3ZG-2

ブール代数に基づく識別とその肝癌診断への応用

荻原 宏是*1　那須 大気*2　藤田 悠介*2　飯塚 徳男*2　岡 正朗*2　浜本 義彦*2
*1 山口大学 工学部知能情報工学科
*2 山口大学 大学院医学系研究科

1. はじめに

肝癌の克服は国民的課題となっている．肝癌の難治性は，その再発の高さにあり，手術ですべての癌を摘出しても１年後には３０％近くの再発が生じている[1]．個別化医療として，この再発予測を簡易な血液検査で行うことが望まれている．本研究では，血液検査として代表的な腫瘍マーカーである AFP や PIVKA とゲノム情報としてメチル化遺伝子データを併用する．これらの血液検査からの２値パターンを対象とした，ブール代数に基づくパターン認識理論を構築し，手術で癌を取り除いた患者を対象にした肝癌の再発予測を行う．

2. ブール識別器

血液検査は，陰性，陽性などの質的データと数値で表される量的データが混在している．質的データと量的データを一括して取り扱うために，量的データを２値化して質的データと組み合わせて２値パターンとする．

再発の予測問題を定式化すると，患者からのデータは２値で，それらを用いて２値パターンベクトルで患者を記述し，特徴選択，識別によって再発の有無を高精度に予測するシステム（図１参照）を構築する問題となる[2]．本研究では特徴は予め与えられたものとし，識別について論じる．

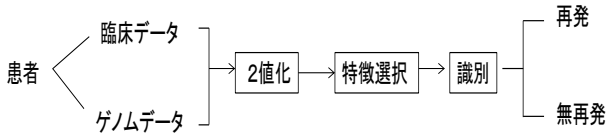

図１　癌の臨床データとゲノムデータの統合による再発予測

いま２個のデータの組があり，それぞれが０と１のいずれかの値を取るものとした，２次元の２値パターンxが与えられたとする．このとき，パターンxは４通り$(=2^2)$通り存在し，再発クラスの２値パターンと無再発クラスの２値パターンを，０と１の組み合わせで分類する．同一分類内で，訓練サンプルが多数を占める２値パターンのクラスを，その分類のクラスとする．表１は２次元の２値パターンの例で，＊は両クラスの訓練サンプル数が同数か，あるいは訓練サンプルが入らない分類を意味し，その分類はラベルなしとする．

表１から再発のクラスω_1を識別する関数f_1の真理値表(表２参照)を作成する．この表２からブール代数に基づき主加法標準形により，再発であれば１，そうでなければ０とする再発の識別関数f_1を定める．表２からf_1の主加法標準形は

$$f_1 = \bar{x}_1\bar{x}_2 + \bar{x}_1x_2$$

となり，識別関数f_1は簡略化されて

$$f_1 = \bar{x}_1$$

となる．表３からは，無再発ω_2の識別関数f_2は

$$f_2 = x_1\bar{x}_2$$

となる．

表１　多数決によるクラスラベル付け

x_1	x_2	ω_1/ω_2	クラスラベル
0	0	$5/3$	ω_1
0	1	$10/1$	ω_1
1	0	$2/9$	ω_2
1	1	$8/8$	＊

表２　f_1の真理値表

x_1	x_2	$f_1(x)$
0	0	1
0	1	1
1	0	0
1	1	0

表３　f_2の真理値表

x_1	x_2	$f_2(x)$
0	0	0
0	1	0
1	0	1
1	1	0

Classification based on Boolean Algebra and its Application to Diagnosis of Liver Cancer
Hiroyuki OGIHARA*1
Daiki NASU*2 Yusuke FUJITA*2 Norio IIDUKA*2
Masaaki OKA*2 Yoshihiko HAMAMOTO*2
*1 Department of Information Science and Engineering, Yamaguchi University
*2 Graduate School of Medicine, Yamaguchi University

3．実験

3-1．データ

手術で癌を取り除いた患者の中で肝癌が２年以内に再発した患者３８サンプル，無再発の患者３５サンプルを対象とし，臨床データである AFP と PIVKA，ゲノムデータであるメチル化遺伝子 SPINT，SRD を特徴として実験に用いる[3]．

