

平成 30 年 5 月 21 日現在

機関番号：15501

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00238

研究課題名(和文) 離散Bayes識別理論による個別化医療のための肝癌再発予測

研究課題名(英文) Prediction of Recurrence of Liver Cancer by a Discrete Bayes Decision Rule for Personalized Medicine

研究代表者

浜本 義彦 (HAMAMOTO, YOSHIHIKO)

山口大学・大学院創成科学研究科・教授

研究者番号：90198820

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：肝癌の怖さは高い再発率にあり、術後1年以内に約30%が再発する。本研究では、質的データを取り扱える独自の離散Bayes識別則を用いて、高精度で肝癌の再発を予測することを目的とする。テストサンプルに対する肝癌再発の予測性能は、感度86%、特異度49%であった。更に、早期胃癌のリンパ節転移の予測問題にも離散Bayes識別則を適用し、テストサンプルに対して感度100%、特異度86%の予測性能を達成した。以上の結果から、離散Bayes識別則の臨床応用の見通しを得た。

研究成果の概要(英文)：The refractory problem lies in the high percentage of recurrence of liver cancer: about 30% of postoperative patients experience recurrence within one year. The purpose of this study is to distinguish between the presence and absence of recurrence with a high degree of accuracy by using our discrete Bayes decision rule that can handle discrete data. Discrimination performance of the discrete Bayes decision rule using the optimal combination of markers against test samples showed a sensitivity of 86% and specificity of 49%. Moreover, the discrete Bayes decision rule was applied to diagnose lymph node metastasis for early gastric cancer. The proposed method was able to differentiate lymph node metastasis with 100% sensitivity and 86% specificity. These results suggest that the proposed method might lead to clinical applications.

研究分野：情報工学

キーワード：パターン認識 癌 診断

1. 研究開始当初の背景

肝癌の難治性は、その再発の高さにある。たとえ手術で全ての癌を完全に摘出して1年後には30%近くの再発が生じている。再発予測の意義は、仮に再発する可能性が低ければ、患者は見えない癌の恐怖から開放され、副作用の危険性のある抗癌剤が不要となり、患者はこの上ない恩恵を受けることができる。また医療費の高騰を抑制することもできる。最近、抗癌剤は効く患者を選んで投与する方向が主流となり、再発予測は抗癌剤投与の判断に用いられている。一方再発の危険性が高ければ、先制医療として適切な医療が行われる。このように、患者個々に応じた医療が、個別化医療である。この個別化医療を実現するために、再発予測が必要なのである。

これまで取り組んできた肝癌の再発予測では、まず手術で切除した癌の組織から得られた遺伝子発現データを用いて再発の有無を予測した。このアプローチでは93%の高精度で再発を予測することができたが、何らかの手段で患者から癌の組織を抽出しなければならぬという欠点がある。これは患者にとって大きな負担である。また、新規の遺伝子検査は、薬事法の承認を受けなければならない。これには膨大な治験が必要とされ、実用化の大きな壁となって、患者への適用には時間がかかるという分子標的アプローチには致命的とも言える問題がある。

一方、保険適用されている検査による肝癌の再発予測の現状では、肝癌の代表的検査としてAFPとPIVKAがあるが、特異度はある一定のレベルがあるものの、感度が低く、癌の見逃しがあり、十分な機能を果たしているとは言いがたい。しかしAFPとPIVKAと他のデータを組合すことにより精度の向上がみられた。これは、単一の臨床データではなく、臨床データの組合せによる予測の有効性を示唆している。

本研究では、成果の社会への還元を促進するため、現在保険適用されている臨床データを当面の対象とする。臨床データには量的データと質的データがある。量的データはGOTなどのような数値データであり、一方質的データとは、検査結果の陽性あるいは陰性などの非数値データである。これらの組合せには、量的データと質的データの統合問題がある。更に量的データは検査会社によって、正常の範囲が異なり、同じ検査会社の製品であっても、その更新によって正常の範囲が異なる場合がある。また量的データには検査内容によっては数値に不安定な面もある。重要なのは、厳密に臨床データの値そのものではなく、そのデータが医学的に正常か否かの2値である。そこで、量的データを医学的に決められたしきい値により2値化する。質的データと

2値化された量的データからなる「こと」を特徴とし、それにより患者を記述して疾患データベースを整備する。以上の準備のもと、「こと」を取り扱う離散Bayes識別理論の研究を開始した。

