



マルチメディアデータの自己組織化と検索に関する研究

波多野, 賢治

(Degree)

博士 (工学)

(Date of Degree)

1999-03-31

(Date of Publication)

2008-03-11

(Resource Type)

doctoral thesis

(Report Number)

甲1925

(JaLCD0I)

<https://doi.org/10.11501/3156326>

(URL)

<https://hdl.handle.net/20.500.14094/D1001925>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



博士論文

マルチメディアデータの
自己組織化と検索に関する研究

1999年2月

神戸大学大学院
自然科学研究科(情報メディア科学専攻)

波多野 賢治

目次

1 序論	1
1.1 本論文の背景と目的	1
1.2 本論文の構成	2
2 基本的事項および関連研究	5
2.1 自己組織化マップ (Self-Organizing Map: SOM)	5
2.2 離散コサイン変換 (Discrete Cosine Transform: DCT)	6
2.3 検索サービス (検索エンジン)	7
2.4 適合率と再現率	8
2.5 関連研究	10
2.5.1 文書の分類に関する研究	10
2.5.2 画像・映像の分類に関する研究	11
2.5.3 検索インタフェースに関する研究	12
3 文書およびビデオデータの自己組織化	15
3.1 緒言	15
3.2 文書の自己組織化	17
3.2.1 文書の特徴ベクトルおよび 3D-SOM の生成	17
3.2.2 3D-SOM 上におけるユーザの操作	20
3.2.3 本システムの実験および評価	21
3.3 カット映像の自己組織化	25
3.3.1 カット映像の特徴ベクトルの生成	25
3.3.2 システムの実装	28
3.3.3 本システムの実験および評価	29
3.4 結言	32

4	映像自己組織化機構に基づく内容記述と類似シーン検索	35
4.1	緒言	35
4.2	ハイブリッド型 3D-SOM の生成	37
4.2.1	ハイブリッド型カット映像特徴ベクトルの生成	38
4.2.2	ハイブリッド型 3D-SOM の生成	39
4.2.3	本システムの評価実験	40
4.3	結言	44
5	多段階自己組織化マップによるビデオ映像記述支援と類似シーン検索	47
5.1	緒言	47
5.2	シーンへの記述および類似シーン検索	49
5.2.1	シーン特徴ベクトルの生成	49
5.2.2	シーン映像 3D-SOM の生成	52
5.2.3	本システムの実装・評価実験	53
5.3	結言	56
6	自己組織化マップと検索エンジンを用いた Web 文書の分類ビュー機構	59
6.1	緒言	59
6.2	適合フィードバック (Relevance Feedback)	61
6.3	Web 文書の分類のための情報単位と特徴ベクトル生成	61
6.3.1	Web 文書の特徴ベクトルの生成	63
6.4	ユーザとの対話による分類マップの段階的修正	63
6.5	Web 文書分類ビュー機構の実行例とその評価	67
6.5.1	分類ビュー機構の実行例	67
6.5.2	分類ビュー機構の評価	71
6.6	Web 文書の情報単位	73
6.6.1	極小部分マッチグラフ	74
6.6.2	極小部分マッチグラフの評価尺度	75
6.6.3	極小部分マッチグラフの抽出	77
6.6.4	極小部分マッチグラフの自動分類	78
6.6.5	極小部分マッチグラフの予備実験	79
6.7	結言	80
7	結論	83

目次	iii
参考文献	85
謝辭	93
研究業績	95

目次

2.1	適合率と再現率	9
3.1	文書の自己組織化機構	17
3.2	3D-SOM とある円柱に含まれている文書 (点検索)	19
3.3	ある領域の第 2 レベル分類マップ	20
3.4	文書から 3D-SOM に対する操作の一例	22
3.5	フレーム画像から DCT 成分の抽出	26
3.6	カット映像の 3D-SOM の一例	29
3.7	3D-SOM からカット映像へのナビゲーション	30
4.1	ハイブリッド型 3D-SOM の生成手順	37
4.2	コンテンツ情報 3D-SOM © 東映株式会社	38
4.3	ハイブリッド型 3D-SOM とカット映像の分類結果の閲覧 © 東映株式会社	40
5.1	コンテンツ情報 3D-SOM © 愛企画センター	50
5.2	シーン映像 3D-SOM	53
5.3	3D-SOM の詳細化機能	54
6.1	システムの全体構成	62
6.2	分類マップ概覧	68
6.3	キーワード選択画面	69
6.4	削除の適合フィールドバック	70
6.5	合併の適合フィールドバック	71
6.6	検索法による検索可能な範囲	74
6.7	検索モデルの事例	76
6.8	評価の一例	77
6.9	極小部分マッチグラフの抽出	78

6.10 極小部分グラフの構造	81
---------------------------	----

表 目 次

3.1	分類対象によるクラスタ数とラベル数 (出現頻度)	23
3.2	分類対象によるクラスタ数とラベル数 (<i>tf/idf</i> 法)	23
3.3	分類マップレベル別のクラスタ数とラベル数 (出現頻度)	24
3.4	分類マップレベル別のクラスタ数とラベル数 (<i>tf/idf</i> 法)	24
3.5	近傍距離 1 の場合のカット映像の適合率と再現率	31
3.6	近傍距離が 2 の場合のカット映像の適合率と再現率	31
4.1	コンテンツ情報 3D-SOM のカット映像の分類結果に対する適合率と再現率 (ニュー ス映像)	42
4.2	コンテンツ情報 3D-SOM のカット映像の分類結果に対する適合率と再現率 (テレ ビアニメ)	42
4.3	コンテンツ情報を分類基準とした ハイブリッド型 3D-SOM のカット映像の分類 結果に対する適合率と再現率 (テレビアニメ)	42
4.4	コンテンツ情報および記述情報によるコンテンツ情報 3D-SOM のカット映像の分 類結果に対する適合率と再現率	43
4.5	コンテンツ情報および記述情報によるハイブリッド型 3D-SOM のカット映像の分 類結果に対する適合率と再現率 ($w = 100$)	43
4.6	コンテンツ情報および記述情報によるハイブリッド型 3D-SOM のカット映像の分 類結果に対する適合率と再現率 ($w = 200$)	43
5.1	シーン映像 3D-SOM の適合率と再現率	55
6.1	拡大による 3D-SOM 上領域数の変化	69
6.2	適合フィードバックによる適合率と再現率の変化	72
6.3	分類・検索単位の違いによる適合率と再現率	80

第 1 章

序 論

1.1 本論文の背景と目的

近年、計算機ネットワークや蓄積メディアの進歩により、World Wide Web (WWW) やオンライン情報サービスなどに注目が集まっている。また、これらのサービスで扱われている情報も文字や数値などの文書を始めとして、静止画像、映像、グラフィックスおよび音声など多岐にわたる。これらマルチメディアデータは情報化社会において日を追うごとに増加を続けており、WWW 上に存在する Web 文書のページ数だけでも 1997 年 12 月の時点で 3 億 2 千万ページにもものぼると推定されている [32]。このように増え続けるマルチメディアデータを設計・管理するための枠組みとしてマルチメディア・データベースが注目されているが、汎用的なマルチメディアデータベースモデルやメディア間の制約を表現できるデータモデル、および一貫性を保持するための構成管理手法は確立されていない。こうした背景から、マルチメディアデータの分類方法や自己組織化の技術、それらデータを効果的かつ効率的にブラウズする方法について様々な研究がなされている。

まず、文書の分類検索機構においては、文書の構造を解析しその集約的な構造の把握を行ったり、WWW 上を周期的に巡回して情報を収集しインデックスを生成する検索エンジンとそのインデックスを利用したデータベースを連動したシステムの構築が必要となってくる。しかし、これらの対処を行ってもなお以下のような問題が存在する。

- 対象となるデータが膨大であるため、それらのデータをデータベースに格納する作業に大変な労力と手間を要する。したがって、計算機を利用したそれぞれの情報から適切な特徴量を抽出するシステムが必要となる。
- 分類・検索されるデータ数が多く、1 度の問い合わせだけでは的確な情報を得ることが難しい。したがって、言葉による問い合わせ表現に関する枠組みの確立、もしくは検索される

データを効果的に扱うことができるユーザインタフェースの仕様を確立する必要がある。

一方、映像データの分類・検索に関する研究においては主に2種類の方法が行われている。1つは情報に付属しているインデックスなどの2次情報に基づく分類・検索の方法、もう1つは、映像全体の色やそれらに映っている物体などの特徴量に基づく分類・検索の方法である。それぞれの研究の中には商用化レベルにまで到達した精度の高いものも存在するが、それでもなお、以下のような問題が存在する。

- それぞれの分野に限って見れば商用化レベルにまで到達しているが、それぞれの分野の技術に関してその性能および検索精度の観点から考慮すれば限界点に到達しているように思われる。よって、さらなる進歩のためにはそれぞれの長所を生かすことが可能な分類・検索基準を新たに確立する必要がある。
- 文書の場合と同様、分類・検索されるデータの数が多く、1度の問い合わせだけでは的確な情報を得ることが難しい。したがって、問い合わせ表現に関する枠組みの確立、もしくは検索されるデータを効果的に扱うことができるユーザインタフェースの仕様を確立する必要がある。

本論文では、以上のような問題点を考慮した手法を提案しその議論を行う。また、実装したプロトタイプシステムの評価についても述べる。

1.2 本論文の構成

本論文では、前節で述べたマルチメディアデータの分類・検索手法について、現在議論されている問題点の改良のための手法について論じる。本論文は以下の7章から構成される。

第2章では、本研究を遂行するにあたり利用した基本的な事項および関連研究について述べる。ここでは、近年、情報の分類に利用されているニューラルネットワークの1つである自己組織化マップ (Self-Organizing Map: 以下 SOM と記す) および映像の特徴量抽出の際に利用した Discrete Cosine Transform (DCT) について簡単に説明する。また、システムの実装、評価の際に利用している検索エンジン、評価尺度および本論文に関連する研究について簡単に述べる。

第3章では、データ数が膨大であるマルチメディアデータ、特に文書および映像データ (カット映像) から計算機で扱える特徴量を抽出し、その特徴量 (特徴ベクトル) を基に SOM を利用してデータを分類・検索するシステムの実装方法およびその評価実験について述べる。ここでは、ベクトル空間法を利用したそれぞれのデータの特徴ベクトルの生成法および overview マップを利用した分類されたデータの可視化の観点から実現した3次元インタフェースの説明を行う。

第4章ではカット映像の分類に焦点をあて、映像の色や映っている物体の形といったコンテンツのみを特徴量として分類する方法、つまり内容検索技術の利用だけでは不十分であるため、映像にその意味内容を記述するための映像記述支援ツールを実装し、記述された意味内容情報とコンテンツ情報を共に利用したハイブリッド型の映像の特徴ベクトルを利用したカット映像3D-SOMの実装およびその評価実験について述べる。ここでは特に、カット映像のハイブリッド型の特徴量の生成法の説明を行う。

第5章では、シーン映像の分類に焦点を当てており、第4章まで行ってきたカット映像に与えられた記述情報を特徴ベクトルとして利用したシーン映像の3D-SOMの実装およびその評価実験について述べる。ここでは、SOMを多段階に利用することにより、初期の段階では見いだすことが出来なかった新たな分類尺度を発見でき、その分類尺度により情報を分類する方法の提案とその有効性について述べる。

第6章では、分類された文書データの分類結果に対してユーザがインタラクティブに操作を行い、一種のユーザビューを生成する分類機構の実装およびその評価実験について述べる。ここでは、ユーザのフィードバック操作をどのように特徴ベクトルに反映させてSOMを学習させるかについて述べ、さらに現在のWeb文書の検索単位をページという概念ではなく関連のあるページ群を検索単位とするためのモデルの提案を行う。

最後に、第7章では、本研究で得られた研究成果をまとめ、さらに、得られた問題点に対する今後の展開について述べる。

第 2 章

基本的事項および関連研究

2.1 自己組織化マップ (Self-Organizing Map: SOM)

ニューラルネットワークの一種である SOM[26, 27, 28] は教師無し競合強化学習モデルである。出力層の各セルが層の中で位置を持つという点が他の学習モデルと異なる。データに隠されているトポロジカルな構造を学習アルゴリズムにより発見し、通常 2 次元空間で表示するという特徴を持っているため、特徴のよく似たデータ同士は出力マップ上の近い位置に配置されるようになっている。生成されたマップはそれぞれのデータの位置関係によって、類似しているデータかどうか直観的に理解しやすいという点からシステムの視覚化に利用できる。SOM には様々な種類があるが、ここでは最も基本的なものについて述べる。

SOM で用いられるネットワークは、セルを 2 次元に六角格子状に配置したものである。それぞれのセル i はセルの特徴ベクトル $\mathbf{m}_i(t) \in R^n$ (R は実数) を持っており (t は時間を表し、 $\mathbf{m}_i(0)$ は適切な方法で初期化されている)、これらのセルの特徴ベクトルを入力である特徴ベクトル $\mathbf{x}_j \in R^n$ ($j = 1, 2, \dots, d$) に選択的に近づけることによって学習は進行する。このとき、SOM では入力となる特徴ベクトルに一番近いパターンを持つ出力セルおよびその近傍のセルの集合のみが入力ベクトルに近づくことができるようなアルゴリズムをとる。

SOM のアルゴリズムを以下に示す。

1. 各入力特徴ベクトルを生成し、その集合を X とする。

$$X = \{\mathbf{x}_j \mid \mathbf{x}_j \in R^n, j = 1, 2, \dots, d\} \quad (2.1)$$

2. 出力層にある各セルの持つ特徴ベクトルを初期化し，入力特徴ベクトルの集合 X に基づき各セルの持つ特徴ベクトルを乱数によりランダムに決定する．

$$M = \{\mathbf{m}_i \mid \mathbf{m}_i \in R^n, i = 1, 2, \dots, k\} \quad (2.2)$$

(ただし， $\mathbf{m}_i(0) = (0, 0, \dots, 0)$ とした)

3. T をあらかじめ設定された学習回数とする．このとき， $t = 0, 1, \dots, T$ について以下を繰り返す．

- (a) $j = 1, 2, \dots, d$ について以下を繰り返す．
 (b) x_j に最も近いセル c を探す．つまり， $\|\mathbf{x}_j - \mathbf{m}_c(t)\|$ を最小にするセル c を求める．
 (c) 探し出したセル c の特徴ベクトル \mathbf{m}_c を更新し，さらにその近傍のセルの集合 N_c も入力パターンに近づける．

$$\mathbf{m}_i(t+1) = \begin{cases} \mathbf{m}_i(t) + \alpha(t)[\mathbf{x}_k(t) - \mathbf{m}_i(t)] & (i \in N_c(t)) \\ \mathbf{m}_i(t) & (i \notin N_c(t)) \end{cases} \quad (2.3)$$

N_c の中央はセル c である． N_c の半径は，学習の初期段階では大抵大きく，学習を繰り返していくうちに単調に減少させる．また， $\alpha(t) \in (0, 1)$ は「学習率」を表し，これもまた時間と共に単調に減少させる．

$$\alpha(t) = \alpha_0(t) \exp\left(-\frac{\|\mathbf{r}_c - \mathbf{r}_i\|^2}{\sigma(t)^2}\right) \quad (2.4)$$

ただし， \mathbf{r}_c と \mathbf{r}_i はそれぞれセル c とセル i の持つベクトルを表す． $\alpha_0(t)$ や $\sigma(t)$ には単調減少の一次関数や指数関数がよく用いられる．

2.2 離散コサイン変換 (Discrete Cosine Transform: DCT)

DCT とは，Joint Photographic Expert Group (JPEG) や Moving Picture Experts Group (MPEG) とよばれる静止画像圧縮技術で用いられている画像の変換符合化方式である [71, 73]．1 枚の自然画像を $N \times N$ 画素の正方形の領域 (ブロック) に分割し，各ブロックに対して変換処理を行うと，領域内の平均的な画像 (領域全体が一様) に始まり，徐々に精細さを表現する画像へと段階的な画像に分解することができる．この分解操作を直交変換と言い，精細さが高いことを別

の言い方では周波数が高いと言う。自然画像は、第 1 低周波項 (平均値画像) から順に、高周波項へと分解した画像の重ね合わせの表現になる。

DCT のメリットは、変換前にランダムに分布していた画素値 (輝度など) が、変換後には低周波項に大きな値が集中する性質があるということである。したがって、高周波項を落とす操作 (量子化) をすれば情報圧縮を行うことができる。

1枚の画像から分割された画素ブロックの大きさが $N \times N$ 画素のとき、画素信号を $f(x, y)$ 、変換によって得られる係数 (DCT 係数) を $F(u, v)$ とすると、 $F(u, v)$ は次のように求まる。

$$F(u, v) = \frac{2}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} C(u)C(v) \cdot f(x, y) \cdot \cos \frac{(2x+1)u\pi}{2N} \cos \frac{(2y+1)v\pi}{2N} \quad (2.5)$$

ただし、

$$C(u), C(v) = \begin{cases} \frac{1}{\sqrt{2}} & (u, v) = (0, 0) \\ 1 & (u, v) \neq (0, 0) \end{cases}$$

このようにして得られた DCT 係数のうち $F(0, 0)$ を DC (Direct Current, 直流) 係数といい、それ以外の DCT 係数を AC (Alternate Current, 交流) 係数と呼ぶ。DCT 係数は、画素数と同じ $N \times N$ 個求まり、低周波成分に集中する。 DC 係数はブロック内の画素値の平均値を表し、 AC 係数はその周波数の活性度を示す。また次式に示す逆変換でブロック画像の再生画素値が求まる。

$$f(x, y) = \frac{2}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)C(v) \cdot F(u, v) \cdot \cos \frac{(2x+1)u\pi}{2N} \cos \frac{(2y+1)v\pi}{2N} \quad (2.6)$$

ただし、

$$C(u), C(v) = \begin{cases} \frac{1}{\sqrt{2}} & (u, v) = (0, 0) \\ 1 & (u, v) \neq (0, 0) \end{cases}$$

2.3 検索サービス (検索エンジン)

WWW 上の検索サービスは、仕事や娯楽のための情報収集にとって不可欠な道具となっている。検索サービスは、ユーザが得たい情報に関するジャンルやキーワードを入力することにより、希望のページの Uniform Resource Locators (URL) や時には多少の内容をも表示してくれる。その情報を利用して、ユーザは表示された URL をクリックすることにより、直接そのページにたどり着くことができる。

さて、WWW 上で利用できる検索サービスは、大きく分けて 2 つの種類に分類される。1 つは yahoo! [55] などに代表されるディレクトリ型検索エンジンと呼ばれるタイプ、もう 1 つは AltaVista [2] や goo [17] に代表されるロボット型検索エンジンと呼ばれるタイプである。

以下にそれぞれの検索サービスについて説明する。

- ディレクトリ型検索エンジン

利用者の登録により成り立つ検索サービスである。Web ページを登録する際に宣伝文やキーワード、カテゴリなどを求められ、それを参考にして検索エンジン上のカテゴリごとのディレクトリに登録される。つまり、希望のジャンルに関する情報を検索する場合に適する検索エンジンである。

検索エンジンによって分類の細かさに違いがあるが、登録者の増加に伴い過不足なくカテゴリを増やしていくなど、しっかりとした管理がなされているサイトは、カテゴリでもキーワードでも検索が行うことができる。

- ロボット型検索エンジン

WWW 上に存在する Web 文書を読み込み、その構造を解析してその Web 文書から参照されている文書すべてを読み込むことを再帰的に繰り返して情報を収集する WWW ロボット [53] を用いて、インデックスを生成し検索を行う検索サービスのことである。WWW ロボットは定期的に各 Web サーバを巡回してインデックスを更新するために、最新情報を網羅できる。

しかし、仕組みは簡単であるが利用価値がある検索エンジンを実装するには、種々のプログラミングの工夫やテクニック、そして強力なハードウェア群による並列 / 並行処理や高速ネットワークへの接続などの投資が必要であるが、それでも 3 億 2 千万ページともいわれる Web 文書をカバーしきれないのが現状である。

2.4 適合率と再現率

情報検索では、検索結果は情報要求に常に合致したものになるのに対して、主題検索の場合には、検索結果が検索質問あるいは情報要求を完全に満たすとは限らず部分的にしか満たさないことがある。つまり、本来検索されるべき情報が検索されなかったり、不必要な情報が得られたりするるのである。前者は検索もれ、後者は検索ノイズと呼ばれる。このような事態が生じるのは、検索対象とそれを表現するインデックスとの間にずれが存在するためであり、また、同じインデックスを使用している場合でも、それで意図する概念が必ず同じではないからである。したがって、この種の検索では検索結果の評価が必要不可欠である。

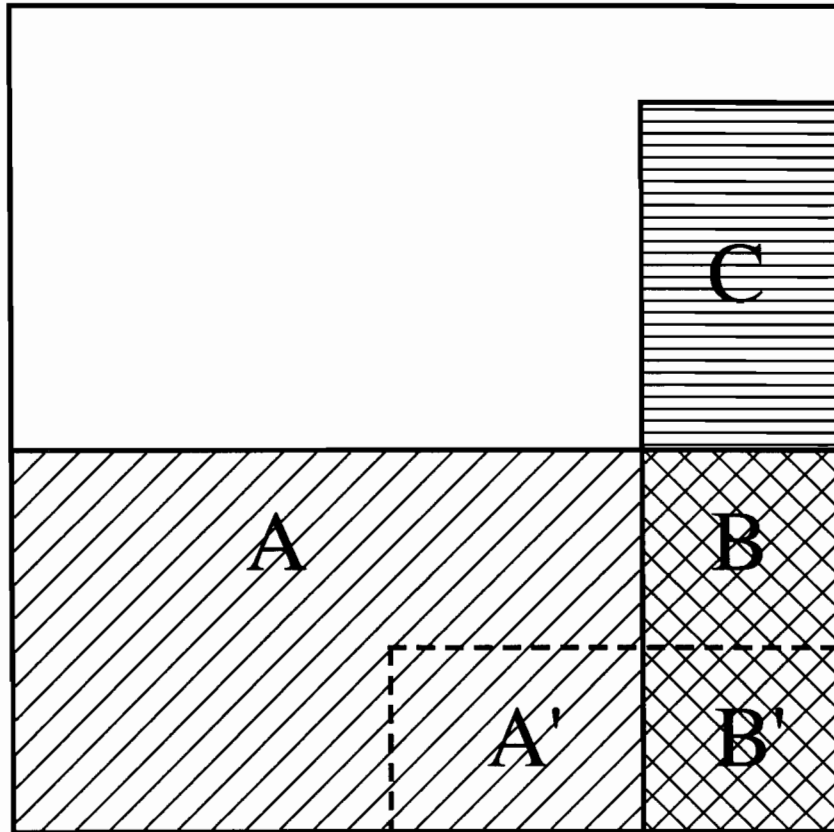


図 2.1: 適合率と再現率

検索結果がどの程度満足できるものであるかを示す概念は適合性と呼ばれ、検索結果と情報要求あるいは検索質問との間の一致の度合いで表される。情報検索の目的が情報要求を満たす情報の提供であるため、適合性の概念は極めて重要である。現在、適合性に関する普遍的な定義は存在せず、「適合する」とはどのような状況をさすのか、適合性の判定者、判定に影響する要因、適合性判定の過程、測定尺度などに関して共通の基盤が確立されていない。しかし、測定尺度に関しては使用されているのは、Aslib Cranfield 実験によって確立された「適合率」と「再現率」である [7, 8].

