

分布の類似性に着目したメッシュ地図からの意味検索システム

山島 一浩[†]

石塚 英弘^{††}

本論文では、土地利用の分布を表現したメッシュ地図を用いて、全体と構成項目別の2段階の類似度比較による検索の仕組みについて提案する。本システムは、社会科地理教育の教材用として利用されることを意図している。メッシュ地図の情報は、凡例の項目とその項目に対応したメッシュ単位の情報をそれぞれXMLで表現したものを利用している。メッシュ地図の位置情報は、データベースに登録されている。システムは、URIをもとにメッシュ地図の全体のメッシュ画像と凡例別のメッシュ画像を生成し、分類して画像を管理する。検索は、メッシュ地図を画像イメージに変換し、その類似度にもとづいて画像検索を行う。検索結果は、利用者に思惟の素材となる情報として利用されるよう意図する。本研究では、画像検索に2段階の分類方法を提案する。第1段階では、主題図を対象とした検索を行う。第2段階では、主題図を構成する要素項目を対象とした検索を行う。類似度は、ユークリッド距離と重み付きユークリッド距離を用いて結果を比較した。

A Semantic-structure-based IR System comparing Similarities of the Distribution of Land Use

Kazuhiro Yamashima[†]

Hidehiro Ishizuka^{††}

In this paper, we will propose an IR(information retrieval) system Comparing the degree of similarity among the mesh maps and the degree of similarity among their composing items of each mesh map by using the mesh maps expressing the distribution of then land use. We intend the system to be used as the teaching materials in geography under social studies education at junior high school level. The mesh map is written using XML format and expresses the information of the mesh units of the items of explanatory notes and their corresponding items. The URI (Uniform Resource Identifier) which is the location in formation of a mesh map is registered on a database. By using the URI, this system transforms XML-formatted information into a image data. The system converts a mesh map into an image, and can perform the search of an image based on the degree of similarity. We aim this system to be used by a user as a tool for speculation. In this study, we suggest two phases of the classification method of an image search. On the first phase, we perform a search for the subject map. On the second phase, we perform a search for the element items constituting a map of the subject. The degree of similarity is based on the comparison of the result of Euclidean distance and weighted Euclidean distance.

1. はじめに

地図やグラフなどの図的表現は、あらかじめ情報を観測し易くすることを目的にした表現である。人は、この表現された情報の中から意味ある情報を観測によって取り出すことができている。図的表現が、文や命題に比べて問題解決に有効な点を指摘している研究もある[1]。

しかしながら、大量の図的表現が存在し、相互の内容を比較して解釈を試みる状況、例えばWeb上をブラウジングして目的の図的表現を探すなど、人の視覚に依存したデータ入力と個々の知識背景に基づいた分類作業では、多くの時間と労力を要することになる。

本研究は、このような図的表現を対象とした意味検索システムの可能性について検討をすることを目的としている。

特に本稿では、図的表現の中で地図表現を対象とし、「文字」によるキーワード検索と同じような、分布図の面的な情報をキーワードとして、そこから地理的解釈可能な意味要素を取り出せる意味検索システムを開発するのが目的である。

本システムの適用対象として、地理学習で学校周辺の土地利用について他の地域と比較する場面を考えてみる。現場で行われている学習活動では、比較の対象を明確にするために調査する主題に合わせて地形図をもとに対象を強調して編集することで比較し易くする主題図編集を行うことがある。

このように比較する観点を絞ることで、学習者は、「地域Aと地域Bとを比較するとCの点が似ている」、「地域Dは、工業地域が、地域Eは、農村地域が特徴といえる」など、うまく見つけることができる。

しかしながら、多数の地域を比較するとこの方法では、学習者が主題図を比較することに要する時間は膨大となろう。また、学習者

[†]筑波学院大学情報コミュニケーション学部

Tsukuba gakuin University

^{††}筑波大学大学院

Graduate School of Library, University of Tsukuba

の知識背景により、比較しようとする観点がいまいであったりすると、比較することの困難さも付きまとう。

そこで、我々は、これまでの研究の中で開発してきたWeb上でメッシュ地図による主題図編集システムで作成され蓄積されたメッシュ地図を情報源として、その中から類似性に着目して似ていると思われるメッシュ地図をある程度の根拠説明を示し、取り出せるようにするシステムの開発についての研究をすすめてきた。

