

コンピュータ支援診断のための識別器の選択法

堀江 俊介[†] 松原 友子^{††} 笠井 聡^{†††} 内山 良一^{††††}

村松 千左子[†] 周 向荣[†] 原 武史[†] 藤田 広志[†]

[†] 岐阜大学大学院医学系研究科知能イメージ情報分野 〒501-1194 岐阜県岐阜市柳戸 1-1

^{††} 名古屋文理大学情報文化学部 〒492-8520 愛知県稲沢市稲沢町前田365

^{†††} Konica Minolta Medical Imaging USA, INC. 411 Newark Pompton Turnpike Wayne, NJ 07470

^{††††} 大分工業高等専門学校制御情報工学科 〒870-0152 大分県大分市牧 1666

E-mail: [†] {horie,hara,fujita}@fjt.info.gifu-u.ac.jp ^{††} tomoko@nagoya-bunri.ac.jp

あらまし コンピュータ支援診断 (CAD) 開発における偽陽性削除や良悪性鑑別などの処理では、様々な識別器が用いられている。しかし、常に、高い識別性能が得られる識別器は一意には決まることはない。そのため、用いる特徴量の性質に基づき最適な分類を可能にする識別器の選択基準を明らかにする必要がある。また、実験段階において、限られたデータ数を用いて得られた結果から、多量の未知データへ適用した場合に安定した性能を示す識別器を選択することは、CAD 開発において重要であると考えられる。そこで、本研究では、臨床データから抽出した特徴量を解析し、良悪性群間にどのような差が見られるのかを調査した。そして、その特徴量を用いて、5種類の識別器(線形判別: LDA, 二次判別: QDA, ニューラルネットワーク: ANN, サポートベクターマシン: SVM, AdaBoost) の良悪性の鑑別能を比較する実験を行った。その結果より、特徴量の性質やトレーニングデータ数の違いが与える影響を考慮し、CAD のための識別器選択法の可能性を提案する。

キーワード 識別器, 臨床データ, 特徴量, AUC, コンピュータ支援診断(CAD)

Choice Method of Classifier for Computer Aided Diagnosis

Shunsuke HORIE[†] Tomoko MATSUBARA^{††} Satoshi KASAI^{†††} Yoshikazu UCHIYAMA^{††††}

Chisako MURAMATSU[†] Xiangrong ZHOU[†] Takeshi HARA[†] Hiroshi FUJITA[†]

[†] Graduate School of Medicine, Gifu University 1-1 Yanagido, Gifu-shi, Gifu 501-1194, Japan

^{††} School of Information Culture, Nagoya Bunri University 365 Maeda Inazawa-cho, Inazawa-shi, Aichi 492-8520, Japan

^{†††} Konica Minolta Medical Imaging USA, INC. 411 Newark Pompton Turnpike Wayne, NJ 07470

^{††††} Oita National College of Technology 1666 Maki, Oita-shi, Oita 870-0152, Japan

E-mail: [†] {horie,hara,fujita}@fjt.info.gifu-u.ac.jp ^{††} tomoko@nagoya-bunri.ac.jp

Abstract Various classifiers are used on the elimination of false positives or diagnosis of benign and malignant lesion. However, no classifier shows consistently superior performance regardless of the nature of the data. Therefore, it is necessary to clarify the selection criterion of the classifiers that shows optimal performance based on the nature of the feature used. Moreover, the available data for training is finite-sized in an experiment step of CAD development. Consequently, it is important to select the classifier that indicate steady classification performance applying to unknown large data set, based on one obtained by training. In this paper, we conducted the comparison study for diagnosis performance of five classifiers, namely, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Neural Network, Support Vector Machine, and AdaBoost. Our dataset consists of benign and malignant lesions extracted clinical cases. The distribution analysis of features was performed. As a result, it becomes clear that the effects of nature of the features and number of training data on classifier performance are different each classifier. We propose the possibility of the selection method for classifier based on our results.