適合率は検索された情報に占める検索された適合情報の割合、再現率は情報要求に適合する蓄積情報 (検索対象) に占める検索された適合情報の割合である。図 2.1 で、 A を検索されなかった適合情報、 B を検索された適合情報、 C を検索された不適合情報 (ノイズ) とすれば、適合率は、 $\frac{B}{B+C}$ 、再現率は、 $\frac{B}{A+B}$ と表すことができる。再現率の計算に必要な $A+B$ はシステム中の適合情報の全体であり、本来検索したい未知の部分である。大規模データベースのように蓄積情報が膨大な場合、この部分を明らかにすることは実質上不可能である。したがって、 $A+B$ で表現される再現率はあくまでも理論上の概念であり、現実には $A+B$ に含まれていることが確認され

ている既知の適合情報 $A' + B'$ のうちどの程度検索されたかをしめす $\frac{B}{(A'+B')}$ が、検索結果や情報システムの評価に使用される [31]. なお、適合率と再現率は逆の関係にあり、単純に片方を高めようとする、もう一方の値を下げることになる。したがって、検索システムの性能を片方だけで評価すると不適当な結論を導くことになりかねないため、適合率と再現率の両方を含めて総合的に評価することが必要である。

2.5 関連研究

2.5.1 文書の分類に関する研究

文書 (ここでは HyperText Markup Language (HTML) で記述されている Web 文書も含む) 分類に関する研究は、主に、文書中から文書の持つ集約的な構造の抽出方法に関する研究 [5, 6, 19] や膨大で複雑な文書情報を文書情報空間内に配置し、それらを overview マップなどを利用して効率よくブラウジングする方法 [4, 37, 58] に関する研究が行われている。最近では、大量の文書の自動分類にニューラルネットワーク技術を採用する研究 [77] や Web 文書の内容に WWW サーバのアクセスログの情報などで表される利用者の興味度を利用した Web 文書の分類・ナビゲーション機構に関する研究 [76] や、検索を行う際に適当な問い合わせをシステムに対して与えることが困難であるため、ユーザからの問い合わせを段階的に詳細化していくという研究 [34] や検索事例を利用したガイダンスシステムに関する研究 [11] も行われている。

ニューラルネットワークを利用した情報の分類に関する研究の例を挙げれば、SOM を利用したものが多い。SOM は 2.1 節で述べたように Kohonen が提案した教師なし競合強化学習モデルであるが、情報検索の分野で SOM を利用した研究は Lin らの研究が最初であると思われる [33]。この研究では論文のタイトルやその著者などの文献データに対し SOM を適用して分類を行っている。また SOM を 2 次元検索インタフェースとして利用しており、SOM をタイトルや著者、キーワードなど別レイヤーで扱うことでそれぞれの項目で検索を可能としたものである。また、津高は同じように新聞記事の自動分類システムに SOM を適用している [69]。さらに、銭や仁木らも収集した文書の自動分類に SOM を適用し、分類データのブラウジングが可能なシステムを構築している [39, 77]。SOM を用いてデータのブラウジングを行いながら効果的に検索を行う研究は、この他にも SOM を提案した Kohonen の研究グループで開発が行われている WEBSOM [24, 30] や、ユーザが大量文書を鳥瞰しながらアクセスできる情報散策システム 3D Stroller [3, 68] などが挙げられる。これらシステムでは検索キーワードを与えて、これにマッチする文書をキーワードの共起などの統計情報に基づき SOM 上で自動分類を行っている。

本研究の特徴も同様に SOM を 3 次元に拡張した 3D-SOM と呼ばれる overview マップを利用して、3D-SOM 内をウォークスルーしデータの分類傾向を把握しながらインタラクティブな操

作・検索ができる点にある。さらに、得られた 3D-SOM の分類状態に対して、ユーザの視点や分類目的に添うように 3D-SOM を再生成したり、検索対象データの再収集を行ったりするユーザの操作をインタラクティブに反映させた情報の分類が可能である点も特徴の 1 つである。こういったインタラクションの実現は、WWW のナビゲーション操作自身が WWW に対する問い合わせ操作の一部であるという考え [1, 35] を採り入れることを可能としており、文書の分類・検索の分野のみならず情報検索一般にユーザビューやユーザのインタラクションの実現に重要な要素であると考えられる。

このようなインタラクションに関する研究は情報検索の分野で数多く行われており、Web 文書の検索に限って見た場合は伝統的な適合フィードバック [42] を用いて、ユーザとのインタラクションを通じてユーザの興味の変化を抽出し、それに応じて検索結果を変化させる研究が行われている [65]。しかし、本論文では 3D-SOM のある領域というデータ集合に対して行うものであるのに対し、これらの研究は分類データ 1 つ 1 つに対して行うもので、フィードバック作業そのものが大変な手間となるという問題が生じる。

2.5.2 画像・映像の分類に関する研究

画像・映像の分類・検索を行う場合は主に 2 種類の方法がある。1 つは情報に付属している画像・映像の内容が記述されたインデックスなどのタグ情報に基づく分類・検索であり、特に映像に関しては記述モデルの提案やデータベースの構築例なども盛んに行われている。もう 1 つは、画像・映像全体の色やそれらに映っている物体などのコンテンツの特徴量や映像セグメンテーションを基にインデキシングを行い、そのインデックスを利用した分類・検索の方法である。近年のマルチメディア情報の爆発的な増加に伴い、前者のようなデータにキーワードや説明文などのインデックスとなる情報を付加し映像の分類や検索を行うといった方法 [25] では、記述の際にかかる手間を軽減できる工夫を行っていても、その正確さ、つまり、言葉による表現の限界や記述者によるばらつきがあることを考慮すれば、すでに限界を越えているように思われる。よって、後者のような情報の特徴量をコンピュータを利用して自動抽出できるコンテンツによる検索技術が必要であり、これに関する様々な研究が行われている。

静止画像の分類・検索システムでよく知られているのは IBM で開発された QBIC [15] や Virage 社で開発された Illustra VIR [18] である。共に色そのものだけでなく画像全体の色の構成、画像内のオブジェクトの形や模様などの特徴量を基に、キーワード検索や例示画の表示による検索のサポートを行っている。また、このような画像検索を WWW へ適用するといった研究も行われている [48, 70]。しかし、この方法でも、画像や映像の特徴量の抽出法の限界問題を抱えており、現在、様々な模索が行われている。

映像の分類・検索システムでは、テレビ映像や映画など普段我々の身近にあるような映像には

ストーリーが存在するケースが多く、映像データベースに利用されている映像のセグメンテーションやインデキシングの手法ではそれを扱うことができない。また、従来の映像の記述モデルでも同様に、ストーリーを扱うと記述された内容が主観に左右されるような複雑な内容を表現できないといった問題が生じる。こうした問題点を考慮して、異なる視点からの映像の意味記述をプロトタイプ概念に基づき統合する方式[38]や知識表現に基づいた内容記述形式と概念依存表現を利用したストーリーや動作を表現したシステム[51, 52]、記述の際に断片的な内容記述をサポートするシステム[57, 66]の開発も行われている。

本論文では、映像を静止画像の集合として扱い、コンテンツによる映像の自動分類に関する研究を行っている。文書の場合と同じように SOM を映像の情報組織化のツールとして利用し、その 3D-SOM を分類ビューだけではなく検索のためのインタフェースとしても利用している。ここで扱っているコンテンツ情報は 2.2 節で述べた画像の DCT 成分であるが、先に述べたようにコンテンツ情報のみの分類にも限界があり、映像の内容記述情報とのハイブリッドな形での分類・検索システムの実装を行っているのが本論文の特徴である。このハイブリッドな形で静止画像や映像を捉えるという考え方は、現在、MPEG7[36]に取り込まれるという方向で標準化が進められている。

また、映像の分類という形ではなく、SOM を利用した映像に関する研究として、映像のモデル化を SOM のネットワーク表現という形でモデル化し、人や物体など映っているオブジェクトの動きを可視化する研究[74, 75]も行われている。

2.5.3 検索インタフェースに関する研究

最近のコンピュータ能力の向上により、ヒューマンインタフェースに利用できるコンピュータが増加し、文書以外にも静止画像や映像などのマルチメディアデータの取り扱いが容易になってきている。そして、それらを対象にした処理やそれらをヒューマンインタフェースとして活用する処理が実現出来るようになってきた。

情報検索の分野では、文書、静止画像、映像、音声などのマルチメディアデータを情報の表現メディアとして捉え、検索対象として処理するものとマルチメディアを人間との情報交換の媒体として捉え活用・処理するものがある。前者の立場では、各メディアごとの検索手法の研究が行われているが、文書の場合と同様、画像理解や音声認識という問題にぶつかっており、様々な模索が続けられている。後者については、情報の可視化という観点から研究が進められており、Xerox の Information Visualizer[16]を始めとして fisheye view[47]や納豆ビュー[67]などのような情報構造の可視化の手法がいろいろ提案されている。これらの手法の共通の概念は、3次元空間を効果的に利用して情報構造を可視化し、その空間のウォークスルーにより情報の探索が可能であるという点であり、映像などを利用して探索を支援する機能も実現されている。しかし、情報の探

索という観点からの構造化については考慮されていないものも多い。

本論文で提案している我々の SOM を利用した情報組織化ツールは、Virtual Reality Modeling Language (VRML)[20] を用いて実現しているシステムである。分類結果を表示している 3D-SOM と呼ばれる overview マップにユーザがアクセスすることで、分類されているデータのブラウジングが行えるようになっている。すなわち、3次元空間を利用しその空間をウォークスルーすることが可能であり、ユーザは分類対象の分類結果の全景を把握しながら、情報の検索を行うことができる。また、ユーザの分類目的に反した 3D-SOM が生成された場合は、分類目的に合うように 3D-SOM を再生成できるようにユーザの操作をインタラクティブに反映することができ、さらに、そういった操作を反映した検索対象データを収集できるシステムであるため、よりユーザに親和性の高いシステムとなっている。

第 3 章

文書およびビデオデータの自己組織化

3.1 緒言

近年、ハイパーテキストデータ [9] はインターネットなどに代表されるネットワーク上に多数分散する膨大な情報を効果的かつ効率的に閲覧するための標準となっている。この技術は計算機のハードディスク内に蓄えられた文書や映像データに代表されるマルチメディアデータが増加すればするほど重要なものとなってきている。ハイパーテキストで用いられているノード—リンクネットワークモデルは分散した情報を組織化するための簡単なモデルであるが、以下のようないくつかの問題が存在する。

- ユーザの情報迷子状態

情報を閲覧しているユーザが、巨大なネットワーク内で自分が何を閲覧したいのかを見失ってしまう。

- ハイパーテキストのリンクのメンテナンスの問題

ハイパーテキストのリンクは静的なリンクであるため、例えば Web 文書の URL が変化したり、新しい Web 文書が作られたりして情報量が増加すると、そのリンクメンテナンスに多くのコストを必要とする。

これらの問題点を克服するために用いられている方法の 1 つが、周期的に各 Web サーバを巡回して得られた文書からインデックスを作るといった検索エンジン [2, 13, 17] の利用である。もう 1 つの方法は、WWW とデータベース管理システム (DBMS) を連動させてユーザの問い合わせによる情報閲覧の方法である [50]。これらの方法はある程度上記のような問題点を解決するために有用であるが、これらは共にユーザの問い合わせに対するシステムの回答量が膨大となり、検索結果を動的に分類・組織化し視覚化するような効果的な機能を有していない。さらに、

現在の検索エンジンでは静止画像や映像を組織化したり分類したりするには不十分である。現在、行われている研究は文書の集約的な構造を把握したり [5, 6, 19], 膨大で複雑な情報からその特徴を効率よく抽出し overview マップを利用して視覚的に情報を分類する [4, 37, 58] といったものである。

本章では、Web 文書に代表されるハイパーテキスト文書および映像のカット映像を動的に組織化する SOM を利用した 3 次元情報オーガナイザーについて述べる。我々が情報の分類に SOM を利用したのは、SOM の学習結果が 2 次元マップで出力されることにより、分類された情報の overview マップを生成する際に有用であったからである。

本章で開発したシステムの特長を以下に挙げる。

- コンテンツを利用した分類

文書およびカット映像は共にそれらのコンテンツを¹利用して動的に分類される。本システムでは、文書またはカット映像から特徴量を抽出し SOM を用いて学習させることで、3 次元 overview マップ (以下 3D-SOM と記す) を生成する。特徴量を表現するために、ここでは Salton により提案されたベクトル空間モデル [43, 44] を適用し、それぞれの特徴ベクトルの生成を行った。文書の特徴量として文書中に出現する単語の出現頻度、カット映像の特徴量としてその DCT 係数を利用している。本章ではこの特徴ベクトル生成の方法の提案を行い、その有効性の検証のために評価実験を行った。

- 3D-SOM 上における検索とナビゲーション

生成された 3D-SOM は VRML を用いて 3 次元表示されている。この 3 次元化に伴い、ユーザは 3D-SOM 上でウォークスルーを行うことができ、さらに 3D-SOM と分類データ間のナビゲーション作業や類似検索、情報の詳細を閲覧することも可能である。特筆すべき情報はキーワードや代表フレーム画像という形で 3D-SOM 上に表示されており、分類されたデータ間の関係はマップ上の位置により表現されている。文書の場合は内容の類似度、カット映像の場合は利用されている色や映っている物体などの類似度により分類されており、分類マップの 3 次元化による情報分類の視覚的効果をユーザに与えることができる。

本章の構成は以下の通りである。3.2 節では、文書を自己組織化するために行う文書の特徴ベクトルの生成法とその分類マップである 3D-SOM の生成方法の説明とその評価について、3.3 節では、カット映像を自己組織化するためのカット映像特徴ベクトルの生成方法とその 3D-SOM の生成手順の説明とその評価について述べ、3.4 節では本章のまとめと今後の課題について述べる。

¹本論文では、コンテンツを、文書の場合は文書中に出現する単語の出現頻度、映像の場合は映像を構成するフレーム画像系列情報のことを指し、JPEG や MPEG 形式に変換して得られる情報もこれに含むものとする。

3.2 文書の自己組織化

図 3.1に、我々の 3D-SOM による文書群の自己組織化とそれに基づくブラウジング・検索機構の流れを示す。未分類状態の文書からキーワード検索などによって興味の対象となる文書集合を設定し、この文書集合から 3D-SOM 生成する。以下では、文書の特徴ベクトルの生成法および 3D-SOM の生成法について述べる。

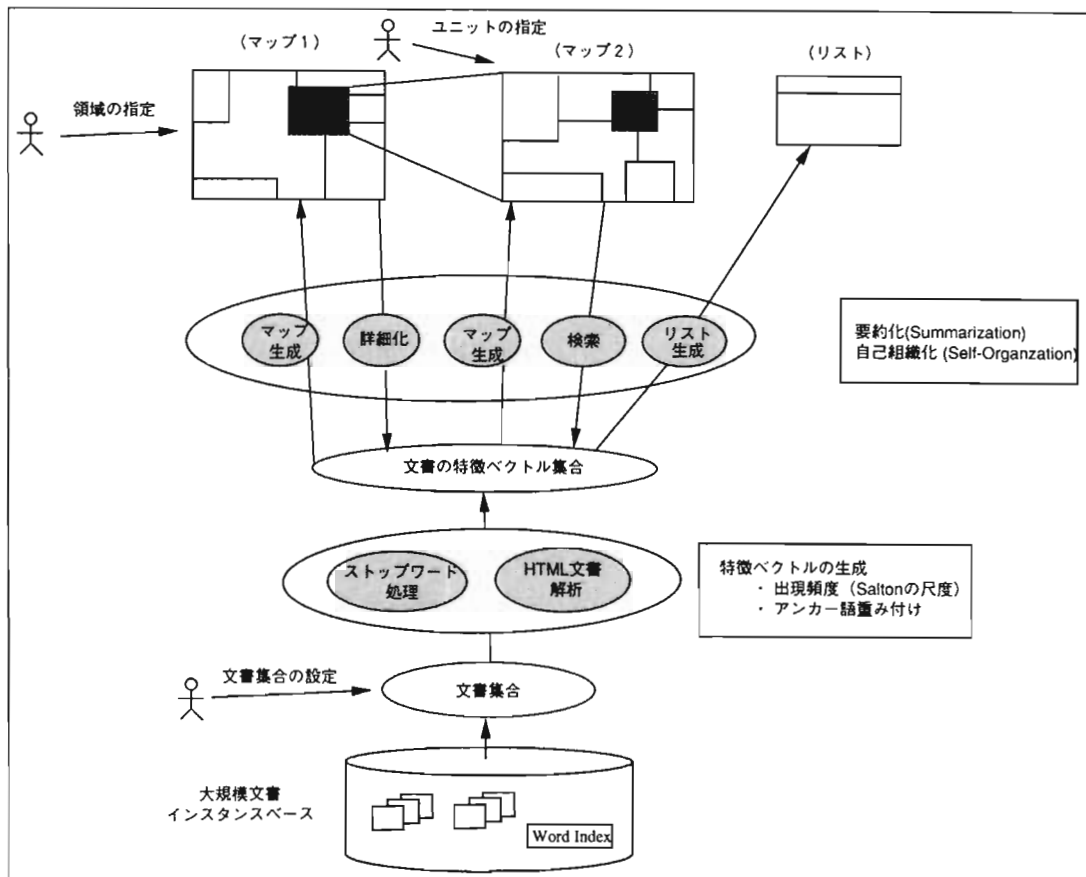


図 3.1: 文書の自己組織化機構

3.2.1 文書の特徴ベクトルおよび 3D-SOM の生成

文書の特徴を表す特徴ベクトルを生成し 3D-SOM を得るために、本システムでは以下の過程に従う。

1. 文書データベースから分類対象となる文書集合を決定し、その文書集合から文書中に現れる単語を抽出する。この際、辞書を用いてあらかじめストップワード²を取り除いておき、ま

²1 文字の単語、数字、冠詞など

た、複数形は単数形に、同義語も可能な限り変換しておく。

2. 抽出された単語の中からユーザが不必要と判断した単語を取り除いた上で、それぞれの単語に対して 2 通りの方法で重みづけを行う。ただし、分類対象が Web 文書などのようにリンクが存在する場合は、最初にアンカー文字列に対する重みづけを行う必要がある。
3. システムの仕様上、特徴ベクトルのベクトルの基底は 100 個以内にしなければならないため、単語の重みの大きな順に 100 個単語を抽出しそれをキーワードと定義する。このキーワードを特徴ベクトルの基底として各文書の特徴ベクトルを生成する。
4. 生成した特徴ベクトルを SOM に入力して学習させ、3D-SOM を得る。このとき通常の SOM の学習では、入力される特徴ベクトルの集合を基にランダムに計算されたベクトルが初期値として出力セルに与えられるが、本研究においては、初期値を与えた後、さらに入力される特徴ベクトルの集合を用いて初期学習 (荒い学習) を行うことで出力セルの学習精度の向上を図っている。

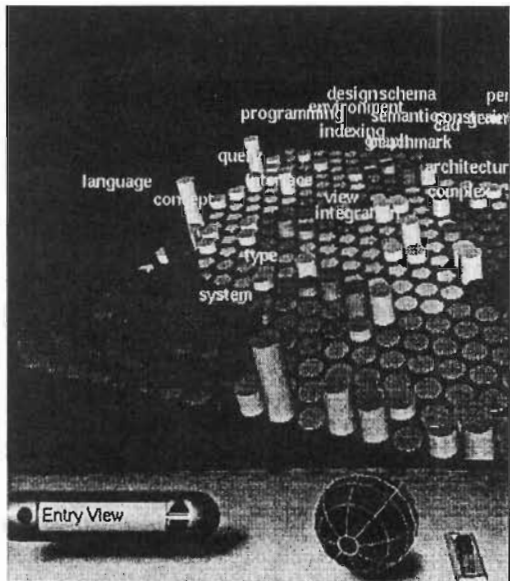
ステップ 2 において、我々は 2 つのアルゴリズムを採用している。1 つは単純に文書中の単語の出現頻度に基づいたもの、もう 1 つが *tf/idf* 法 [45, 46] である。前者は文書の主題はその中に含まれている単語の出現頻度により表現されているという仮定に基づいている。よって、文書 D_i の各単語 $t_k^i (k = 1, \dots, 100)$ の重み w_k^i は以下のように定義される。

$$w_k^i = tf_k^i \quad (3.1)$$

一方、Salton は、ある特定の文書にのみ出現数が多い単語に対して大きな重みを割り当てるような方法を用いたほうが他の文書との区別をつけやすいというアルゴリズムの提案を行っている。この後者のアルゴリズムを用いると、重み w_k^i は以下のように定義できる。

$$w_k^i = \frac{tf_k^i \cdot \log(N/n_k^i)}{\sqrt{\sum_{j=1}^t (tf_j^i)^2 \cdot (\log(N/n_j^i))^2}} \quad (3.2)$$

ただし、 tf_k^i は文書 D_i 中に出現する単語 t_k^i の出現頻度、 N は文書 D_i の数、 n_k^i は単語 t_k^i を含んでいる文書の数を表す。3.2.3 小節に示す実験から分かるように、単語の出現頻度による特徴ベクトル生成のアルゴリズムは文書群の大まかな内容をユーザが把握するために有用であるため、3D-SOM を生成する場合に役に立つ。また他方、*tf/idf* 法による特徴ベクトル生成のアルゴリズムは各文書の違いを区別するために有用であるため、情報検索に利用したり検索精度を上昇させる効果がある。



Bulk Types for Large Scale Programming

- Atkinson, M., Richard, P., and Trinder, P.
- Proceedings of the 1st International East/West Database Workshop, Kiev, USSR, Lecture Notes in Computer Science 504, Springer-Verlag, (Schmidt, J.W. and Stogny, A.A. Ed.), pp.228-250, October 1990
- Work on the design of constructors for bulk data types is reported. It introduces highly parametric constructors, parameterised both by types and by properties other than types. Such constructors we call type-quarks. The two examples of bulk types provided via type-quarks, maps and quads, are discussed. Several important questions about this approach to bulk types are identified.

Understanding Object-Oriented Database Systems

- Bancilhon, F.
- Proceedings of the 3rd International Conference on Extending Database Technology Vienna, Austria, Lecture Notes in Computer Science 500, Springer-Verlag, 1990

図 3.2: 3D-SOM とある円柱に含まれている文書 (点検索)

ステップ 4 において、ユーザは図 3.2 のような 3D-SOM を得ることができる。3D-SOM は円柱を六角形の格子状に並べた基本的に 2 次元のマップであり、それぞれの円柱が SOM におけるセルを表している。円柱はそれぞれ特徴ベクトルを持っており、SOM を用いて学習を行う時に円柱が持つ特徴ベクトルと文書が持つ特徴ベクトルの間の「距離」が最小となるように、各文書は各円柱に配置される。その結果、通常、ゼロまたは複数の文書が円柱に配置されることとなる。円柱の高さはその円柱に配置された文書の数に比例している。全ての文書が 3D-SOM 中に配置された後、システムはそれぞれの円柱に対してラベルづけを行う。例えば、ある円柱 c の特徴ベクトルを $\mathbf{V}(c) = (v_1, \dots, v_n)$ とした場合、 v_k ($k = 1, \dots, n$) は特徴ベクトルの基底を表す単語 t_k の各成分、すなわち重みを表している。このとき、円柱 c を最も象徴するラベル (以下第 1 キーワード、 $1stRepWord(c)$ と記す) は、 $\mathbf{V}(c)$ の各成分の中で最も値の大きな基底を表す単語 t_c であると定義される。第 2 キーワードや第 3 キーワード (以下それぞれ $2ndRepWord(c)$, $3rdRepWord(c)$ と記す) も同様に定義される。こうして定義されたこれらのラベルは 3D-SOM の領域形成に利用される。つまり、得られた 3D-SOM のある領域は、それぞれの円柱の持つ第 1 キーワードが同じである円柱の集合である。これらは 3D-SOM の詳細化機能を利用した場合も同様で、1 階層詳細化機能を利用して得られた 3D-SOM は第 2 キーワードを利用して領域形成が行われている。3D-SOM 内では同じ領域である円柱は全て同じ色で塗り分けられており、図 3.2 では、それぞれの領域は第 1 キーワードによって色分けすることで領域分けがなされている。

こうした 3D-SOM 内における文書の配置は、ユーザが多くの文書を分類するために効果的であり直感的に理解されやすい。関連のあるトピックは 3D-SOM 上で互いに近くに配置され、また、ラベルづけされた領域の大きさによって各領域に配置された文書がどれほど重要かどうかを判断することもできる。

3.2.2 3D-SOM 上におけるユーザの操作

ユーザのインタラクティブな情報の閲覧をサポートした検索機能の実現のために、我々は 3D-SOM に段階的詳細化機能と文書 3D-SOM 間のナビゲーション機能の実装を行った。

- 再計算のない詳細化機能

SOM の学習後に得られた 3D-SOM は第 1 レベルの分類マップである。これらの領域はこの機能を利用することで、図 3.3 のようにさらに細かな分類マップ、すなわち第 2 レベルの分類マップを得ることができる。もちろん、こうして得られた第 2 レベルの分類マップにおいても詳細化機能は利用することができる。この操作はマップ上に存在しているラベルには表現されていない情報をユーザに提供することができ、ユーザが欲している情報を逐次的に提供することができる。

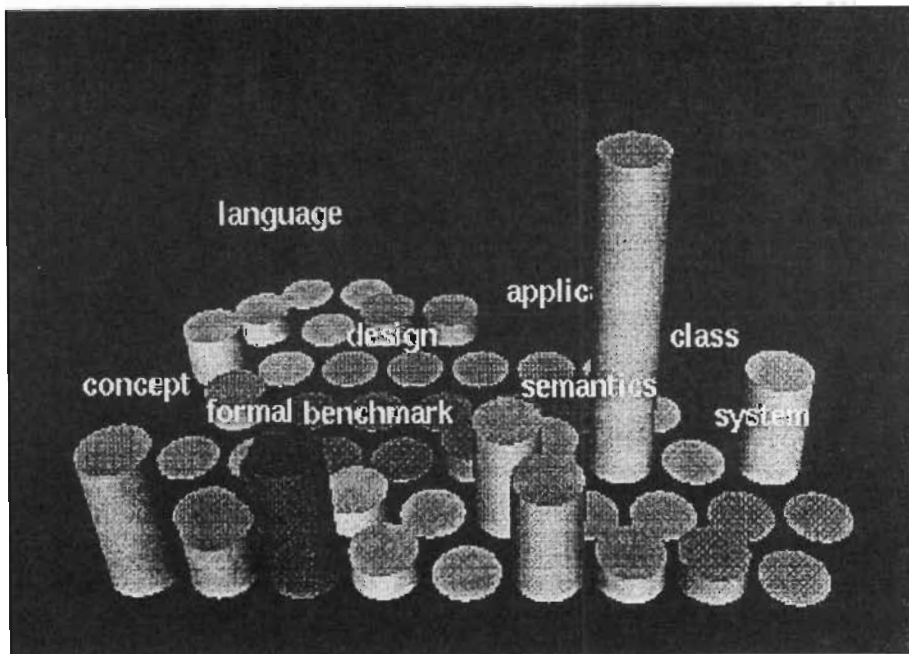


図 3.3: ある領域の第 2 レベル分類マップ

- 再計算のある詳細化機能

得られた 3D-SOM には、様々な要因によりユーザが望んでいるような情報が表現できない場合も起こりうる。この対策として、生成された 3D-SOM に入力した文書の特徴ベクトルをユーザの分類目的を反映させるよう再生成し、再度 SOM を用いて学習を行い新たな 3D-SOM を生成する。もちろん、このような操作は元の 3D-SOM の一部分の領域のみに限定され、ユーザが多くの情報の中から欲しい情報を絞り込む際に利用されることが多い。

● 文書と 3D-SOM 間のナビゲーション

3D-SOM はマップと文書間をつなぐ一種の対話的ユーザインタフェースとなっている。3D-SOM から文書へのナビゲーションには 2 種類あり、1 つは 3D-SOM 上のラベルを選択することによって行う「領域検索」であり、その領域中に配置されている文書を返すというものである。もう 1 つは、3D-SOM 上の円柱を選択することによって行う「点検索」であり、その円柱に配置されている文書を返すというものである。図 3.2 は、ある円柱に配置されている文書群をリスト表示したものである。この点検索機能は曖昧な情報検索に利用できる。すなわち、同じ領域に配置されている文書であってもそれらが配置されている 3D-SOM 中の位置によって、文書の持つ特徴は異なってくるからである。つまり、領域の境界線付近に配置されている文書というのは、各領域の特徴を含んだ文書であると考えることができるからである。

また、逆に文書からマップへのナビゲーション、すなわち、ユーザが指定したキーワードから、それらのキーワードを含んでいる文書および配置されている円柱を返すこともできる。この機能により、返される文書の総数の文書全体に対する相対的な数をユーザは把握することができる。図 3.4 に本ナビゲーションの例を示す。ここでは、ユーザが“model”というキーワードを指定することで、このキーワードを含む全ての文書のリストと 3D-SOM 中の位置が表示されている。

3.2.3 本システムの実験および評価

本章で試作した 3D-SOM のプロトタイプシステムは、米・シリコングラフィックス社により開発された Open Inventor 環境および VRML 環境で実現されている。本システムは図 3.1 に示されているように、主に次の 3 つの部分から成り立っている。

- 特徴ベクトル生成部
- SOM による学習部
- 3D-SOM 生成部



図 3.4: 文書から 3D-SOM に対する操作の一例

実験の対象として、オブジェクト思考データベース関連の英語論文 200 編を利用しそれらのタイトルおよびアブストラクトからマップサイズ 20×20 、円柱数 400 の 3D-SOM の生成を行った。特徴ベクトルの学習には Kohonen の研究グループで開発された SOM_PAK[29] を改良したものを利用し、米・シリコングラフィックス社のワークステーション Indy 上で 11,000 回学習させると、特徴ベクトルの基底が 72 の場合 4 分程度で 3D-SOM が生成された。

ここで、いくつかの実験結果をまとめる。この実験の目的は我々が提案した文書特徴ベクトル生成アルゴリズムの品質を評価することである。具体的には、3.2.1 小節に記述したように 2 種類のアルゴリズムを適用して論文のタイトルとアブストラクトを用いてそれぞれの 3D-SOM を生成し、マップ上のラベル数と生成されたクラスタの数を比較する。すなわち、同じラベルを持つ

