

## インタラクティブな情報空間の形成法に関する提案

赤羽喜治 山田達司 中村太一  
NTTデータ通信株式会社 技術開発本部

ネットワークの発展と接続端末の機能向上により、従来よりはるかに大量の情報が広域ネットワーク上を流通するようになりつつある。こうした情報流通システムはコミュニケーションの場、ビジネスの場として社会生活の一部を担うことが期待されている。

一方、情報量の巨大さとネットワークの広さ、アクセスツールの不十分さのために有用な情報の獲得や提供が困難な場合が多い。

本研究は流通する情報に対し、情報の自己組織化技術とリンクの自動生成技術を用いることにより、利用者の行動や情報の生成に応じて変動し、利用者のニーズにあった見え方を提供するインタラクティブな情報空間の実現を目指す。これにより従来の情報流通システムに比べて短時間でより多くの有用な情報を得、また、多くのユーザに情報を提供させることができると考えている。

これらの実現手法として、本稿では自己組織化マップ手法と遺伝的アルゴリズムを適用することを提案する。

### Organization Schemes of Interactive Infosphere

Yoshiharu Akahane Tatsushi Yamada Taichi Nakamura

NTT Data Communications Systems Corporation

66-2 Horikawa-cho, Saiwai-ku, Kawasaki-shi 210

According to the growth of computer networks and evolution of computer power , information distribution systems will be widespread and process a large amount of information and are expected to play an important role in social activities

On the other hand, it is difficult to get or to serve useful information because of huge quantity of information, vast computer networks and poor navigation tools on infosphere.

Our purpose is to construct "Interactive Infosphere". It serves each user his own view of infosphere, this view keep on changing along with user activities on infosphere. We expect this makes users more positive in getting or serving information and communicating on infosphere.

In this paper, we propose to apply "Self-Organization Map Method" and "Genetic Algorithm" in realizing organization schemes of interactive infosphere.

## 1. 背景

近年、パソコン通信やインターネット上のニュースシステム、WWWシステム等のような情報流通システムが急成長を遂げており、これからも飛躍的な発展をすと期待されている。こうしたシステムでは、ある程度分類はされているものの、利用者間の情報のやり取りがほぼ無秩序に集積されている。

こうした情報流通システムは、人と知り合う、情報を交換し合う、商業活動をする、等といった社会活動の場として使われている。これは、距離や時間といった物理的制約に関係なく情報を交換・共有できるという点に依るところが大きい。

しかし、これを情報を集約・検索するという側面から眺めてみると大きな問題を抱えている事がわかる。情報を集約するシステムとして対極に位置するデータベースシステム(DBS)と比較すると、情報の有用性の低さ、検索稼働の大きななどが問題点として浮かび上がってくる。「距離・時間に関係なく情報を共有・交換することができる」という情報流通システムの長所も、こうした障害があるために、十分活かされているとは言いがたい。毎日大量に流れてくる情報の中に有用な情報が埋もれてしまったり、内容の同じ議論が異なる場所、異なる時間で行われ、互いにそのことに気付かないでいるといったことは日常茶飯事になっている。

## 2. 問題点

現行の情報流通システムの問題点として、以下の項目が挙げられる。

### (1) 見通しの悪さ

類似した議論が過去、あるいは異なる場所で行われている場合、これを知ることは困難である。BBSやニュースシステムではタイトル一覧や、未読数の増え方等で、どのあ

Date	No	Creator	Res	Title
4/14	99			9 千葉J/Tの電話番号を教えて。
4/14	100			12 【教えて】WinかMacの視覚処理ソフト
4/14	101			28 アスキーのInInternet権限サービス
4/14	102			70 Test/Mine: 北野坂音響映像実験室
4/14	103			48 test/yonezaki: ティオール・スヴェルト 逆に日本発売
4/14	104			11 ATLANTIS : ごみ処理
4/14	105			3 ish upload test
4/14	106			27 もし 4/15 に関東地方に大地震が起きたら

