

修士論文

医療文書検索のための  
シェーマ認識における  
誤認識検出法に関する検討

平成 28 年度修了

三重大学大学院工学研究科

博士前期課程 電気電子工学専攻

森山 浩平

# 目次

第1章	はじめに.....	3
1.1	研究の背景.....	3
1.2	本研究の目的及び概要.....	6
第2章	実験試料.....	7
2.1	院内で使用されている医療文書.....	7
2.2	本研究で用いたシェーマ画像.....	7
第3章	シェーマ認識.....	10
3.1	シェーマ認識の概要.....	10
3.2	前処理.....	11
3.3	特徴量抽出法.....	12
3.4	識別関数.....	14
第4章	誤認識の抽出手法.....	15
4.1	誤認識結果の特徴.....	15
4.2	用いる相違度に関する事前検討.....	17
4.3	提案手法.....	18
第5章	評価実験.....	19
5.1	実験の概要.....	19
5.2	実験結果と考察.....	20
第6章	まとめと今後の課題.....	23
6.1	本研究のまとめ.....	23
6.2	今後の課題.....	23
	謝辞.....	24
	参考文献.....	25
	発表論文リスト.....	28

## 図一覧

1.1	シェーマを含んだ医療文書の一例.....	5
2.1	シェーマとアノテーション.....	8
2.2	本研究にて使用したシェーマ.....	8
2.3	アノテーションの例.....	9
2.4	実験試料.....	9
3.1	シェーマ認識のタグ付けの一例.....	10
3.2	シェーマ認識の流れ.....	11
3.3	前処理の結果.....	12
3.4	輪郭線の方法.....	13
3.5	方向指数ヒストグラムの作成.....	13
3.6	ガウス形分布の空間フィルタ.....	13
4.1	誤認識したシェーマの一例.....	15
4.2	正しく認識された場合の相違度.....	16
4.3	誤認識された場合の相違度.....	16
4.4	しきい値を用いた誤認識抽出法.....	16
4.5	SVM を用いた誤認識抽出法.....	18
5.1	パラメータ $k$ を変化させた時の認識率の変化.....	19
5.2	提案手法における認識率と誤認識抽出率（カテゴリ 1～10） .....	21
5.3	提案手法における認識率と誤認識抽出率（カテゴリ 11～20） .....	21
5.4	提案手法における認識率と誤認識抽出率（カテゴリ 21～30） .....	22
5.5	誤認識検出において SVM の構築に失敗したカテゴリの例.....	22
5.6	提案手法によって誤認識が検出されたシェーマの例.....	22

## 表一覧

4.1	しきい値を用いた誤認識抽出の結果.....	17
-----	-----------------------	----

# 第 1 章

## はじめに

### 1.1 研究の背景

近年、病院情報システム(Hospital Information System : HIS)の普及に伴い、カルテやレポートをはじめとした様々な診療記録が電子化されつつある[1-5]。しかしその一方で、病院情報システムの導入以前の医療文書は、主に紙媒体のまま保管されている。これらの過去の医療文書は、症例や治療内容、またその結果や患者の予後まで記載されているため、治療方針の決定や効果の確認等の臨床研究に役立てることができる貴重なデータベースである[6]。しかしながら、紙媒体であることから、類似症例検索などの二次活用が難しい。一方、これらの文書を電子データに変換する試みも報告されているが、文書量が膨大であるために手作業による入力では時間と費用がかかり、結果として紙文書の電子化は積極的に進められていない。そのため現在、院内の紙文書を利活用するためのシステムの開発が熱望されている。

このような要望に応えるため、多くの医療機関では診療記録のスキャニングや診療情報の利活用に関する研究や試みが数多く報告されている[7-11]。例えば文献[7]では、業務プロセスを通じて発生する多種多様で多量なデータのソースからデータ収集・統合・蓄積し、分析・可視化するデータウェアハウスの必要性について詳しく紹介されている。また文献[10,11]では、電子化した退院サマリからテキストマイニングによって患者別の重要語を抽出し、これをもとに退院サマリの文書から疾患名を特定できるかについて検討している。これまでに筆者の研究室においても、紙媒体で保存されている表形式医療文書を対象とした文書構造認識と文書内容の XML 化、ならびに類似症例検索システムの構築に関する研究を進めてきた[12-17]。例えば文献[17]では、文字認識において広く用いられている加重方向指数ヒストグラム法と疑似ベイズ識別関数を用いて、図 1.1 のような医療文書に含まれるシェーマ画像（身体の部位をかたどった線画）の認識手法を提案した。これら従来研究では、実際の病院にて使用されているシェーマ画像を用いた評価実験の結果、認識率を正しく認識した数を認識に用いた総数で除したものと定義した場合、90%を超える認識率が得られており、シェーマ情報の利活用の可能性についても示唆されている。しかしながら、シェーマ認識の際、識別機に用いる疑似ベイズ識別関数では、辞書画像との比較を行い、最も相違度

が低いカテゴリに分類するという特徴があるため、誤認識された場合でも必ずどこかのカテゴリに属してしまうという問題が存在している。そのため、利用者は求めているシェーマに関連する診療記録のデータを確実に得ることができない。その結果、ユーザが実際に求めている診療記録とは関係のない診療記録が検索結果として提示されてしまう可能性がある。医療現場における文書検索において、このような本来の目的とは異なる文書が抽出されることは、システムの信頼性の観点からも避けるべきものであるとともに、ユーザにとっても不便である。このような背景から、現場からの声として、認識率が落ちることを許容しつつ、このような誤認識をなくす、言い換えれば高い確度で誤認識を検出したいという要望が高い。

5

## 1.2 本研究の目的及び概要

そこで本研究では、シェーマ認識における現場からの声を考慮したシステムのために、誤認識を検出する方法について検討する。ここでは、従来法にて用いられている擬似ベイズ識別関数から得られる相違度を用いて、サポートベクターマシン（Support Vector Machine : SVM）による誤認識検出法を提案する。実際の医療現場にて用いられているシェーマを用いた評価実験を行い、提案法の有効性について検討する。

本論文の構成を以下に示す。本論文ではまず、第2章で実験試料となる医療文書におけるシェーマ及びアノテーション概要について述べた後、第3章にて、従来研究におけるシェーマ認識法について簡単に紹介する。第4章では、第3章で述べたシェーマ認識の結果を用いた誤認識の検出方法を提案する。第5章では実際の病院で用いているシェーマ画像を用いた評価実験の評価・考察を行う。最後に第6章にて、本論文の内容のまとめと今後の課題について述べる。

## 第2章

# 実験試料

### 2.1 院内で使用されている医療文書

前章にて示した図1.1は院内で用いられている文書の一例であり，退院サマリと呼ばれる．退院サマリは退院患者の入院中の状況を正確，明瞭，簡潔にまとめた情報が記載されており，今後の診療や疾患に関する情報の管理，及び医学教育などに用いられる文書である．また，入院担当医と退院後の外来診療担当医との情報伝達媒体としても機能する重要な医療文書である．例えば，入院担当医は退院サマリに患者の入院理由や治療経過，検査結果などに加えて患者の退院時診断や退院後の継続治療，経過観察期間，在宅ケアなどの指示についても退院サマリに記録する．一方，外来診療の担当医は退院サマリから患者の入院診療経過と病状を把握することにより，入院から外来に至る継続治療が円滑に行われる．