Keywords Classifier, Clinical Data, nature of the data, AUC, Computer-aided diagnosis (CAD)

1. まえがき

近年、さまざまなモダリティの医用画像が普及し、病変の早期発見に期待が寄せられている。しかし、同時に読影枚数の増加から、医師にかかる負担が増加し、読影ミスが生じる危険性も考えられる。そこで、現在、このような医師の負担を軽減させるため、計算機を用いた画像解析を行うことにより、病変と疑われる部位を示すコンピュータ支援診断(Computer-aided diagnosis: CAD)システムの開発が注目されている。

CAD システムでは、画像解析により多くの特徴量を算出し、識別器を用いて、病変部の検出、良悪性鑑別などを行う。CAD システムの開発において、患者の情報に基づき、正常であるか、異常であるかを正しく区別できる識別器の設計は、重要な役割を担っていると考えられる。

識別器には、線形判別や二次判別を代表として様々なものがある[1]。しかし、常に高い識別性能が得られる識別器は一意には決まってはいない。よって、有効な特徴量の選択基準、特徴量の数やその分布、また、トレーニングのデータ数が識別器に与える影響が明らかになれば、CAD における最適な識別器の選択が可能となる。

Chan らは、人工データを用いて、線形判別分析(Linear Discriminant Analysis : LDA)、二次判別(Quadratic Discriminant Analysis : QDA)、ニューラルネットワーク(Neural Network : ANN)におけるサンプルサイズや次元数の違いが識別性能に与える影響を報告している[2]。また、我々はその研究の識別器にサポートベクターマシン(Support Vector Machine : SVM)[3]と AdaBoost を比較候補として追加し、同様の実験を行い、識別性能に与える影響を調査した。

1. 濃度値の標準偏差 最大値
2. 重心回り慣性主軸 θ 最大値
3. 縦横比 最大値
4. 円形度 最小値
5. 周囲長比率 最小値
6. 後方エコー欠損状況 最小値
7. 分岐輪郭イレギュラー 最大値
8. 修正なし凹面面積比率 最小値
9. 修正なし凹面辺縁比率 最大値
10. 角ライン比率 最大値
11. 角ライン個数 最大値
12. 分散の差 最大値

表 1 実験に用いた特徴量 12 個の内容

本研究では、臨床データから抽出した特徴量の解析を行い、各識別器において、有効な特徴量の組み合わせを調査する。そして、そのいくつかの組み合わせ手法で選択された特徴量を用いて、LDA, QDA, ANN, SVM と AdaBoost の識別性能を比較する実験を行った。これらの結果より、CAD のための識別器選択法の可能性を提案することを目的とする。

2. 実験

2.1 データ群の詳細

本研究では、全乳房超音波画像における良性 30 個、悪性 33 個の腫瘤像から抽出した特徴量 12 個を用いて、2つのデータ群(Class Benign, Class Malignant 以下 Class B, Class M)を作成した。12 個の特徴量は、診断基準に従い、エコー情報と、形状に関するものである。用いた各特徴量を表 1 に示す。

2.2 特徴量の正規性検定

正規分布に仮定することができる分布形状を持つ特徴量は、ベイズ理論より 2 群に分類する際、誤り確率が予測可能であり、その場合、先行研究等より識別器の性能は明らかとなっている[4]。しかし、CAD で用いられる特徴量の分布形状は明らかではなく、それらを用いた識別性能も明らかとなっていない。よって、2.1 で述べた特徴量の Class B, Class M に対して Shapiro-Wilk (シャピロ・ウィルク) 検定を行い、その結果を表 2 に示す。

12 個の特徴量における正規分布と仮定できる特徴量は、有意水 5% の時、良性データで 7 個、悪性データで 10 個であった。12 個中、6 個が正規分布に従う結果となり、臨床データにおいては、正規分布を仮定できるものばかりではないことが明示された。正規分布と仮定できないものとして、慣性主軸 θ や円形度など形状に関する特徴量であった。