42:	[ 10:k@nn.csse.yam ]	I'm looking for the TSP library.
43:	[ 14:zengoku@gcd.b ]	Re: I'm looking for the TSP library.
44:	[ 01:naga@cs1.son ]	CFP: Workshop on New Trends in Dialogue Research
45:	[ 272:yamano@etl.go ]	CFP - S/W'95 in Beppu [2nd]
46:	[ 49:jyunichi@ta.n ]	ILP workshop
47:	[ 1273:watsubar@etl ]	IJERI-95 program
48:	[ 1438:watsubar@etl ]	IJERI-95 registration info.
49:	[ 38:hiraga@llis.a ]	LECTURE at ULIS (5/19): Computer Game Playing
50:	[ 86:dohsak@oton ]	MLC schedule (May 12)
51:	[ 51:dohsak@oton ]	CFP: MLC (July 28, 21)
52:	[ 22:hattori@etec ]	Modeling and Making Algorithm
53:	[ 25:yashino@etlc ]	Re: Modeling and Making Algorithm
54:	[ 27:kuroda@skydog ]	Re: Modeling and Making Algorithm
55:	[ 29:yashiki@etl.g ]	Re: Modeling and Making Algorithm
57:	[ 64:yasuaki@walcke ]	Re: Modeling and Making Algorithm
64:	[ 27:shingo@keu4 ]	Re: Modeling and Making Algorithm
56:	[ 48:nakano@ssun ]	[ICONS] Advance Program (not electronic)
58:	[ 346:chisato@csl.a ]	The 2nd announcement of BIES'95
59:	[ 112:slrau-l@iz.a ]	1995 Symposium on a Comprehensive Approach to Visual
60:	[ 6:sun@coe.dalk ]	help
61:	[ 17:harada-k@jed ]	Re: help
62:	[ 21:koike@asc.co ]	Re: help

図1：現行のBBS、ニュースシステム

たりでどんな話題が盛り上がっているかというのを見ることが出来る。(図1)。だが現状ではそれは一つのフォーラム、ニュースグループといった枠の中でしか調べることができない。例えばPC-9801にSCSI機器を接続する際のトラブルについて知りたくても、様々なBBSやニュースグループを渡り歩いてタイトル一覧、過去の書き込みが圧縮保存されていればこれをダウンロードをし、一通り調べるといった作業が必要になる。

### (2) 硬直したリンク

WWWシステムでは基本的にリンクは固定である。情報の作成者によってリンク付けられるため必ずしも利用者のニーズにあった情報へたどり着けるわけではない。更に、リンクの更新はマニュアルで行われることが多いため、更新期間が長くなりがちである。また、リンクの向きも片方向であるため、その情報にたどり着いた他の利用者がどのような情報経路を経たのか、参考にしにくくも知ることができない。

また、図2のようにある特定の情報に対して、互いに関連の強い2つの情報A、A'が独立にBに対してリンクを張っていたとする。現行のシステムではいくらこの2つの情報の間の関連性が高くても、誰かがこれに気付かない限り2つの間にリンクが張られることはない。利用者は遠回りなリンクを辿ってA'に気付くか、あるいはその存在を一生知らずに過ごすことになる。

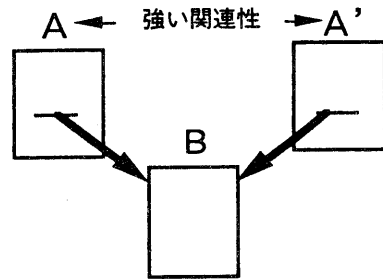


図2：硬直したリンク

### (3) privateな場の小ささ

情報の発信を疎外する要因として、現行流通システムではpublicな場所が大きな位置を占めていることが挙げられる。これはニュースシステムにおいて特に顕著であり、投稿者に多大なプレッシャーを与える原因になっている。同時に、一部のプライベートなやり取りがpublicな空間で行われるために衝突と混乱を招く原因にもなっている。

こうした、特定の話題ごと、嗜好の似通った利用者同士で情報をやり取りしたいという要求に一部応えているものとして、各種BBSのCUG (Closed User Group) や、ASCII Netのハイパー・ノーツシステム、NiftyServeのホームパーティなどがある。

### (4) 得られる情報の有用性の低さ

の際の取りこぼしや余分な情報の出力を避けることができない。無秩序に流通する情報の中でタイトルやキーワード等による情報の検索は、利用者にDBSよりも大きな検索稼働を強いることになる。また、フルテキストサーチや意味理解等の技術も情報流通ネットワークに対しては情報量の巨大さのために適用することは困難である。