本稿は、まず対象とする図的表現のメッシュ地図の情報表現について述べる。次に、類似性の尺度として用いた手法について説明する。そして、本システムについて説明する中で、我々が開発したメッシュ地図編集システムの概要について述べ、そのデータが、本システムの中でどのように利用されているのかを説明する。最後に、実験としてサンプルデータを作成して、試行した検索結果について述べる。

2. メッシュ地図の情報表現

地図表現には、地形など地表面の情報を網羅した地形図とある特定の主題にもとづいて表現した主題図がある。主題図は、地形図をもとに主題にそった情報を観測したものを強調して表現したものである。地形図は、点・線・面の幾何表現で記述されているが、主題図表現においては、本来点で表現されたものが、線や面で表現されることがある。

浮田らは、地図表現に伴うデータの性格の変化を「データの転換」と呼び、主題図における「データの転換」の例として、ドットマップ、等値線図、メッシュ地図を紹介している [2]。

これによるとメッシュ地図は、面データを点データに転換する手法であり、面の形状や広狭を捨ててメッシュの位置だけを考察の対象として情報を残した表現である。これは、ドットマップや等値線図が点データを面に表現するのと対照的な表現である。

従って、メッシュ地図から意味的な解釈を行うことは、凡例で定義されている項目をもとにして、メッシュ単位に登録された情報が意味解釈を図る対象となり、強調された分布の広がり具合が一つの意味解釈をすすめる際の説明要因となり得る図的表現であるといえる。

3. 類似度と画像検索課題の定式化

本稿では、メッシュ地図の意味検索を行うために画像を検索対象のメディアとすることにし

た。メッシュ地図を画像で表現し、画像認識分野で用いられている手法を用いることを試みた。

人間が、大量の画像から比較する作業を行うには困難が伴うために、その支援策として、計算機を用いて機械的に効率よく抽出するための方法論がある[3][4]。

支援策としては、次に述べる点が挙げられる。

- 1) パターン認識技術を用いて画像解析プログラムによる類似度を計算する仕組みを準備する。
- 2) 類似度が算出された画像をあらかじめ意味あるまとまりとして分類できるようにインデキシングして蓄積しておく。

類似性を測る尺度である類似度は、画像認識の分野で、画像の中に埋め込まれた色・濃淡の情報を解析によって意味的な解釈が判断できるようにする処理である。

その識別方法は、特徴抽出した結果から、入力パターンと標準パターンとを比較する。そして最もよく類似したものを標準パターンと同じカテゴリーと見なすことで識別を行うものである。中でもパターンマッチングは、入力画像の中に比較するパターンと同じか、それに近いものを検出する方法である。

類似度の算出には、パターン認識で用いられる式(1)で示すユークリッド距離(Euclidean distance)と式(2)で示す重み付きユークリッド距離(weighted Euclidean distance)がある。

ユークリッド距離では、 f を未知ベクトル、 g を標準ベクトルとして、対象とするイメージとのパターンマッチングを行い、その距離 d を算出して得られた値から、最も近いイメージを類似性が高い画像とし、値を0~1の間で正規化したものを類似度として表す。

$$d^2(f, g) = \sum_{m=1}^K (f_m - g_m)^2 \quad \dots\dots\text{式(1)}$$

$$d^2(f, g) = \sum_{m=1}^K w_m (f_m - g_m)^2 \quad \dots\dots\text{式(2)}$$

w : 重み

本稿では、主題図間の類似度として、ユークリッド距離と重み付きユークリッド距離を使って図全体の類似度計算を行った。その他、後述するがハミング距離も用いている。

ここで課題の定式化について整理をする。

観測できる情報が画像データの場合、その画像の情報は、「画素(ピクセル)」からなる平面格子状の点と捉え、 n 次元のユークリッド空間と考えることができる。

そして、それぞれの点には、計測尺度として分類された土地利用等の凡例に基づく情報が付与されている。これらの点集合の部分集合を領域の意味空間として捉え、画像間の類似度計算を行い、類似性を説明因子として使用する。利用者には、類似度の高い情報順に結果を提示することで、利用者の比較と解釈の支援を行う。

