハイパーメディアのリンク構造と、履歴H、および参照中のデータd_Nが与えられ、d'_{k+1}が指定されたとき、d'_{k+1}に関するd_Nの類似データを検索するためのアルゴリズムを次に示す。但しd⁺(u)はdの直後のリンク先データの中の u 番目のデータである。

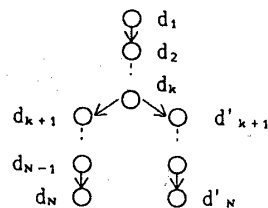


図3 d'_{k+1}に関するd_Nの類似データd'_{k+N}

- ① i=Nにする。
- ② もし iが1であれば終了する。そうでなければH(i)とd'_{k+1}のクラスが一致するか調べる。一致すれば③に進み、もしそうでなければ iの値を1減らして②を繰り返す。
- ③ d_sにd'_{k+1}を代入する。
- ④ r=1とする。

- ⑤ $t=1$ とする。
- ⑥ もし t が ds のリンク先のデータの総数より大きければ②へ進む。そうでなければ、 $ds^*(t)$ が $H(i+r)$ と同じクラスであるか調べる。同じクラスであれば⑦に進む。そうでなければ t の値を1増やして⑥を繰り返す。
- ⑦ ds に $ds^*(t)$ を代入し、もし r の値が $N-i$ であれば終了し、そうでなければ r の値を1増やして⑤へ進む。

この結果 ds が得られれば、 ds が d_N に対する類似データである。そうでなければ類似データは存在しない。

以上の類似検索アルゴリズムを図4のリンク構成を持つハイパーメディアシステムに適用した例を示す。現在参照中のデータを”部品図1”とする。このデータは図5の履歴から”A市地図1”を経由してナビゲートされたものと分かる。ここで検索したいデータは、このデータと同じクラスのデータ”B市地図1”に関しての、”部品図1”に対応するデータである。以後データは”

”で、クラスは「」で表記する。
 ”B市地図1”は表1から「大項目地図」に属することがわかる。現在参照中のデータから履歴を遡ることにより、「大項目地図」に属したデータは”A市地図1”であり、この次にナビゲーションしたデータは、”A市地図2”と分かる。このデータが属するクラスが「小項目地図」であることから、”B市地図1”のリンク先のデータの中に「小項目地図」に属したデータがあるかを調べる。その結果、”B市地図3”と”B市地図2”が得られる。先と同様に履歴から、次に検索された”ポンプ1”は「ポンプ」に属しているので、先に得られた2つのデータの中から「ポンプ」に属するデータをリンク先に持つ”B市地図2”を選択する。このデータのリンク先にある「ポンプ」に属するデータは”ポンプ2”である。最後に検索するクラスとして、同様に履歴から「図面」が得られるので、”ポンプ2”のリンク先のデ

ータの中から、「図面」に属するデータを検索し、”部品図3”が得られる。こうして”部品図1”と類似したデータとして、”部品図3”を検索できる。

4. まとめ

ハイパーメディアシステムにおけるデータ検索機構について提唱した。これはデータをクラスと呼ばれる集合によって分類し、参照されたデータと類似したデータをナビゲーションの履歴を用いて検索することを特徴とする。現在、この検索機構を用いたハイパーメディアシステムを実装中である。

今後は、この手法によって検索される類似データがユーザの意思をどれだけ反映したものを検証をする。またデータの分類方法を拡張して、データの前後のリンク構成が多少変わっても類似データとして検索できる柔軟な分類方法を検討する予定である。

<参考文献>

- (1)宗像 浩一他、”オブジェクト指向ハイパーメディアサーバ「Omnilinker」の開発”、第46回情報処理学会全国大会講演論文集,1993 投稿中
- (2)原 良憲、”ハイパーテキストデータにおける抽象化操作とクラスタリング手法”、Obase シンポジウム, pp 3-10,1991

表1 各データの属するクラス

クラス名	データ名
大項目地図	A市地図1 B市地図1
小項目地図	A市地図2 A市地図3 B市地図2 B市地図3
図面	部品図1 部品図2 部品図3 部品図4
ポンプ	ポンプ1 ポンプ2
バルブ	バルブ1 バルブ2
メニュー	地図選択メニュー

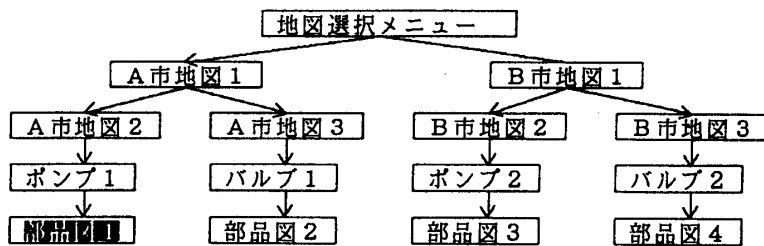


図4 データのリンク構成図（反転文字は現在参照中のデータを示す）

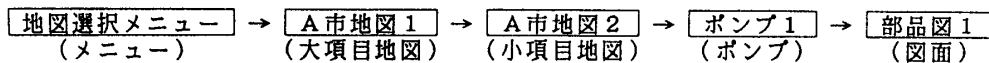


図5 検索の履歴（括弧内は上記データの属するクラスである）