また、動画像や音声などのデータは、人間の感性に大きく依存しており、画一的なキーワード検索や意味解釈では利用者の要求に応えることは困難である。

さらに将来的には以下の様な要因により、こうした問題はさらに深刻さと増すと思われる。

- (1) 情報量の爆発的增加  
接続端末数の増加と、情報の作成・発信作業の簡易化に伴う端末当たりの情報発信量の増加との相乗効果。
- (2) 検索範囲の広域化  
ネットワークの拡大
- (3) マルチメディア化  
動画像・静止画・音楽等のマルチメディアデータの流通量の増加

### 3. 本研究のアプローチ

こうした問題点に対して、これまで文書の階層構造や参照構造を解析してリンクを自動生成する機構や、双方向リンクで情報間を結合するシステム等、様々な研究が行われている。[註1,2,3]

本研究では以下のアプローチをとることにする。

- (1) 情報空間の中での活動の視覚化
- (2) 利用者毎に各々の関心を反映した視覚化
- (3) 利用者の行動を反映した情報間のリンクの自動生成

本研究における「情報空間」とは、社会活動の場としての情報流通ネットワークを意味する。

こうした方法を実現する手法として、本研究では次章の2つの方策を採用した。

## 4. 対処策

### 4. 1 自己組織化

自分にとって有用な情報へのポイントを得る方法は、フルテキストサーチ、意味理解による分類等の手法に限る必要はない。実生活を省みると、任意の分野において他者、特に自分と類似の嗜好を持つ集団の行動によって、有用な情報へのポイントを得ることはよくあることである。街を歩いている時に人だかりがあれば何事かと興味を抱くし、それが自分が属する集団であれば（デート中にカップルが列をなしている店を見かける等）さらに強く興味を引かれる、などというのはその一例である。

そこで、情報空間において、その中でどのような活動が行われているかを示す視覚化ツールが存在すれば情報の円滑な流通が促進されることになる。ニュースシステムにお

いては話題のマップを提供することで見通しの悪さを解消できる。また、適用する情報が仮想現実の街の中のオブジェクトであれば街の構造の最適化に応用することも可能となる。

さらに、すべての利用者に対して同じように視覚化するのではなく、利用者毎に嗜好にあったマップを提供する必要がある。これには単なる情報の自己組織化だけではなく、利用者の行動を反映した自己組織化を行わなければならない。

### 4. 2 リンク

実生活における情報行動（例えば書籍を買う、テレビを見る、会話をする等）では興味深いこと、楽しいこと、思いも寄らぬ方向への話題の転回等多種多様な思考の連なりを享受することが多い。ハイパーテキストシステムは、思考の連なりを情報間のリンクという形でコンピュータ上に実装したものである。コンピュータネットワーク上ではWWWシステムがこれにあたる。

本研究では、利用者を満足させるのには一つ一つの情報それ自体だけではなく、「思考の連なり=リンク」も重要であるとする。

利用者が大量のリンクの中を渡り歩く際、必要のない情報でもそこにしか次の情報へのポイントがないため、仕方なくその情報を見なければならないという事態は頻繁に起っていると考えられる。これは、以下の理由による。

- (1) リンクの提供は情報提供者の一存に任されている
- (2) 一つの情報から張られたリンクは次にどの情報へいくべきかということしか示さない。

有効な情報行動のためには、図3の様に情報Aから情報Bを経て情報CやDに至るリンクが張られていたとき、利用者にとってBが不要であるならば、Aから直接Cへ行ける仕組みがなければならない。また、そのようにして得られたリンクの集合こそが利用者を満足させるのに最も重要な要素である。