4. メッシュ編集システムの概要

ここでは、メッシュ地図を作成する編集システム[5]について概要を述べる。ここで用いられているデータが、本システムで扱う画像データに関連があるからである。

編集システムが想定している利用目的は、小中学校の学習教材である。このため学校周辺の地域間比較という観点から、作成するメッシュ地図は、学校を中心としている。小学校の学習指導要領によれば、地域学習のような地理的な観点を養う段階として、学校周辺から段々と広い地域に目を向けるような授業展開が図られている。

ここでは、図1に示すような、日本の小中学校の位置を中心とした、3 km四方を網羅したメッシュ地図を作成することができる。図1の右側に凡例が表示されており、この凡例は利用者が自由に定義することができる。メッシュ地図の作成には、まずこの凡例を定義し、その凡例に従って地形図から読図を行い、メッシュに該当する色を塗る作業によって完成する。その成果として、図1の左側のようなメッシュ地図が表示できる。

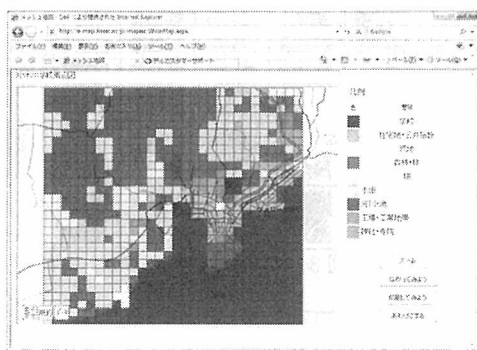


図1 メッシュ地図の例

このメッシュ地図のもとになる地形図は、国土地理院の電子国土Webシステムを利用して、このシステムで生成されるメッシュデータは、学校の所在地がほぼ中心となるように、計算されている。1区画が100mの長さで、

全体として30×30の3 km四方を網羅している。

このシステムで作成される凡例データとメッシュの色情報は、XML形式で登録されている。

表1がメッシュデータ、表2が凡例データの構造を示したものである。

表1 メッシュの基本構造

要素	属性
凡例番号	メッシュ番号

表2 凡例データ項目

要素	属性
経度・緯度	
タイトル	
凡例名	凡例番号
	R
	G
	B

本システムでは、この凡例データとメッシュデータを読み込んで画像データを生成する。

5. 意味検索システムの概要

本稿で述べる意味検索システムの概要について述べる。

本システムの構成を図2に、システムの実装環境と開発ツールを表3に示す。

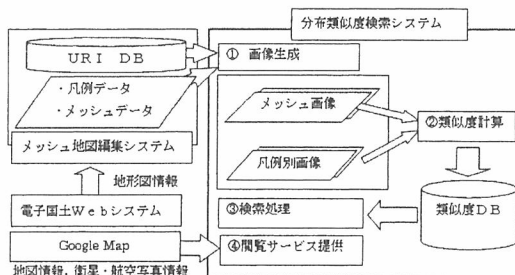


図2 意味検索システムの構成

表3 システムの実装環境と開発ツール

Hardware Spec	Dell PowerEdge600SC	
	CPU	Pentium®4 1.8GHz
	HDD	Raid-0 330G
	Memory	1G
	OS	Windows2000Server
Development Environment	Microsoft VB.Net	

システムでは、まずメッシュ地図編集システムに蓄積されているXML書式の凡例データおよびメッシュデータから画像検索の対象となる画像を作成する。メッシュ地図の所在情報であるURIは、メッシュ地図編集システムで編集

の際にURIが登録されるデータベースを利用した。

作成する画像は、Jpeg形式の画像で、サイズは、30×30ピクセルとした。メッシュ地図が30×30のメッシュで構成されているため、画像の1ピクセルが1メッシュと対応している。つまり、1ピクセルに対して、凡例の「畑」等の意味情報があらかじめ付与されており、画像全体または、凡例の中の特定の分布領域の類似度を算出することで、比較の解釈を試みる。