特徴量	p値(class B)	p値(class M)
1. 濃度値の標準偏差 最大値	0.7241	0.3946
2. 重心回り慣性主軸 θ 最大値	0.0002	0.0030
3. 縦横比 最大値	0.0811	0.1498
4. 円形度 最小値	0.8642	0.0487
5. 周囲長比率 最小値	0.7597	0.6421
6. 後方エコー欠損状況 最小値	0.0234	0.1005
7. 分岐輪郭イレギュラー 最大値	0.0006	0.1282
8. 修正なし凹面面積比率 最小値	0.0258	0.3697
9. 修正なし凹面辺縁比率 最大値	0.0225	0.0740
10. 角ライン比率 最大値	0.1954	0.9737
11. 角ライン個数 最大値	0.0934	0.0870
12. 分散の差 最大値	0.0898	0.0541

表 2 Shapiro-Wilk 検定の結果(有意水準以下をグレーで表示)

逆に、濃度値など輝度に関するものは正規分布と仮定できる特徴量であった。このことは、他のモダリティから抽出される特徴量に対しても同様の傾向が見られそうである。

2.3 一元配置分散分析

使用する特徴の良悪性群間に有意な差があるかを調査するため、一元配置分散分析を行った。その結果を表3に示す。この結果より一元配置で見たときは、12個の特徴量中5個の特徴量では良悪性群間に有意な差が見られないため、有用な特徴量ではないように考えられる。

2.4 特徴量選択方法

- 本研究では、特徴量選択を以下の4種類の方法で行った。
- 選択方法1：総当たりすることにより、最も良い組み合わせを選択する方法。
 - 選択方法2：各次元でROC解析をし、最も高いAUC値であった特徴量を逐次追加する選択方法。
 - 選択方法3：2.3で行った一元分散分析の結果、2つの母集団が分離度合の指標であるウィルクスの Λ が小さい順番に追加する方法。
 - 選択方法4：無作為に特徴量を選択し、追加する方法。今回は表1にある特徴量を上に記されている順番に追加していく。

2.5 選択方法比較実験

2.4で選択された特徴量の組み合わせを用いて、今回、我々が行った実験の流れを図1に示す。

識別能の評価は、自己学習の結果をROC曲線下の面積AUC(area under the curve)を用いて検討する。

また、識別器には、LDA, QDA, ANN, SVM, および AdaBoost を用いる。特徴量の組み合わせは識別器ごとで選択されるものが異なっている。なお、ANNの中間層のノードの数は次元数の半分の数と等しい。SVMのカーネルはradial, gammaは

特徴量	p値
1. 濃度値の標準偏差 最大値	2.66E-01
2. 重心回り慣性主軸 θ 最大値	1.97E-05
3. 縦横比 最大値	2.40E-05
4. 円形度 最小値	5.23E-02
5. 周囲長比率 最小値	6.29E-02
6. 後方エコー欠損状況 最小値	9.40E-06
7. 分岐輪郭イレギュラー 最大値	1.27E-01
8. 修正なし凹面面積比率 最小値	1.09E-04
9. 修正なし凹面辺縁比率 最大値	1.94E-06
10. 角ライン比率 最大値	1.88E-02
11. 角ライン個数 最大値	2.59E-05
12. 分散の差 最大値	1.07E-01