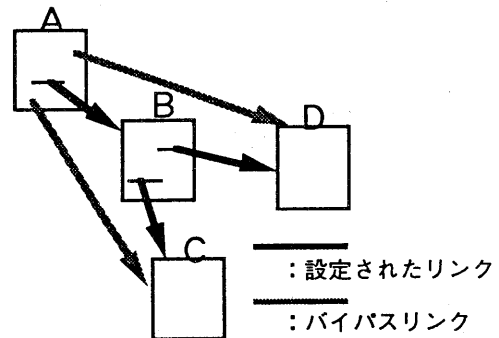


図3：情報のリンクとバイパス

現状のWWWシステムを例に挙げると、Mosaicの持つ「Hot List登録機能」が手動でのバイパスを可能としている点でこれに近い機能といえるが、決して十分なものではない。

本研究ではリンクの生成機能は以下の条件を満たすものとする。

- (1) 利用者に有用なリンクを作成する。
- (2) 利用者の情報行動や情報の生成に伴ってリンクの集合を変化させる。
- (3) 生成されるリンクは双方向である。

こうしたリンクの自動生成機能は、利用者の行動をもとにした学習アルゴリズムを備えることが必須となる。

## 5. 実現手法

### 5.1 自己組織化

4.1の自己組織化の実現には自己組織化マップ手法を用いる。

#### 5.1.1 自己組織化マップ

自己組織化マップ (Self Organizing (Feature) Map ; 以下SOM) 法は、T.Kohonenによって提案された中間層のない2層型の競合学習モデルである。

SOM法の特徴として、トポロジカルマッピングが挙げられる。これは入力パターンに隠されたトポロジカルな構造を学習によって発見し、マップに表現することである。例えば入力パターンとしてNewsの記事を考えると、学習の結果関連性の高い記事はマップの上では近くに集まり、クラスタリングが生じる事が期待される。[註4,5,6]

SOM法には様々なバリエーションがあるが、基本的には以下のような手順で学習が進んでいく。ここで、入力パターンを $x(t)$ 、出力ユニット $i$ の持つパターンを $m_i(t)$ 、出力平面上での位置を $n$ とする。

- (1) 各ユニットのパターンを初期化する。
- (2) 入力パターンに一番近いパターンを持つ出力ユニット $c$ を探す。距離の計算にはEuclidean距離等を用いる。
- (3) 出力平面上で $c$ の近傍の出力ユニットの集合 $N_c(t)$ を求める。
- (4) ユニット $N_c(t)$ の持つパターンを入力パターンに近づける。ここでは以下の計算をする。

$$m_i(t+1) = \begin{cases} i \in N_c(t) : m_i(t) + \alpha(t) [x(t) - m_i(t)] \\ \text{上記以外} : m_i(t) \end{cases}$$

- (5)  $N_c(t)$ と $\alpha$ を次第に小さくしながら(1)～(4)を繰り返す。 $\alpha$ は学習率を表わし、以下の様なガウス関数等を用いる。

$$\alpha(t) = \alpha_0(t) \exp(-||rc - r_i||^2 / \sigma(t)^2)$$

$N_c(t)$ 、 $\alpha_0(t)$ 、 $\sigma(t)$ は時間とともに単調減少する関数が用いられる。

#### 5.1.2 SOMの実装方式

##### (1) SOM法の適用対象

- ・News等の記事
- ・情報行動履歴

##### (2) 実装方式

5.1で述べたSOM法のアルゴリズムを適用対象の性質に応じて最適なパラメータを求め実装する。

##### (3) 評価

###### (3-1) 評価対象

- ・テキストのベクトル化手法
- ・履歴データのベクトル化手法

###### (3-2) パラメータ

- ・学習率
- ・初期マップ

###### (3-3) 評価項目

- ・情報空間の視認性→満足度、必要とする情報を入手するまでの稼働時間
- ・情報の流量と有用性の相関
- ・計算量
- ・レスポンス

### 5.2 自動リンク生成

4.2の自動リンク生成の学習アルゴリズムとして、遺伝的アルゴリズム (Genetic Algorithm:GA) を応用することとする。これは、対象とする情報空間が非常に大規模で常に変動しており、GAがこうした領域において優れた最適化能力を備えているからである。

#### 5.2.1 遺伝的アルゴリズム

GAは、Hollandにより考え出された計算理論であり、工学的なシステムに対して非常に強力な最適化能力を持ち、多くの分野で利用され始めている。

#### 5.2.2 実装方式

##### (1) GAの適用対象

- ・情報間のリンクを要素とする集合 (以下リンク構造体と呼ぶ)