2. 研究の目的

癌再発の診断において癌は手術後のCTや超音波検査では見えず、診断が困難となっている。この見えない癌が再発するか否かは、その重要性にも関わらず、極めて困難な問題となっている。本研究では、手術で癌を完全に取り除いた患者を対象にした肝癌の1年以内の早期再発予測問題を解決するための離散Bayes識別理論を確立することを目的とする。

3. 研究の方法

まず疾患データベースを整備し、離散Bayes識別の基礎理論を構築した。臨床データの「こと」に関する情報を格納した疾患データベースから条件付き確率を求め、肝癌再発予測の予備検討を行った。次に「こと」を取り扱う特徴選択法を確立し、予測に最適な「こと」の組合せを求めた。また特許を出願し、事業化の検討も行った。更に臨床応用を目指して他の医学問題へも本方法を適用し、その有効性を医学的に評価した。

まず、肝臓専門医の指導のもと、疾患データベースを整備した。離散Bayes識別理論は、統計的パターン認識の中心的な識別則であるBayes識別則にその理論的基盤をおく。Bayes識別則では、各クラスに対する事後確率を計算し、事後確率が最大のクラスへパターンを識別する。これを「こと」に対して行うのが離散Bayes識別理論である。

事後確率を計算するためには、事前確率とクラス条件付き確率が必要である。本研究では、2クラスを対等に扱うため事前確率は等しいものと仮定する。

いま、最適な「こと」の組合せが $x_{i_1}, x_{i_2}, \dots, x_{i_d}$ とすると、患者は d 次元パターン $x = [x_{i_1}, x_{i_2}, \dots, x_{i_d}]$ と表現される。クラス条件付き確率を $P(x | \omega_i)$ とすると、Bayesの定理により事後確率 $P(\omega_i | x)$ は

$$P(\omega_i | x) = \frac{P(x | \omega_i)}{P(x | \omega_1) + P(x | \omega_2)}$$

で与えられる。ここでクラス ω_1 は再発のクラス、クラス ω_2 は無再発のクラスであり、互いに排反事象となっている。一つの「こと」 x_{i_j} に対するクラス条件付き確率を $P(x_{i_j} | \omega_i)$ とし、各「こと」が独立であると仮定することによってクラス条件付き確率 $P(x | \omega_i)$ は

$$P(x | \omega_i) = \prod_{j=1}^d P(x_{ij} | \omega_i)$$

で与えられる。これを上式に代入することにより、各クラスの事後確率を計算することができる。計算上で重要なことは、上記は全てスカラー計算であり、極めて高速に計算できる点である。医学的に有望視されている数種類の臨床データを用いて、離散 Bayes 識別理論の予備的評価を行った。

次に「こと」を取り扱う特徴選択法を確立した。特徴の良さは、最終的には識別性能で評価されるものとする。そこで、識別実験を行うために再標本化手法（リサンプリング法）を用い、それによって特徴選択問題を解く。再標本化として、訓練サンプル集合をランダムに2分割し、一方を仮想訓練サンプル集合、他方を仮想テストサンプル集合とする。仮想訓練サンプル集合からクラス条件付き確率を求め、仮想テストサンプル集合内の仮想テストサンプルを識別した。

p個の「こと」を要素とする候補特徴集合に対して、総当たり法により部分集合それぞれを感度と特異度で評価する。以上を再標本化から複数回、独立に繰り返して各部分集合を評価する。得られた感度と特異度の平均から最適な「こと」集合を求めた。

総当たり法には計算コストの問題があるが、前述したように事後確率の計算を極めて高速に行うことができるため、実行可能である。予備的検討として、医学的に評価されて用いられてきた10種類くらいの有望な候補を決め、その中から特徴選択法により最適な「こと」の組合せを求め、その妥当性を医学的観点から行った。

以上の成果を肝癌の早期再発予測問題や早期胃癌のリンパ節転移予測問題へ適用し、論文投稿した。更に患者層別化等の臨床上重要なテーマに向けて本研究成果を展開するとともに、事業化へ向けて、主に製薬企業との折衝を行った。

4. 研究成果

肝癌の早期再発予測問題

前述したように、肝癌の怖さはその高い再発率にある。完全に手術で癌を切除しても術後1年以内に約30%が再発する。再発を予測できれば効果的な先制医療を実施でき、もし無再発なら副作用のある抗癌剤やCT検査が不要となり、医療費の抑制にもなる。