表3 一元配置分散分析の結果（有意水準以下をグレーで表示）

1/3, costは1となっている。そして、AdaBoostのboostingタイプはdiscrete, 縮小パラメータは1となっている。

3. 結果と考察

3.1 Classifier 1 : LDA

識別器としてLDAを利用したときのAUCのグラフを図2(a)に示す。

特徴量12個を用いた際のAUC値は0.959である。総当たりによる選択では6個の組み合わせ時に、同等の値0.959を得られている。また、逐次選択と総当たり選択では6個以降の組み合わせでは、AUC値は0.95付近ではほぼ横ばいとなった。また、ランダム選択では5個目の特徴量選択で急激にAUC値が上昇した。この時、追加された特徴量「周囲長比率」であり、総当たり選択では3組以上の組み合わせでは常に「縦横比」「円形度」「周囲長比率」が組み込まれており、組み合わせの効果が期待できる特徴量だったことが分かる。しかし、逐次選択では10個目に追加された特徴量であった。また、分散分析による選択では、8個目に追加されたとき、ようやく0.95付近に上昇した。この特徴量は、「円形度」であり、分散分析では、群間に有意な差が見られないとされた特徴量であった。

3.2 Classifier 2 : QDA

識別器としてQDAを利用したときのAUCのグラフを図2(b)に示す。

特徴量を12個用いた際のAUC値は0.981である。LDAと同様に、総当たりによる選択では6個の組み合わせ時に、同等の値0.983を得られている。また、逐次選択でも同等の値0.972を得られている。7個目以降の組み合わせにおいて、逐次選択は総当たり選択に匹敵するAUC値であった。分散分析

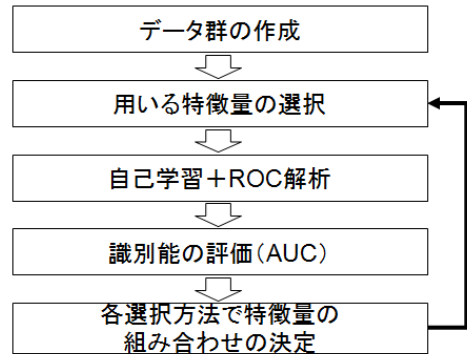


図1 実験の流れ

による選択は、LDA では AUC 値が停滞する 때가 あったが、QDA では、追加する度に AUC 値が上昇した。また、ランダム選択では、LDA と同様に、5 個目の特徴量選択で急激に AUC 値が上昇した。やはり「縦横比」「円形度」「周囲長比率」が組み合わされた時に効果を発揮する特徴量であることが分かる。しかし、QDA での総当たり選択では、「周囲長比率」は 8 個目で選択される特徴量であり、QDA と LDA で最も効果のある組み合わせは異なることが示唆されている。

3.3 Classifier 3 : ANN

識別器として ANN を利用したときの AUC のグラフを図 2 (c) に示す。結果を安定させるため、50 回の試行の平均を結果としている。

特徴量を 12 個用いた際の AUC 値は約 0.967 である。総当たりによる選択では 4 個の組み合わせ時に、同等の値 0.963 を得られている。逐次選択では、7 個目を追加した時に 0.967 を得ている。ANN は他の識別器とは異なり、特徴量選択方法によって、違いが顕著に表れた。ANN では、特徴量の組み合わせに敏感に反応することが明らかとなった。よって、ANN は、他の識別器よりも特徴量選択に注意が必要であることが示されている。

3.4 Classifier 4 : SVM

識別器として SVM を利用したときの AUC のグラフを図 2 (d) に示す。

特徴量を 12 個用いた際の AUC 値は 0.991 である。総当たりによる選択では 5 個の組み合わせ時に、同等の値 0.993 を得ている。また、逐次選択では 6 個目の組み合わせ時に、0.982 を得ている。グラフの推移は LDA と似た結果となっている。また、どの選択方法でも、比較的、AUC 値は高く、識別器ごとの比較をしても、今回、用いた識別器の中で最も高い結果となった。

3.5 Classifier 5 : AdaBoost

識別器として AdaBoost を利用したときの AUC のグラフを図 2 (e) に示す。

特徴量を 12 個用いた際の AUC 値は約 0.987 である。総当たりによる選択では 4 個の組み合わせ時に、同等の値 0.988 を得ている。逐次選択でも 5 個組み合わせ時、0.981 を得ている。また、ランダム選択においても、選択された特徴量が少ない時点で他の識別器よりも高い AUC 値を示した。