リンク構造体は利用者がどの情報を辿ってきたのか、経由したリンク情報から構成される。

##### (2) 実装方式

本研究では実装に当たって以下のような機構を設けた。(図4)

1. 情報から他の情報へのリンクを1単位とし、利用者が辿っていったリンクの履歴で構成される任意の長さの

構造体を定義する。(図4-(1))

2. 参照された情報は、利用者のそれまでのリンクの履歴を利用者から得て、これをリンク構造体として保管。(図4-(2))
3. 上記の結果、複数のリンク構造体が一つの情報に対して付属することになる。これらのリンク構造体間で配列の組み替えなどの操作を行い、さらに多くのリンク構造体を作成する。(図4-(3))
4. 上記2、3で作成されたリンク構造体は、その構成要素であるリンクの発呼元の情報へ、発呼元の情報からリンク構造体を作成した情報までのリンク構造体の形に区切られて、送信される。(図4-(4))
5. ここまで、各情報は上記2、3で作成されたリンク構造体と、上記4で受信した構造体とを管理することとなる。
6. 各情報は各々が保管するリンク構造体を何らかの評価関数で評価し、評価値の高いものについて、その情報から関連情報へのリンクとして採用し、利用者がこれを選択することができるように自身を更新する。(図4-(5))
7. 上記2~6の過程は常時繰り返す。したがって、各情報からのリンク先は常に変化し、次々に有用な情報が増え、あるいは移り変わっていても、自動的に、かつ利用者の負担を増加させることなく速やかに有用な情報を得られるようになる。

### (3) 評価

現在、リンク構造体のデータ取得用兼資料共有システムとして、WWWベースのペーパーレス回覧システム(PLESS)を作成、稼働させている。これに加えて上記の動的リンク構造体生成機構を作成した後、以下の項目について、測定を行い、機構の改善を図っていく。