ている円柱は、同じクラスタを形成するが、入力する特徴ベクトルによっては、同じラベルを持った別のクラスタを形成する可能性があるためである³。理想的にはクラスタの数とラベルの数は同じであることが望ましい。つまり、本実験では以下の項目の調査を行っている。

- 単純な出現頻度による特徴ベクトル生成アルゴリズムと *tf/idf* 法による特徴ベクトル生成アルゴリズムの比較
- 分類対象に論文のタイトルを利用した場合とアブストラクトを利用した場合の 3D-SOM 上に現れるラベルの数とクラスタの数の比較
- 分類対象に論文のタイトルを利用した場合の各詳細レベルの分類マップ上に現れるラベルの数とクラスタの数の比較

表3.1は、単純に文書中の単語の出現頻度を基にして 3D-SOM を生成し、そのマップ上に現れたクラスタの数と第 1 キーワードの個数を表したものである。また、表3.2は *tf/idf* 法を利用して 3D-SOM を生成した場合のクラスタ数と第 1 キーワードの個数を表しており、それぞれ論文のタイトルとアブストラクトを利用した場合について比較している。

	クラスタ数	第 1 キーワード数
タイトル	25	25
アブストラクト	76	36

表 3.1: 分類対象によるクラスタ数とラベル数 (出現頻度)

	クラスタ数	第 1 キーワード数
タイトル	39	39
アブストラクト	183	55

表 3.2: 分類対象によるクラスタ数とラベル数 (*tf/idf* 法)

また、表3.3と表3.4は、分類マップの各詳細レベルにおける領域数とラベル数を比較したものである。表3.3は単純な出現頻度による特徴ベクトル生成アルゴリズムを用いた場合を、表3.4は *tf/idf* 法を用いた場合である。この場合、検索対象データとして論文のタイトルのみを扱った。これらの結果から、以下のようなことが容易に解釈できる。

³ここでは「領域」を同じラベルのついた 1 つ以上のクラスタと定義している。つまり領域の数とラベルの数は同じである。

	クラスタ数	ラベル数
level 1	25	25
level 2	87	42
level 3	254	51

表 3.3: 分類マップレベル別のクラスタ数とラベル数 (出現頻度)

	クラスタ数	ラベル数
level 1	39	39
level 2	158	54
level 3	328	56

表 3.4: 分類マップレベル別のクラスタ数とラベル数 (*tf/idf* 法)

- 分類対象に論文のタイトルのみを利用した場合、どちらのベクトル生成アルゴリズムを用いた場合も 3D-SOM 上におけるクラスタ数とラベル数が同じ値となった。このことは、論文のタイトルを利用するような情報量が少ない場合はどちらのアルゴリズムを用いた場合もクラスタが完全に生成されているため、我々の提案したアルゴリズムは正しいものであるといえる。
- 論文のアブストラクトのみを利用した場合、*tf/idf*法を利用したベクトル生成アルゴリズムではクラスタ数がラベル数の 2 ~ 3 倍となっている (表 3.1, 表 3.2 参照)。よって、*tf/idf*法を利用したベクトル生成アルゴリズムでは、同じラベルづけがされた異なるクラスタが生成されたと考えられ、文書のクラスタを完全には行うことができなかった。
- 出現頻度によるベクトル生成アルゴリズムと *tf/idf*法によるベクトル生成アルゴリズムを比較した場合、*tf/idf*法を利用した場合のほうがクラスタ数、ラベル数は多くなり、またその領域の大きさは小さくなる (表 3.3, 表 3.4 参照)。
- 一般的に *tf/idf*法を利用したベクトル生成アルゴリズムのほうがその文書に特化したラベルが生成されていた。例えば、*tf/idf*法を利用したベクトル生成アルゴリズムでは、オブジェクト思考データベースの名前である “VLSI” や “ODE” といったその文書のみ出現する非常に特定の単語がラベルに利用されているが、出現頻度によるベクトル生成アルゴリズムでは、“generalization” や “integration” などといった一般的な単語がラベルとして利用さ

れていた。

以上のことから、出現頻度によるベクトル生成アルゴリズムは、文書を分類し分類対象全体の分類傾向を把握するための overview マップを生成するのに有用なアルゴリズムであると考えられる。一方、*tf/idf*法によるベクトル生成アルゴリズムは、情報の絞り込み、すなわち、情報検索を行う際に検索精度を上昇させたり、それぞれの文書を区別するために効果的であると考えられる。

3.3 カット映像の自己組織化

映像というものには、映像自体の色や映っている物体の形、また、そのセリフの内容や BGM など多くの情報が含まれている。しかし、それらの情報の関連づけが複雑で、また内容や構成に関する情報が明示的に含まれていないため、映像のデータベース化は非常に困難である。そこで、我々は 3.2 節で開発した 3D-SOM を映像データのコンテンツ情報による分類システムとして拡張を行った。この拡張の目的は、映像データの分類ツールを提供することで、ユーザが分類された映像データを再利用することが可能となると考えられる。

本節で開発するシステムの実現のためには、以下の問題にうまく対処する必要がある。

- 入力データの単位決定

時系列データである映像データを再利用可能な素材として活用できる形に断片化する必要がある。本システムでは、カット映像をこのための単位とする。従って、1 つのシーン映像 (例えば 1 つのニュース記事) は通常複数のカット映像に断片化される。

- 映像データからの特徴ベクトルの生成

静止画フレームの時系列データであるカット映像に対する特徴ベクトルをいかに生成するかが問題である。本システムでは各静止画フレームの DCT 係数を集約させる形で特徴ベクトルを生成する方式を提案する。

- 映像データの 3D-SOM の生成と表示

SOM を用いたカット映像の分類結果を基に、素材検索やブラウジングを効果的に行うためのインタフェースの開発が必要である。本システムでは、VRML を用いて 3 次元インタフェースの開発を行う。

3.3.1 カット映像の特徴ベクトルの生成

カット映像の特徴ベクトルを生成するために、本システムでは以下の過程に従う。

1. 映像データからのカット映像分割

通常、映像の自動分割は連続したフレーム画像間の特徴の変化を利用して行われる。本システムにおいてもその方式を採用し、フレーム画像の特徴ベクトル間の距離が大きく変化した部分で行われる。つまり、フレーム画像特徴ベクトルは、多次元空間の 1 点として表されるため、同じカット映像内の隣接したフレームのフレーム画像特徴ベクトルは、この多次元空間において近いところに存在しクラスタを形成すると考えられる。ところがカットが変わると、多次元空間中の点はこのクラスタから逸脱しはじめるので、この逸脱をもってカット点の検出を行うことができる。

2. フレーム画像の特徴ベクトルの生成

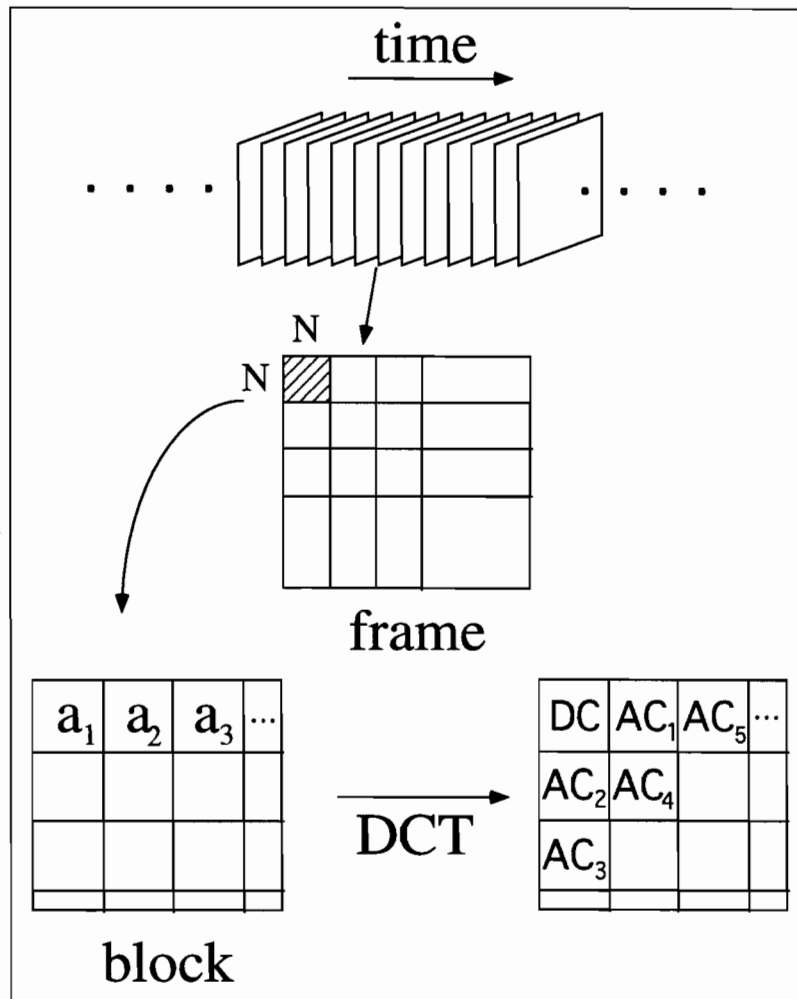


図 3.5: フレーム画像から DCT 成分の抽出

映像を構成している各フレーム画像を、サイズ $N \times N$ (通常は 16×16 または 8×8) ブロック分割し、そのブロック単位で2.2節で述べた2次元DCTにより周波数成分に変換する。変換された成分は図3.5のように左上を直流成分とし、右下に向かって周波数が上がっていくようにジグザグに成分が並ぶ。フレーム画像の特徴ベクトルを生成する際には、このうち DC , AC_1 および AC_2 の3つの成分を用いて、以下のように定義する。

$$(DC(1), AC_1(1), AC_2(1), \dots, DC(n), AC_1(n), AC_2(n)) \quad (3.3)$$

ただし、 $DC(k), AC_1(k), AC_2(k) (k = 1, 2, \dots, n)$ は分割されたフレーム画像の第 k ブロックにおける DC 成分と AC 成分である。

3. カット映像の特徴ベクトルの生成

カット映像は時間的に連続したフレーム画像の集合である。また、シーン映像も複数のカット映像の集合である。ここで問題となるのは、カット特徴ベクトルを生成するために、ベクトルの基底をどう定義するかが問題となる。そこで本システムでは、カット映像の特徴ベクトルの生成方法として、カット映像を構成するフレームのフレーム特徴ベクトルの重みつき平均値を用いる方法を提案する。この提案では、DCTを行うと低い周波数成分にその画像の特徴が集中するという特徴を利用することにより、情報量が膨大な映像の特徴をできるだけ低次元のベクトルに反映させることができる。

以下にカットの特徴ベクトル生成のアルゴリズムを示す。

- (a) カット映像を $M_i = \{f_1^i, \dots, f_m^i\} (m \geq 1)$ とする。ここで、 $f_j^i (j = 1, \dots, m)$ はカット映像 M_i を構成するフレーム画像を表す。
- (b) カット映像 M_i を構成するフレーム画像 f_j^i を $n = L \times M$ 個のブロックに分割し、各ブロックに対し3つのDCT成分 DC, AC_1, AC_2 を抽出する。フレーム全体で $n = 3 \times L \times M$ 個のDCT成分を得ることになるので、フレーム画像特徴ベクトル $V_{DCT}(f_j^i)$ は以下のように定義される。

$$V_{DCT}(f_j^i) = (DC^j(1), AC_1^j(1), AC_2^j(1), \dots, DC^j(n), AC_1^j(n), AC_2^j(n)) \quad (3.4)$$

- (c) カット映像 M_i の特徴ベクトルには重みづけを行う。先ほど抽出したフレーム特徴ベクトルの3つの成分、 DC, AC_1, AC_2 成分に、 DC 成分には重み w_1 を、 AC_1, AC_2

成分にはそれぞれ重み w_2 と w_3 をかける。よって、抽出されたカットに含まれる全てのフレームの特徴ベクトルの重みつき平均を求めると以下ようになる。

$$\begin{aligned} V_{DCT}(M_i) &= \left(\frac{w_1}{m} \cdot \sum_{j=1}^m DC^j(1), \frac{w_2}{m} \cdot \sum_{j=1}^m AC_1^j(1), \frac{w_3}{m} \cdot \sum_{j=1}^m AC_2^j(1), \right. \\ &\quad \dots, \\ &\quad \left. \frac{w_1}{m} \cdot \sum_{j=1}^m DC^j(n), \frac{w_2}{m} \cdot \sum_{j=1}^m AC_1^j(n), \frac{w_3}{m} \cdot \sum_{j=1}^m AC_2^j(n) \right) \end{aligned} \quad (3.5)$$

この $V_{DCT}(M_i)$ で表されるベクトルを本システムではカット特徴ベクトルと定義する。ここで、重み w_1 , w_2 , w_3 を用いた理由は、重みを考慮せずにカット映像の分類を行った場合、カット映像の DC 成分の影響が大きく映像の色で分類されるのではなく、むしろ映像の色合い (明暗) のみで分類されるという結果が得られたためである。

3.3.2 システムの実装

本システムは以下の3つの構成から成り立っている。

- カット映像分割部

龍谷大学の有木研究室で開発されたカット検出ソフトウェア [61] を改良して利用しているこのソフトウェアは米・シリコングラフィックス社のワークステーション Indy 上に実装されている Cosmo Compressor と呼ばれる画像圧縮・伸張ハードウェア上で実現される。このとき、各フレーム画像の DCT 成分が抽出されるため、同時にカット映像の特徴ベクトルの生成も行っている。

- SOM による学習部

文書の分類システムの場合と同様、Kohonen の研究グループで開発された **SOM_PAK**[28] を改良し SOM の学習に利用している。

- 3D-SOM 生成部

SOM によるカット映像の分類結果を VRML を利用して3次元表示を行い、カット映像の分類およびブラウジングを可能としている。図3.6はカット映像の分類結果を表示した3D-SOM の一例であり、VRML ブラウザを利用して3D-SOM 内をウォークスルーしたりカット映像の検索を行ったりすることができる。

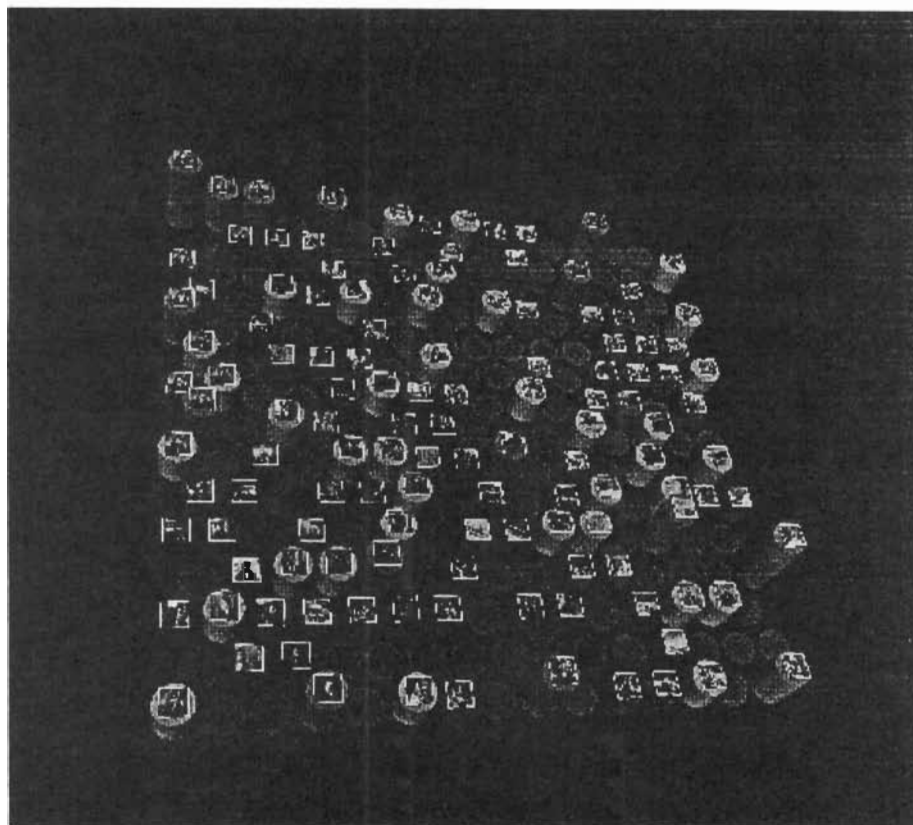


図 3.6: カット映像の 3D-SOM の一例

3.3.3 本システムの実験および評価

3D-SOM の生成に利用した映像は NHK のニュース映像 30 分 (10 分 \times 3, 計 56,392 フレーム) である。このとき、検出されたカット映像数は 192, カット特徴ベクトルの基底数は 144 であった。米・シリコングラフィックス社のワークステーション Indigo2 Extreme を利用し, SOM の学習回数を初期学習が 1,000 回, 本学習を 10,000 回の計 11,000 回とした場合, マップサイズが 15×15 の 225 セルの 3D-SOM の生成に 3 分程度の時間を要した。

3D-SOM は, 図 3.6 に示されているように VRML ブラウザ上に表現される。図中において円柱の高さはそれらに分類されたカット映像の数を表しており, またそれぞれの円柱上に貼りつけられた代表画像は, それぞれの円柱に分類されたカット映像の中でその円柱の持つ特徴ベクトルの値に最も近いカット映像の代表フレーム画像 (カット映像の最初のフレーム画像) である。図 3.6 上で, 任意の円柱を選択すると, 図 3.7 のようにその円柱に分類された代表フレーム画像がブラウザ上に表示される。ここに表示されている 3 つのカット映像は「類似している」と判断できる。

カット映像の分布状況を円柱に張りつけられた代表フレーム画像から判断すると, 3D-SOM 上でカット映像が集中している部分は雰囲気や明暗, 背景, 場面状況などといった視覚的に似通っ

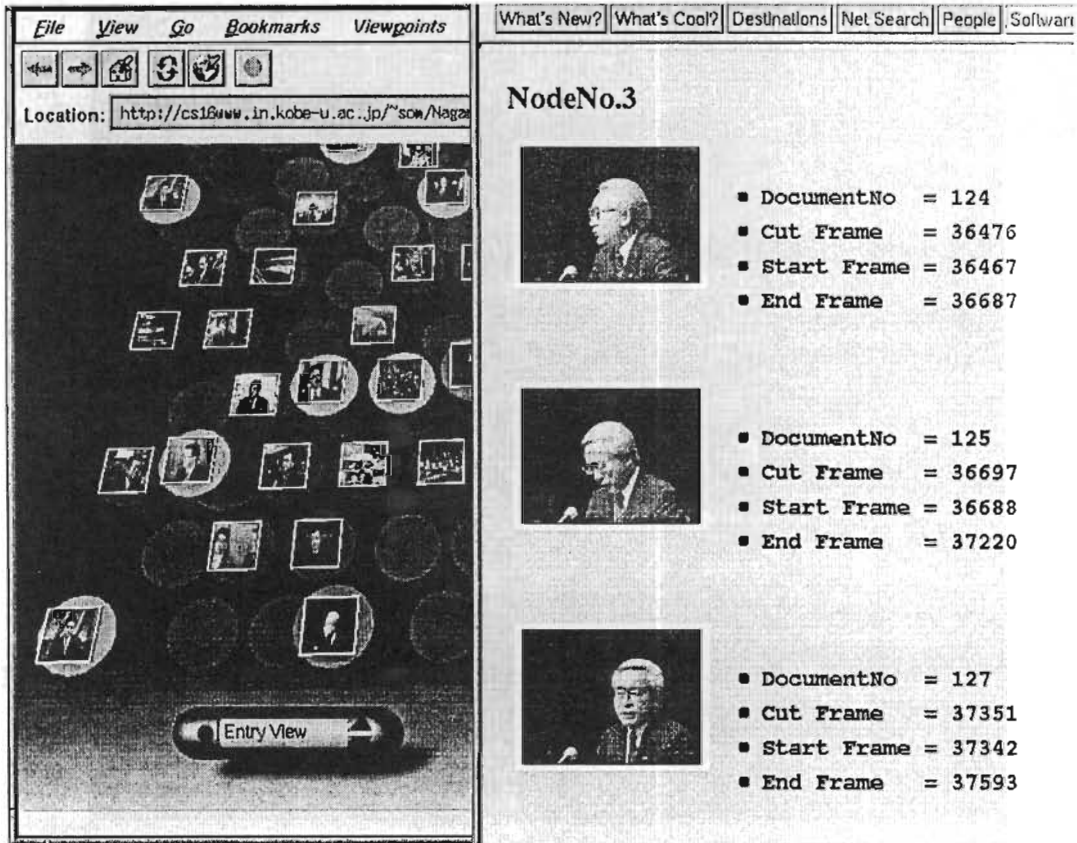


図 3.7: 3D-SOM からカット映像へのナビゲーション

たカット多く、これらの多くはカット映像中に映っている物体の動きやカメラワークが比較的静かなものが多い。さて、本システムを評価するために2.4節で述べたように、カット映像の分類結果を情報検索の分野でしばしば用いられているで述べた「適合率」と「再現率」を用いる。本評価実験では適合率および再現率の定義を多少改良し、あらかじめ円柱の上部に貼られた各カット映像に類似した特徴を持っているカット映像の集合を決めておく。つまり、円柱上に表示されている代表フレーム画像を持つカット映像 M に対して、我々はこの M に類似したカット映像が得られているとみなし、これを $similar(M)$ と表現することにする。よって、ここで定義される適合率および再現率は、

$$\text{適合率} = \frac{|neighbour(M) \cap similar(M)|}{|neighbour(M)|} \quad (3.6)$$

$$\text{再現率} = \frac{|neighbour(M) \cap similar(M)|}{|similar(M)|} \quad (3.7)$$

と定義される。

さて、我々はこの評価実験に際しいくつかの条件の元で適合率および再現率の計算を行っている。具体的には、3.5式で与えられている DCT 係数に与えられている重みの値を様々に変化させて重み間の関係を発見するというものである。 $similar(M)$ の決定は本論文の著者の主観であらかじめ行っている。

表 3.5 は、測定範囲が基準となる円柱から近傍距離が 1 である円柱である場合の、表 3.6 は近傍距離が 2 である場合の適合率と再現率の平均値である。

w_1	w_2	w_3	適合率 (%)	再現率 (%)
1	1	1	44.45	36.08
1/DC	1	1	46.99	50.50
1/DC	10	20	-	-
1/DC	20	10	-	-
1	10	20	53.71	43.52
1	20	10	52.66	47.93
1	10	100	40.22	51.53

表 3.5: 近傍距離 1 の場合のカット映像の適合率と再現率

w_1	w_2	w_3	適合率 (%)	再現率 (%)
1	1	1	47.42	45.75
1/DC	1	1	48.32	53.55
1	10	20	52.17	54.73
1	20	10	52.87	53.26
1	10	100	47.28	52.64

表 3.6: 近傍距離が 2 の場合のカット映像の適合率と再現率

表 3.5 と表 3.6 を見ると適合率、再現率ともに 40 ~ 60 % 前後と実験結果としてはそれほど良い値とは言い難い。しかし、類似しているカット映像の分類方法が、本論文の著者の主観に基づいたものであり、また、その結果算出された $similar(M)$ 、つまり類似したカット映像として利用されたカット映像の数が全カット映像の数の半数程度しか存在しなかった (その他のカット映像はカット映像自身以外類似カットがない) ことを考慮すれば、この実験結果は十分に納得のいく結果であろうと思われる。また、実験結果からカット映像の特徴ベクトルの生成には重み $w_1 < w_2 < w_3$

という関係が満たされていればよいことが経験的に判明した。これらの重みの間にどういった大小関係や相関関係があるのか厳密な部分は未だ解析ができていないが、本システムのカット映像の分類精度をを上げるには大変重要なものであると考えられ、さらに実験を行い最適な重みを設定できるようにする必要があると考えられる。

以上よりカット映像の視覚的な分類にはある程度の成果を得ることが出来たが、映像を符合化するという点でカット映像の内容に対する意味的なものをとらえられたかどうかは疑問である。この意味的な部分を取り出すには、映像の DCT 係数の利用などのようなコンテンツ情報を利用した分類方法ではすでに限界の域に達しており、映像の内容に対する断片的な記述、すなわち映像のそれぞれのカット映像に対する印象や感想といった内容記述情報を基に映像のインデックスづけを行う方法 [57, 66] とのハイブリッドな構成によって行って補えるものであると考えられる。

3.4 結言

本章では、3D-SOM をデータのブラウジングツールおよびデータの問い合わせのインタフェースと見なして文書および映像データをそのコンテンツを利用して自動分類するシステムの試作を行った。ここでは、文書や映像データからそれぞれの特徴量を抽出し、それぞれの特徴ベクトル生成を行う方法の提案を行い、その提案の妥当性を実験により評価した。

試作したシステムの特長として次のものが挙げられる。

- 大量のデータをコンテンツ情報のみを利用して自動的に分類でき、その overview マップは自己組織的に生成される。
- 円柱の高さを分類データの数に反映させるなど、分類マップを 3 次元で表現することによりユーザは 3D-SOM の全景からデータの分類傾向の把握が可能である。
- 3D-SOM の領域分割を行い、さらに 3D-SOM の点検索機能をサポートすることで、曖昧検索への対応が行われている。
- VRML を利用したアプリケーションであるため、WWW を利用したアプリケーション開発のための汎用性が確保されている。

また、今後の課題として、

- SOM の学習アルゴリズムやそれぞれの特徴ベクトル生成アルゴリズムは、本システムでは経験的に定められている。したがって、これらのパラメータの自動設定を行うことが可能なシステムの試作を行う。

- ユーザに対して親和性の高いユーザインタフェースの開発を行う。
- プロトタイプシステムの性能向上のための十分な評価実験を行う。
- 映像の意味内容記述システムとのハイブリッドな構成による映像分類検索システムの開発およびシーン映像の分類システムに対応できるような機能拡張を行う。

ことが挙げられる。

第 4 章

映像自己組織化機構に基づく内容記述と 類似シーン検索

4.1 緒言

WWW の急速な普及と発展や、計算機性能の向上、2 次記憶媒体の大容量化などに伴い、文書データのみならず、映像や音声などのマルチメディアデータを取り扱う機会が増加している。そのため、爆発的に増加するマルチメディアデータをデータベース化する技術はますます重要になってきている [12]。しかし、マルチメディアデータ、特にここでは映像データのことであるが、意味情報が複雑であり、かつ内容や構成に関する情報が明示的には含まれていないなどの理由から、現在の映像の分類方法や自己組織化の技術ではまだまだ不十分である。

さて、マルチメディアデータを動的に要約または自己組織化するためには、一般的にデータ全体を見渡すことができるよう overview マップを用いることが良いとされている。これに関しては様々な研究がなされているが、このうちのいくつかのアルゴリズムによって、データの複雑さや情報スペースの大きさを軽減できることが知られている [37]。これはまた、overview マップにも言えることで、overview マップをうまく生成、利用することで、データの複雑さ、また困難な構造をユーザに対して理解可能な環境を提供できる [4, 58]。