3.6 5つの識別器の比較

当然の結果として、総当たりによる選択法が、どの識別器においても選択される特量数に依存せず、常に最も高い AUC 値であった。

5つの識別器の中で、LDA は最も AUC 値が低い結果となった。CAD で扱う臨床データは、正規性を持たない特徴量が存在することから、そのような特徴量を有効に活用することができなかったことが原因だと思われる。総当たりで選択された特徴量は、正規性を持たない特徴量の占める割合が LDA 以外は 33% から 75% であったが、LDA では、常に 50% 以下であった。この結果よりも、他の識別器と比べ、正規性を持たない特徴量が有効なものとして選択されにくい傾向にあった。

また、QDA は、多くの特徴量を追加した時には、他の識別器と比べても、高い判別能であったが、特徴量が少ない段階では、他のノンパラメトリックな識別器に劣る結果となった。

総当たりにより、最良の特徴量選択が行われるのであれば、ANN は有効であることが分かった。しかし、特徴量選択にかなり依存があるため、特徴量選択が正しく行われない場合には注意が必要であると考えられる。

SVM は 5つの識別器の中で最も安定して、高い判別能を示していたことから、特徴量の性質が十分に分析されていない試験段階でも、最も適切な結果を示すと考えられる。

最後に AdaBoost であるが、こちらも、SVM と同様に安定して高い性能を示した。しかし、特徴量が 2 個の時にも、他の識別器より異常に高い判別能を示していたことから、過学習の可能性がある。よって、試験データに用いるときは注意が必要であると考えられる。

特徴量を 12 個用いた際の AUC 値は 0.987 程度である。総当たりによる選択では 4 個の組み合わせ時に、同等の値 0.988 を得ている。逐次選択でも 5 個組み合わせ時、0.981 を得ている。また、ランダム選択においても、選択された特徴量が少ない時点で他の識別器よりも高い AUC 値であることから、少ない特徴量でも、高い判別能を示すことが分かる。

また、識別器によっても、選択数によっても、選択される特徴量の組み合わせが異なった。このことから、最適な特徴量選択は、識別器や扱う特徴量の個数の決定が必要であるとも示唆された。

