

遺伝子型セットの組合せ最適化による薬剤の作用・副作用予測

山本 郁[†] 藤田 悠介[†] 浜本 義彦[†] 恒富 亮一[†] 碓 彰一[†] 岡 正朗[†]

[†]山口大学 大学院医学系研究科

1. 緒言

がん治療の困難性はその多様性にあると言われており、がん治療のための個別化医療が求められている。がん治療のために抗がん剤を投与する際、個人によって作用は異なり、最悪の場合には効果がなく副作用のみ現れることがある。したがって抗がん剤などの薬剤を投与する際に作用や副作用を予測することは極めて重要である。

抗がん剤の副作用と遺伝子型との関係に注目し、バイオマーカーの発見を目的とした研究が多くなされている[1]。しかし、従来の研究の多くでは、単独あるいは2つの遺伝子型との組合せしか考慮されていない。また、作用・副作用のメカニズムは複雑なため、従来の単独あるいは少数の遺伝子型セットでは予測が困難である。

本研究では複数の因子からなる条件式（遺伝子型セット）を組合せて適用性と信頼性の高い予測式を生成する方法を開発している。候補となる因子の中から有用な因子の組合せのみを抽出し、それらを組合せて用いることが可能である。また、遺伝子型だけでなく年齢や性別などの因子を含めた条件式も予測に用いることができる。

2. 予測システム

2.1. 概要

本システムは被験者の遺伝子型や性別などを含む症例データを入力とし、識別器によって「副作用有」、「副作用無」各クラスの予測信頼度をつけ、信頼度が高いクラスへと識別する。システムの概要を図1に示す。

本システムは、クラスごとに異なる信頼度をもつ複数の識別器から構成される。識別器は学習用の症例データにより設計する。予測時には、予測対象の症例データに合致する条件を持つ識別器のうち信頼度が最も高い結果を採用する。各クラスの信頼度が同値の場合やいずれの識別器にも合致しなかった場合には予測不可とする。

2.2. 設計

設計の流れを図2に示す。条件式は遺伝子型や

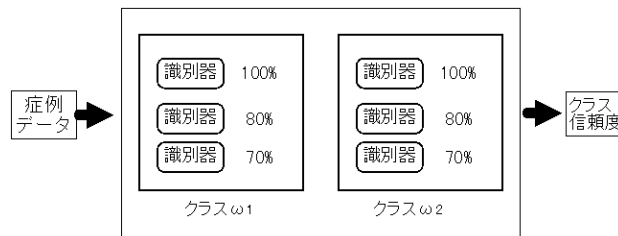


図1 システムの概要

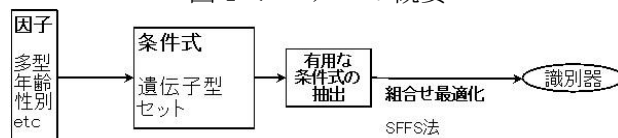


図2 設計の概要

年齢、性別などの因子から構成される。

はじめに、因子の組合せを網羅した条件式を生成する。組合せに用いる因子数は $1 \sim n$ (n : 因子の総数) である。次に予め設定しておいた該当数および信頼度の下限値未満の条件式を削除し、また冗長な条件式の削除を行う。SFFS法による条件式の組合せ最適化により識別器が設計される。

識別器の設計は各クラスで、信頼度下限値の異なる持つ複数の識別器を設計する。高い信頼度を持つ識別器は該当数が少なくなるため、複数の識別器を設計する。条件式の組合せ最適化では条件式を追加しても性能が変化しない場合に最適化を終了する。

2.3. 組合せ最適化

SFFS法は特徴選択のための最適化技法の一種である[2]。SFFS法は特徴の追加および削除を繰り返し最適な組合せを探索する手法である。

組合せ最適化の評価関数には該当数と占有率を用いる。該当数は、その条件式に合致する症例数を指し、適用性の高さを示す指標である。占有率は、該当する症例のうち「副作用有」あるいは「副作用無」の症例が占める割合であり、予測時の信頼性の高さを表す指標である。

2.4. 予測

予測では、識別器に含まれる条件式に症例が合致するかを調べる。合致した場合には識別器の信頼度を予測の信頼度とする。また、複数の識別器に合致した場合には、最も高い信頼度を持つ結果を予測結果とする。全ての識別器で合致するかを調べ、最終的に信頼度の高いクラスへ予測する。

Prediction of severe toxicities and tumor responses using combinational optimization of genotype subset