そこで、第2章で述べたように、SOM を用いて、カット映像の動的な分類および組織化の方法を提案した [21]。このシステムは VRML を用いて 3 次元 overview マップ (3D-SOM) を生成しているが、映像の意味・内容的な情報が無視されるという欠点を持っていた。そこで、本章ではカット映像の意味的な情報をそれらの分類や検索に反映できるようシステムの改良を行っている [22]。

本章における要点を以下に挙げる。

- SOM を用いたコンテンツと記述情報を共に用いたカット映像の分類

我々が映像データを扱う場合に考慮すべき点は、ユーザの問い合わせというものは、データのコンテンツ情報を基にしたものというものよりも、むしろ、その意味内容を基にしたものであるということである。つまり、映像データに対して、その意味内容についての記述を施す必要があるということである。また、3.3節で述べたような SOM を用いたカット映像のコンテンツ情報のみによる分類システムにおいては、ある程度の分類精度は出るものの満足いく結果が得ることが出来なかった。このような背景から、映像のコンテンツ情報を基にした分類結果を、映像の内容記述を行う際の補助的な手段として用いることにする。そして、実際にカット映像に対し記述を行い、その記述情報と映像のコンテンツ情報を共に取り入れたハイブリッド型の映像の特徴ベクトルを生成し、それを入力として SOM による学習およびカット映像の分類を行う。さらに、この手法が適切かどうかを検証するためにその評価実験も併せて行う。

- コンテンツと記述情報を基に生成した 3D-SOM による、効果的な内容記述環境の整備

我々が映像データに対して意味情報を与える場合、多かれ少なかれ全ての映像データに対してその内容の記述を施さなくてはならない。しかし、これらの作業は人手に依るところが大きいため映像データ量が増えれば増えるほど、記述を施す手間がかかるという大きな問題がある。そこで、我々のハイブリッド型 3D-SOM のシステムでは、映像データをコンテンツ情報のみで分類した overview マップをあらかじめ生成しておき、そのマップ上において同じセルに分類されたカット映像には同じ記述を施すという方法をとることで、ユーザが全てのカット映像に対して記述を行わなければならないという手間を省き、効率的に記述を行なうことが出来るように工夫した。

- Web-Based SOM ツール

3D-SOM は VRML 形式で生成しているため、VRML ブラウザ上で 3 次元で表示される。VRML とは Web 環境において 3 次元環境を提供するものである。よって、この 3D-SOM を使うユーザはマウスを用いて VRML 空間内を容易にウォークスルーすることができ、またマップ中のセルを選択することでセル内に分類されたカット映像をブラウジングすることができる。この点から、本システムは Web-Based SOM ツールと呼ぶことができる。

本章の構成は以下の通りである。4.2節ではカット映像のコンテンツ情報および記述情報を共に用いたハイブリッド型のカット映像分類システムの実装およびその評価実験について述べる。ここでは、ハイブリッド型 3D-SOM の生成手順およびその機能について説明し、システムの評価実験に対して議論を行う。そして、最後に 4.3節で結論および本システムの問題点について述べる。

4.2 ハイブリッド型 3D-SOM の生成

映像データに対して映像の意味情報を付加する場合、多かれ少なかれ全てのカット映像に対してその内容の記述を施さなくてはならない。しかし、これらの作業は人手に依るところが大きいため、映像データが増えれば増えるほど、記述を施す手間が増えるという大きな問題が生じる。そこで3.3節で述べたように、映像データをカット映像の DCT 係数などのコンテンツ情報によりカット映像を分類した 3D-SOM をあらかじめ生成しておき、これを利用してカット映像に対して記述を行う方法を提案し、ユーザのそのような手間を省き効率的に記述を行うことができるように工夫している。

カット映像のハイブリッド型 3D-SOM を生成するためには、図 4.1 に示すような各手順を順にふんでゆく。

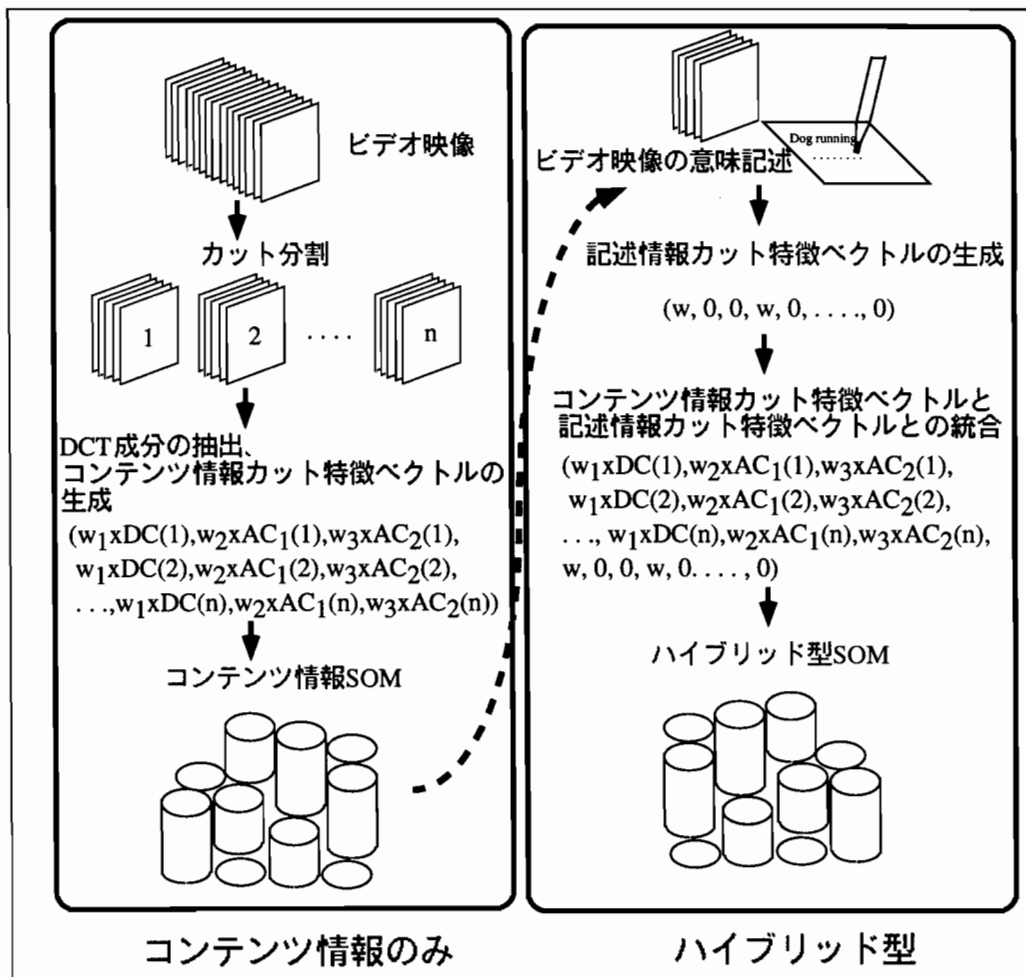


図 4.1: ハイブリッド型 3D-SOM の生成手順

4.2.1 ハイブリッド型カット映像特徴ベクトルの生成

1. コンテンツ情報を用いたカット映像の分類

3.3節で述べたように、まず、カット映像のコンテンツ情報を用いてコンテンツ情報によるカット映像特徴ベクトルを生成する。次に、こうして生成されたカット映像特徴ベクトルを SOM に入力して学習させ、コンテンツ情報 3D-SOM の生成を行う (図4.2参照)。この 3D-SOM はカット映像のコンテンツ情報、すなわち、映像の色情報や映っている物体の形などのよって分類されている。

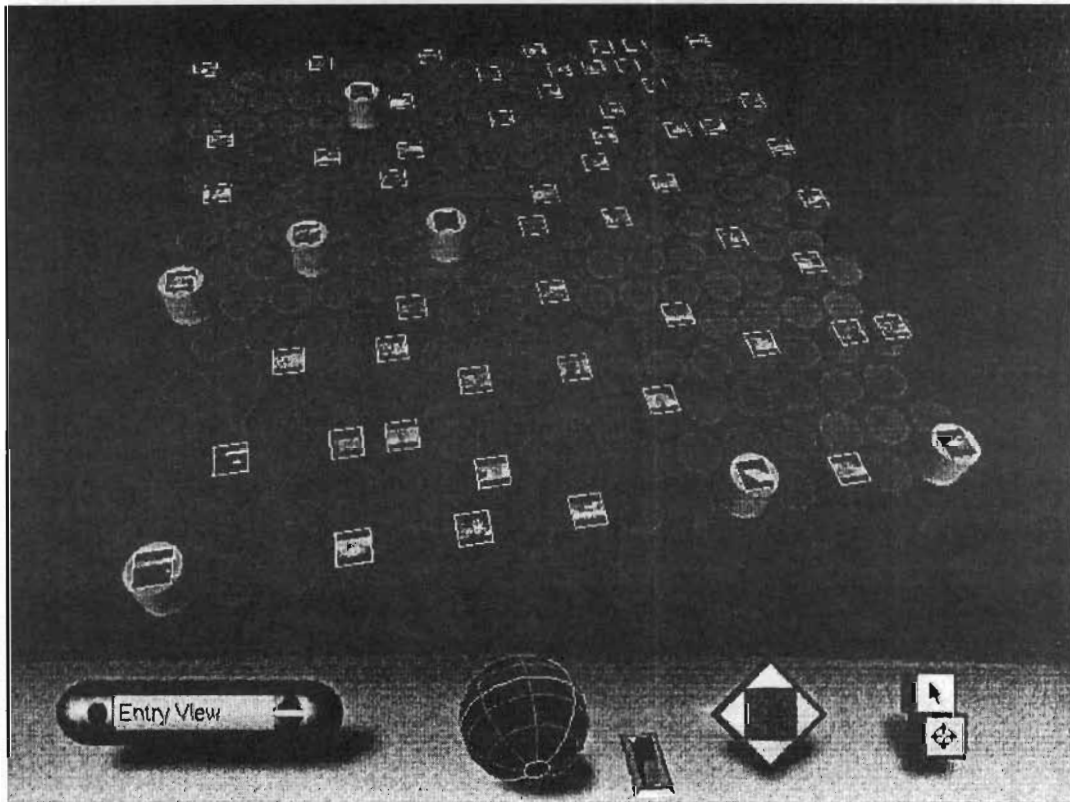


図 4.2: コンテンツ情報 3D-SOM © 東映株式会社

2. カット映像に対する意味記述

コンテンツ情報 3D-SOM において、同じ円柱 (セル) に分類されたカット映像に対して同じキーワードを与えカット映像に意味情報を付加する。映像データに対する内容記述の作業自体がデータ量が多ければ多いほど大変な作業となるため、コンテンツ情報 3D-SOM を用いたこのような方法は、全てのカット映像に対して内容記述をする手間を省くのに有効である。ただし、当然のことながら、各円柱に分類されたカット映像にはノイズが含まれている

場合があり、与えたキーワードに当てはまらないものが存在する可能性がある。本システムの実装においては、こういったノイズは手動で取り除くことにしている。

3. 記述情報カット映像特徴ベクトルの生成

カット映像に対して付加されたキーワード情報を基に記述情報カット映像特徴ベクトルを生成する。ここでは、カット映像 M_i の記述情報特徴ベクトルを以下のように定義する。

$$\mathbf{V}_{keyword}(M_i) = (k_1^i, k_2^i, \dots, k_n^i) \quad (4.1)$$

ここで、

$$k_j^i = \begin{cases} 0 & (j = 1, 2, \dots, n) \\ 1 & \end{cases}$$

ただし、全てのカット映像の内容記述の際に用いたキーワード集合を $\{K_1, K_2, \dots, K_n\}$ ($i \geq 1$) とする。

4. ハイブリッド型カット映像特徴ベクトルの生成

ハイブリッド型のカット映像特徴ベクトルを生成するために、コンテンツ情報 3D-SOM の生成時に利用したコンテンツ情報カット映像特徴ベクトルと記述情報カット映像特徴ベクトルを統合する。3.5式で定義したコンテンツ情報カット映像特徴ベクトルを、

$$\mathbf{V}_{DCT}(M_i) = (c_1^i, c_2^i, \dots, c_m^i) \quad (4.2)$$

のように与えた場合、ハイブリッド型カット映像特徴ベクトル $\mathbf{V}_{hybrid}(M_i)$ を、4.1式および4.2式を結合した以下のような形で定義する。

$$\mathbf{V}_{hybrid}(M_i) = (c_1^i, \dots, c_m^i, wk_1^i, \dots, wk_n^i) \quad (4.3)$$

ここで、 w は正規化のために用いられる重みである。

4.2.2 ハイブリッド型 3D-SOM の生成

得られたハイブリッド型カット映像特徴ベクトルを SOM の入力として学習を行い、ハイブリッド型 3D-SOM の生成を行う。SOM の学習部分は Kohonen の研究グループが開発した SOM.PAK[29]

を用い、その出力結果を VRML 形式に表示させることで、カット映像の分類およびブラウジングが可能となっている。これらのソフトウェアは全て米・シリコングラフィックス社の Indy および Indigo2 IMPACT ワークステーション上で実装されている。図 4.3 はカット映像をコンテンツ情報と後に記述された内容記述情報を共に利用して分類した分類結果であり、VRML ブラウザによりブラウズされユーザが分類されたカット映像を閲覧することができる。

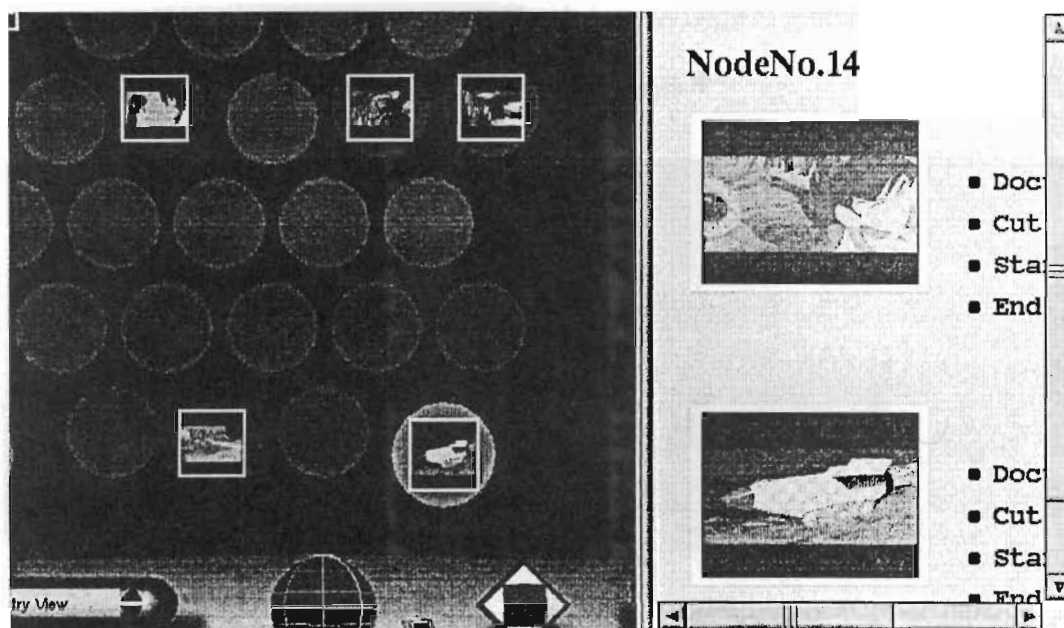


図 4.3: ハイブリッド型 3D-SOM とカット映像の分類結果の閲覧 © 東映株式会社

4.2.3 本システムの評価実験

4.2.1 小節で述べた方法により生成されたハイブリッド型カット映像特徴ベクトルを SOM を用いて学習させ、その分類結果を VRML によって出力した。分類に使用した映像データは 18,999 枚のフレーム画像を持つ 15 分のテレビアニメーションであり、龍谷大学の有木研究室により開発されたカット検出ソフトウェア [61] を利用して映像をカット分割すると、125 個のカット映像に分割された。また、同時に基底数が 144 のコンテンツ情報カット映像特徴ベクトルと基底数 29 の記述情報カット映像特徴ベクトルを生成し、マップサイズが 15×15 のマップを用いてハイブリッド型 3D-SOM の生成を行った。この時、学習回数は 12,000 回で、セル数 225 個のマップは約 1 分で生成された。ハイブリッド型 3D-SOM の円柱の高さは、その円柱内に分類されているカット映像の数に比例しており、また、円柱の上部に貼られているフレーム画像は、図 4.3 に示されているように、その円柱が持つ特徴ベクトルに最も近いハイブリッド型カット映像特徴ベクトルを持つカット映像の最初のフレーム画像 (代表フレーム画像) である。VRML 空間内にある円

柱をマウスで選択すると、その円柱に分類されているカット映像の一覧が Webブラウザ上に表示される。

ここで得られたハイブリッド型 3D-SOM を評価するため、多少改良を施した適合率、再現率を用いる。あらかじめ、円柱の上部に貼られた各カット映像に類似した特徴を持つカット映像の集合を決めておく。つまり円柱の上に表示されているカット映像 M に対して、我々はこの M に類似したカット映像が得られているとみなし、これを $similar(M)$ と表現することにする。つまり、3.3.3小節で述べたように適合率は、

$$\frac{|neighbour(M) \cap similar(M)|}{|neighbour(M)|} \quad (4.4)$$

また、再現率は、

$$\frac{|neighbour(M) \cap similar(M)|}{|similar(M)|} \quad (4.5)$$

と定義することができる。

さて、我々はこの評価実験に際しいくつかの条件の元で適合率および再現率の計算を行った。具体的には、 $similar(M)$ の値を 2 つの異なる基準で評価を行った。1 つはコンテンツ情報のみを利用した類似度、もう 1 つはコンテンツ情報と記述情報を共に考慮した類似度を基準として $similar(M)$ を求めている。前者の類似度判定では M に類似したカット映像が映像のコンテンツ情報のみから決定されることを意味する。従って、カット映像の意味的な情報は全く考慮されていないといえる。また、後者の類似度判定では $similar(M)$ の値が、映像のコンテンツ情報と意味情報の両方の類似度により算出されている。ただし、 $similar(M)$ は本論文の著者の主観により決定されている。図 4.3はテレビアニメーションを入力データとした場合のハイブリッド型 3D-SOM で、あるセルに分類された 2 つのカット映像が表示されている。つまり、これら 2 つのカット映像はコンテンツ情報と記述情報の両情報から類似していると判断されたものである。具体的には、図 4.3上のカット映像は、上のカット映像は“battle”と“009”，下のカット映像は“battle”と“fighter”というキーワードによって記述されているため、このハイブリッド型 3D-SOM は確かに各カット映像がコンテンツ情報のみではなく記述情報によっても分類されているといえる。

表 4.1、表 4.2では、それぞれニュース映像とテレビアニメーションをコンテンツ情報 3D-SOM を利用してカット映像を分類した場合の適合率と再現率を示している。この場合、類似度判定の基準はコンテンツ情報のみを扱い、各表には近傍距離を中心のセルからの距離が 1 の場合と 2 の場合に関して示した。

コンテンツ情報による分類結果を比較すると、本システムはニュース映像よりもテレビアニメー

近傍距離	適合率 (%)	再現率 (%)
1	53.71	43.52
2	52.17	54.73

表 4.1: コンテンツ情報 3D-SOM のカット映像の分類結果に対する適合率と再現率 (ニュース映像)

近傍距離	適合率 (%)	再現率 (%)
1	73.35	45.08
2	33.75	61.93

表 4.2: コンテンツ情報 3D-SOM のカット映像の分類結果に対する適合率と再現率 (テレビアニメ)

ションを題材として使用した場合のほうが良い結果が得られることがわかる。この理由として考えられるのは、テレビアニメーションの方がニュース映像に比べて映像内で使用されている色数が少なく、カット映像内のオブジェクトの動きもより単純なものであるということが挙げられる。

さらにコンテンツ情報カット映像特徴ベクトルとハイブリッド型カット映像特徴ベクトルを用いた場合において、類似度判定の基準をコンテンツ情報のみを扱った場合の適合率および再現率を表 4.2, 表 4.3 に示す。このとき我々は、 $similar(M)$ の値をカット映像のコンテンツ情報のみで算出しているため、カット映像の内容記述情報はノイズとして扱われる。したがって、ハイブリッド型カット映像特徴ベクトルを用いた場合の適合率、再現率は低くなるという結果が得られている。

近傍距離	適合率 (%)	再現率 (%)
1	64.59	47.67
2	37.95	60.69

表 4.3: コンテンツ情報を分類基準としたハイブリッド型 3D-SOM のカット映像の分類結果に対する適合率と再現率 (テレビアニメ)

また、分類対象をテレビアニメーションとし、コンテンツ情報およびハイブリッド型の 2 種類の 3D-SOM において、類似度判定の基準にコンテンツ情報と記述情報共に利用した場合の適合率、再現率を表 4.4～表 4.6 に示した。表 4.5 と表 4.6 の違いは 4.3 式における重み w の値の違いに

よる.

近傍距離	適合率 (%)	再現率 (%)
1	61.22	41.74
2	39.98	52.36

表 4.4: コンテンツ情報および記述情報によるコンテンツ情報 3D-SOM のカット映像の分類結果に対する適合率と再現率

近傍距離	適合率 (%)	再現率 (%)
1	65.03	47.67
2	37.95	60.69

表 4.5: コンテンツ情報および記述情報によるハイブリッド型 3D-SOM のカット映像の分類結果に対する適合率と再現率 ($w = 100$)

近傍距離	適合率 (%)	再現率 (%)
1	67.82	48.88
2	40.47	65.70

表 4.6: コンテンツ情報および記述情報によるハイブリッド型 3D-SOM のカット映像の分類結果に対する適合率と再現率 ($w = 200$)

これらの表からは、ハイブリッド型カット映像特徴ベクトルを用いる利点として以下のことが挙げられる。

- コンテンツ情報カット映像特徴ベクトルを用いるよりハイブリッド型カット映像特徴ベクトルを用いた方が分類結果の適合率、再現率は良い値となる。つまり、コンテンツ情報による分類にある程度誤差があっても、内容記述情報による分類で分類精度をある程度カバーすることが可能であり、その逆もいえる。
- カット映像の内容記述情報に対する重み w を変化させると適合率、再現率の値が変化するため、 w の最適値が定めればある程度の分類精度が得られる。つまり、ハイブリッド型 3D-SOM を利用する目的であった分類精度の向上に貢献できる。
- ユーザがカット映像の内容で検索する傾向が強い以上、ハイブリッド型 3D-SOM を利用し

てカット映像の記述情報を分類結果に反映可能であれば、従来のようなコンテンツ情報を基にした映像の検索に比べ、検索時における手間や労力が軽減される。

4.3 結言

本章では SOM を用いて、DCT 係数のような映像のコンテンツ情報およびその記述情報を組み合わせたカット映像の自動分類を行うシステムの試作とその評価実験を行った。本章で提案した映像のコンテンツ情報および記述情報を共に用いたこのアプローチ法をハイブリッド型アプローチと呼ぶ。

本章のまとめを以下に挙げる。

- **ハイブリッド型カット映像特徴ベクトルの生成**

ハイブリッド型 3D-SOM を実現するため、DCT 係数と重みつきキーワード情報の組み合わせとなるハイブリッド型カット映像特徴ベクトルを生成する方法を提案し、コンテンツ情報および内容記述情報によるカット映像の分類システムの実現を行い、各々のカット映像特徴ベクトルを利用した分類における欠点を補えるカット映像の分類システムの実現を図った。

- **VRML によるハイブリッド型 3D-SOM の実装**

VRML によりシステムを実装しているため、複雑なデータ構造を 3 次元化することによる視認性の良さ、また Web アプリケーションとして利用できるという汎用性の確保を行っている。

- **映像のオーサリング環境をサポートするハイブリッド型 3D-SOM**

我々が開発したハイブリッド型 3D-SOM は、映像データに対して効果的に内容記述できるように工夫されている。ユーザは 3D-SOM に表示されている代表的なカット映像に対してのみ、その内容記述を与えるだけでよく、全てのカット映像に対してに記述を行う必要がないため、記述者の手間や労力を軽減できる。

- **プロトタイプシステムの実験およびその評価**

今回提案したハイブリッド型 3D-SOM は、従来のコンテンツ情報 3D-SOM よりも分類精度の向上が図られた。つまり、マルチメディアデータの多レベル性を考慮すれば、今後、あらゆる情報検索に必要なアプローチであると考えられる。

今後の課題としては以下の問題が挙げられる。

- 現在の記述情報カット映像特徴ベクトルは、ベクトルの各基底の意味空間が曖昧である。つまり、抽出した単語をそのままベクトルの基底として利用している。こういったベクトルの基底は本来は避けるべきであり、この問題点を改良するためのベクトル生成アルゴリズムの提案を行う。また、同時にカット映像の内容記述として与えられたキーワード間の相関関係(共起)を考慮した記述情報カット映像特徴ベクトルの生成方法の提案を行う。
- ハイブリッド型カット映像特徴ベクトルを生成する際に現在は単純にベクトル同士の結合を行っている。しかし、数学的にはこういった合成の仕方は誤っており、コンテンツ情報カット映像特徴ベクトルと記述情報カット映像特徴ベクトルの数学的統合方法による合成方法の提案を行う [62]。
- 公開型の映像オーサリングシステムの場合、1つの映像に対して複数の人間が異なる記述を行う可能性がある。そうした場合、それら記述内容をどのように統合していくのか、その統合方法の提案を行う。

第 5 章

多段階自己組織化マップによるビデオ映像 記述支援と類似シーン検索

5.1 緒言

WWW の急速な普及と発展，計算機性能の向上，2 次記憶媒体の大容量化などに伴い，文書データのみならず，映像や音声などのマルチメディアデータを取り扱う機会が増加している．そのため，爆発的に増加するマルチメディアデータをデータベース化する技術はますます重要になってきている [12]．しかし，マルチメディアデータ，特に映像データは，その意味情報表現が複雑でありかつ意味情報や構成に関する情報が明示的には含まれていないなどの理由から，映像データの現在の分類方法や自己組織化の技術ではまだまだ不十分な点が残されている．従来より映像データベースの構築には，映像データにキーワードや説明文など索引となる情報を付加し，この 2 次情報を利用して映像の分類や検索を行うといった方法が採られている [25, 52, 57, 66]．この場合，映像の分類体系が静的に規定されていることが多く，これらの分類や 2 次情報の付加は手作業であるため，データベースの構築と維持に巨大なコストがかかるという問題や，分類作業の効果的な支援機能が不十分であるという問題が生じている．