従って、画像は、メッシュ地図全体イメージを表現した画像と凡例別に記録した画像を作成している。

画像が生成された次の段階では、画像の類似度を計算し、データベースに登録する。登録した類似度の値は、75%～100%までを対象として残りは切り落とした。

データベースに登録されたデータと画像との識別には、画像のURI情報と凡例番号にもとづいてインデックスを付与した。

最後に利用者への提供サービスとして、画像の類似度に基づいた検索処理と、相互の画像を比較できるようにした閲覧サービスを作成した。

6. 実装と実験について

システムを稼働させ、編集システムを利用して、メッシュ地図を実験用に作成した。作成した場所は42ヶ所である。メッシュ地図は、地形図から判読できる特定の地目を10個選択した凡例を設定し、それに基づいて色情報を付与した。表4は、実験に用いた各図の凡例項目別のメッシュ数の一覧である。

図3は、表1と表2のXMLデータから生成した30×30ピクセルのJpeg画像である。

この画像1枚ごとに、図4に示す凡例別の2値画像が存在している。

2値画像を使ったのは、次の理由からである。

メッシュ地図を説明する際の意味情報は、凡例である。凡例の意味に基づいて比較する分には、相互の識別は容易であるが、ここでは、例えば「田」と「宅地」など、凡例間の総当たりマッチングを行うことにした。

この場合、色情報を用いていると、「田」と「住宅」等凡例間の領域比較に類似度が行えない。そこで、凡例別に図4のような2値画像を作成した。

色のパターンが白と黒の2種類に分かれる場合に定義される距離に、ハミング距離(Hamming distance)がある。これは、2つのパターンを対応させた際に、同位置にある成分が等価でない個数を求めたものである。凡例間

の類似度は、ハミング距離によって求めている。

表4 実験データの地目構成別メッシュ数

図	凡例1	凡例2	凡例3	凡例4	凡例5	凡例6	凡例7	凡例8	凡例9	凡例10
1	175	5	20	19	120	203	216	16	66	0
2	32	11	142	0	0	0	703	0	0	0
3	165	19	503	55	75	3	35	24	8	12
4	162	8	97	155	72	318	24	3	47	14
5	79	13	474	12	18	6	171	92	17	18
6	160	9	385	130	63	0	56	15	7	0
7	86	7	199	159	57	36	282	40	4	20
8	159	44	247	254	22	24	150	0	0	0
9	20	4	122	58	0	504	66	33	10	0
10	0	1	10	7	428	325	56	52	21	0
11	9	7	81	155	329	12	275	19	8	5
12	3	20	288	54	75	336	111	0	15	0
13	8	1	57	319	29	328	124	30	4	0
14	38	2	33	30	272	171	198	139	4	7
15	6	11	150	82	143	27	160	260	33	19
16	9	0	167	35	241	25	309	82	8	19
17	7	7	97	57	237	05	347	67	9	7
18	3	10	320	2	0	0	441	77	17	30
19	8	8	64	21	282	70	151	276	4	14
20	0	17	472	56	1	100	38	26	47	44
21	3	6	155	45	59	4	507	80	12	29
22	3	17	496	19	36	15	231	32	39	12
23	33	27	568	4	4	32	7	45	161	18
24	0	6	132	280	0	3	371	66	22	20
25	1	9	174	23	155	05	338	55	50	10
26	0	15	386	0	0	8	323	32	100	33
27	1	15	236	15	242	2	278	54	23	34
28	7	15	267	8	74	33	359	36	05	35
29	0	5	87	15	208	82	88	419	0	26
30	1	37	617	3	0	14	196	0	10	22
31	8	8	93	29	147	285	215	113	0	4
32	4	13	320	21	282	26	88	111	8	7
33	0	4	102	0	368	4	282	131	0	9
34	0	4	95	4	404	108	184	101	0	0
35	0	3	9	13	633	1	0	221	0	6
36	23	23	343	12	416	0	0	49	15	19
37	16	2	12	47	398	28	0	396	6	0
38	18	3	12	8	611	16	58	170	0	4
39	5	16	274	5	0	4	389	148	17	42
40	5	12	268	20	180	26	242	116	7	24
41	5	34	440	0	143	0	0	68	10	200
42	24	6	171	12	373	15	158	85	32	24