また，退院サマリをはじめとする医療文書にはシェーマと呼ばれる人体・骨格・臓器などの身体部位を象った線画が含まれていることがある．一般的に，シェーマは視覚的に患者の疾患部位の詳しい位置や形状を記録することができるため，多くの医療文書で用いられている．また，シェーマ中には図2.1に示すように，アノテーションと呼ばれる診療担当医による注釈が加筆されており，そこには疾患の部位やその症状など，患者の疾患に関する詳細な内容が加筆される．三重大学医学部附属病院では，月に2400枚の医療文書がスキャンされているが，その中にシェーマ付きの医療文書は800枚程度である．

### 2.2 本研究に用いたシェーマ画像

本研究では，三重大学医学部附属病院にて使用されているシェーマを用いることとした．ここでは，これらのシェーマから典型的かつ頻繁に使用されている30種類(図2.2)を対象とし，図2.3のように，対象となるシェーマ画像にそれぞれ50種類のアノテーションを加筆した．ここでは，個人情報保護の観点から，実際の文書中に現れるシェーマ画像は使用せず，擬似的に作成したものを用いた．具体的には，学生15人に対して実際の医療文書に書かれているシェーマ画像を提示し，それらの画像を参考にア



ノテーションを記入してもらった。なお、実験材料の作成に関わった学生はカルテ記載を日常的には行っていないが、アノテーションの書き方について知識のある者の指導のもとで作成を行っているため、実験には大きな影響はない。また、画像を作成した学生はアルゴリズムの開発には関与していない。スキャン作業は病院職員が全て手作業で行っているため、スキャン時に発生する画像の傾きを考慮し、アノテーションを加筆したシェーマ画像に対してそれぞれ-2, -1, 1, 2度の傾きを加えたものを用いることとした（図2.4）。本論文では、上述の処理により作成されたアノテーション付きのシェーマ画像計7500枚を実験材料とした。これらは光学スキャナを用いて、解像度300dpi, 24bit/pixelの画像ファイルとした。

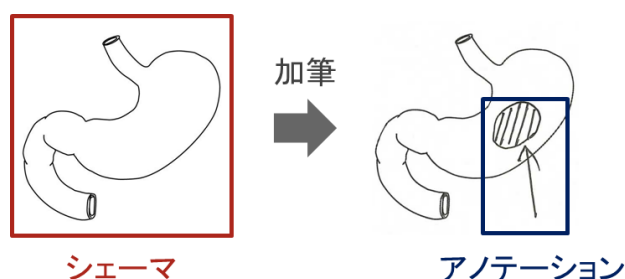


図2.1 シェーマとアノテーション



図2.2 本研究に用いたシェーマ

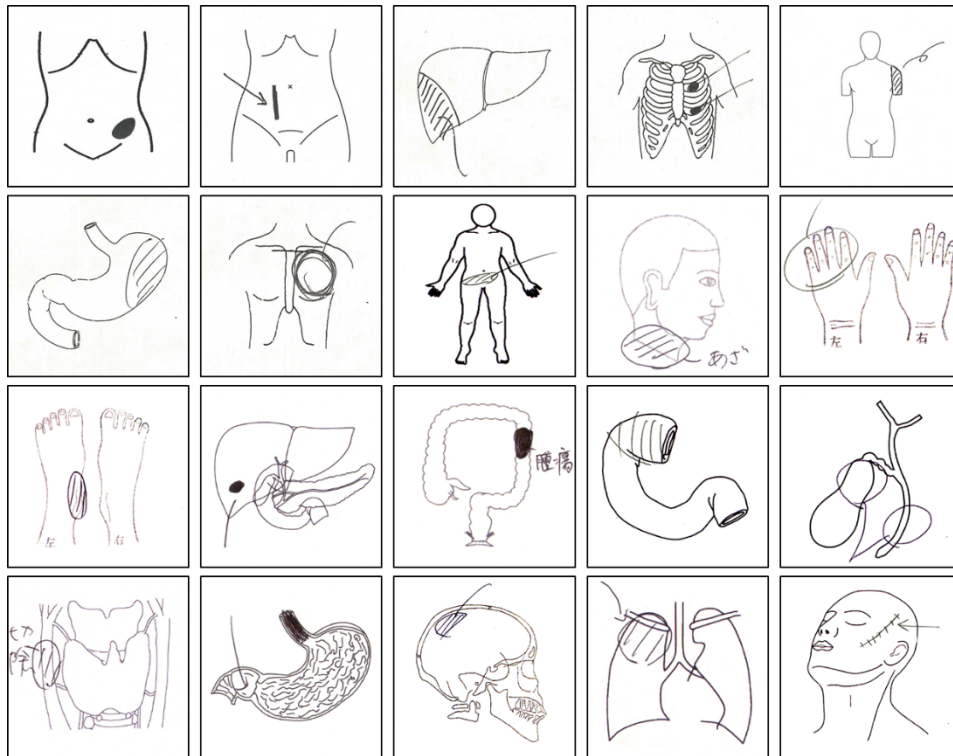


図2.3 アノテーションの例

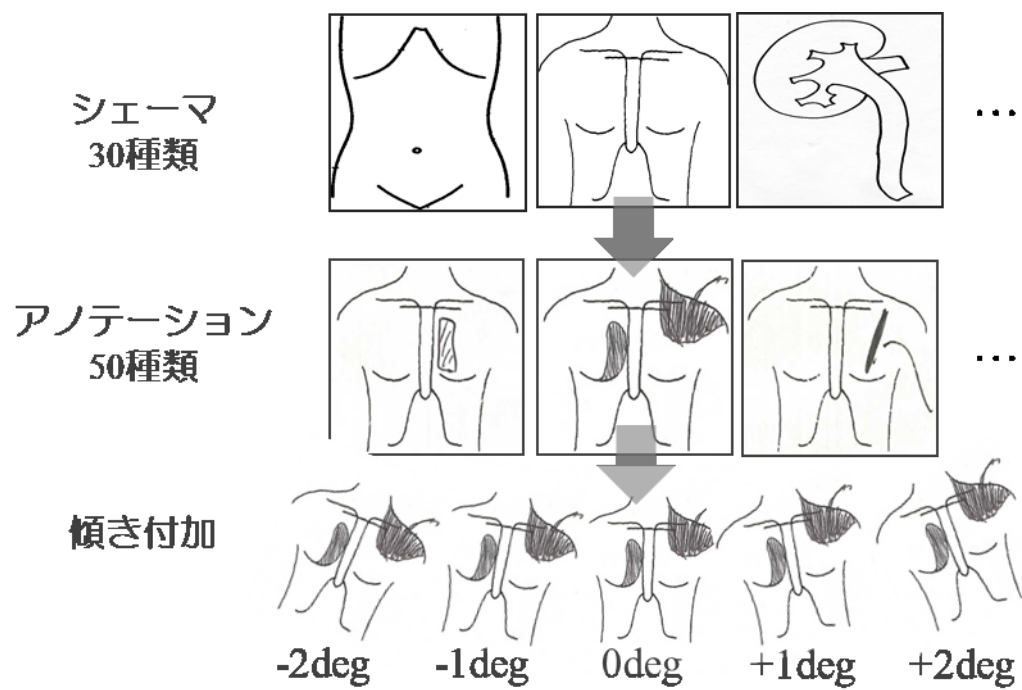


図2.4 実験試料

## 第3章

# シェーマ認識

### 3.1 シェーマ認識の概要

前章で述べたように，シェーマには疾患部の位置情報や注釈・コメントなどがアノテーションとして医師により加筆される．また，シェーマやアノテーションは視覚的に患者の疾患部位に関する情報が記録することができるため，図 3.1 に示すようにシェーマの持つ情報をタグとして文書に付与することができれば，類似症例検索の際の有効なキーとして活用できる可能性が高い．本章では，シェーマのタグ付けの第一段階として，医療文書から切り出されたシェーマ画像が身体のどの部位を表した種類のシェーマであるかを認識し，タグ付けを行う手法について述べる．