3-2．従来法との比較

AFP と PIVKA それぞれ単体で識別する方法を従来法とする．AFP，PIVKA の Cut off 値をそれぞれ 20,40 と設定して２値化した後に，それぞれ単独で識別に用いた．例えば，AFP の場合，AFP の検査値が２０以上であれば，再発する，また２０未満であれば再発しないと識別する．

次に AFP と PIVKA を用いたブール識別器を考える．これは AFP と PIVKA を組み合わせて，２次元２値パターンの特徴として用いるものである．

最後に AFP,PIVKA に SPINT，SRD を組み合わせたブール識別器を提案手法とする．これは，AFP,PIVKA にゲノムデータである SPINT,SRD を加え，４次元２値パターンの特徴として用いるものである．

ブール識別器の識別では，識別関数としてクラス毎に定義された論理関数f_1とf_2を用意している（図２参照）．

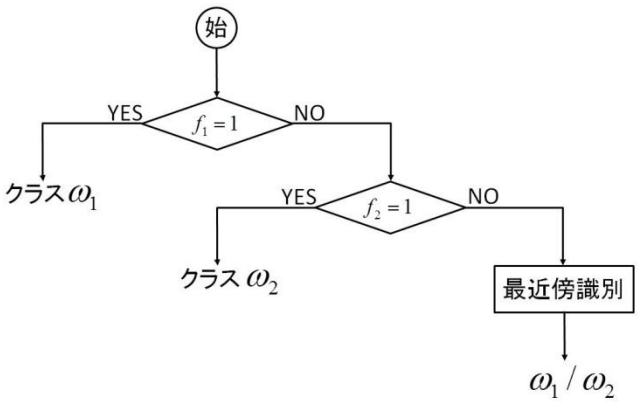

図２　識別の流れ

4．結果と考察

識別の評価には Youden_index[4]（感度＋特異度−1.0）を用いる．この Youden_index の値が高いほど識別性能は高いといえる．表４に識別結果を示す．

既存の AFP や PIVKA では十分な識別性能は出ないが，それらをブール識別器に用いることにより性能は改善され，メチル化遺伝子と組み合わせることによりさらに性能が改善された．従って，ゲノムデータと臨床データを組み合わせたブール識別器による肝癌の再発予測は有効であると考えられる．

最後に，AFP，PIVKA，SPINT，SRD の４次元２値パターンに対する再発の識別関数f_1は，

$$f_1 = \bar{x}_1\bar{x}_2\bar{x}_3x_4 + \bar{x}_1x_2\bar{x}_3\bar{x}_4 + x_1\bar{x}_2x_3 + x_1\bar{x}_2\bar{x}_4 + \bar{x}_2x_3\bar{x}_4$$

となった．一方，無再発の識別関数f_2は

$$f_2 = \bar{x}_1\bar{x}_2\bar{x}_3\bar{x}_4 + \bar{x}_1\bar{x}_2x_3x_4 + \bar{x}_1x_2x_3\bar{x}_4 + x_1\bar{x}_3x_4$$

となった．

5. おわりに

本研究では臨床データにゲノムデータを組み合わせたブール識別器が肝癌の再発予測の精度向上に役立つことが明らかになった．今後，用いる臨床データ，ゲノムデータの組合せを増やして実験を行う予定である．

文献

[1]N. Iizuka, M. Oka, and Y. Hamamoto et al., Oligonucleotide Microarray for Prediction of Early Intrahepatic Recurrence of Hepatocellular Carcinoma after Curative Resection, Lancet, 361, pp.923-929, 2003.

[2]浜本義彦　統計的パターン認識入門，森北出版 2009.

[3]N.Iizuka et al., Efficient Detection of Hepatocellular Carcinoma by Hybrid Blood Test of Epigenetic and Classical Protein Markers, Clinica Chimica Acta, pp.152-158, 2011.

[4]W.J.Youden, Index for Rating Diagnostic Tests, Cancer, pp.32-35, 1950.

表４　各手法の識別結果

識別手法	感度	特異度	Youden_index
AFP 単体での識別	0.57	0.50	0.07
PIVKA 単体での識別	0.00	1.00	0.00
AFP と PIVKA を用いたブール識別器	0.57	0.68	0.25
AFP，PIVKA，SPINT，SRD を用いたブール識別器	0.77	0.74	0.51