最近，データベース全体の把握や曖昧検索および分類作業の支援などの目的を達成するため，ニューラルネットワーク技術を用いた研究が行われている．我々も，ニューラルネットワークの一種である SOM を用いて文書群の動的な分類，検索システムの開発 [39, 77] やカット映像のコンテンツ情報による分類システムの開発 [22] を行ってきている．このような従来のコンテンツ情報によるカット映像の分類システムは，映像の意味的な情報が無視されるという欠点を持っている．本章では，映像のコンテンツ情報による SOM の機能拡張として，映像データに意味的な情報を効果的に付与できる機能を併せ持った類似シーン検索システムの試作を行ったのでそれにつ

いて報告する。

本章で提案する手法の要点を以下に挙げる。

- **カット映像単位の内容記述**

一般に、映像データは、シーン映像と呼ばれる映像単位 (例えば、ニュース映像で言えば記事と呼ばれるもの) から構成されている。ここでカット映像とはカメラの切り替えや素早いカメラワークなどによって分割される 1 つ 1 つの映像のことをいうが、シーン映像はこのカット映像から構成されている。本章では各カット映像ごとに意味記述を行う方法を採用する。なぜなら、シーン映像に対して意味記述を行う場合、1 つのシーン映像に多くのカット映像が含まれているため記述すべき情報が多く、記述者が映像の内容を容易に記述することができないと考えられるからである。また、逆にフレーム画像に対して意味記述を行うという方法は 1 秒間の映像が普通 30 枚のフレーム画像から構成されていることを考慮すれば非現実的である。このような理由から、本システムにおいて記述される映像の最小単位カット映像と定めた。

- **3D-SOM によるカット映像分類マップによる記述情報の共有 [21]**

第 4 章で述べたように、映像データに対して意味情報を付加する場合は、全てのカット映像に対してその内容の記述を施さなくてはならない。しかし、これらの作業は人手に依るところが大きく、依然、記述作業に手間を要するという大きな問題を生じる。そこで、映像データを映像そのものの色情報や色合いの情報であるコンテンツ情報 (ここでは DCT 係数) で分類した SOM をあらかじめ用意しておき、そのマップ中の同じセルに分類されたカット映像に対しては同じ記述を施すという方法をとることで、ユーザが全てのカット映像に対して意味記述するという労力を省き、効率的に記述が行えるように工夫した。

- **カット映像の意味記述からのシーン映像の意味記述の生成**

一般ユーザが映像データに対して検索を行うことを想定した場合、考えられる検索データの単位はシーン映像である場合が多い。しかし、シーン映像はカット映像に比べ含まれている情報量が多く、それに対して記述を行うには、映像データの内容の知識がある程度必要であるなど、記述者が簡単に意味内容を記述するのは困難である。そこで、描写されている対象物やその動作内容などが少なく比較的記述の行いやずいカット映像に対し意味記述を行い、その記述内容をそのカット映像から構成されているシーン映像に継承させることで、シーンの記述情報が得られる。

- 多段階自己組織化マップ

本章において SOM は、カット映像のコンテンツ情報によるカット映像の分類の際と、シーン映像に対して与えられた意味記述情報によるシーン映像の分類の際の計 2 段階に用いられている。このように、SOM を多段階で使用するという研究は他にも始まっている。Kohonen の研究グループでは、文書データに対しこの多段階 SOM を用いており、出現した単語のカテゴリ分類 [40, 41] に第 1 段階の SOM、カテゴリ分類された単語を基にした文書データの分類に第 2 段階の SOM [24, 30] を用いている。扱うデータの違いがあるにせよ、この多段階 SOM の利用は情報検索の際に有効であるものと考えられる。

本章の構成は以下の通りである。5.2 節においてカット映像に与えられた記述情報をシーン映像に継承させる方法について述べる。ここでは、カット映像に与えられた記述情報をシーン映像に継承させる方法およびその記述情報からシーン映像の特徴ベクトルの生成方法について述べ、実装したプロトタイプシステムの評価実験を行いそれについて議論する。そして、最後に 5.3 節で結論および問題点について議論する。

5.2 シーンへの記述および類似シーン検索

分類対象がシーン映像の場合はカット映像の場合とは異なり、登場人物や場面のストーリーなど含まれている情報が多いため、シーン映像に対して簡単な内容記述を行う際には何らかの工夫を施す必要があると考えられる。

そこで、3.3 節に示したように、最初にカット映像をコンテンツ情報や意味情報を利用して分類した 3D-SOM をあらかじめ生成しておき、3D-SOM の各セル、すなわち、そのセル中に分類されているカット映像に対して与えられたキーワードを利用する方法を提案する。具体的には、シーン映像に含まれているカット映像を求めておき、そのカット映像に対して与えられている記述情報をそのままシーン映像の記述として継承する方法である。このとき、シーン映像の長さに対するカット映像の長さを考慮した重みづけを行う必要がある。

5.2.1 シーン特徴ベクトルの生成

以下にシーン映像に対する特徴ベクトルの生成法を示す。

シーン映像はカット映像とは異なり、その映像における話の流れ、場面の展開などから決定されるものである。したがって、カット映像のようにコンテンツ情報の変化から自動的にシーンを検出することは困難である。よって、ここではシーン映像の分割は手動で行うことにした。

また分割された各シーン映像に対して、記述者が直接キーワードを付与することも考えられるが、一般に 1 つのシーンには複数の人物、物体やその動作の情報が含まれており、これらの情報

をその場で理解し、記述を行うことは素人では困難であると考えられる。これに対して、カット映像に描写されている対象物や動作はシーン映像の場合と比較すると比較的少ないため、シーン映像に対して直接記述を行う代わりにカット映像に対して記述を行うことにした。さらに、そのカット映像に与えられた記述情報をそのカット映像が属しているシーンに継承させてシーン映像の記述情報として利用する。

シーン映像の特徴ベクトルを生成する際の手順を以下に示す。

1. カット映像の 3D-SOM 生成

3.3節で述べたように、まず、カット映像をコンテンツ情報を用いてコンテンツ情報によるカット映像特徴ベクトルを生成する。次に、こうして生成されたカット映像特徴ベクトルを SOM に入力して学習させ、コンテンツ情報 3D-SOM の生成を行う (図5.1参照)。この 3D-SOM はカット映像のコンテンツ情報、すなわち、映像の色情報や映っている物体の形などのよって分類されている。

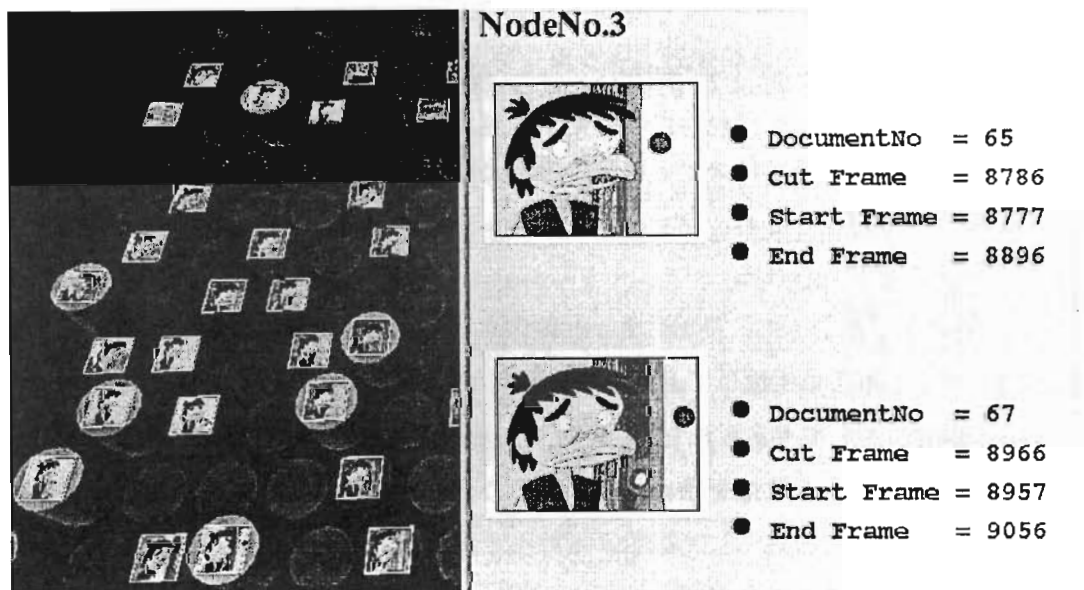


図 5.1: コンテンツ情報 3D-SOM © 愛企画センター

2. カット映像へのキーワード付与による内容記述

映像データに対してその内容の記述を施すという作業は、自動化しにくいと記述者に非常に大きな手間を要するという問題が生じる。そこで、その手間を最小限に押さえるために、ここでは次のような 2 つの方式を提案する。

- カット映像のコンテンツ情報 3D-SOM 上での記述

カット映像とは、場面の変わり目で切られる映像の1つの単位であるので、映像に映っているものに動きがあったとしても、映っている物体はカット映像内ではさほど変わらないと考えられる。つまり、3.3節で述べたように、コンテンツ情報の類似度が高ければ、記述する情報も類似性があると考えられることができるため、あらかじめコンテンツ情報 3D-SOM を生成しておき、そのマップ上において同じセルに分類されたカット映像に対して同じ記述を施すという、コンテンツ情報 3D-SOM を用いてカット映像に内容記述を行う方法である。これによって、記述者が全てのカット映像に対してその映像の意味内容を記述しなければならないという労力を省くことができると思われる。また、コンテンツ情報では類似しているが、意味記述が異なるカット映像も存在するため、意味記述が異なるカット映像は手動で取り除くことにしている¹。

内容記述はキーワードによって行うが、これをシーンへ継承する際に適切なキーワードとして、登場人物や場所などが挙げられる。

- ハイブリッド型 3D-SOM 上での記述

我々が映像データの分類システムを構築する場合に考慮した点は、ユーザが映像に対する問い合わせを行う場合は、データのコンテンツ情報を基にした分類よりも、その映像の意味内容を基にした分類が必要であるという点である。こうして、開発されたカット映像の分類システムが4章で述べたハイブリッド型 3D-SOM である。ハイブリッド型 3D-SOM は、4.2.1小節で述べた4.1式で与えられる記述情報カット映像特徴ベクトル $V_{keyword}(M_i)$ と4.2式によって与えられたコンテンツ情報カット映像特徴ベクトル $V_{DCT}(M_i)$ とを結合したハイブリッド型カット映像特徴ベクトル

$$V_{hybrid}(M_i) = (c_1^i, \dots, c_n^i, w \times k_1^i, \dots, w \times k_h^i) \quad (5.1)$$

を生成し、これを SOM に学習させることで生成されるマップである (生成手順は図4.1を参照)。このマップは、カット映像のコンテンツ情報と人手により記述されたキーワードによる意味情報を用いてカット映像を分類しており、分類精度はコンテンツ情報のみの分類システムと比較しても、比較的よい結果が得られている。

このハイブリッド型 3D-SOM 自身もまた多段階 SOM と位置づけられ、コンテンツ情報によるカット映像の分類を第1段階に、コンテンツ情報と意味情報によるカット映像

¹ただし、扱うデータによっては取り除くべきカット映像の数も多くなると考えられるが、第一段階の SOM によるカット映像の分類の時点でクラスタリングの精度を上げれば対処できると考えられる。例えば、DCT 係数だけでなく他のコンテンツ情報 (カラーヒストグラムなど) を組み合わせるなど。

の分類を第 2 段階に SOM を利用している。こうしてできたハイブリッド型 3D-SOM を用いて、コンテンツ情報 3D-SOM と同様にカット映像に対して記述を行えば、計 3 段階の SOM の学習を行うことになる。

3. シーンに含まれるカット映像を求める

話の流れや場面展開、登場人物の変化などを考慮し手作業でシーン映像の分割を手動で行っている。このシーン映像の分割は、本研究では本論文の著者により主観的に行われている。また、後にカット映像の記述情報をシーン映像に継承させるために、分割された各シーン映像の中に含まれている複数のカット映像もあらかじめ求めておく。

4. カット映像の記述情報をシーン映像に継承させる

各カット映像に付与された映像の意味記述情報を基に、シーン映像の特徴ベクトルを生成する。以下に本システムの実装において用いたシーン映像 $S_i = \{M_1^i, M_2^i, \dots, M_l^i\}$ の特徴ベクトルの定義を示す。

$$\mathbf{V}(S_i) = \sum_{M_j^i \in S_i} \frac{n(M_j^i)}{N} \mathbf{V}_{keyword}(M_j^i) \quad (5.2)$$

ただし、 $\mathbf{V}_{keyword}(M_j^i)$ はカット映像 M_j^i の記述情報カット映像特徴ベクトル、 N はそのシーン映像の全フレーム数、 $n(M_j^i)$ はカット映像 M_j^i に含まれるフレーム数とする。

ここでは、シーン映像のフレーム数に対するカット映像のフレーム数の割合を算出し、その値をカット映像キーワード特徴ベクトルへの重みとして加えている。これは、各カット映像の時間的な長さが互いに異なることから、カット映像の持つ記述情報をシーンに継承させる際に、キーワード情報に対してこの映像の長さ按比例しただけの重みを与えることで正規化処理を行うためである。これから、シーンに含まれている各カット映像に対応する重みづけ済みの各カット映像キーワード特徴ベクトルの和をとることでそのシーンの特徴ベクトルの生成を行っている。

5.2.2 シーン映像 3D-SOM の生成

このようにして得られたシーン特徴ベクトルを入力ベクトルとして再度 SOM に学習させ、この結果をマップ表示させる。ここでは、カット映像の分類の際の SOM の学習を第 1 段階、シーン映像の記述情報による分類の際の SOM の学習を第 2 段階の計 2 段階の多段階 SOM を使用している。

5.2.3 本システムの実装・評価実験

生成されたシーン特徴ベクトルを入力として SOM による学習を行い、その結果を VRML によって出力する。出力結果を表示する際には、 6×6 のマップを用いた。また、今回、実験において用いた映像データは、人間に助けられた動物が恩返しをするというアニメーション映像 (約 30 分) であったが、その映像のカット映像の総数は 250 個、さらにシーン分割を手作業で行った結果、検出されたシーン映像の総数は 22 個であった。また、1 シーン当りの平均カット数は 11.36 であり、意味記述が異なるという理由でノイズとされたカット映像は 6 個であった。セル数 36 個のマップは学習回数 15,000 回の場合、約 5 秒で生成された。分類結果を図 5.2、図 5.3 に示す。このマップ生成においては、3.2 節で述べた文書分類・検索システムで用いたマップ生成プログラムを使用した。

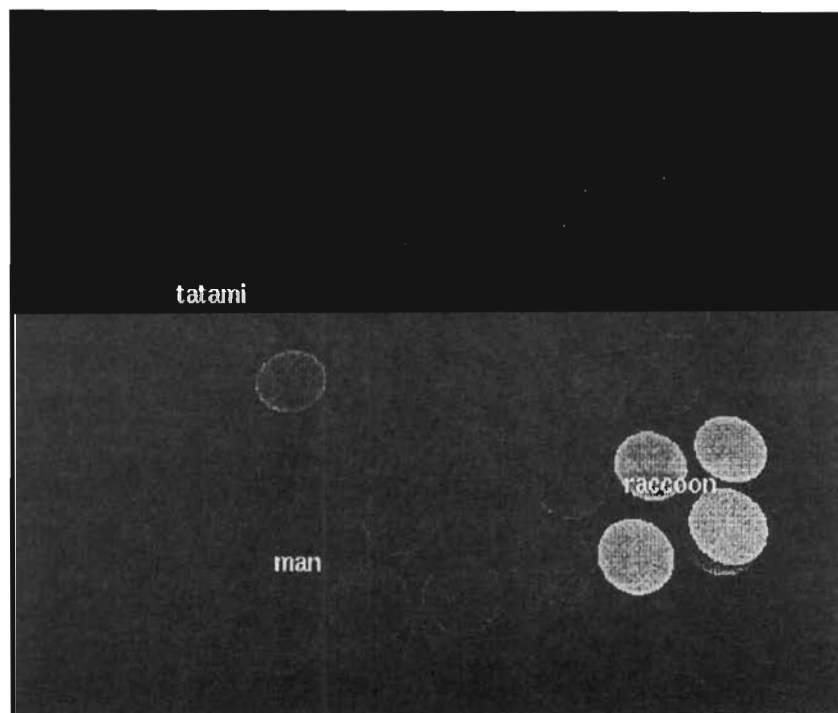


図 5.2: シーン映像 3D-SOM

図 5.2 上で “man” という領域内の高さを持ったあるセルを選択すれば、そのセル内に分類されているシーン映像がブラウザ上に現れるようになっている。

また、“man” 自身を選択すれば図 5.3 のように、1 つの領域がさらに細かい複数の領域に分けられた詳細図が現れる (3D-SOM の詳細化機能)。

この映像から得られたシーンの分類状況を見てみると、例えば、人および動物と一緒に登場しているシーンが多かったが、そういったシーンはマップ上で互いに近くに集まっている様子が目

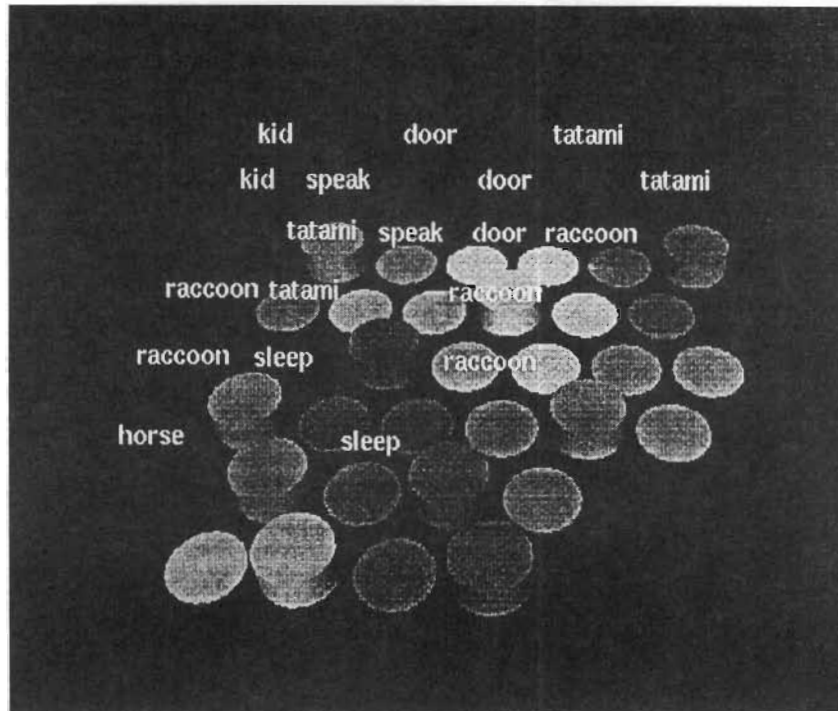


図 5.3: 3D-SOM の詳細化機能

について。

そこで、本システムがどれほどの分類精度を持っているかを定量的に判断するために、適合率と再現率を用いて評価を行う。本実験で利用した適合率および再現率の定義は、2.4節で述べたような定義とは若干異なり、以下のように定義される。

すべてのカット映像の中から、マップ上の円柱の上に張りつけられたあるカット映像 M_i に類似しているカット映像の数を $similar(M_i)$ とし、カット映像 M_i が張りつけられている円柱付近にあるカット映像の数を $neighbour(M_i)$ で表すものとする。このとき、適合率は、

$$\frac{|neighbour(M_i) \cap similar(M_i)|}{|neighbour(M_i)|} \quad (5.3)$$

と定義され、また再現率は、

$$\frac{|neighbour(M_i) \cap similar(M_i)|}{|similar(M_i)|} \quad (5.4)$$

で定義される。

表 5.1 は、 $neighbour(M_i)$ に含まれる範囲が中心の円柱から近傍距離 1 の場合と近傍距離 2 の

場合の適合率と再現率である。ただし、コンテンツ情報 3D-SOM を生成する際に利用するカット映像特徴ベクトル (3.5式参照) で定義する重み w_1 , w_2 , w_3 の値はそれぞれ 1, 10, 20 を用いている [64]。またこの時、22 個あるシーンは 8 つの類似シーンとして分類されている。

近傍距離	適合率 (%)	再現率 (%)
1	85.00	37.50
2	66.50	62.50

表 5.1: シーン映像 3D-SOM の適合率と再現率

表 5.1を見ると適合率、再現率ともに 40 ~ 80% 前後と分類結果としては十分満足のいく結果であると思われる。しかし、カット映像およびシーン映像の類似判定は、分類を行う人物の主観により異なるため、その評価も異なるという問題点を考える必要がある。よって、この評価実験の結果から一概に本システムが有効であるということは言い難い。また、本システムでは、多段階 SOM を利用するため、1 つのデータに対して数種類の特徴ベクトルを生成する。この際、ベクトルの各成分に対する重みづけや統合のパラメータをうまく設定できれば、本システムの性能が上昇すると考えられる。ただし、これらのパラメータは経験的に定まるものであるため、最適なパラメータを設定可能とするために更なる評価実験を行う必要がある。

映像データに対する意味記述にあたっては、本評価実験では映像に登場してくる人物の名前、物体、およびそれらのとりうる動作や状態を表わす単語を付与した。この意味記述を利用したシーン映像の分類結果を見る限りでは、我々が行った意味記述情報を反映してある程度クラスタリングされていることが視覚的にもまた実験評価表 5.1を見てもある程度理解できる。しかし、シーン映像の総数が 22 個と評価実験を行うには極端に少ないことや、記述者の主観が多少影響していることなどから、中にはそれほど類似性の高くないデータが分類されている例も見受けられた。また、シーン映像の分割は話の流れや場面の展開をもとに我々の判断によって行ったため、各シーンを構成するカット数は当然ながらまちまちになっている。シーン映像の意味記述をどの程度のカット数から構成するかという問題は非常に重要であるが、もしシーン映像を構成するカット数に制限を設けてしまえば意味的なつながりを持つにもかかわらず異なるシーン映像に分類されてしまうカットが生じる可能性も考えられるため、今後は、このような不都合をいかにして解消していくのか検討していく必要があると思われる。

5.3 結言

我々は、SOM を用いたカット映像のコンテンツ情報またはハイブリッド情報による分類システムをシーンの意味記述支援ツールとして応用する手法を提案した。また、カット映像に与えられた記述情報をシーン映像に継承させてシーン映像特徴ベクトルを生成し、これを再度 SOM に学習させるという 2 段階 SOM を用いたシーン映像の分類システムの提案およびそのプロトタイプシステムの試作を行い、そのシステムの実現およびその評価実験を行った。

本章のまとめを以下に挙げる。

- シーンの意味記述支援ツールとしてカット映像の分類結果を利用

全てのカット映像に対して人が手作業で記述する手間を軽減するために、SOM を利用したカット映像のコンテンツ情報またはハイブリッド情報を基に生成したマップを生成しシーン映像の意味記述支援ツールとして用いた。ここでは、生成したマップにおいて同じセルに分類されたカット映像に対しては同じ記述を与え、全てのカット映像に対し記述を行うという手間を省いている。

- カットに対する記述情報のシーンへの継承

一般にシーンには複数の人物、物、出来事等が描写されているため、これらの情報を一度に把握、記述することはかなり困難な作業となり、ユーザへの負担も大きくなる。そこで意味記述はカット映像単位で行い、その記述情報をそれらカット映像により構成されているシーン映像に継承させることで、シーン映像への意味記述を行った。

- VRML による 3D-SOM の実装

VRML によりシステムを実装しているので、Web アプリケーションとして利用できるという汎用性の確保を行った。

- 実験・評価および考察

今回試作したシステムによって、実際にカット映像への内容記述とそれを利用したシーン映像に対する内容記述およびその記述情報を基にしたシーン分類を行い、その考察を行った。

今後の課題としては以下の問題が挙げられる。

- 与えられたキーワード間の相関関係や共起関係を考慮した、カット映像キーワード特徴ベクトルの生成法を確立する。
- 手動で行っているシーン映像の分割過程を、映像の意味情報を含めた自動分割システムにより行う。

- 内容記述をより効率的に行うためのインタフェースの開発や内容記述方法の提案を行う。
- 1つのカット映像に対して複数の人間が異なる記述を行った場合、その記述情報を1つに統合する方法の提案を行う。
- システムにより生成された3D-SOMに対してユーザがインタラクションを行い、ユーザの分類目的に合った3D-SOMを生成するフィードバック機能の追加を行う。

第 6 章

自己組織化マップと検索エンジンを用いた Web 文書の分類ビュー機構

6.1 緒言

WWW の発展は我々に多様な情報を提供しており、近年の情報化社会においては必要不可欠なものとなってきている。しかし、WWW で公開される情報量は膨大であり、このため我々が必要とする情報を効率よく検索することを困難にしている。

通常は必要とする情報を収集するために AltaVista[2] に代表されるロボット型の検索エンジンや yahoo![55] に代表されるディレクトリ型検索エンジンを利用してユーザがキーワードを入力し、現在数千万ともいわれる Web 文書に対してフィルタリングを行うことで必要な情報に該当するページを得ている。この手法は WWW と DBMS を連動させたシステムの一つであり、膨大な量の情報に対して問い合わせを行い、その結果を表示することができる点で評価できる。しかし、この手法を利用しても問い合わせの結果は依然膨大であり、結果的に検索結果を 1 つ 1 つ閲覧していかなばならずユーザにとって大変な手間と労力を要する。さらに、WWW は、膨大な量の情報資源からなるハイパーテキスト構造のネットワークであり、それぞれの情報資源はリンクにより結び付けられ、HTML により半構造化されている。つまり、情報提供者がある話題について Web 文書を作成する際に、リンクでつながった複数の文書で表現する場合も考えられる。このとき、検索エンジンは、通常、各文書ごとにキーワード索引を生成するために、ユーザの問い合わせキーワード全てにマッチする文書のみを答えとして返すため、そういった文書群は現在の検索エンジンの問い合わせ結果には現れない場合が多い。

これらの問題点を解決するために、従来から Web 文書群全体の把握や曖昧検索、分類作業の支援などを目的として、Web 文書群の集約的な構造を把握する研究 [5, 6, 19, 37] や大量の文書群

の自動分類等にニューラルネットワーク技術を採用入るといった研究 [22, 24, 30, 77] が行われている。しかし、いずれの場合においても収集した文書群の分類にユーザの視点や目的を反映させるような機構や、分類する文書の単位を選択できるような機構については十分な研究がなされているとは言い難い。

そこで、本章ではロボット型検索エンジンと 3.2 節で述べた SOM を用いた文書群の動的な分類・検索システムを統合・拡張し、ユーザの多様な検索目的を支援した Web 文書の分類ビュー機構を提案する。ここで提案する手法は、ユーザの多様な分類・検索意図をインタラクティブに支援するユーザビューを利用し、先に述べた問題点を解決するものである。ここで挙げるユーザビューとは以下のようなものである。

- ユーザの視点を取り入れた Web 文書の分類 [59, 63]

3.2 節で述べた文書の分類・検索システムには、ユーザ側の意図や興味を分類マップに反映することができないという問題点があった。そこで、ユーザ自身が分類結果を把握しやすいように、システムが返す分類結果に対してフィードバックを行い、ユーザの視点を分類結果に反映させることでユーザ独自のビューを生成する。

- 分類対象である Web 文書の基本単位の決定 [60]

検索エンジンを利用した Web 文書の検索は、ユーザがあらかじめ明確な検索目標を持っているわけではないため、関連のある Web 文書群を検索の基本単位と考えたほうがよい場合がある。そこで、Web 文書群を一種のグラフ構造とみなして [49]、問い合わせに対する Web 文書の極小部分グラフを生成し、これを基本単位として Web 文書の分類・検索を行う方式を提案する。この場合も、Web 文書の情報単位をユーザ自身が選択できるという点で、一種のユーザビューと考えることができる。

本章の構成は以下の通りである。6.2 節では、ユーザのインタラクションを反映するために採り入れた伝統的手法である適合フィードバックについて簡単に述べる。次に 6.3 節では Web 文書を分類するために利用する特徴ベクトルの生成法について、6.4 節ではユーザとの対話による 3D-SOM の分類結果に対する段階的修正について、6.5 節では本システムの実装およびその性能評価について説明および議論を行う。さらに、6.6 節では、Web 文書の意味情報単位として我々が提案する極小部分マッチンググラフについての説明とその予備実験を行い、6.7 節で本章の結論および問題点について述べる。

6.2 適合フィードバック (Relevance Feedback)

情報検索システムを構築する上での大きな問題の 1 つは、ユーザがどのような検索意図を持っているときにどのような問い合わせがなされるのかをあらかじめ予測できないため、ユーザが入力した問い合わせに対して検索意図に沿った検索結果が返ってくる保証がないという点である。

このような状況を解決するには、システムがユーザの視点や分類目的をインタラクティブに支援すべきである。このインタラクティブな情報検索の伝統的な手法の 1 つに適合フィードバック (Relevance Feedback)[42] という手法が存在する。この手法は初期の検索結果に対してユーザが適合または不適合と判断した検索結果を基にシステムが新たな問い合わせを生成し、検索結果にフィードバックをかけながら問い合わせを繰り返すことで、徐々にユーザの情報要求に近づけていくというものである。近年の検索エンジンにはこのような機能が付加されたものも存在する [13, 14].