図 3.2 に従来研究におけるシェーマ認識手法の流れを示す．従来手法では，あらかじめアノテーションが加筆されたシェーマ画像から特徴量を抽出し，認識用辞書画像を作成する．シェーマ認識における特徴量抽出には，加重方向指数ヒストグラム法を用いる．加重方向指数ヒストグラム法により入力画像の特徴ベクトルを作成し，疑似ベイズ識別関数を用いて認識用辞書画像の特徴量との相違度を得る．得られた相違度の中で最も相違度が小さい，すなわち類似しているシェーマの種類を認識結果として出力する．次節以降では，認識手法の詳細について述べる．

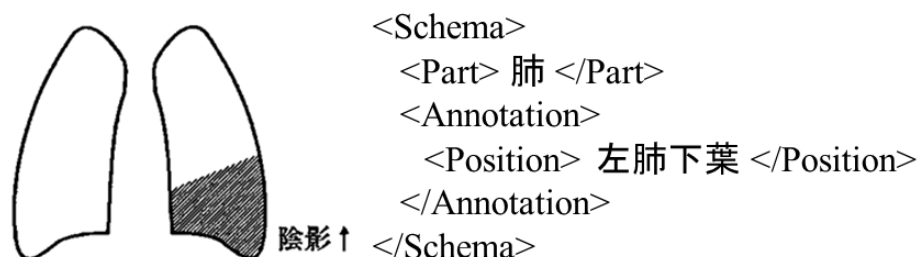


図3.1 シェーマ認識のタグ付けの一例

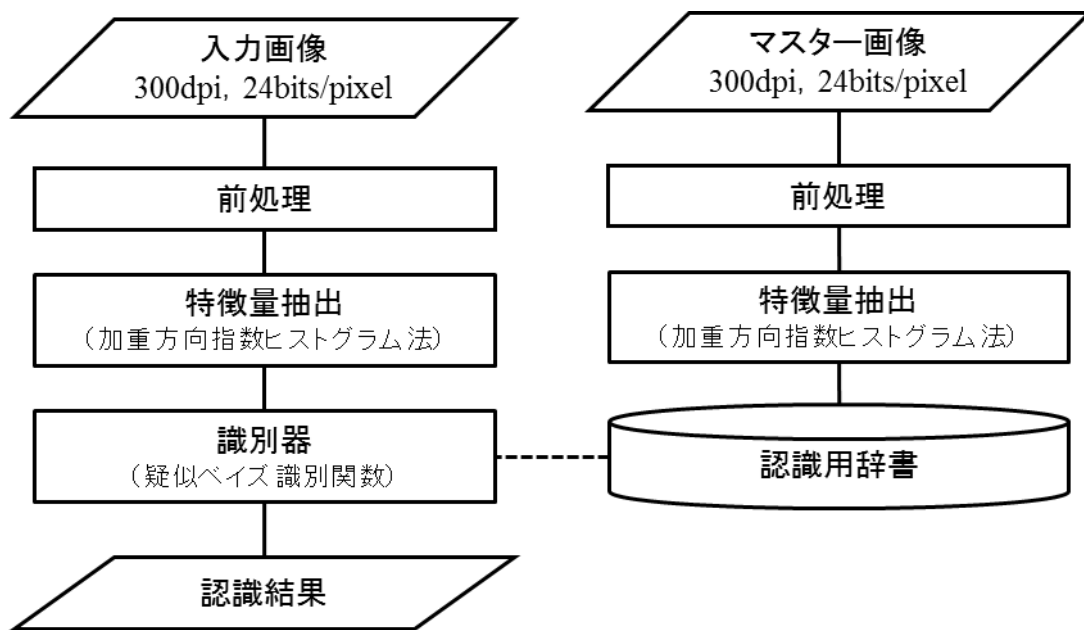


図3.2 シェーマ認識の流れ

## 3.2 前処理

### 3.1.1 ノイズ除去

従来法ではまず、使用するすべての画像に対して前処理として二値化、ノイズ除去及び正規化を行う。ここでは、入力画像に対して動的に二値化のしきい値を決定することが可能な大津の二値化を用いることにより二値化処理を行う[19][20]。この方法を用いることにより、固定しきい値を割り当てることなく、各画像に応じて自動的にしきい値を決定することが可能となる。次に、スキャナで画像を取り込む際にノイズとして現れるゴマ粒状のノイズを除去するために、メディアンフィルタを用いてノイズを除去する。ここでは、フィルタ処理時間を短縮させるために、対象となる画像が二値画像であることを利用したノイズ除去を行った。具体的には、 $3 \times 3$ の局所領域内の黒画素がその領域の画素数の半分以上であれば中央値を黒画素とし、その条件が満たされない場合は中央値を白画素とした。

### 3.1.2 位置及び大きさの正規化

次に特徴ベクトルを作成するために、入力画像を正規化する。ここではまず、入力画像をシェーマ本体とアノテーションを含む外接矩形でトリミングし、それをあらかじめ決定した大きさの正方形に拡大・縮小する。これにより、入力画像を一定の大きさに揃え、大きさの違いによる認識結果への影響を軽減することが可能となる。なお本論文では、一般的なシェーマ画像をスキャン時の大きさを考慮し、 $350 \times 350$ [pixel]の

大きさに正規化する．図3.3に，以上の前処理（二値化，ノイズ除去及び正規化）を行った結果を示す．

### 3.3 特徴量抽出

従来法では，前処理を施した画像から特徴量を抽出するために，加重方向指数ヒストグラム法を用いた．加重方向指数ヒストグラム法は文字認識における特徴量抽出法の一つであり，手書き文字認識エンジンの基本アルゴリズムとして用いられている[21,22]．同法ではまず，正規化された入力画像に対して輪郭部を左回り（穴部は右回り）に追跡し，各境界点で輪郭線の方角を16方向に量子化した方向指数を求める．次に16方向を8方向に集約する(図3.4)．このときの重み係数は同一方向のとき1，両隣のとき0.5，それ以外のとき0である．さらに互いに反対方向の向きを同一とみなし，最終的に4方向に集約する．次に，図3.5に示すように，シェーマ画像を7×7の小領域に等分割し，それぞれの小領域における方向指数ヒストグラムを作成する．その後，入力画像による，特徴量の変動を吸収するために，図3.6に示すような空間フィルタを用いて，求めた方向指数ヒストグラムを空間的および方向的にぼかす．以上の手順により，64（4×4×4）次元の特徴ベクトル（加重方向指数ヒストグラム）を得る．