この問題に対して山口大学医学部附属病院から提供された患者を対象に、現在保険適用されている臨床データだけを用いて術後1年以内の早期再発を予測した。まず消化器腫瘍外科学を専門とする医師により肝癌に係りのある臨床因子として考えられる11のマーカー候補をリストし、この中から独自の特徴選択法により5マーカーを選択した。次にそれら5マーカーを用いて離散 Bayes 識別則

によりテストサンプルに対して感度86%、特異度49%を達成した。この水準は、良く知られているTNM分類(ステージ)やModified Jisよりも高感度で、ROC解析においても、より有効であることが示された。

早期胃癌のリンパ節転移予測問題

早期胃癌であれば内視鏡治療で助かる時代となったが、リンパ節転移が疑われると外科手術が必要となる。実際にはリンパ節転移のない患者にも、結果的には不要な外科手術が行われている。

そこで、山口大学医学部附属病院から提供された患者を対象に、内視鏡治療後の現在保険適用されている臨床データだけを用いてリンパ節転移を予測した。まず、消化器内科学を専門とする医師によりリンパ節転移に関わる臨床因子として考えられる8マーカー候補をリストし、この中から3マーカーを独自の特徴選択法により選択した。それらの3マーカーを用いて離散 Bayes 識別則によりテストサンプルに対して感度100%、特異度86%を得た。ここで感度100%とはリンパ節転移の見逃しがゼロであることを意味する。

上記2つの研究成果はいずれも査読のある専門雑誌に掲載された(下記参照)。今後の展望として離散 Bayes 識別則を活用した患者層別化による創薬がある。現在、創薬の成功確率は3万分の1と極めて低い。その原因は、薬効不足である。最新の分子標的薬剤であっても、せいぜい効果があるのは20%程度で、ほとんどが効かない。そのため、効く患者を選んで効く薬剤を投与することが求められている。この効く患者を選ぶことが患者層別化である。上記の医学問題はいずれも患者層別化問題の一つであり、本研究成果を創薬に適用することを計画している。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計2件)

A. Goto, J. Nishikawa, E. Hideura, R. Ogawa, M. Nagao, S. Sasaki, R. Kawasato, S. Hashimoto, T. Okamoto, H. Ogihara, Y. Hamamoto and I. Sakaida, Lymph node metastasis can be determined by just tumor depth and lymphovascular invasion in early gastric cancer patients after endoscopic submucosal dissection, European Journal of Gastroenterology & Hepatology, 査読有, DOI 10.1097/MEG.0000000000000987 Vol.29, pp.1346-1350, 2017.

H. Ogihara, N. Iizuka, Y. Hamamoto, Prediction of Early Recurrence of Liver Cancer by a Novel Discrete Bayes Decision Rule for Personalized Medicine, BioMed Research International, 査読有, Vol. 2016,

8567479, 2016.

〔学会発表〕(計 0 件)

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 1 件)

名称：情報処理装置と情報処理プログラム並びに情報処理方法
発明者：浜本義彦、荻原是宏、飯塚徳男、爲佐卓夫、岡 正朗
権利者：山口大学
種類：特許
番号：特願 2016-35605
出願年月日：平成 28 年 2 月 26 日
国内外の別：国内

取得状況(計 1 件)

名称：情報処理装置と情報処理プログラム並びに情報処理方法
発明者：浜本義彦、荻原是宏、飯塚徳男、爲佐卓夫、岡 正朗
権利者：山口大学
種類：特許
番号：特許第 6041331 号
取得年月日：平成 28 年 11 月 18 日
国内外の別：国内

〔その他〕

ホームページ等

6. 研究組織

(1) 研究代表者

浜本 義彦 (HAMAMOTO YOSHIHIKO)
山口大学・大学院創成科学研究科・教授
研究者番号：9 0 1 9 8 8 2 0

(2) 研究分担者

飯塚 徳男 (IZUKA NORIO)
広島大学・医歯薬保健学研究科(薬)・教授
研究者番号：8 0 3 3 2 8 0 7
藤田 悠介(FUJITA YUSUKE)
山口大学・大学院創成科学研究科・准教授
研究者番号：4 0 5 0 9 5 2 7

(3) 連携研究者

なし