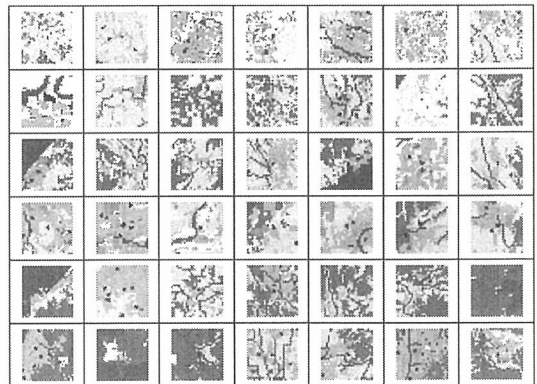


図3 実験画像データ

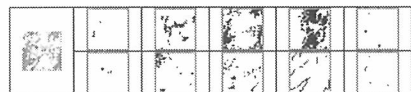


図4 凡例別に分解した2値画像

6. 1. メッシュ地図画像間における検索

主題図対象42ファイルに対する類似度を計

算した結果、ユークリッド距離では172組、重み付きユークリッド距離では205組の組み合わせが出力された。

その特徴として、重み付きユークリッド距離では75.0%から77.8%の間で多く存在している。

図5が、左側の画像と比較した中で類似性が高い画像の一覧である。この場合3つの画像が選択されたことになる。





	1	2	3
			
	79.2	78.6	77.8

図5 類似度の例

しかし、この情報だけでは、どのような点で似ているのかを示す情報が提供されていない。

そこで、凡例別の類似度結果をもとに、メッシュ地図画像の類似度に対して、凡例別の類似度順を示す図6のような一覧表を作成し、検索結果を凡例別に比較できるようにした。

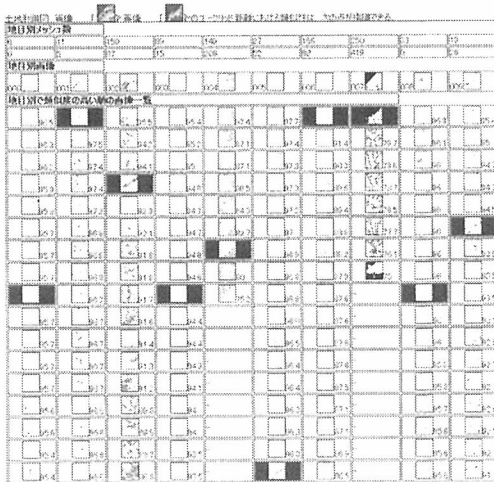


図6 凡例別の類似度一覧表示画面

この例では、最も高い79.2%の画像との比較例である。地目別で類似度が高い順に並んでいる一覧で、塗りつぶされた個所が比較する画像との地目別類似度の順位となる。ここでは、3つの凡例で他の画像より高い値を示していることを示している。これらの地目は、「学校」「田」「海」である。

「学校」の位置は、メッシュの中央の位置には必ず配置されるようになっていることや学区という区分により、公立、私立、種別の違いによる配置もあろうが、凡例の中でも類似度一覧に

出現し易くなっている。二つ目は、「田」である。「海」の配置が似ている点から、少し内陸側に同じように分布していることがわかる。最後に「海」であるが、視覚的にもその類似性が確認できる例である。

6. 2. 凡例間における検索

2値画像に対して、凡例間の総当たりのマッチングを行い、類似度計算を行った結果54017組が検出された。

例えば、図7の上段の図に類似した図を検索対象しようとする場合、地目別に分解された画像を利用して類似度計算を行い、中段に示された類似性の高い一覧が出力される。類似度計算結果の最上段が検索対象の凡例別画像であり、左側が「田」、真中が「海」、右側が「神社・寺院」である。画像の右側に出ている数値が類似度である。

その結果、下段の検索結果として3つの画像がその類似性の高いメッシュ地図として出力されたことになる。

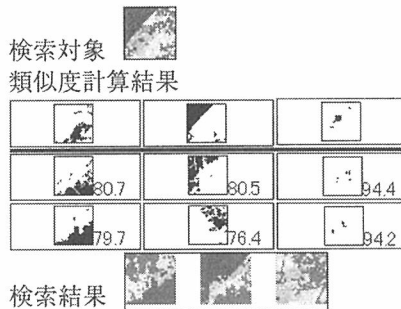


図7 凡例別の類似度計算結果

検索結果で、類似性が一番高い左側の画像を図8で詳細に表示してみる。視覚的には「海」がある点で類似しているということがわかるが、全体をもとにした検索のみで凡例別の検索を行わない場合には、2つの画像の類似性は確認できなかった。しかしながら、凡例別で検索してみると、「田」と「海」との類似性があるとの結果が出ている。