このように，加重方向指数ヒストグラム法は文字線分の局所的な位置変動に対してロバストであるとともに，文字線分の輪郭線（線素構造）を考慮した手法である．カルテをはじめとする医療文書に用いられているシェーマは，文字同様に様々な方向成分を持つ線分の集合である．そのため，パターン認識の観点から推察すると，これらの画像は互いに類似した特性を持つものと考えられる．そのためここでは，抽出されたシェーマ画像を認識するために，加重方向指数ヒストグラム法が用いられている．



図3.3 前処理の結果

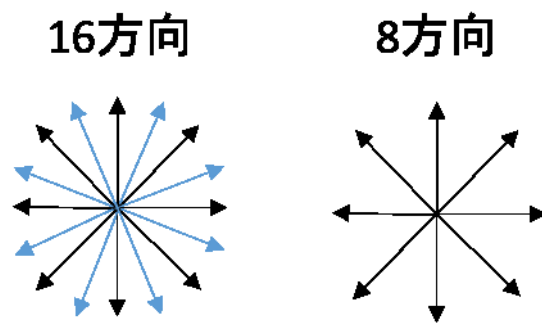


図3.4 輪郭線の方角

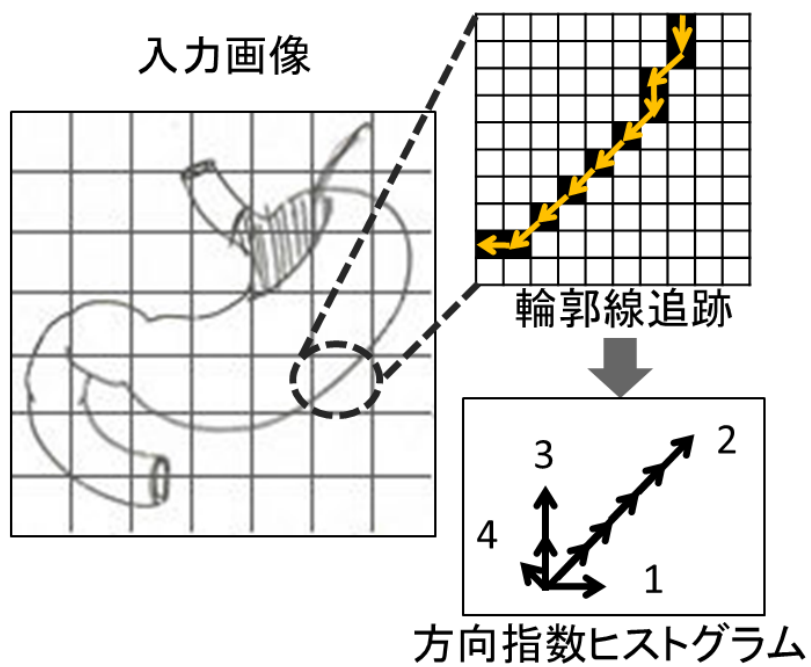


図3.5 加重方向ヒストグラムの作成

.0	.009	.017	.009	.0
.009	.057	.105	.057	.009
.017	.105	.194	.105	.017
.009	.057	.105	.057	.009
.0	.009	.017	.009	.0

図3.6 ガウス形分布の空間フィルタ

### 3.4 識別関数

3.3 節の方法により得られた特徴ベクトルは，識別関数を用いて認識用辞書の各シェーマカテゴリとの相違度が計算され，相違度が最小となるカテゴリが認識結果として出力される．従来法では識別関数として，以下の式で与えられる疑似ベイズ識別関数を用いている[22].

$$g(x) = \sum_{i=1}^k \frac{\{\Phi_i^T(x - \mu)\}^2}{\lambda_i} + \sum_{i=k+1}^n \frac{\{\Phi_i^T(x - \mu)\}^2}{\lambda_{k+1}} + \ln \left( \prod_{i=1}^k \lambda_i \cdot \prod_{i=k+1}^n \lambda_{k+1} \right)$$

ここで， $x$  は入力画像の特徴ベクトル， $\mu$  は辞書の平均ベクトル， $\lambda_i$  と  $\Phi_i$  は辞書の共分散行列の固有値と固有ベクトル， $k$  は使用する固有値の数， $n$  は特徴ベクトルの次元数， $x$  は入力文字の特徴ベクトル ( $n$  次元) である．

疑似ベイズ識別関数は，従来から様々なパターン認識の分野で使用されているベイズ識別関数（二次識別関数）を改良したものであり，加重方向指数ヒストグラム法と共に使われることが多い．ベイズ識別関数は対象が正規分布に従い，特徴ベクトルの平均，共分散行列が既知の場合，最適識別関数となる．梶原らはこれらのシェーマ認識手法を用いることにより，90%を超える認識精度を得ている[23].

## 第4章

# 誤認識の抽出手法

### 4.1 誤認識結果の特徴

本研究で対象としているシェーマの形状はあらかじめ決まっているため、多くの点で文字と類似した特徴を持つと考えられる。しかしながら、医師によりアノテーションが加筆されているという点において、文字とは異なる。パターン認識の観点から考えると、アノテーションが加筆されるということは、シェーマの形状を大きく変えることを意味するため、シェーマ認識結果に悪影響を与える。つまり、アノテーションが大きく加筆されるほどシェーマの形状が大きく変化するため、認識に与える影響が大きくなる可能性は高い。図4.1に従来法において誤認識されたシェーマ画像の一例を示す。図からも分かるように、これらのシェーマ画像には比較的大きなアノテーションが加筆されており、これらアノテーションにより、特徴量が大きく変化し、その結果、誤認識されたものと考えることができる。言い換えれば、従来法における識別関数により算出される相違度を何らかの形で利用することにより、誤認識を検出できると考えられる。そこで本研究では、事前実験として、従来手法により得られた実験結果の辞書画像と入力画像の比較により得られる相違度を算出した。図4.2と図4.3に算出した相違度をプロットしたものを示す。図4.2は正しく認識された場合、図4.3は誤認識