本章におけるフィードバックとは、伝統的な適合フィードバックのような検索結果に対してのフィードバックではない。むしろ検索エンジンの検索結果を自動分類した 3D-SOM の各領域に対するフィードバックのことであり、3D-SOM の Web 文書の分類結果に対してユーザの視点や目的をインタラクティブに反映させるために利用している。

6.3 Web 文書の分類のための情報単位と特徴ベクトル生成

3.2節で述べたように、我々は従来より SOM を文書データベースのブラウジングツールおよび問い合わせインタフェースと見なして、文書の自動的な構造化や分類を支援するシステムを開発してきた。しかし、このシステムでは、

- マップ上のキーワード同士の関連を見ながらの検索だけでは、目的の情報にたどり着くまでに時間がかかる。
- ユーザ側の意図や興味をマップに反映することができないため、ユーザ本位のインタフェースを提供できない。

といった問題点があった。

このような問題点を解決するために、システムに対してユーザのインタラクションを取り入れる研究が始まっている。Kohonen の研究グループでは、文書中に現れる単語のカテゴリ分類を行った上で文書の分類に SOM を利用し、カテゴリ分類の際に単語を選択したり、データをブラウズする際にマップ上でデータの詳細表示を行うといったユーザの操作がシステムに反映されるような工夫がなされている [24, 30]。また、江口らは伝統的な適合フィードバックを利用し、ユーザ

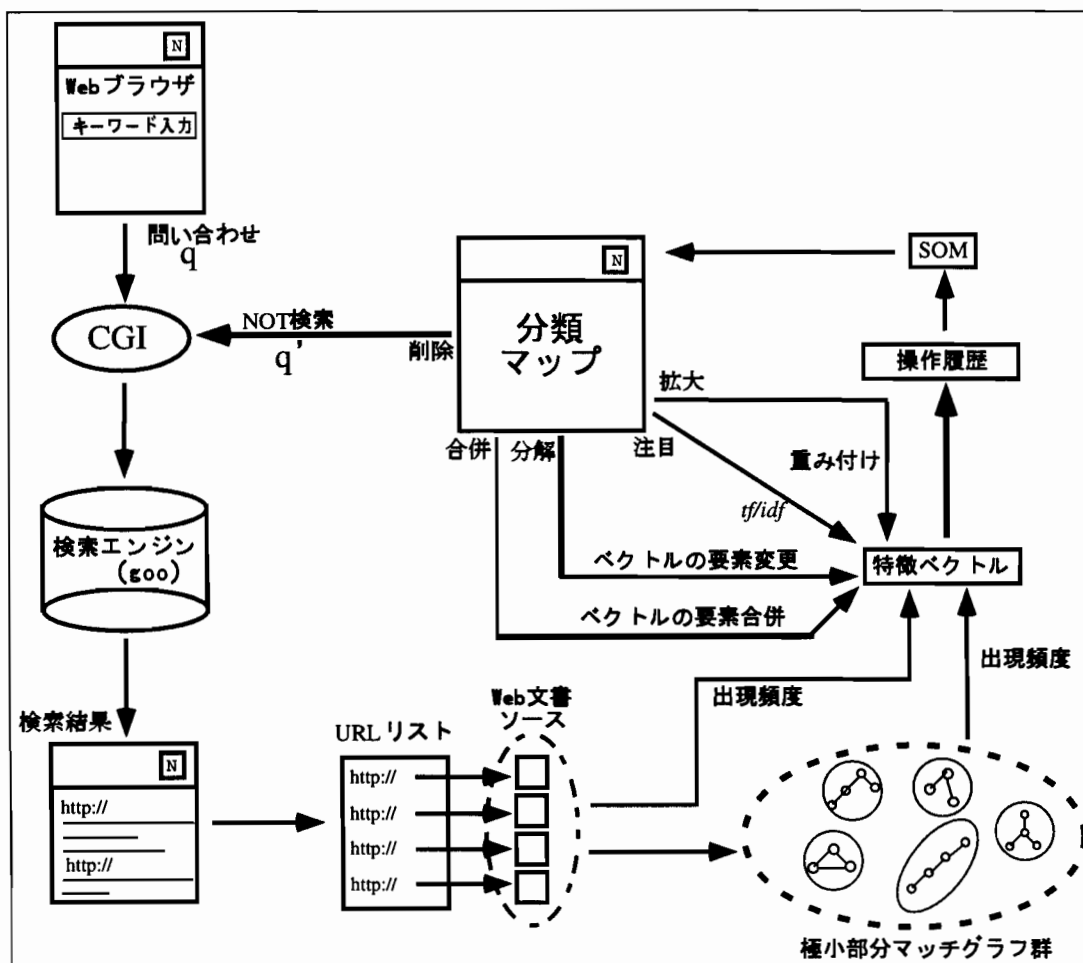


図 6.1: システムの全体構成

の興味の変化を検出しシステムがユーザの検索目標に適応できるような方式を提案している [65]. さらに Xerox Palo Alto Research Center においても収集した情報に対してユーザのインタラクションを分類に反映することができる Scatter/Gather を開発している [10, 23].

これに対し、我々はこれらの研究では行われていない収集した情報の分類ビュー、すなわち 3D-SOM を利用して情報の分類傾向を把握しながら、3D-SOM に対してユーザのインタラクションをサポートする機能を実現することで問題点を解決する。すなわち分類・検索対象となる Web 文書の情報単位の決定やシステムによって返された分類結果に対して、ユーザの視点や目的をシステムにインタラクティブに反映させる機能を持った Web 文書分類ビュー機構を実現する。

本システムの全体構成を図 6.1 に示す。

6.3.1 Web 文書の特徴ベクトルの生成

検索エンジンの検索結果に対して SOM を利用した情報の組織化を行うには、検索エンジンに対する問い合わせの結果から各 Web 文書の URL を抜き出し、それにしたがって Web 文書自体を解析しその特徴を抽出しなければならない。本システムでは、検索エンジンによって検索された各 Web 文書に対して単語抽出を行い、これを基に各 Web 文書の特徴ベクトルを構成する。すなわち、特徴ベクトルの基底は Web 文書から抽出された単語であり、また、ベクトルの各成分はそれらの重要度を表す。

以下に、Web 文書の特徴ベクトルの生成過程を示す。

1. 検索エンジンの検索結果の獲得

通常検索エンジンで検索を行う場合と同様にユーザは Web ブラウザ上でキーワード k_1, k_2, \dots, k_m を入力する。つまり、検索エンジンに対して検索質問 $q = k_1 \wedge k_2 \wedge \dots \wedge k_m$ ($m \geq 1$) を問い合わせると、検索結果である URL が返され、それを利用して検索された Web 文書のソースを得る。

2. Web 文書からの単語の抽出

得られた Web 文書の HTML ソースから単語を抽出しそれぞれの総出現頻度を求める。このとき、システムの仕様上、総出現頻度の高いものから順に上位 500 の単語 w_1, \dots, w_{500} を特徴ベクトルの基底とする。

3. Web 文書の特徴ベクトルの生成

Web 文書 D_i における各基底 w_1, \dots, w_{500} の重要度を求め、各文書の特徴ベクトル $F(D_i)$ を生成する。

$$F(D_i) = \left(\frac{f_1^i}{N_i}, \frac{f_2^i}{N_i}, \dots, \frac{f_{500}^i}{N_i} \right) \quad (6.1)$$

ここで、 $f_j^i (j = 1, \dots, 500)$ は Web 文書 D_i における単語 w_j の出現頻度、 N_i は D_i 内に現れる単語の総出現頻度を表す。

6.4 ユーザとの対話による分類マップの段階的修正

SOM は、入力データ群をトポロジカルマッピングの性質により 2 次元空間に表示することができるため、overview マップを生成するには大変有効な手段である。そこで、我々は SOM に VRML ベースの 3 次元インタフェースを適用した 3D-SOM を提案・実装しており、ユーザが

マップ中の円柱 (SOM におけるセルを表現) を選択することでその円柱に分類された Web 文書の URL と円柱が持っている重要な 3 つのラベル (第 1 キーワード, 第 2 キーワード, 第 3 キーワード) を Web ブラウザで参照できるように設計している。また, 分類結果に対して適合フィードバックの機能を実現することで, 検索された Web 文書の分類結果をユーザ自身でインタラクティブに変更可能となっているため, Web 文書の分類状況を把握しやすいようになっている。

以下にユーザインタフェースにおける各機能を説明する。

- **マップの 3 次元表示**

3D-SOM の出力セルは円柱になっており, その高さはそのセルに分類された Web 文書の数に比例している。したがって, 3D-SOM の外観から分類されている Web 文書の数を把握できる。

- **各出力セルのキーワード**

各出力セルの特徴ベクトルの成分中, 最も大きな値を持つ成分, すなわちその出力セルのベクトルの成分の中で最も値の大きな単語をそのセルのキーワードとする。同様に, 成分が 2 番目に高いもの, 3 番目に高いものを選び各セルを 3 つのキーワードでラベルづけを行う。以下, これらのキーワードをそれぞれ第 1 キーワード, 第 2 キーワード, 第 3 キーワードと呼ぶ。

- **3D-SOM の領域分割**

生成された 3D-SOM は, 各セルにつけられた第 1 キーワードを利用して色分けされる。SOM の性質上, 第 1 キーワードが等しいセルはマップ中の近い位置に配置されるため, 3D-SOM は第 1 キーワードによる領域形成がなされる。

- **3D-SOM の詳細表示**

第 1 キーワードによって色分けされた領域を選択すると, その領域をさらに細かく表現した詳細マップが表示される。この詳細マップは第 2 キーワードを利用して色分けされたものであり, 関連した Web 文書を検索する場合に利用できる。同様に, この詳細マップをさらに詳細に見ることも可能である。

- **適合フィードバック機構 [59]**

得られた分類マップに対して, ユーザの意図や興味を反映させる機能は, ユーザが欲している情報をフィードバックを行うことで絞り込み, 効率的に検索する上で大変重要な機能である。Scatter/Gather[10, 23]においてはこの機能を利用して検索の高速化に役立てている。本システムにおいてもこれにならない, 適合フィードバック機構を実現するために以下の機能

を付加している。ただし、ここでは特徴ベクトルを Web 文書単体から生成される $F(D_i)$ として扱っているが、6.6節で後述するように極小部分マッチグラフと呼ばれる Web 文書の意味情報の単位から生成される特徴ベクトルも利用できる。

1. 領域の拡大

ユーザがある領域 K に関する Web 文書を欲している場合その領域を指定すると、その領域内の Web 文書のうち特徴ベクトルの成分 $\frac{f_j^i}{N_i}$ が最も大きいものを探し出し、その文書が分類されているセル c を中心にすべての $F(D_i)$ の成分 $\frac{f_j^i}{N_i}$ に対して同心円状に重み付けの再計算を行うものである。具体的には Web 文書 D_i の特徴ベクトルは以下のように再定義される。

$$F(D_i) = \left(\frac{f_1^i}{N_i}, \dots, \alpha(r) \cdot \frac{f_j^i}{N_i}, \dots, \frac{f_{500}^i}{N_i} \right) \quad (6.2)$$

$$\alpha(r) = \frac{\alpha(r-1) + 1}{2}, \quad r \geq 1, \quad \alpha(0) = 5$$

ここで、 r はセル c から D_i が含まれているセルまでの距離 (半径) を表す。この特徴ベクトルを用いて再度 SOM によって学習させることで、キーワード K の重要度が高い Web 文書を 3D-SOM 上に多く浮かび上がらせることができる。

2. 領域の削除

ユーザが 3D-SOM 上のある領域 K は全く必要ないと判断した場合、その領域を指定するとキーワード K を含んだ Web 文書自体を検索対象から削除する、すなわちユーザビューが検索エンジンに対して NOT 検索を行う機能を実現している。この場合、ユーザは意識することなく以下のように定義される新たな検索質問 q' を生成し検索エンジンに対して問い合わせを発行できるようになっており、新たな 3D-SOM を得ることができる。

$$q' = k_1 \wedge k_2 \wedge \dots \wedge k_m \wedge \neg K$$

3. 領域の合併

3D-SOM 上の複数の領域 (キーワード) $K = \{w_l \mid l \in \{1, \dots, 500\}\}$ を 1 つに合併したい場合に用いる方法である。例えばキーワードに “red” や “white” など色に関するキーワードが数多く出ている場合、これらを “color” などと 1 つのキーワードにまとめてユーザにとって視認性のよい 3D-SOM を再構成することを目的としている。この機能は Web 文書 D_i の特徴ベクトルを以下のように再定義する。

$$F(D_i) = \left(\frac{f_1^i}{N_i}, \dots, \sum_l \frac{f_{w_l}^i}{N_i}, \dots, \frac{f_{500}^i}{N_i} \right) \quad (6.3)$$

4. 領域の注目

ユーザが 3D-SOM 上のある領域群に注目したい場合、それらの領域内にある Web 文書だけで特徴ベクトルを再生成し、分類マップを再構成するものである。その際に、単純に単語の出現頻度を利用した特徴ベクトルを生成するのではなく、Salton により提案されている *tf/idf* 法 [45, 46] を用いる。*tf/idf* 法は情報検索の分野で用いられている方法であるが、WWW における情報検索においてもその有効性が確認されている [56]。実際に特徴ベクトルを生成する際には、6.3.1 小節と同様に行うが、重要度の大きさは出現頻度を基に計算する場合とは異なり *tf/idf* 法を基に重要度が算出される。値の大きい順に上位 500 の単語 w'_1, \dots, w'_{500} を抽出した場合、Web 文書 $D_i (i = 1, \dots, M)$ の *tf* および *idf* は

$$tf_j^i = \frac{f_j^i}{N_i}, \quad idf_j^i = \log\left(\frac{M}{df_j^i}\right)$$

のように計算される。ただし、 $f_j^i (j = 1, \dots, 500)$ は Web 文書 D_i に含まれている単語 w'_j の出現頻度、 df_j^i は単語 w'_j を含む Web 文書 D_i の数を表す。したがって、 D_i の特徴ベクトルは以下のように再定義される。

$$F(D_i) = (f_1^{i'}, f_2^{i'}, \dots, f_{500}^{i'}), \quad f_j^{i'} = tf_j^i \cdot idf_j^i \quad (6.4)$$

この方式を採用した理由は、3D-SOM 全体を概観する際には、単純に出現頻度による特徴ベクトルを利用することがよく、目的の情報を絞り込んでいくという際には *tf/idf* 法を利用することがよいことがこれまでの実験により判明しているためである [22]。よって、この機能はユーザの欲している文書を絞り込む際に有効な手段であると考えられることができる。

5. 領域の分解

Web 文書は世界中の Web サーバに分散していることを考えると、各国の言語の問題が生じてくる。つまり、ユーザに馴染みのない単語が 3D-SOM 上に現れることも考えられる。このフィードバック機能は、3D-SOM 上でそういったユーザにとって意味不明な単語による領域形成を防ぐために用いられる機能である。この場合もまた、重要度の高い単語が変化するため上位 500 の単語 w'_1, \dots, w'_{500} を抽出しなおす必要があるため、Web 文書 D_i の特徴ベクトルは以下のように再定義される。

$$F(D_i) = \left(\frac{f_1^i}{N_i}, \frac{f_2^i}{N_i}, \dots, \frac{f_{500}^i}{N_i} \right) \quad (6.5)$$

これらの適合フィードバックの機能の他に、過去に行ったフィードバック操作の履歴を利用することで、同類の問い合わせに対して過去の適合フィードバック操作を利用した分類ビューもユーザに提供でき、ユーザ独自のビューを生成することに応用できる [63].

6.5 Web 文書分類ビュー機構の実行例とその評価

6.5.1 分類ビュー機構の実行例

実際に 20×20 の大きさのマップを用いて Web 文書の 3D-SOM の生成を行った。本システムでは、Web 文書を検索する検索エンジンとしてロボット型検索エンジンである goo[17] を利用した。検索エンジンに対して “worldcup” という検索キーワードで得られた検索結果から、検索エンジンの検索結果の高スコア順に 400 件を抽出し、6.3.1 小節の方式に従って Web 文書単体から特徴ベクトルの生成を行い 3D-SOM を生成する。この時、特徴ベクトルの学習には Kohonen の研究グループで開発された SOM_PAK[29] を改良したものを利用しており、米・シリコングラフィックス社のワークステーション Origin 200 上で学習させると、セル数 400 の 3D-SOM はユーザがキーワードを発行してから 5 分程度で生成された¹。

生成された分類マップの様子を図 6.2 に示す。この図を見ればわかるようにマップは、“world”, “football”, “france”, “soccer”, “description”, “apr”, “coupe”, “french”, “communication”, “paris”, “jun”, “english”, “ski”, “korea”, “main”, “oct”, “dec”, “nov”, “sep” のキーワードによって、19 の領域に分割されている。

6.4 節で述べたようなユーザインタフェース機能により、ユーザはこの 3D-SOM を検索エンジンの検索結果のブラウジングツールおよび問い合わせツールとして利用できる。3D-SOM 上の円柱を選択するとすると、3 つのキーワード (図 6.2 の場合は、“football”, “finale”, “europe”), 適合フィードバックへのリンク、および円柱に振り分けられた URL のデータが Web ブラウザに表示される。また、マップ上に表示されているラベルを選択すると 3D-SOM の段階的詳細化機能によってその領域の詳細マップを見ることができる。ここで、適合フィードバック機能が選択されると、マップ上のどの領域 (キーワード) に対して適合フィードバックを施すかどうか、図 6.3 のような選択画面が表示され、フィードバックを行う領域をユーザが選択することでそれぞれの領域に対して適合フィードバックが行われるようになっている。

¹検索エンジンが検索する対象を海外のサイトかつ検索キーワードを英語とした場合の数値。検索対象を国内のサイトかつ検索キーワードを日本語とした場合はさらに日本語の形態素解析 [72] を行う必要があるため 10 分程度の時間を要する。

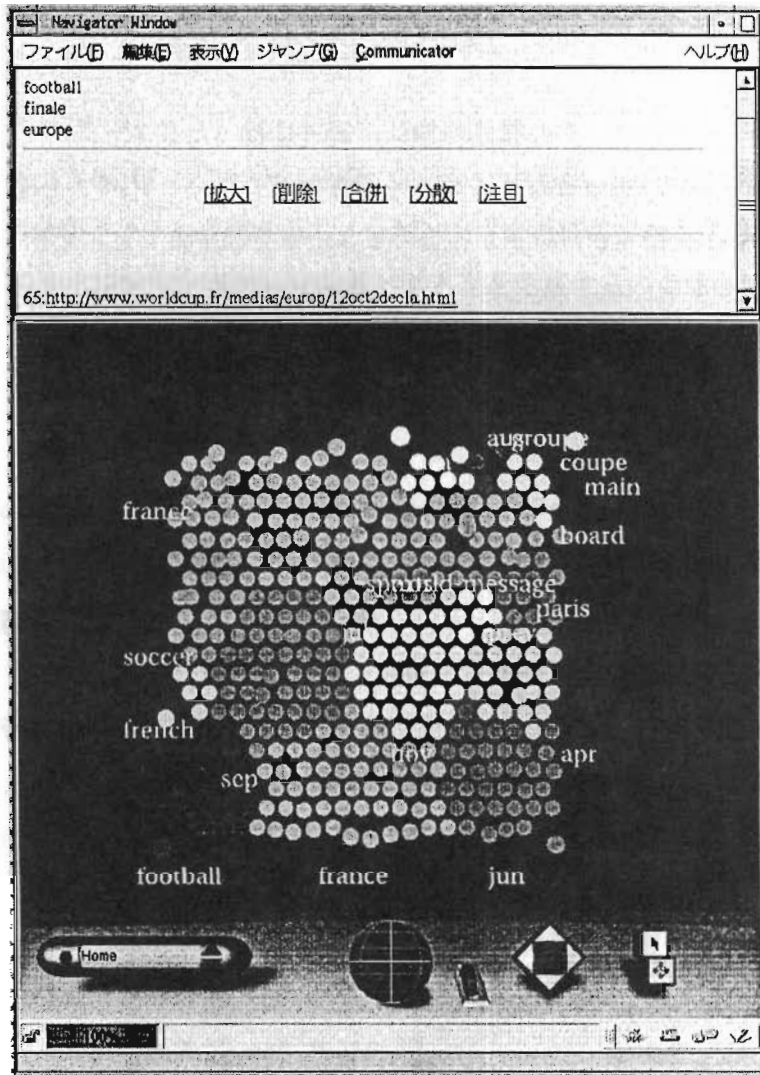


図 6.2: 分類マップ概覧

以下にそれぞれの適合フィードバックの実装例を示す。

1. 領域の拡大

図 6.2 に対して，“france” というキーワードで、「拡大」の適合フィードバックを行い、フィードバックを施す前後で 3D-SOM 上の “france” を第 1 のキーワード，第 2 のキーワード，第 3 のキーワードとして分類されている Web 文書数の変化を比較した。表 6.1 を見ればわかるように，“france” に関連する Web 文書をマップ上に数多く浮かび上がらせるという機能が実現できている。

2. 領域の削除

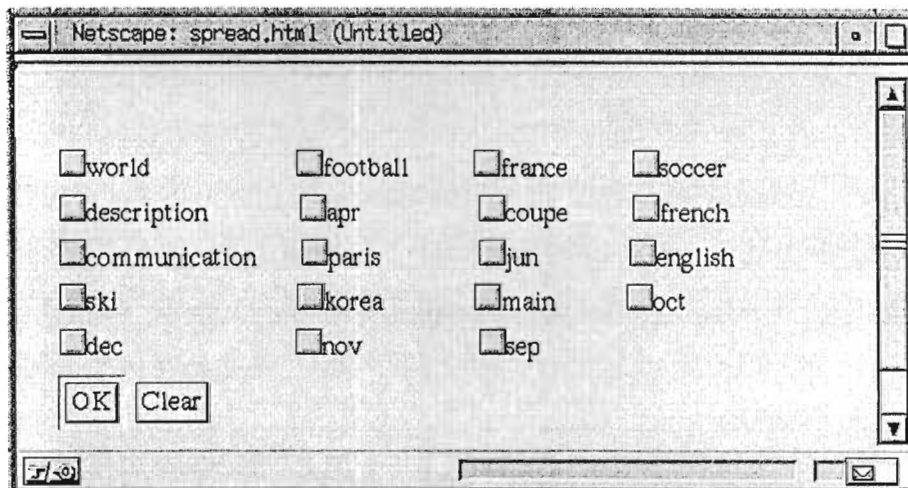


図 6.3: キーワード選択画面

	拡大前	拡大後
第 1 キーワード	27	32
第 2 キーワード	15	17
第 3 キーワード	1	9
合計	43	58

表 6.1: 拡大による 3D-SOM 上領域数の変化

図 6.2 に対して，“ski” というキーワードで「削除」の適合フィードバックを行ってみた結果，図 6.4 のようなマップが生成された．3D-SOM には，“description”，“dec”，“jun”，“sep”，“nov”，“apr”，“french”，“soccer”，“jul”，“prev”，“paris”，“sports”，“world”，“message”，“board”，“france”，“main”，“usa”，“coupe”，“aus”，“groupe” のキーワードによって，21 の領域に分割され，“ski” というキーワードがマップ上には表れていない．また，検索エンジンに対して NOT 検索を行っているため検索対象となるデータ群の中にも“ski”に関連した文献は見当たらなかった．

3. 領域の合併

「合併」は，複数のキーワードの特徴ベクトルの成分を加えて，ある 1 つのキーワードに置き換える機能である．最初に生成された図 6.2 のマップ上の，“apr”，“jun”，“oct”，“dec”，“nov”，“sep” のキーワードを“month”というキーワードに置き換えて，特徴ベクトルを再生成し SOM に学習させてみたところ，“description”，“french”，“month”，“soccer”，“korea”，

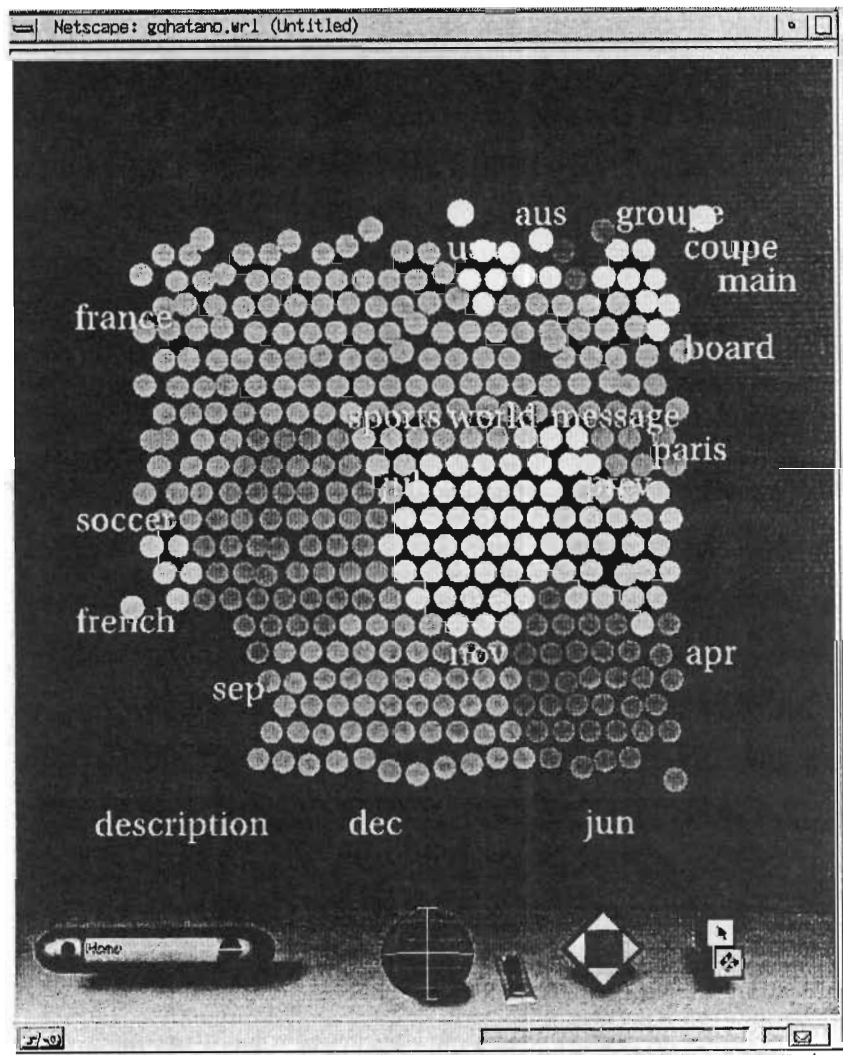


図 6.4: 削除の適合フィードバック

“paris”, “ital”, “usa”, “main”, “world”, “green”, “france”, “groupe”, “communication”, “coupe”, “ski”, “english” のキーワードによって, 17 の領域に分割された. 図 6.5 のマップにおいて, 合併された “month” という領域がマップの右下の 4 分の 1 を占める大きな領域を形成していることが分かる.