人が視覚的に確認できる「海」があるとの結論にはならないが、2段階検索を行うことで、1段階では、見落としそうな類似性を欠落させないことに成功している。

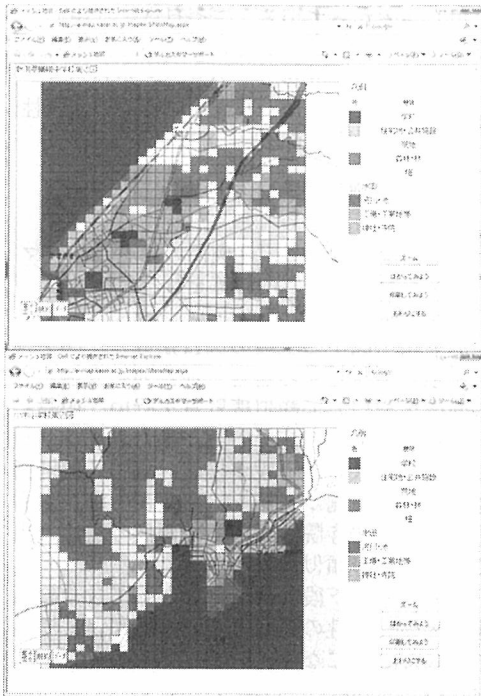


図6 メッシュ地図上での比較

7. 閲覧サービス用のコンテンツ

この実験をもとに、利用者向けの類似性を説明する閲覧サービスとして実装したコンテンツについて述べる。

図7がその初期画面である。比較を試みる2つの画像を選択し、クエリ送信ボタンを押す。



図7 閲覧サービス画面

図8は、その結果としてGoogleの地図表示APIを利用して、2つの地域間の位置の比較と類似度による結果説明を表示した画面である。ここでは、2つの地図を表示させてそれぞれの位置が確認できるようにしている。

Googleで提供されているAPIを使い、

衛星画像と切り替えながら、現況を確認できる。

その下には、全体画像間の類似度を表示させており、類似度が75%以上の値の場合には、その値が表示される。

その下が、凡例別の類似度を一覧にして表示したものである。その際に、表4の凡例別のメッシュ数も比較して、比較的近い値のペアを左に、個数に違いのあるものを右側に表示するようにした。

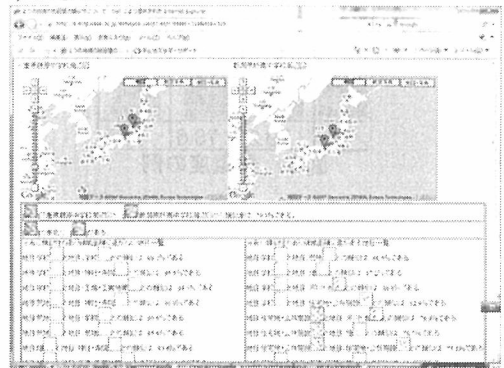


図8. 検索結果表示画面

類似性のある項目について複数の見せ方を提供することにより、閲覧者の比較の支援を行うことをねらう。

8. まとめ

類似性にもとづくメッシュ地図間の比較によって、地理的なコンテキスト生成を行う支援策について論じた。画像の類似性に着目し、図と凡例の表現領域の2段階検索を行うシステムを開発し、実装した。

本研究は科研費(19500816)の助成を受けたものである。

参考文献

- [1] Larkin, J.H. and Simon, H.A., "Why a Diagram is (sometimes) Worth Ten Thousand Words," *Cognitive Science*, Vol.11, 1987, pp.65-99.
- [2] 浮田典良, 森三紀: 地図表現ガイドブック, ナカニシヤ出版, 2004
- [3] 末松良一, 山田宏尚: 画像処理工学, コロナ社, 2000.
- [4] 鳥脇純一郎, 認識工学—パターン認識とその応用, コロナ社, 1993
- [5] Kazuhiro Yamashima, Hidehiro Ishizuka: "Development of an e-learning system using Cyber Japan Map System for an elementary school education", *Proc. of the 7th China-Japan Joint Conf. on Graphic Education*, p.234-240, 2005.7, Xi'an.