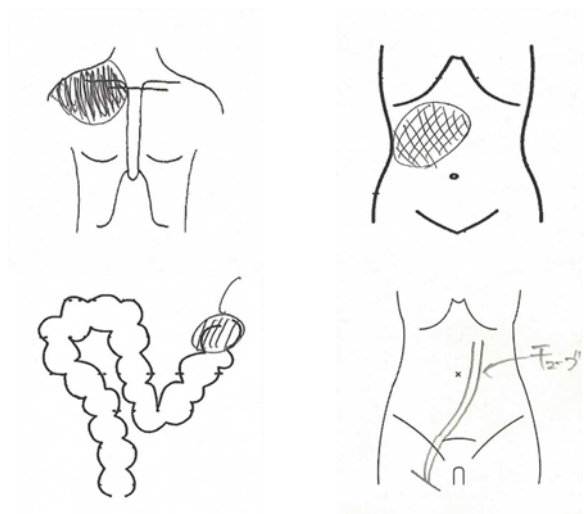


図4.1 誤認識したシェーマの一例



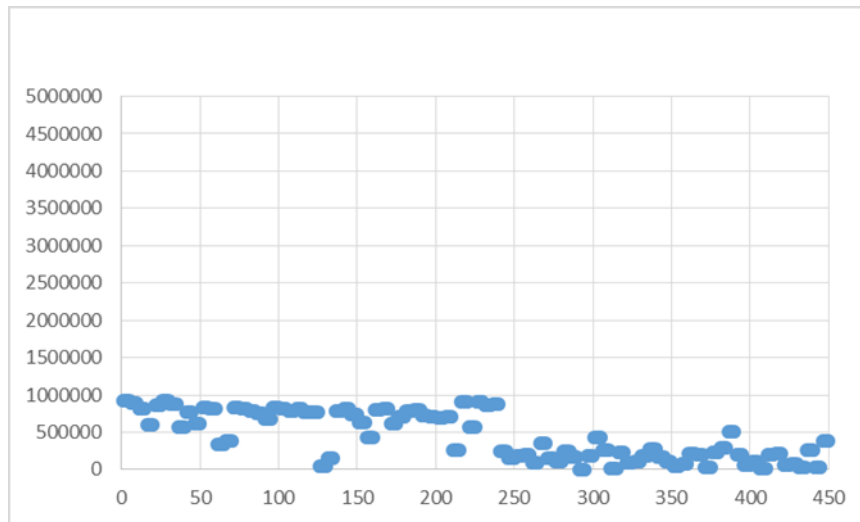


図4.2 正しく認識された場合の相違度の分布

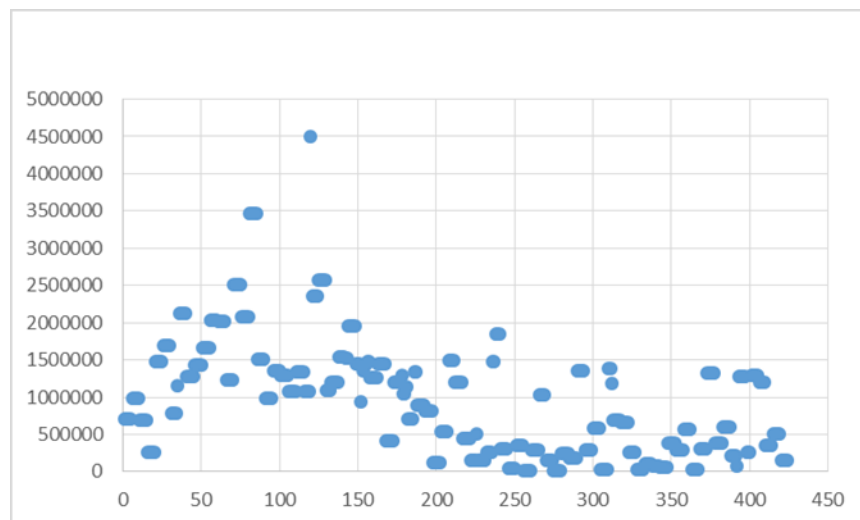


図4.3 誤認識された場合の相違度の分布

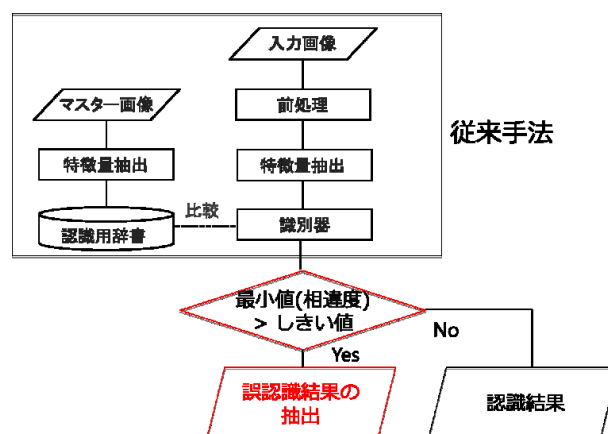


図4.3 しきい値を用いた誤認識抽出法

された場合の相違度の分布を表している．図から分かるように，正しく認識された場合と比べて誤認識された場合では全体的に相違度が高い傾向にあることが分かる．上記の内容を考慮し，次節では，従来の疑似ベイズ識別関数により得られる相違度を用いることとした．

## 4.2 用いる相違度に関する事前検討

前節にて示したように，誤認識の場合には画像の形状が大きく変わる傾向があることから，疑似ベイズ識別関数により求められる相違度も高くなる．そこで本論文ではまず事前検討として，検証実験で疑似ベイズ識別関数により得られる相違度に対してしきい値を設定して誤認識を抽出する試みを行った．図4.3に誤認識検出の流れを示す．ここでは，従来手法によって得られた認識結果の辞書画像と入力画像を再比較して相違度を算出し，カテゴリ毎にしきい値を設定して誤認識を抽出した．本節ではまず，事前検討として病院で用いられているシェーマの中から特に使用頻度が高いと思われる10種類のシェーマ画像2500枚を用いた．その後，使用するシェーマ画像を30種類に増やし，計7500枚の画像を用いた．

表4.1にしきい値を用いた誤認識抽出実験によって得られた誤認識抽出率，認識率を示す．本論文では，誤認識抽出率と認識率を以下のように定義する．

- 認識率 = (誤認識を誤認識とした数 + 正認識を正認識とした数) / (全認識数)
- 誤認識抽出率 = (誤認識を誤認識とした数) / (誤認識の総数)

また，実験では誤認識の検出精度に関する検討という観点から，各カテゴリ内において誤認識の抽出率が最も高くなる相違度の値をしきい値として手動で設定した．

表4.1の結果から分かるように，各カテゴリにおいて低めのしきい値を設定することにより誤認識を100%検出することができる一方，枚数やカテゴリ数が多くなると認識精度が著しく低下してしまう．このような結果が得られる原因として，全ての誤認識

用いた枚数	2500枚	7500枚
誤認識検出率	100%	100%
認識率	87.7%	76.9%

表4.1 しきい値を用いた誤認識抽出の結果

結果において相違度が高くなるとは一概に言えず、その値を一意に設定することは難しいためであると考えられる。その結果、サンプル数を多くすると認識精度が落ちてしまったものと推察される。これらの結果を勘案すると、認識率の何らかの方法を用いて動的にしきい値を設定する等のアプローチが必要となる。

### 4.3 提案手法

前節で述べたしきい値を用いた手法では、サンプル数が増えていく現場において活用するには難しく、認識率をより向上させる必要がある。今回の検証実験では、疑似ベイズ識別関数によって得られる最も相違度が低い値を用いていたが、認識結果を精査した結果、誤認識しているものは、上位1～10位に正しい認識結果が含まれており、それらの類似度の分布が特徴的である等の理由から、これらを10次元データとして誤認識の抽出に用いることとした。これらの調査内容を踏まえ、図4.4に提案する誤認識検出法の概要を示す。今回の提案手法における誤認識検出部では、一般的に高次元の二値分類に多く用いられているSVMを用いることとした。また、誤認識の場合の相違度が各カテゴリによって大きくレンジが異なるなどの特徴があることから、検証実験と同様に、各カテゴリにおいて正誤認識を検出するための識別器を作成することとした。