4. 領域の注目

「注目」は, ユーザの欲している情報を最終的に絞り込むために使われる機能である. ここで, 出現頻度によって「注目」の機能を実現した場合と tf/idf 法によって実現した場合のマップの領域数を調べてみると, 単純に文書中の単語の出現頻度を利用した場合は 25 個, tf/idf 法を利用した場合は 35 個の領域に分割された. つまり, tf/idf 法を利用したほうが

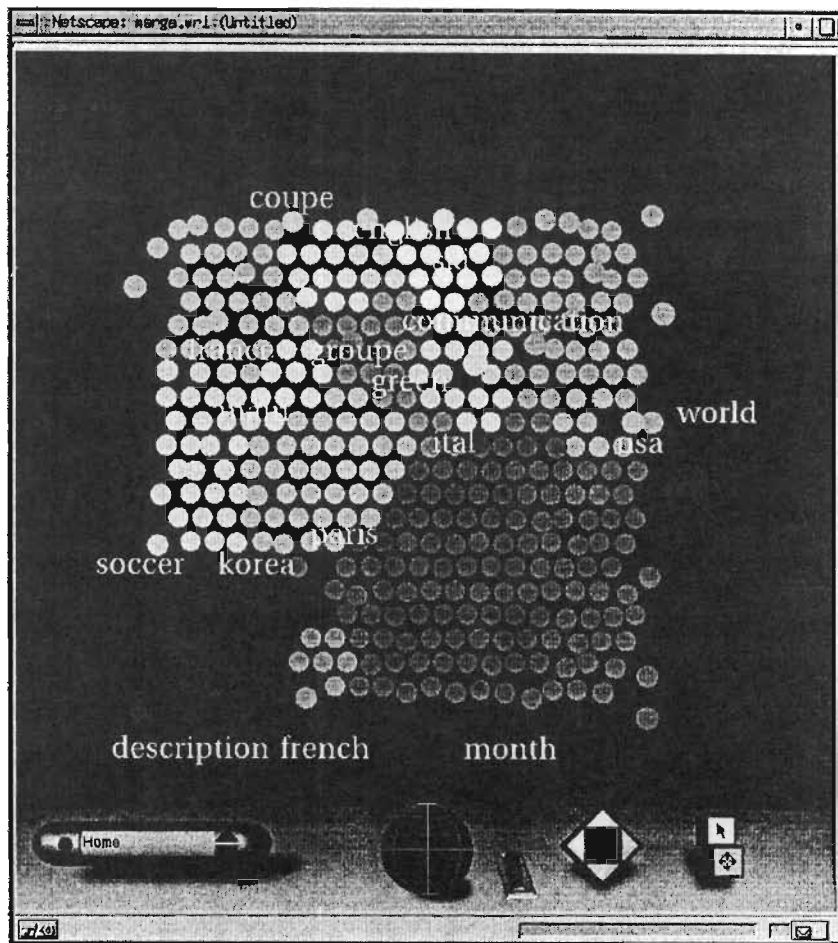


図 6.5: 合併の適合フィードバック

それぞれの文書特有のキーワードにより領域が細かく分類されるため、ユーザが欲する文書
を特定する際に非常に有用であることがわかる。

5. 領域の分解

図 6.2において，“coupe” や “main” といったキーワードは、何を表現しているものなのか
ユーザには非常にわかりにくい。こうした意味不明の単語による領域形成を防ぐための操作
は「領域の削除」のフィードバック機能と似ているが、検索エンジンに再問い合わせを行う
過程がない点が異なる。

6.5.2 分類ビュー機構の評価

Web 文書分類ビュー機構の評価を、分類マップ上に分類されたデータの適合率と再現率を用い
て行った。さて、本評価実験における適合率と再現率は以下のように定義される。

全ての Web 文書の中から、マップ上において n 個の領域を作っているキーワード $k_i (i = 1, \dots, n)$ を特徴として表す Web 文書の数 $similar(k_i)$ とし、 k_i の領域に分類されている Web 文書の数 $neighbour(k_i)$ で表わすとすると、適合率は、

$$\frac{1}{n} \cdot \sum_{i=1}^n \frac{|neighbour(k_i) \cap similar(k_i)|}{|neighbour(k_i)|}$$

で表わされ、また再現率は、

$$\frac{1}{n} \cdot \sum_{i=1}^n \frac{|neighbour(k_i) \cap similar(k_i)|}{|similar(k_i)|}$$

で表わされる。

今回の様に検索エンジンの検索結果というものは、「情報が適合する、しない」という評価は主観的であり、評価者によって異なるものであるため、ここでは第一著者の基準でページの内容の評価を行い、 $similar(k_i)$ を定めた。検索エンジンの出力結果に適合フィードバックを施す前と施した後の適合率と再現率の変化を表 6.2 に示す。

表 6.2: 適合フィードバックによる適合率と再現率の変化

	適合率 (%)	再現率 (%)
前	37.44	34.50
後	48.82	44.84

表 6.2 が示しているように、適合フィードバックを行うことにより適合率および再現率が上昇しているため、検索時に目的の情報にたどり着くまでのユーザの労力が軽減されており、これらのフィードバック機能の有効性が確かめられている。

しかし、現在のところ以下のような問題点を抱えており、それぞれの問題点に対処していく必要があると思われる。

- 評価の際に、適合フィードバック前後の分類マップ各領域の適合率および再現率を用いているが、フィードバックによる検索対象データ群の変化や、各 Web 文書の特徴ベクトルの変化について考慮していないため、これらの数値の評価がそのままシステムの評価とはなり得ない。つまり、新しい評価尺度を提案すべきである。
- 検索エンジン固有のランキング手法について全く考慮せず、検索対象データ 400 件を選定している。したがって検索エンジンのランキングに影響されないような検索対象データの選定方法を提案すべきである。

- ユーザがシステムに対して問い合わせを発行してから検索結果が返ってくるまで 5 分とユーザのインタラクションを反映した検索システムである Scatter/Gather に比べ実用的ではない。したがって、現在のように問い合わせの度に検索対象データをネットワーク上に収集していくようなシステム構成ではなく、Web ロボットを利用してあらかじめデータを収集しデータベースに格納されたデータに対して、問い合わせを行うようなシステムを構築し時間短縮を図るべきである。これにより、Scatter/Gather と同様なインタラクションの機能をサポートする Web 文書の極小部分マッチグラフの分類ビュー機構の実現が図られる。さらに、SOM の学習速度を高速化するために、特徴ベクトルの低次元化や学習パラメータの自動設定なども考慮すべきである。
- Web 文書の分類に関しては、ユーザのインタラクションを反映できるインタフェースとなっているが、検索に関してはインタラクションをサポートする機能は存在しない。したがって、検索作業に関するフィードバック機能もサポートすべきである。
- 本評価における適合率、再現率の値がそれほど良い値ではない。これに関しては、これまでの研究で SOM の学習パラメータや特徴ベクトル生成アルゴリズムを適切に設定すればよいことが分っている。ただし、これらパラメータの設定は、経験則的な部分があるため最適値を設定するのが困難であるという問題がある。

6.6 Web 文書の情報単位

Web 文書はハイパーテキストの一種であり、互いにリンクによってつながれた構造を形成している。こうした Web 文書の構造的特徴を利用した Web 文書の検索に関する研究が近年行われており、Weiss らは Web 文書中に含まれている単語の類似度と各々のリンクの関係を共に利用した Web 文書の分類に関する研究を行っている [54]。しかし、この研究は各 Web 文書を単にハイパーテキスト中のノードとして取り扱っているため、検索時にユーザがいくつかのキーワードを用いて問い合わせを行った場合は、それら全てのキーワードを含んだノード、すなわち Web 文書のみが答えとして返ってくるだけで、そのうちいくつかのキーワードを含んでいる Web 文書や Web 文書制作者がある話題についていくつかの Web 文書にまたがって記述した場合の関連のある文書の検索に問題を生じることとなる。また、Web 文書の検索はユーザがあらかじめ欲している情報について明確にその回答を持っているわけではないため、AND 検索や OR 検索の単なる組み合わせでは検索されない情報が多く、また新たな検索目標を発見しにくい。

したがって、本節ではこういった問題を解決するために、SOM に入力する特徴ベクトルを Web 文書単位で生成するのではなく、ある話題を基に Web 文書のハイパーテキスト構造を分割して生

成される極小部分グラフを単位として特徴ベクトルを生成し，Web 文書を意味情報単位で分類・検索を行う方法の提案を行う．この提案は図 6.6 に示すように AND 検索および OR 検索の中間的な検索²を行うことを目的としている．つまり，従来の AND や OR 検索では検索が不可能であった関連のある Web 文書群を 1 つの情報単位として検索が可能となる．また，Web 文書の分類・検索の単位を Web 文書単体か極小部分グラフかを定めるのはユーザであるため，分類・検索するデータの単位に関する一種のユーザビューの生成がなされていると考えることができる．以下に，その概要を示す．

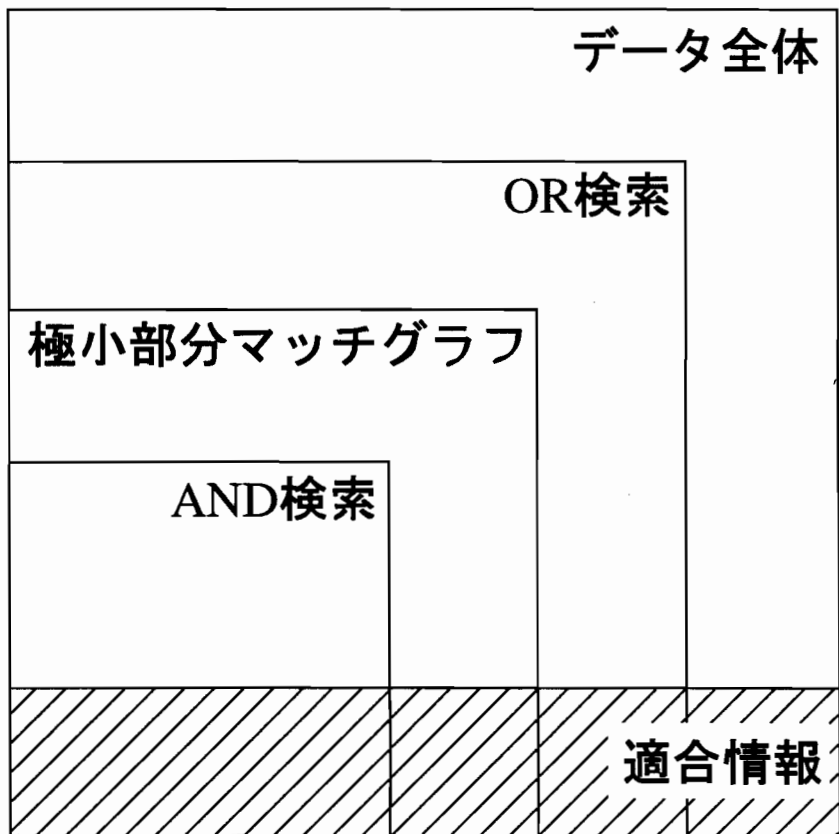


図 6.6: 検索法による検索可能な範囲

6.6.1 極小部分マッチグラフ

本節では，極小部分マッチグラフの定義を行う．

問い合わせを行う Web 文書のデータベースを

²極小部分マッチグラフはデータの基本単位が他の場合と異なり複数の Web 文書群であるが，極小部分マッチグラフを構成している Web 文書に置き換えて考えると，OR 検索では検索されるが AND 検索では検索できない Web 文書と考えることができる．

$$DB = \{U_1, U_2, \dots, U_l\}, \quad (l \geq 1)$$

と定義する。ただし、各 $U_i (i = 1, \dots, l)$ は各 Web 文書をノード、文書間をつなぐリンクを有向枝とする互いに素な連結グラフであり、 $U_j \cap U_k = \phi (j, k \in \{1, \dots, l\}, j \neq k)$ となる。

次に、このデータベースに対する検索質問 $Q = k_1 \wedge \dots \wedge k_m (m \geq 1)$ の解として得られる各連結グラフ U の極小部分マッチグラフ $G = (V, E)$ を以下のように定義する (図 6.7 参照)。

1. G は U の部分グラフである。
2. 各キーワード $k_j (j = 1, \dots, m)$ に対し、 G 内の少なくとも 1 つの節点 (に該当する Web 文書) がこれを含む。
3. G のいかなる部分グラフ G' も上記 1, 2 を満足しない。

6.6.2 極小部分マッチグラフの評価尺度

極小部分マッチグラフをデータベースに対する問い合わせの解と考えた場合も、従来の検索方法の場合と同様、解として膨大な量の極小部分マッチグラフが返される。よって、検索エンジンの場合と同様、それら極小部分マッチグラフのスコアリングを行うための評価式を定義し、極小部分マッチグラフを評価する必要がある。ここで、極小部分マッチグラフを評価する際には、以下の点を考慮する必要がある。

- 極小部分マッチグラフの「直径」。

ここでいう「直径」とは部分グラフの最長パスのノード数を表す。極小部分マッチグラフの直径が大きくなると、どのキーワードにもマッチしないノードを内部に含む可能性が高くなり、質問に無関係な Web 文書を多く含むことになる。つまり、このような極小部分マッチグラフのランク付けを行う必要があると考えられる。

- 各節点のキーワード数。

極小部分マッチグラフの各節点が、質問で指定されたキーワードをどの程度有しているかも重要な尺度となる。

これらを表現するために本研究では評価式として以下の式を提案する。

$$\text{評価尺度 1} = \frac{1}{d} \tag{6.6}$$

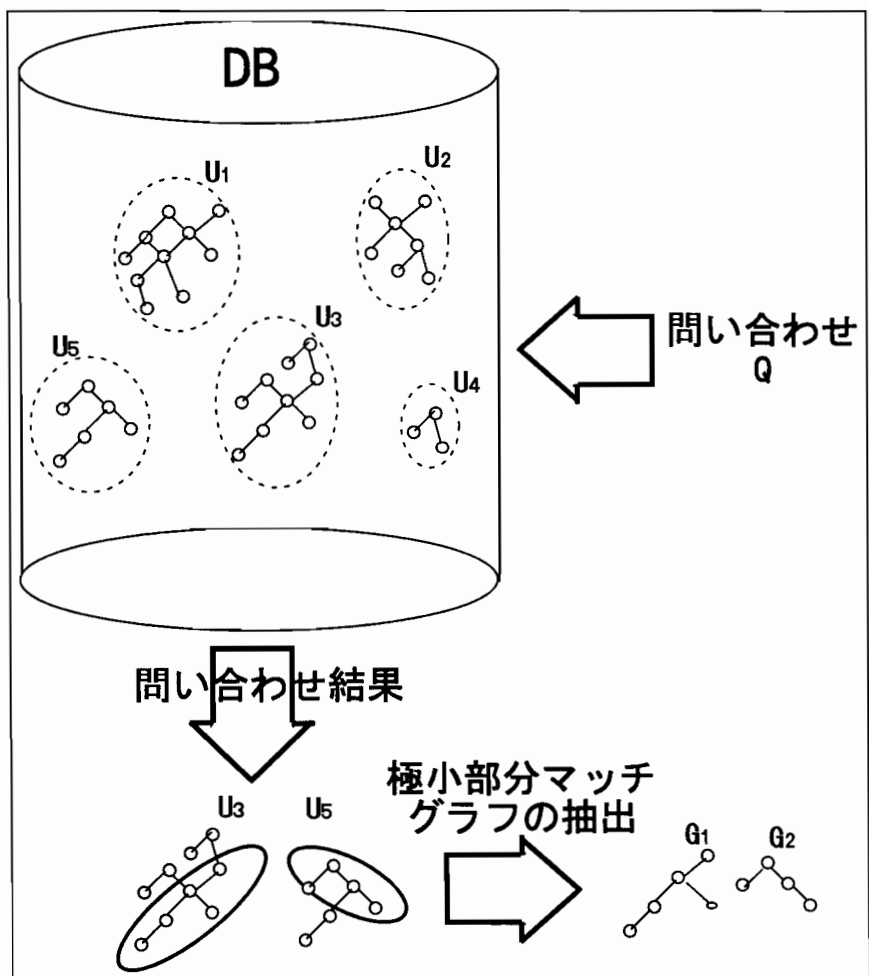


図 6.7: 検索モデルの事例

$$\text{評価尺度 2} = \frac{1}{\sum_{v \in V_i} \frac{m}{n \cdot K(v)}} \quad (6.7)$$

ただし、 n は極小部分マッチグラフ中のノード数、 d はグラフの直径、 m は問い合わせ Q のキーワード数、 v は直径上のノード、 V_i は直径上のノード集合、 $K(v)$ は極小部分マッチグラフの各節点 v における Q にマッチするキーワード数を表す。ただし、式 6.7 において、 $K(v) = 0$ の場合は $\frac{m}{n \cdot K(v)} = C$ ($C \geq 1$ である定数) とする。

式 6.6 は、グラフの直径のみを考慮した評価式であり、部分グラフの大きさが大きくなればなるほど、どのキーワードにもマッチしないノードを含む確率が高くなるためスコアが悪くなる。また、式 6.7 は、グラフの直径だけではなくグラフ中の各ノードが問い合わせ Q にどれほどマッチしているかについて考慮した評価式である。例を挙げて説明すると、それぞれの評価式により図 6.8 のようなグラフは図中のスコアのようになる。式 6.6 の場合、図中の (3)、(4)、(5) のようにグ

ラフの構造が同じであるが、ノードにマッチするキーワードが異なるような場合でも同じスコアが与えられてしまうという欠点を持つ。そこで、式6.7は問い合わせ Q に対する各ノードのマッチキーワード数を考慮した。ただし、図中の (4), (5) のようにノイズのノードが含まれている場合も考えられ定数 C ($C \geq 1$) を定めている。 C はこれらノイズのノードをどう扱うかにより変化させることができ、 C を大きくすればノイズが数多く含まれるとスコアが悪くなる。図6.8の例では、 $C = 1$ として多少のノイズに対しては無視している。

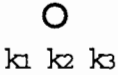
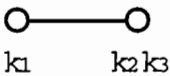
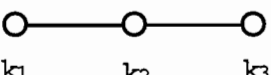
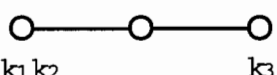
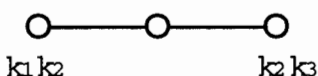
極小部分 マッチグラフ	評価尺度1 のスコア	評価尺度2 のスコア
(1) 	1	1
(2) 	1/2	4/9
(3) 	1/3	1/3
(4) 	1/3	2/5
(5) 	1/3	1/2

図 6.8: 評価の一例

6.6.3 極小部分マッチグラフの抽出

極小部分マッチグラフを抽出する方法を以下に述べる。

1. 問い合わせ Q を $Q = k_1 \wedge \dots \wedge k_m$ とする。本モデルでは、答えの Web 文書には全てのキーワード k_1, \dots, k_m が含まれている必要がないため、検索エンジンに対する問い合わせは OR 検索である $q = k_1 \vee \dots \vee k_m$ となる。

2. 問い合わせ q により返された Web 文書それぞれのリンク構造を調べ、各連結グラフ U_i を生成する。
3. 生成された各連結グラフの親ノードから順に、各ノードのマッチするキーワードを調べ、 U_i の中から極小部分マッチグラフを抽出する。例えば、図 6.9 のように、問い合わせが $Q = k_1 \wedge k_2$ により、部分グラフ U が与えられた場合、(1) ~ (5) のような極小部分マッチグラフが得られる。

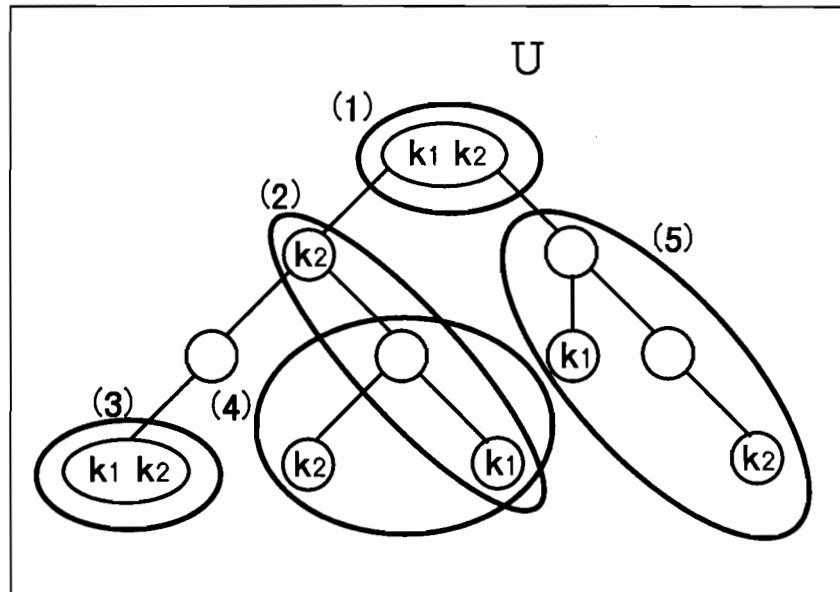


図 6.9: 極小部分マッチグラフの抽出

6.6.4 極小部分マッチグラフの自動分類

抽出された極小部分マッチグラフを Web 文書単体の場合と同様に自動分類するためには、極小部分マッチグラフの特徴ベクトルを生成しなければならない。

以下に、極小部分マッチグラフ G_i の特徴ベクトルを生成する方法について説明する。

1. 極小部分マッチグラフの各ノードは Web 文書 D_j であるため、6.3.1 節に述べたように、各ノードの特徴ベクトル $F(D_j)$ を求める。この $F(D_j)$ の生成方法には 2 通りあり、単純にノード内に出現する単語の出現頻度を基に計算する方法と、 tf/idf 法を基に計算する方法が考えられる。
2. 本論文では、極小部分マッチグラフは各ノードの集合であるため、極小部分マッチグラフの特徴ベクトル $F(G_i)$ も各ノードの特徴ベクトル $F(D_j)$ の和であると考え。ただし、各

ノードは

- 問い合わせキーワードに対するマッチ率.
- グラフの形状 (直径上のノードかどうか).
- グラフ内における親子関係.