[†]Iku YAMAMOTO [†]Yusuke FUJITA

[†]Yoshihiko HAMAMOTO [†]Ryouichi TSUNEDOMI

[†]Syoichi HAZAMA [†]Masaaki OKA

[†]Graduate School of Medicine, Yamaguchi University

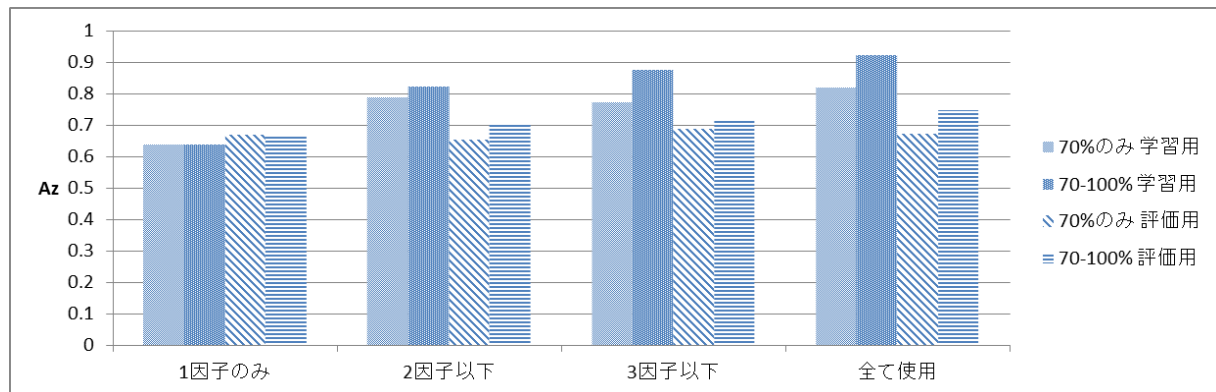


図3 予測結果のAzによる比較

表1 症例データの内訳

	副作用有	副作用無	合計
学習用	35	40	75
評価用	17	33	50

3. 実験

大腸がんの抗がん剤イリノテカンを投与された症例データ[1]を用いて副作用の予測を行い評価した. 評価にはROC解析のAzを用いた. 本手法の最適化において, 組合せに用いる条件式の因子数の影響を調べるために, 1因子の条件式のみを用いる場合, 2因子以下の条件式のみを用いる場合, 3因子以下の条件式のみを用いる場合, さらに全ての条件式を用いる場合の4通りを評価した. 該当数3以上かつ占有率70%以上の条件式を抽出し, 条件式の組合せを最適化した.

3.1. 症例データ

評価実験には抗がん剤イリノテカンを投与された125例の症例データを用いた. 設計用データと評価データの数の内訳を表1に示す. また, 症例データ中の因子は以下のとおりである.

- UGT1A1*6(211G/A): Wild, Hetero, Homo
- UGT1A9*22(-118T): Wild, Hetero, Homo
- UGT1A7(N129K): Wild, Hetero, Homo
- UGT1A7(W208R): Wild, Hetero, Homo
- UGT1A1*60(-3279 T/G): Wild, Hetero, Homo
- 性別: 男性 or 女性
- 年齢: 60歳以上 or 未満
- 投与回数: 1回目 or 2回目

条件式の総数は $27647 = 4^5 \cdot 3^3 - 1$ である.

3.2. 実験結果

ROC解析により得られたAzを図3に示す. 1因子の条件式のみ, 2因子以下の条件式のみ, 3因子以下の条件式のみ, および全ての条件式を候補として用いた場合を示す. 各候補を用いて70%の識別器のみを構築した場合と70%, 80%, 100%の識別器(以下, 70-100%とする)を構築した場合の学

習データおよび評価データに対するAzを示す. 評価用データに対して全ての候補条件式から70-100%の最適化を行ったときにAzは最大の0.746となった. また, いずれの場合にも70%最適化のみよりも70-100%最適化の方がAzは高い.

3.3. 該当数の下限値の影響

該当数の下限値の影響を調べるために, 該当数の下限値を2および5として同様の実験を行った. 全ての条件式を候補して70-100%最適化を行い, 評価用の症例データを予測した. 該当数2以上のときのAzは0.650, 該当数5以上のときのAzは0.675であり, 該当数3以上の場合より低い.

該当数の下限値を2とした場合には, 学習データに対してのみ有用である条件式が選択されている. また, 該当数5という下限値は, 本実験の学習データ数に対して大きいため, 本来予測に有用である条件式を抽出できていないと考えられる.

4. 結言

本研究では複数の因子からなる条件を組合せて適用性と信頼性の高い予測式を生成する方法を提案した. また, 評価実験によりその有効性を示した.

今後は, 過学習への対策やブースティングによる学習などの他手法による組合せ最適化を検討する予定である.

参考文献

- [1] S. Hazama, et al, "UGT1A1*6, 1A7*3, and 1A9*22 genotypes predict severe neutropenia in FOLFIRI-treated metastatic colorectal cancer in two prospective studies in Japan", Cancer Sci, Vol.104, No.12, pp.1662-1669, 2013.
- [2] 浜本義彦, "統計的パターン認識入門", 森北出版, 2009.