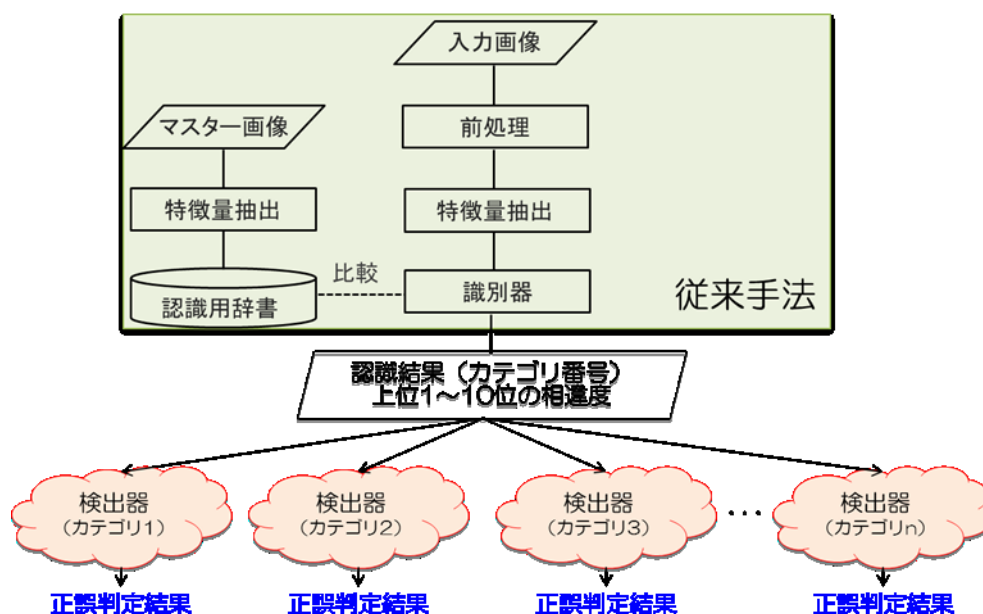


図 4.4 SVM を用いた誤認識抽出法

## 第5章

# 評価実験

### 5.1 実験の概要

本論文では、疑似ベイズ識別関数を用いて求められた入力画像と辞書画像との相違度を特徴量とし、SVMを用いた誤認識検出に関する評価実験を行った。実験材料には、2章にて述べた30種類、7500枚のシェーマ画像を用いた。なお、提案法では加重方向ヒストグラム法における領域分割数は従来手法により16 (4×4) が最も認識に適していることが分かっているため、本論文でもこの分割数を用いた。また、疑似ベイズ識別関数におけるハイパーパラメータ $k$ については、事前実験の結果から最適値は $k=11$ であることが判明しているため、この値を用いた(図5.1)。

本実験では、これらの7500枚のシェーマ認識の結果より生成された特徴ベクトルに対し、SVM (Support Vector Machine) を適用した。実験では、各カテゴリにおいて全データの10%をテスト用とし、残り90%をモデル構築とパラメータフィッティング(学習用)に用いた。学習用データを3分割し、そのうちの一つをモデル構築に、残りの二つをテスト用と評価用とし、3-fold Cross Validationにより、モデル構築およびパラメータフィッティングを行った。本論文では上述の処理を10回行い、その平均を用いて全30カテゴリ分の検出精度を評価した。なお実験用のプログラム作成にはR(version3.3.2)を用いた。

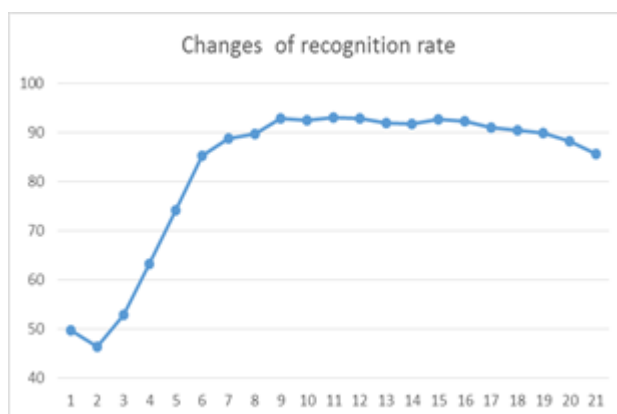


図5.1 パラメータ $k$ を変化させた時の認識率の変化

## 5.2 実験結果と考察

図5.2から図5.4に、上述の評価実験の結果を示す。本実験において、各カテゴリ毎の認識率、誤認識抽出率は前章と同様に定義する。

今回、図5.2ならびに5.4において、カテゴリ3, 22, 24, 30については認識率と誤認識抽出率が表示されていない。これらのカテゴリについては、誤認識検出用のSVMが構築できなかったことを意味している。図5.5にSVMのモデルが構築できなかったカテゴリのシェーマを示す。図からもわかるように、これらのシェーマは複雑な線分により構成されているとともに、その形状が特徴的である。その結果、擬似ベイズ識別関数を用いたシェーマ認識において誤認識が発生しなかったため、モデルを構築することができなかった。

また図5.6に、提案手法によって検出されたシェーマの一例を示す。図において(a)は検出率が低い場合の例を、(b)は検出率が高い場合の例を示している。これらの例では、双方の場合において特徴的な違いは見られなかった。提案手法では、擬似ベイズ識別関数により得られた1位から10位までの相違度を用いているため、シェーマ画像から検出率の差異を直接見出すことは難しいと考えられる。そのため、今後はこのようなシェーマに対して、擬似ベイズ識別関数により得られる相違度の分布の特徴などを調査し、さらなる検出率向上を目指す予定である。

本実験の結果、全てのシェーマを用いた場合、97.1%の認識率と94.6%の誤認識抽出率を得ることができた。今回得られた結果から分かる通り、提案法を用いることにより、従来手法の認識率（約90%[23]）より大幅に向上することができている。また、誤認識抽出率についても、実運用にあたって十分な結果となっている。これらの結果から、提案手法を用いることにより、現場からのニーズにあった認識機構の利便性を向上することができたと考えられる。今後は、誤認識抽出ができなかったカテゴリについても適用できるよう、提案法を改善策について検討していく予定である。