によって各々重要度が異なっているため、重み $\alpha(D_i)$ を考慮に入れる必要がある。 $\alpha(D_i)$ には、極小部分マッチグラフを評価するために提案した式 6.7 を利用する。

$$\alpha(D_i) = \frac{1}{\sum_{v \in V_i} \frac{m}{n \cdot K(v)}} \quad (6.8)$$

したがって、極小部分マッチグラフの特徴ベクトル $F(G_i)$ は以下のように求められる。

$$F(G_i) = \frac{\sum_j \alpha(D_j) \cdot F(D_j)}{n} \quad (6.9)$$

ここで、 n は極小部分マッチグラフのノード数である。

こうして得られた極小部分マッチグラフの特徴ベクトル $F(G_i)$ を Web 文書単体で特徴ベクトルを生成した場合と同様に、SOM の入力として学習させ、分類マップを生成することで自動分類が可能となる。これらの機能は適合フィードバックの機能というよりは、分類・検索を行う初期の段階で、ユーザが問い合わせ時に Web 文書単体または極小部分マッチグラフで分類・検索を行うのかを選択するという形で実現する予定である。

6.6.5 極小部分マッチグラフの予備実験

現段階では極小部分マッチグラフを扱うことが可能な分類ビュー機構が未実装であるため、ここでは予備実験を行うことで極小部分マッチグラフの有用性を示す。予備実験では、3つの検索キーワードを与えて検索されたデータに対しての評価は適合率と再現率を用いた。さて、適合率・再現率を算出する際、全ての Web 文書を閲覧して利用者の意図に合致した Web 文書 (以後この Web 文書を「適合文書」と呼ぶ) を得るのは実質上不可能である。そこで本予備実験では、OR 検索で検索された Web 文書の中からランダムに抽出し分類・検索の母集団を定めてから、その母集団に対して我々が適合文書を定める。この適合文書の割合から OR 検索で検索された Web 文書の中に存在する適合文書を算出し、さらに検索エンジンによって検索された Web 文書、されなかった Web 文書を算出することで、適合率・再現率を計算する。

表 6.3 のおける「文書数」とは検索エンジンによって検索された Web 文書数を表している。これらの値を見れば分かるように、分類・検索の単位を Web 文書単体にするよりは極小部分マッ

表 6.3: 分類・検索単位の違いによる適合率と再現率

検索対象	文書数	適合率 (%)	再現率 (%)
Web 文書 (AND)	9	11.11	6.605×10^{-4}
極小部分グラフ	294	21.43	4.161
Web 文書 (OR)	20531	7.373	100.0

グラフを利用した方が、AND 検索を行う場合に比べ検索されるデータ数が多くかつ OR 検索のように膨大なデータ数とはならない。また、対象データを関連のある Web 文書群である極小部分マッチグラフとしているため、AND 検索では検索することができなかつたような関連のある Web 文書も同時に検索することが可能となっている。つまり、AND 検索および OR 検索の中間的な検索を行うという当初の目的を達成することができた。しかし、この条件下で分類精度を評価を行っても実質的な評価とはなりえない。なぜなら、評価対象のデータ母集団が通常の AND 検索や OR 検索は Web 文書単体であるのに対し極小部分マッチグラフの場合は Web 文書群であるため、表 6.3における極小部分マッチグラフの評価はグラフ中に 1 つでも適合情報にマッチしている Web 文書があれば適合したグラフであると評価しており「評価の場」が異なるからである。また、関連ある Web 文書群には、本来、図 6.10のようにトピックに関する並列構造や階層構造を持ったものが考えられるため、これらの構造を考慮して極小マッチ部分グラフの評価を行う必要もあると考えられるからである。今後はこれらのことを踏まえた評価実験を行い極小部分マッチグラフの有用性を証明する必要があると思われる。

6.7 結言

本章では、ニューラルネットワークの一種である SOM を用いて、ユーザの多様な分類・検索意図をインタラクティブに支援した Web 文書の自動分類ビュー機構を実現し、その一部機能について実装を行った。

本分類ビュー機構の利点は、

- 検索エンジンに対する問い合わせの答えが、従来のようにリスト表示ではなく 3 次元分類マップにより提供されるので、ユーザは分類状況全体を把握しながら検索することができる。つまり、分類を行っていくにしたがって、ユーザ自身がどのような情報を要求しているのかを理解しながら検索を行うことが可能となる。
- システムがユーザの分類意図や検索目的をインタラクティブに支援することができるため、

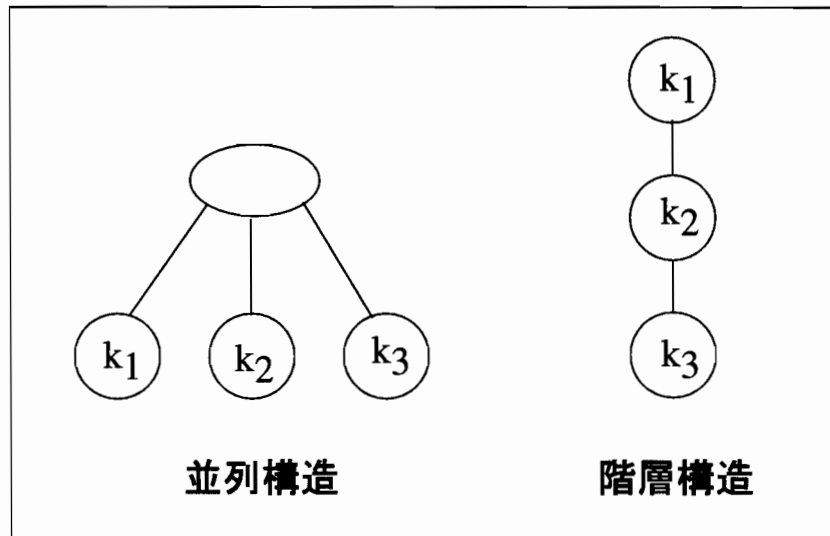


図 6.10: 極小部分グラフの構造

ユーザ本位の分類マップを生成することができる。つまり、検索効率の向上を図ることが可能である。

- 検索対象に極小部分マッチンググラフ，すなわちある話題に関する Web 文書群を検索出来るようになれば，従来では検索できなかった Web 文書を検索できる可能性がある。

ことが挙げられる。さらに，今後の課題として以下のような問題が挙げられる。

- 検索エンジンで算出されているスコアの利用および高速化を考慮した，Web 文書の特徴ベクトルの生成および SOM 学習アルゴリズムの改良。また，SOM 学習パラメータの動的最適設定法の確立。
- 適合フィードバック機能による検索対象データ母集団の変化や Web 文書特徴ベクトルの変化を考慮した，本システムの評価法の提案。
- 検索対象データを極小部分マッチンググラフとした，Web 文書分類ビュー機構の実装とその評価。
- 検索の高速化を考慮してロボット型の検索エンジンの機能を本システムに付加し極小部分マッチンググラフを扱うことができる Web 文書分類ビュー機構を実装。
- 検索作業に関するフィードバック操作機能の充実。

第 7 章

結 論

本論文では、マルチメディアデータの分類・検索を行う際に生じる問題点を解決するために、まず、データのコンテンツ情報および意味情報を用いて特徴づけを行い、ニューラルネットワークの一種である SOM を用いて分類を行った。さらに、データの分類状況を概観でき分類結果に対してユーザの分類目的を反映させることができる、3D-SOM と呼ばれるインタフェースの開発を行い、システムの評価実験およびその問題点について議論を行った。

以下、本論文の統括を行う。

第3章では、データ数が膨大であるマルチメディアデータ、特に文書および映像データ (カット映像) から計算機で扱える特徴量を抽出し、その特徴量を基に SOM を利用してデータを分類・検索するシステムの実装方法およびその評価実験について述べた。ここでは、ベクトル空間法を利用したそれぞれのデータの特徴量の生成方法および overview マップを利用した分類されたデータの可視化の観点から実現した 3 次元インタフェースの説明を行い、開発したマルチメディアデータの自動分類機構の実現により、データ 1 つ 1 つに対して行う作業量の軽減を図ることが可能となり、その有効性を確認した。

第4章ではカット映像の分類に焦点をあて、映像の色や映っている物体の形といったコンテンツのみを特徴量として分類する方法だけでは不十分であるため、映像にその意味内容を記述するための映像記述支援ツールを実装し、記述された内容情報とコンテンツ情報を共に利用したハイブリッド型の映像の特徴量を利用したカット映像の分類システムの実装およびその評価実験について述べた。ここでは映像を問わず静止画像に関しても、そのデータの意味記述情報というものは必要であることを述べ、マルチメディアデータの特徴量として最適なのはコンテンツ情報と意味記述情報を合わせたハイブリッド型の情報であることを提案し、その有効性を確認した。

第5章では、シーン映像の分類に焦点をあて、カット映像に与えられた記述情報を特徴量としてシーン映像の分類システムの実装およびその評価実験について述べた。ここでは、SOM を多

段階に利用し、初期の段階では見いだすことが出来なかった新たな分類尺度により分類する方法の提案とその有効性について確認を行った。

第6章では、分類された文書データの分類結果に対して、ユーザがインタラクティブに操作を行い一種のユーザビューを生成する分類機構の実装およびその評価実験について述べた。ここでは、1度の問い合わせを行うだけでは的確な情報を得ることが出来ないという問題に対し、単にAND検索やOR検索などの組み合わせによる問い合わせではなく、システムが返した分類結果に対して、ユーザのフィードバック操作を反映させて再度分類結果を表示する方法を提案し、その有効性を確認した。

今後は、これらの研究で得られた問題点、特に、以下のようなマルチメディアデータの特徴、

- マルチメディアデータは Web 文書のように通常、半構造データである。
- マルチメディアデータは映像データの代表されるように、コンテンツ情報や意味情報、音声情報など多レベルの情報を持っている。

に注目した特徴抽出に関する研究を行っていきたいと考えている。

参考文献

- [1] Abiteboul, S. and Vianu, V.: Queries and Computation on the Web, *Proc. of the 6th International Conference on Data Theory (ICDT'97)*, pp. 262–275 (1997).
- [2] <http://www.altavista.com/>.
- [3] Arita, H., Yasui, T. and Tsudaka, S.: 3D Stroller: Strolling in the self-organized information space, *Proc. of RWC Symposium*, pp. 53–58 (1997).
- [4] Bernstein, M., Garzotto, F., Paolini, P. and Schwabe, D.: Tools for Designing Hyperdocuments, *Hypertext, Hypermedia Handbook*, McGraw Hill, pp. 179–209 (1991).
- [5] Botafogo, A. R., Rivlin, E. and Shneiderman, B.: Structural Analysis of Hypertext: Identifying Hierarchies and Useful Metrics, *ACM Transactions on Information Systems*, Vol. 10, No. 2, pp. 142–180 (1992).
- [6] Botafogo, A. R. and Shneiderman, B.: Identifying Aggregates in Hypertext Structures, *Proc. of the 3rd ACM Conference on Hypertext*, ACM, pp. 63–73 (1991).
- [7] Cleverdon, W. C.: Aslib Cranfield Research Project: Report on the First Stage of an Investigation into the Comparative Efficiency of Indexing Systems, Technical report, College of Aeronautics (1960).
- [8] Cleverdon, W. C., Mills, J. and E., K. M.: Aslib Cranfield Research Project: Factors Determining the Performance of Indexing Systems, Design. Cranfield (1966).
- [9] Conklin, J.: Hypertext: An Introduction and Survey, *IEEE Computer*, pp. 17–41 (1987).
- [10] Cutting, R. D., Karger, R. D., Pedersen, O. J. and Tukey, V. J.: Scatter/Gather: A cluster-based approach to browsing large document collections, *Proc. of the 15th Annual International ACM SIGIR conference on Research and development in information retrieval*, ACM, pp. 318–329 (1992).

- [11] Daniels, J. J. and E., R. L.: A Case-Based Approach to Intelligent Information Retrieval, *Proc. of the 18th annual International ACM Conference on Research and Development in Information Retrieval*, pp. 238–245 (1995).
- [12] Elmagarmid, K. A., Jiang, H., Helal, A., Joshi, A. and Ahmed, M.: *VIDEO DATABASE SYSTEMS: Issues, Products and Applications*, Kluwer Academic Publishers (1997).
- [13] <http://www.excite.com/>.
- [14] <http://www.excite.co.jp/>.
- [15] Flickner, M., Sawhney, H., Niblack, W., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D. and Yanker, P.: Query by Image and Video Content: The QBIC System, *IEEE Computer*, pp. 23–32 (1995).
- [16] George, R. G., Stuart, C. K. and Jock, M. D.: Information Visualization using 3D interactive animation, *Communication of the ACM*, Vol. 36, No. 4 (1993).
- [17] <http://www.goo.ne.jp/>.
- [18] Gupta, A. and Jain, R.: Visual Information Retrieval, *Communication of the ACM*, Vol. 40, No. 5 (1997).
- [19] Hara, Y., Keller, M. A. and Wiederhold, G.: Implementing Hypertext Database Relationships through Aggregations and Exceptions, *Proc. of the 3rd ACM Conference on Hypertext*, ACM, pp. 75–90 (1991).
- [20] Hartman, J. and Wernecke, J.(eds.): *The VRML 2.0 Handbook*, Addison Wesley (1996).
- [21] Hatano, K., Kamei, T. and Tanaka, K.: Clustering and Authoring of Video Shots Using Hybrid-type Self-Organizing Maps, *Proc. of the International Symposium on Digital Media Information Base (DMIB'97)*, pp. 150–158 (1997).
- [22] Hatano, K., Qian, Q. and Tanaka, K.: A SOM-Based Information Organizer for Text and Video Data, *Proc. of the 5th International Conference on Database Systems for Advanced Applications (DASFAA'97)*, World Scientific, pp. 205–214 (1997).
- [23] Hearst, A. M. and J., P. O.: Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results, *Proc. of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, pp. 76–84 (1996).

- [24] Honkela, T., Kaski, S., Lagus, K. and Kohonen, T.: WEBSOM - Self-Organizing Maps of Document Collections, *Proc. of the Workshop on Self-Organizing Maps (WSOM'97)* (1997).
- [25] Kim, Y. and Shibata, M.: Content-Based Video Indexing and Retrieval - A Natural Language Approach, *IEICE Trans. on Inf. and Syst.*, Vol. E79-D, No. 6, pp. 695–704 (1996).
- [26] Kohonen, T.: Self-Organized formation fo topologically correct feature maps, *Biological Cybernetics*, No. 43, pp. 59–69 (1982).
- [27] Kohonen, T.: The Self-Organizing Map, *Proceedings of the IEEE*, Vol. 78, No. 9, pp. 1464–1480 (1990).
- [28] Kohonen, T.: *Self-Organizing Maps*, Springer, Berlin (1995).
- [29] Kohonen, T., Hynninen, J., Kangas, J. and Laaksonen, J.: SOM_PAK: The self-organizing map program package, Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science (1996).
- [30] Lagus, K., Honkela, T., Kaski, S. and Kohonen, T.: Self-organizing Maps of document collections: A new approach to interactive exploration, *Proc. of the 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD-96)*, pp. 238–243 (1996).
- [31] Lancaster, W. F.: *Vocabulary Control for Information Retrieval*, Information Resources Press (1972).
- [32] Lawrence, S. and C., G. L.: Searching the World Wide Web, *Science*, Vol. 280, No. 3, pp. 98–100 (1998).
- [33] Lin, X., Soergel, D. and Marchionini, G.: A Self-organizing Semantic Map for Information Retrieval, *Proc. of the 14th annual international ACM Conference on Research and Development in Information Retrieval*, ACM, pp. 262–269 (1991).
- [34] Liu, Y., Chen, H., Yu, J. X. and Ohbo, N.: Using Stem Rules to Refine Document Retrieval Queries, *Proc. of the 3rd International Conference on Flexible Query Answering System (FQAS'98)*, pp. 248–259 (1998).
- [35] Mendelzon, O. A. and Milo, T.: Formal Models of Web Queries, *Proc. of the 16th ACM Symposium on Principles of Database Systems (PODS'97)*, pp. 134–143 (1997).

- [36] <http://drogo.cselt.stet.it/mpeg/>.
- [37] Mukherjea, S., Foley, D. J. and Hudson, E. S.: Interactive Clustering for Navigating in Hypermedia Systems, *Proc. of European Conference on Hypertext Technology (ECHT'94)*, ACM, pp. 136–145 (1994).
- [38] Pradhan, S., Tajima, K. and Tanaka, K.: Utilizing Prototype Objects and Powerdomains to Support Public Authoring of Video Database, *Proc. of International Symposium on Cooperative Database Systems for Advanced Applications (CODAS'96)*, pp. 210–217 (1996).
- [39] Qing, Q., Shi, X. and Tanaka, K.: Document Browsing and Retrieving based on 3D Self-Organizing Map, *Proc. of Workshop on New Paradigms in Information Visualization and Manipulation in Conjunction with CIKM'95* (1995).
- [40] Ritter, H. and Kohonen, T.: Self-organizing semantic maps, *Biological Cybernetics*, No. 61, pp. 241–254 (1989).
- [41] Ritter, H. and Kohonen, T.: Learning 'semantotopic maps' from context, *Proc. of the Int. Joint Conference on Neural Networks (IJCNN'90)*, Vol. I, pp. 23–26 (1990).
- [42] Rocchio, J. J.: *The SMART Retrieval System: Experiments in Automatic Document Processing*, Prentice-Hall, chapter Relevance feedback in information retrieval, pp. 313–323 (1971).
- [43] Salton, G.: *Automatic Information Organizations and Retrieval*, McGraw-Hill (1968).
- [44] Salton, G.: Recent studies in automatic text analysis and document retrieval, *Journal of the ACM*, Vol. 20, No. 2, pp. 258–278 (1973).
- [45] Salton, G.: *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley (1989).
- [46] Salton, G., Allan, J. and Buckley, C.: Automatic Structuring and Retrieval of Large Text Files, *Communications of the ACM*, Vol. 37, No. 2, pp. 97–108 (1994).
- [47] Sarkar, M. and Brown, H. M.: Graphical fisheye views, *Communication of the ACM*, Vol. 37, No. 12, pp. 73–84 (1994).
- [48] Smith, R. J. and Chang, F. S.: Visually Searching the Web for Content, *IEEE Multimedia*, pp. 12–20 (1997).

- [49] Tajima, K., Mizuuchi, Y., Kitagawa, M. and Tanaka, K.: Cut as a Querying Unit for WWW, Netnews, and E-mail, *Proc. of the 9th ACM Conference on Hypertext and Hypermedia*, ACM, pp. 235–244 (1998).
- [50] Tanaka, K., Nishikawa, N., Hirayama, S. and Nanba, K.: Query-Pairs As Hypertext Links, *Proc. of the 7th IEEE International Conference on Data Engineering (ICDE'91)*, IEEE, pp. 456–463 (1991).
- [51] Uehara, K. and Horiuchi, N.: Semantic Representation of Video Content based on Prototype-Instance, *Proc. of International Symposium on Digital Media Information Base (DMIB'97)*, pp. 167–175 (1997).
- [52] Uehara, K., Oe, M. and Maehara, K.: Knowledge Representation, Concept Acquisition and Retrieval of Video Data, *Proc. of International Symposium on Cooperative Database Systems for Advanced Applications (CODAS'96)*, pp. 527–534 (1996).
- [53] <http://info.webcrawler.com/mak/projects/robots/active/html/index.html>.
- [54] Weiss, R., Vélez, B., Sheldon, A. M., Namprempre, C., Szilagyi, P., Duda, A. and Gifford, K. D.: Hypursuit: A hierarchical network search engine that exploits content - link hypertext clustering, *Proc. of the 7th ACM Conference on Hypertext'96*, ACM, pp. 180–193 (1996).
- [55] <http://www.yahoo.com/>.
- [56] Yuwono, B. and Lee, L. D.: Search and ranking algorithms for locating resources on the World Wide Web, *Proc. of the 12th International Conference on Data Engineering (ICDE'96)*, pp. 164–171 (1996).
- [57] Zettsu, K., Uehara, K. and Tanaka, K.: A time-stamped authoring graph for video databases, *Proc. of the 8th International Conference on Database and Expert Systems Applications (DEXA '97)*, pp. 192–201 (1997).
- [58] Zizi, M. and Beaudouin-Lafon, M.: Accessing Hyperdocuments through Interactive Dynamic Maps, *Proc. of European Conference on Hypertext Technology (ECHT'94)*, ACM, pp. 126–135 (1994).
- [59] 佐野綾一, 波多野賢治, 田中克己: 自己組織化マップを用いた Web 文書の対話的分類とその視覚化, *情報処理学会研究報告*, 98-DBS-116-5, pp. 33–40 (1998).

- [60] 段一為, 佐野綾一, 波多野賢治, 田中克己: 部分グラフを基本単位とする Web 文書群の検索モデルと自動分類について, 情報処理学会第 57 回全国大会論文集分冊 3, pp. 161-162 (1998).
- [61] 岩成英一, 有木康雄: DCT 成分を用いた動画シーンのクラスタリングとカット 検出, 電子情報通信学会技術研究報告, PRMU93-119, pp. 23-30 (1994).
- [62] 島田恭宏, 大倉充, 塩野充, 橋本禮治: 複数特徴部分空間法による手書き類似漢字識別, 電子情報通信学会論文誌 D-II, Vol. J78-D-II, No. 10, pp. 146-1468 (1995).
- [63] 波多野賢治, 佐野綾一, 段一為, 田中克己: 自己組織化マップと適合フィードバック機能を用いた Web 文書群の分類ビュー機構について, 情報処理学会第 57 回全国大会論文集 分冊 3, pp. 181-182 (1998).
- [64] 波多野賢治, 田中克己: 映像データベースの動的クラスタリングと素材検索機構について, 情報処理学会研究報告, 96-DBS-109-18, pp. 105-110 (1996).
- [65] 江口浩二, 伊藤秀隆, 隈元昭: ユーザへの適応性を考慮した適合フィードバックによる WWW 情報検索, 電気学会論文誌 電子・情報・システム部門誌, Vol. 117-C, No. 11, pp. 1643-1649 (1997).
- [66] 是津耕司, 上原邦昭, 田中克己: 時刻印付オーサリンググラフによるビデオ映像のシーン検索, 情報処理学会論文誌, Vol. 39, No. 4, pp. 923-932 (1998).
- [67] 塩澤秀和, 西山晴彦, 松下温: 「納豆ビュー」の対話的な情報視覚化における位置づけ, 情報処理学会論文誌, Vol. 38, No. 11, pp. 2331-2342 (1997).
- [68] 豊浦潤, 岡隆一: テキストの知識ベース化のための自己組織化ネットワークの提案, 電子情報通信学会技術研究報告, NLC96-59, pp. 23-30 (1997).
- [69] 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 回全国大会論文集, Vol. 3, pp. 187-188 (1993).
- [70] 谷田部智之, 高羽洋樹, 佐藤隆, 坂内正夫: 画像情報による情報媒介を目指した公開型画像サーチエンジン GIRLS, *Proc. of Advanced Database Symposium'96*, pp. 139-145 (1996).
- [71] 加藤茂夫: 画像データ圧縮の基礎知識, インターフェース, No. 175, pp. 132-159 (1991).
- [72] 松本裕治, 北内啓, 山下達雄, 今一修, 今村友明: 日本語形態素解析システム「茶釜」 version 1.0 使用説明書, 技術報告 NAIST-IS-MT9551092, NAIST Technical Report (1997).
- [73] 藤原洋: 最新 MPEG 教科書, 株式会社アスキー (1994).

- [74] 向井理朗, 西村拓一, 遠藤隆, 岡隆一: 動画像の自己組織化ネットワークによるモデル化と要素動作の自動抽出, 電子情報通信学会技術研究報告, PRMU97-128, pp. 55-61 (1997).
- [75] 遠藤隆, 高橋裕信, 豊浦潤, 向井理朗, 岡隆一: 動画像の自己組織化ネットワークによるモデル化とその動的特徴の可視化 - Video Intra-structure Visualization -, 電子情報通信学会技術研究報告, PRMU97-78, pp. 49-54 (1997).
- [76] 畑島隆, 元田敏浩: WWW 情報空間ナビゲーション手法について, 情報処理, Vol. 38, No. 9, pp. 848-849 (1997).
- [77] 仁木和久, 田中克己: ニューラルネットワーク技術の情報検索への応用, 人工知能学会誌, Vol. 10, No. 1, pp. 1-7 (1995).

謝辞

本研究の遂行ならびに論文の作成にあたり、懇切なる御指導を賜りました神戸大学大学院自然科学研究科情報メディア科学専攻の田中克己教授，神戸大学工学部情報知能工学科の田島敬史助手に謹んで感謝の意を表します。

本論文をまとめるにあたり、有益な御助言と御教示を賜りました神戸大学都市安全研究センターの上原邦昭教授，ならびに神戸大学工学部情報知能工学科の瀧和男教授に謝意を申し上げます。

本論文の作成にあたり、御助言ならびに御協力を頂きました松下電器産業株式会社の角谷和俊氏，神戸大学大学院自然科学研究科情報メディア科学専攻の木俣豊氏、小磯健吾氏に深く感謝致します。

本研究の遂行にあたり、貴重な映像資料の学術利用を許可して下さった東映株式会社および愛企画センター 川内彩友美代表取締役役に感謝致します。

本研究の遂行ならびに論文の作成にあたり、御協力を頂いた神戸大学工学部情報知能工学科田中研究室の皆様，特に神戸大学大学院自然科学研究科情報知能工学専攻の亀井俊之氏，佐野綾一氏，段一為氏に感謝致します。

最後に、日頃より研究活動を支えてくれた父母 波多野隆弘・まり，および福田あかり氏に心より感謝いたします。

研究業績

主要論文

1. Kenji Hatano, Qing Qian and Katsumi Tanaka
“A SOM-Based Information Organizer for Text and Video Data”
Proceedings of the 5th International Conference on Database Systems for Advanced Applications (DASFAA '97), World Scientific, pp.205-214, Melbourne, Australia, April, 1997.
2. Kenji Hatano, Toshiyuki Kamei and Katsumi Tanaka
“Clustering and Authoring of Video Shots Using Hybrid-type Self-Organizing Maps”
Proceedings of the International Symposium on Digital Media Information Base (DMIB'97), pp.150-158, Nara, Japan, November, 1997.
3. 波多野 賢治, 亀井 俊之, 田中 克己
「多段階自己組織化マップによるビデオ映像記述支援と類似シーン検索」
情報処理学会論文誌, Vol.39, No.4, pp.933-942, 1998 年 4 月.
4. 波多野 賢治, 佐野 綾一, 段 一為, 田中 克己
「自己組織化マップと検索エンジンを用いた Web 文書の分類ビュー機構」
情報処理学会論文誌: データベース, (条件付き採録)

学術講演

1. 波多野 賢治, 亀井 俊之, 田中 克己
「映像データベースの内容記述へのコンテンツ情報の応用」
情報処理学会第 55 回全国大会論文集, 分冊 3, pp.214-215, 福岡市, 1997 年 9 月.

2. 段一為, 佐野 綾一, 波多野 賢治, 田中 克己
「部分グラフを基本単位とする Web 文書群の検索モデルと自動分類について」
情報処理学会第 57 回全国大会論文集, 分冊 3, pp.161-162, 名古屋市, 1998 年 10 月.
3. 波多野 賢治, 佐野 綾一, 田中 克己
「自己組織化マップと適合フィードバック機能を用いた Web 文書群の分類ビュー機構について」
情報処理学会 第 57 回全国大会論文集, 分冊 3, pp.181-182, 名古屋市, 1998 年 10 月.

学術報告

1. 銭 晴, 波多野 賢治, 田中 克己
「3 次元自己組織化マップに基づく文書のブラウジングと検索」
情報処理学会研究報告, Vol.95, 95-DBS-104-6, pp.41-48, 八代市, 1995 年 7 月.
2. Qing Qian, Kenji Hatano, Toshio Sasao and Katsumi Tanaka
“A VRML-Based Information Organizer by 3D Self-Organizing Map”
第 7 回データ工学ワークショップ (DEWS'96) 論文集, 神戸市, pp.85-90, 1996 年 3 月.
3. 波多野 賢治, 田中 克己
「映像データベースの動的クラスタリングと素材検索機構について」
情報処理学会研究報告, Vol.96, 96-DBS-109-18, pp.105-110, 下諏訪町, 1996 年 7 月.
4. Kenji Hatano, Qing Qian and Katsumi Tanaka
“Design and Implementation of A SOM-Based Information Organizer”
平成 8 年度第 2 回重点領域研究集会「高度データベース」松江ワークショップ, pp.185-194,
松江市, 1996 年 9 月.
5. 波多野 賢治, 亀井 俊之, 田中 克己
「映像自己組織化機構に基づく内容記述と類似シーン検索」
情報処理学会研究報告, Vol.97, No.64, 97-DBS-113-29, pp.173-178, 札幌市, 1997 年 7 月.
6. 佐野 綾一, 波多野 賢治, 田中 克己
「自己組織化マップを用いた Web 文書の対話的分類とその視覚化」
情報処理学会研究報告, Vol.98, No.57, 98-DBS-116(1)-5, pp.33-40, 芦原町, 1998 年 7 月.