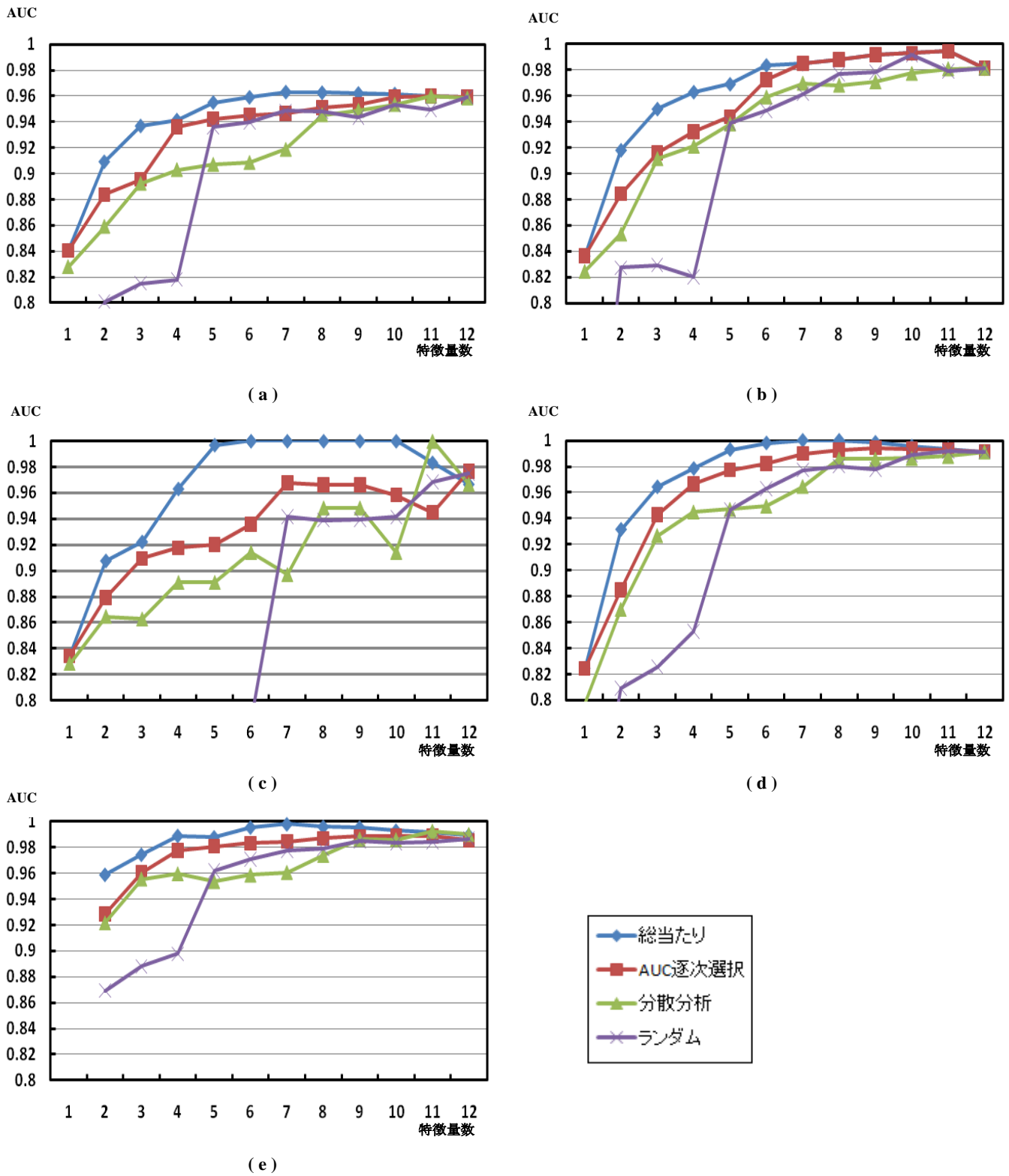


図2 選択方法比較実験の結果 (縦軸: AUC 値, 横軸: ROC 解析に用いた特徴量数)
 (a)LDA の結果 (b)QDA の結果 (c)ANN の結果 (d)SVM の結果 (e)AdaBoost の結果

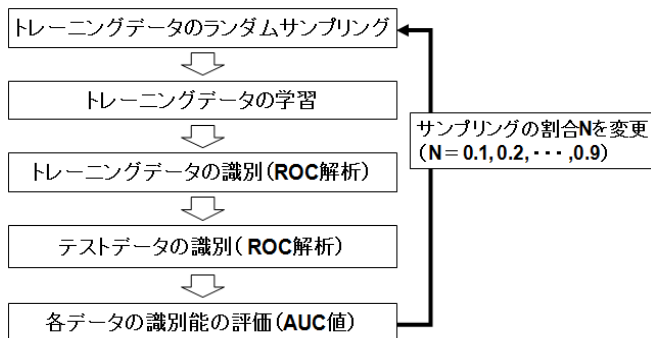


図3 トレーニングデータ数の変化による判別能比較実験の流れ

3.7 トレーニングデータ数と特徴量数の関係性

3.1の結果を例として、総当たりで選択された6個の特徴量と12個全ての特徴量を用いた場合のテストデータへの適応の違いを観察する実験を行った。実験の流れを図3に示す。

トレーニングデータは、ランダムに Class B, Class M から各 class のデータ数×N(sampling-rate : 0.1, 0.2, 0.3, ... 0.9)個ずつサンプリングしたものとす。その後、サンプリングしたトレーニングデータを用いて、自己学習を行い、トレーニングデータとテストデータの識別を行う。識別能の評価はROC解析によって得られたAUC値を用いて行った。また、割合を変更する度、50回のサンプリングをし、実験結果のバイアスを小さくさせた。その結果を図4に示す。