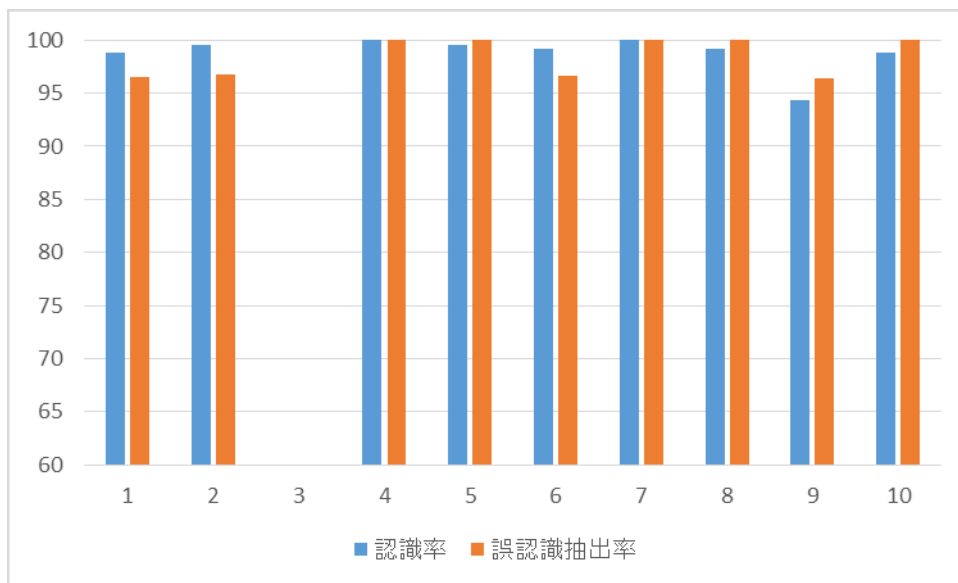


図5.2 提案手法における認識率と誤認識抽出率（カテゴリ1～10）

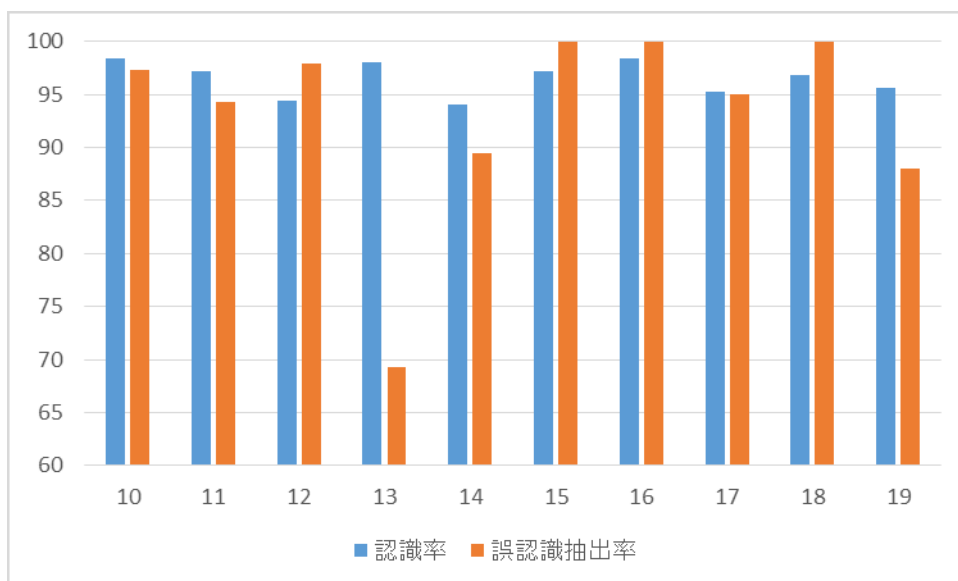


図5.3 提案手法における認識率と誤認識抽出率（カテゴリ11～20）

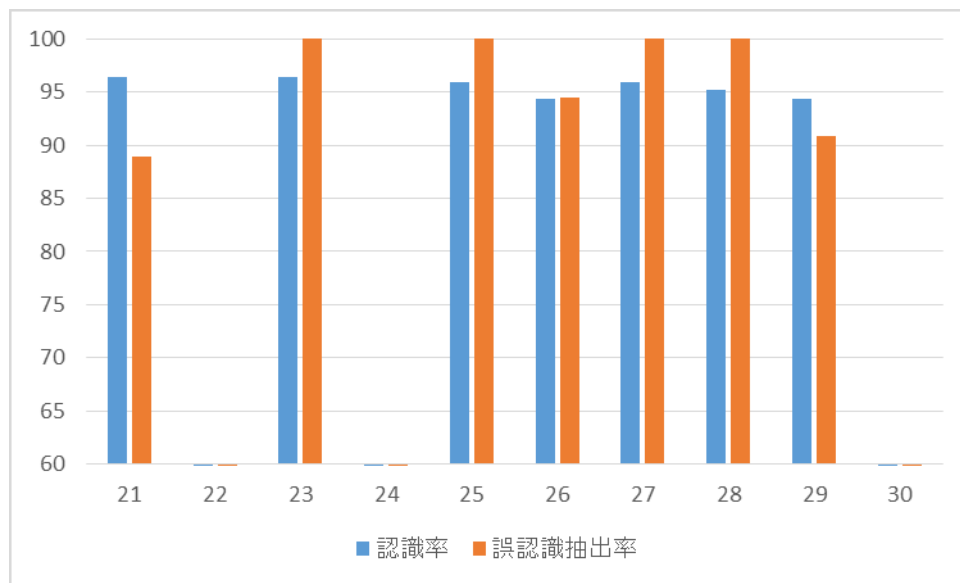


図5.4 提案手法における認識率と誤認識抽出率（カテゴリ21～30）

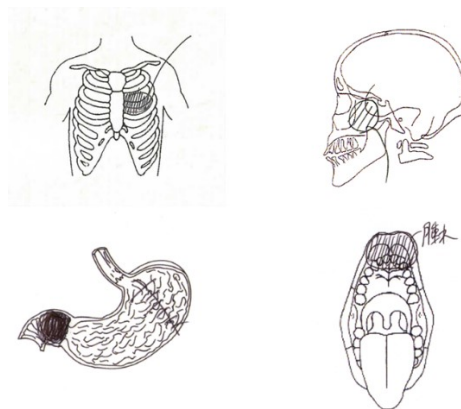
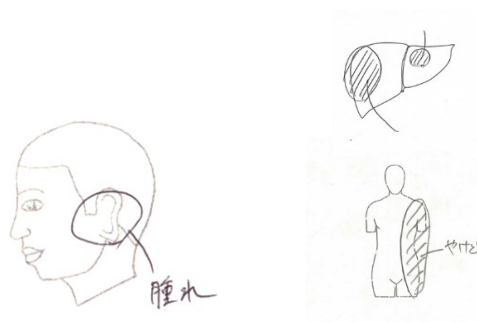


図5.5 誤認識検出においてSVMの構築に失敗したカテゴリの例



(a)検出率が低い場合 (b)検出率が高い場合

図5.6 提案手法によって誤認識が検出されたシェーマの例

## 第6章

### 終わりに

#### 6.1 本研究のまとめ

本研究では、シェーマ認識における現場からの声を考慮したシステムのために、誤認識を検出する方法について検討した。本論文では、従来法にて用いられている擬似ベイズ識別関数から得られる相違度を用いて、サポートベクターマシン（Support Vector Machine : SVM）による誤認識検出法を提案した。実際の医療現場にて用いられているシェーマを用いた評価実験を行い、提案法の有効性について検討した。

本論文では、誤認識結果を考慮したシェーマ認識手法を提案するとともに、その有効性について検討するために評価実験を行った。実験の結果、提案手法を用いることにより、誤認識結果を考慮した上で97.1%のシェーマ画像を認識することができるとともに、94.6%の誤認識抽出率を得ることができた。本研究の当初の目標としては、誤認識結果を完全に抽出することであったため、提案法を用いて全ての誤認識結果を検出することが不可能であったが、実運用に耐える程度の精度は得られており、一定の成果は得られたと言える。

#### 6.2 今後の課題

今後の課題として、提案手法によって失敗したカテゴリにも適用できるようにすることや、辞書にないシェーマ種が入力された場合の検討実験を行う必要性などが挙げられる。また、今回誤認識抽出に用いた以外の特徴量を使った場合どうなるのか、従来手法の認識器である擬似ベイズ識別関数の妥当性などを検討していく必要がある。