#### (3-1) 評価対象

- ・評価関数
- ・組み替えアルゴリズム

#### (3-2) パラメータ

- ・リンク構造体の長さ
- ・リンク構造体更新頻度
- ・交差率
- ・変異率

#### (3-3) 評価項目

- ・リンク構造体の寿命と有用性の相関
- ・アクセス量と更新速度(情報空間としてのレスポンスの良さ)
- ・ネットワークへの負荷
- ・レスポンス

## 6. 実験

現在、5、1、5、2のSOM法をニュースシステムで配布される記事に対して適用し、その有効性の確認を行っ

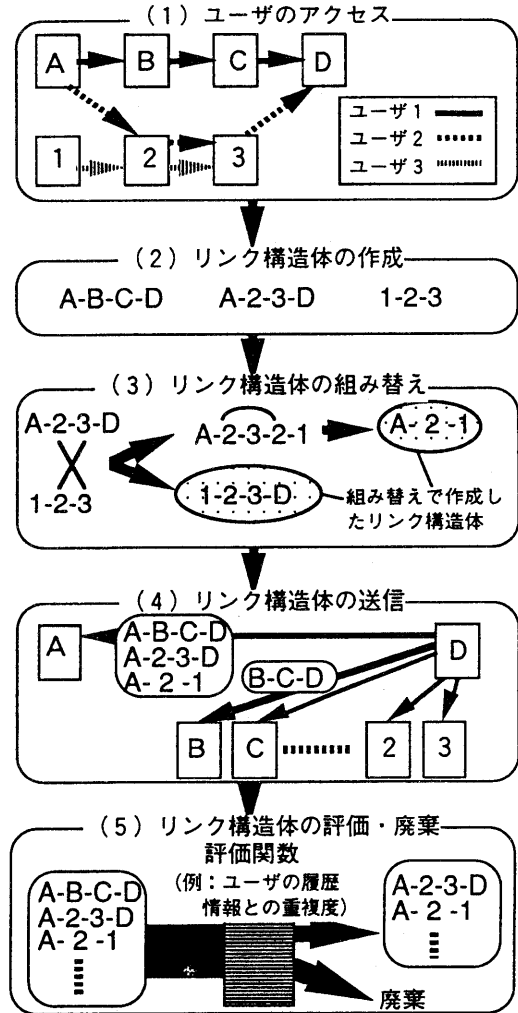


図4：GA応用システムの動作例

ている。このとき、出力されるマップ上には記事へのポイントがクラスタリングされて分布していることが期待される。

### 6. 1 実装方式

#### 6. 1. 1 記事のベクトル化

ニュースシステムを通じて配布される記事についてベクトル化を行った。

記事は以下の3つに分け、本文のみをベクトル化の対象とした。

- ・ヘッダ部：タイトル、ニュースグループ等
- ・コメント部：前の記事から引用部分

・本文

対象となる一連の記事群について単語の頻度情報を作成し、高頻度の単語で構成される辞書を作成する。次に各記事について、辞書の単語数を次元数とし、各要素が辞書に登録された単語の出現頻度に比例するようなベクトルを作成する。これを正規化したものを入力パターンとする。

#### 6. 1. 2 学習アルゴリズム

5. 2で述べた手法で学習を行う。各パラメータは以下のように設定する。

- (1) 出力ユニットの初期値  
マップ上の出力ユニットの値は乱数によって定める。
- (2) マップの大きさ  
学習させる情報量、テキストベクトルの次元数などのパラメータとの相関を調べ、最適な大きさの決定法を求める。
- (3) 近傍  
出力ユニット $c$ を中心とする一辺 $4\sigma(t)$ の正方形に含まれるユニットを $c$ の近傍とする。
- (4) 関数  
 $\alpha o(t) = 0.5x(1-t/\text{学習回数})$   
 $\sigma(t) = (1-t/\text{学習回数}) \times (\text{マップの長さ}/2)$

#### 6. 1. 3 キーワードマップの作成

出力されたマップを元に、各ユニットで最も強い要素をそのユニットを代表するキーワードとして抽出する。出力マップ全体について抽出を行い、キーワードマップを作成する。

#### 6. 2 マップの評価法

得られた2つのマップの有効性を評価するのに以下の方法を実施し、評価に基づいて実装方法の改善を図っていく。

- (1) クラスタリングの確認  
出力マップについて、話題による記事のクラスタリングが生じているかどうか確認する。
- (2) キーワードマップの純度  
キーワードマップ上で意味のない単語(冠詞、代名詞等)が出現することをノイズとする。ノイズの混入率を純度と定義し、純度を測定する。
- (3) 利用者の主観評価  
生成したマップをISMAP形式(クリックした位置をサーバに送信できる画像形式)に変換し、WWWベースのシステムでニュースにアクセスすることを可能とするシステムを作成する。  
実際にニュースの購読にこのシステムを利用してもらい、従来のニュースリーダーを使用した場合との主観評価を比較する。

### 7. 今後の進め方

現在、実証が始まったばかりであり、大まかな見通ししかできないが、以下のように進めていく予定である。

- (1) アルゴリズム、実装方式の最適化
- (2) 独立した情報アクセスツールの制作
- (3) 自己組織化手法とGAの統合

#### [参考文献]

- [1] Haan, B.J., Kahn, P., Riley, V.A., Coombs, J.H. and Meyrowitz, N.K.: IRIS Hypermedia Services, Comm. ACM, Vol.35, No.1, pp.36-51(1992)
- [2] 土井美和子, 福井美和子, 山口浩志, 竹林洋一, 岩井勇: プレーンテキスト/ハイパーテキスト間の変換, 情報処理学会, 情報額基礎研究会, FI-13-5, pp.1-8(1989)
- [3] Hara, Y., Keller, A.M. and Wiederhold, G.: Implementing Hypertext Database Relations through Aggregations and Exceptions, Hypertext '91, pp.75-90(1991)
- [4] Teuvo Kohonen: "The Self-Organizing Map", Proceedings of the IEEE, Vol.78, No.9, pp.1464-1480(1990)
- [5] 津高新一郎: 自己組織化マップを用いたテキスト分類の試み, 情報処理学会第46回全国大会, 分冊4, pp.187-188, 1993
- [6] 銭晴, 史欣, 田中克己: 自己組織化マップと語彙索引を用いたデータベースの抽象化機構, 情報処理学会, データベースシステム研究会, Vol.99-22pp.163-170(1994)