総当たりで選択された6個の特徴量を用いた場合、全データの3割程度の学習でトレーニングとテストの差を小さくすることが可能であり、トレーニングと近い性能をテストに期待できると考えられる。しかし、12個全ての特徴量を用いた場合、その差は、全データの9割を学習させても、その差は縮小しない上、トレーニングで得られる性能をテストに見込めないことが分かる。この結果より、有効な特徴量選択により、利用する特徴量を絞ることで、未知データへ適用した際の識別能の予測が可能となる。

4. まとめと今後の展望

本研究では、臨床データから得られた特徴量を解析し、特徴量選択と識別器の関係性を明らかにするため、5種類の識別器を用いて、判別能をAUCにより評価し、検討した。

結果、LDAは正規性を持たない特徴量を多く含むデータに対して不向きであることから、扱う際は、データの正規性に注視する必要がある。QDAは有効な特徴量が多く利用できる場合、有効である。しかし、次元数の増加とともに、試験データへの適応はデータ数を必要とすることがわかっていることから、データ数が

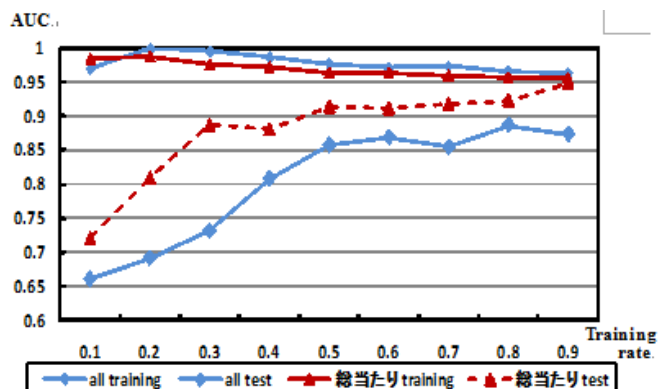


図4 トレーニングデータ数の変化による判別能比較実験の結果 (All training: 特徴量 12 個, 総当たり: 特徴量 6 個)

十分にある時には有効だと考えられる。ANNは特徴量選択が十分に行われている場合には有効である。しかし、用いる特徴量に強く依存することから、特徴量数が多い場合は、特徴量の解析を十分にし、有効な組み合わせの上で利用する必要がある。今回はAdaBoostとSVMは安定して、他の識別器より高い性能を示したことから、他の識別器を利用する条件が整っていないときにも利用でき、汎用性が高いと考えられる。しかし、AdaBoostに関しては、過学習の可能性が考えられるため、未知データへの適用には注意が必要である。

また、特徴量選択は、総当たりによる選択方法が最良の組み合わせであるが、数百もの特徴量がある場合、現実的な方法とは考えられないため、定量的な選択方法の考案が必要であると考えられる。

今回は全乳房超音波画像における良性30個、悪性33個の腫瘍像から抽出した特徴量を用いて実験を行った。データ数や特徴量数も少ないことから、現在のモダリティから得られたデータでの実験をしており、偏りの少ない解析結果の提示により、より定量的な特徴量選択の基準となる結果を得ることが期待できる。

文 献

- [1] 永田靖, 棟近雅彦, “多変量解析法入門”, pp. 99-131, サイエンス社, 2001.
- [2] Heang-Ping Chan, Berkman Sahiner, Robert F. Wagner, and Nicholas Petrick, “Classifier design for computer-aided diagnosis: Effects of finite sample size on the mean performance of classical and neural network classifiers”, Med.Phys., 26, pp.2654-2668, 1999.
- [3] 小野田崇, “サポートベクターマシン (知の科学)”, オーム社, 2007.
- [4] Keinosuke Fukunaga, “Statistical Pattern Recognition”, pp.124-180, Academic Press, 1990