# 謝辞

本研究の遂行および修士論文の作成にあたり，丁寧なご指導とご助言を頂きました本学工学研究科電気電子工学専攻の川中普晴准教授，高瀬治彦准教授，本学副学長兼理事の鶴岡信治教授に感謝致します．また貴重な時間を割いて本論文を査読していただいた本学工学研究科物理工学専攻の野呂雄一准教授に深く感謝いたします．そして，本研究の共同研究者であり、研究方針の検討に協力していただきました同志社大学理工学研究科の大崎美穂教授，三重大学病院医学・病院管理部医事課診療情報管理担当の重岡真実様に深く感謝いたします．また，日頃お世話になった情報処理研究室の皆様に感謝致します．

最後となりましたが，本論文をまとめるにあたり，ご助言，ご討論，その他お世話になりました全ての方々に感謝致します．

## 参考文献

- [1] H. Harold Friedman, Ed., “Problem-Oriented Medical Diagnosis 5th Edition”, Lippincott Williams & Wilkins, 1991
- [2] K. Seto, T. Kamiyama, H. Matsuo, “An Object-Modeling Method for Hospital Information Systems”, The 9th World Congress on Medical Informatics, 52 Pt.2, pp.981-985, 1998
- [3] HJ. Lowe, I. Antipov, W. Hersh, CA Smith, M. Mailhot, “Automated Semantic Indexing of Imaging Reports to Support Retrieval of Medical Images in the Multimedia Electronic Medical Record”, Methods of Information in Medicine, vol.38, no.4, pp.303-307, 1999
- [4] 大谷芳弘, 川中普晴, 吉川大弘他, “自己組織化マップを用いたインシデント報告書の大意による分類に関する一考察”, 第22回ファジィシステムシンポジウム講演論文集, 6B1-4, 2006
- [5] 松村泰志, “ペーパーレス電子カルテ適用のための統合的文書管理システム DACS の構想”, 第29回医療情報学連合大会論文集, 1-B-4-7, 52-53, 2009
- [6] 周書義, 高柳和江, 木村哲彦, “退院サマリーの認識に影響を与える要因に関する研究”, 日医大誌 1999 第66巻第4号, pp.52-60, 1999
- [7] 稲岡則子, 紀ノ定保臣, 宇都由美子, 石原謙, 伏見清秀, “データウェアハウスとデータ利活用”, 医療情報学, vol.27, no.3, pp.261-268, 2007
- [8] 桑田成規, 稲田拓, 大越厚, 浜本政志, 近藤博史, “e-文書法対応スキャンシステムの構築および稼働実績評価”, 日本診療録管理学会誌, vol.20, no.2, p.146, 2008
- [9] 竹村匡正, “A knowledge extraction from medical practical documents on medical terminology and ontology”, 日本放射線技術学会雑誌, vol.65, no.7, pp.962-963. 2009
- [10] 小野大樹, 高林克日己, 鈴木隆弘, 横井英人, 井宮淳, 里村洋一, “テキストマイ

ニングによる退院サマリー自動分類の試み”, 医療情報学, vol.24, no.1, pp.35-44, 2004

- [11] S. Doi, T. Suzuki, G. Shimada, M. Takasaki, S. Fujita, T. Tamura and K. Takabayashi, “Auto-Selection of DPC Codes from Discharge Summaries by Text Mining in Several Hospitals and Analysis of Differences in Discharge Summaries”, Journal of Computational Intelligence and Intelligent Informatics, vol.16, no.1, pp.48-54, 2012
- [12] H. Kawanaka, Y. Otani, T. Yoshikawa, K. Yamamoto, T. Shinogi, S. Tsuruoka, “Tendency Discovery from Incident Reports with Free Format Using Self Organizing Map”, Japan Journal of Medical Informatics, vol.25, no.2, pp.87-96, 2005
- [13] Y. Otani, H. Kawanaka, T. Yoshikawa, K. Yamamoto, T. Shinogi, S. Tsuruoka, “Keyword Extraction from Incident Reports and Keyword Map Generation Method Using Self Organizing Map”, Proc. of 2005 IEEE International Conference on Systems, Man and Cybernetics, pp.1024-1029, 2005
- [14] H. Kawanaka, T. Sumida, K. Yamamoto, T. Shinogi, S. Tsuruoka, “Document Recognition and XML Generation of Tabular Form Discharge Summaries for Analogous Case Search System”, Methods of Information in Medicine (Schattauer), Vol. 46, No. 6, pp.700-708, 2007
- [15] H. Kawanaka, Y. Shiroyama, K. Yamamoto, T. Shinogi, S. Tsuruoka, “A Study on Document Structure Recognition of Discharge Summaries for Analogous Case Search System,” Proc. of International Workshop on Document Analysis Systems (DAS2008), pp.423-430, 2008
- [16] 林裕樹, 川中普晴, 中村峻太, 山本皓二, 高瀬治彦, 鶴岡信治, “文書種ならびに文書構造認識機能を備えたスキヤニングシステムの開発とその評価”, 第30回医療情報学連合大会論文集, pp.895-898, 2010
- [17] H.Kawanaka, K.Yamamoto, H.Takase, S.Tsuruoka. “Document Image Processing for Hospital Information Systems. Modern Information Systems”, Christos Kalloniatis (Ed.), ISBN: 978-953-51-0647-0, InTech, Available from:  
<http://www.intechopen.com/books/modern-informationsystems/document-image-processing-for-hospital-information-systems>.

- [18] 社団法人 全日本病院協会 医療の質向上委員会(DRG・TQM 委員会), “標準的診療記録作成・管理の手引き”, 株式会社 じほう, p.100, 2004
- [19] N.Otsu, “Discriminant and Least Squares Threshold Selection”, Proc. of 4IJCP, pp.592-596, 1978
- [20] N. Otsu, “A Threshold Selection Method from Gray-Level Histograms”, IEEE Transactions. Systems, Man, and Cybernetics, SMC-9, no.1, pp.62-66, 1979
- [21] 栗田昌徳, 鶴岡信治, 三宅康二, 横井茂樹, “加重方向指数ヒストグラムと疑似マハラノビス距離を用いた手書き漢字・ひらがな認識”, 電子情報通信学会技術研究報告, PRL-82-79, pp.105-112, 1983
- [22] 鶴岡信治, 栗田昌徳, 原田智夫, 木村文隆, 三宅康二, “加重方向指数ヒストグラム法による手書き漢字・ひらがな認識”, 電気情報通信学会論文誌 D, vol. J70-D, no.7, pp.1390-1397, 1987
- [23] 梶原宏, 川中普晴, 山本皓二, 高瀬治彦, 鶴岡信治, “医療文書検索を目的としたシェーマ認識手法の提案とその評価”, 医療情報学(日本医療情報学会), pp65-73, 2014

# 発表論文リスト

## 国際会議

- (1) K. Moriyama, H. Kawanaka, H. Takase, and S. Tsuruoka, “Detection of False Recognition for Document Retrieval Using Body Diagrams”, Proc. Of the 8<sup>th</sup> Intl. Workshop on Regional Innovation Studies (IWRIS2016), pp.1-4, 2016

## 国内会議

- (1) 森山浩平, 川中普晴, 高瀬治彦, 鶴岡信治, “シェーマ認識における誤認識の検出に関する一検討”, 平成26年度電気学会東海支部連合大会講演論文集, N1-2,2015
- (2) 森山浩平, 川中普晴, 高瀬治彦, 鶴岡信治, “シェーマ認識における誤認識の識別に関する研究”, 2016年映像メディア学会年次大会, 13B-4, 2016