

| | |
|-------|------------|
| 六物黄芩湯 | ロクモツオウゴントウ |
| 六物附子湯 | ロクモツブシトウ |

| |
|---------------------------------|
| 北里処方集から「烏薬順気湯」を削除 |
| 北里処方集の「(香川)解毒剤」を削除 |
| 北里処方集の「葛根湯加辛夷川芎」を削除 |
| 北里処方集の「加味逍遙散」を削除 |
| 北里処方集から「桂枝加苓朮附湯」を削除 |
| 北里処方集から「桂枝芍薬知母湯」を削除 |
| 北里処方集から「千金内托散料」を削除 |
| 北里処方集の「八味丸料」を削除 |
| 北里処方集の「六味丸料」を削除 |
| 北里処方集の「五淋散料(加味)」を「加味五淋散料」に変更 |
| 北里処方集の「沈香降気湯」のよみがなを「デンコウキトウ」に変更 |

表 4 富山大学による北里処方への追加処方

| 処方名 | カナ |
|-------------|-----------------------|
| 黄芩湯 | オウゴントウ |
| 黄芩加半夏生姜湯 | オウゴンカハンゲシヨウキョウトウ |
| 黄土湯 | オウドトウ |
| 黄連解毒湯(万病回春) | オウレンゲドクトウ |
| 葛根加苓朮附湯 | カッコンカリヨウジュツプトウ |
| 橘皮枳実生姜湯 | キツピキジツシヨウキョウトウ |
| 橘皮竹茹湯 | キツピチクジョウトウ |
| 桂枝加桂湯 | カイシカケイトウ |
| 桂枝加芍薬生姜人参湯 | ケイシカシャクヤクシヨウキョウニンジントウ |
| 桂枝加芍薬大黄湯 | ケイシカシャクヤクダイオウトウ |
| 桂枝加附子湯 | ケイシカブシトウ |
| 桂枝二麻黄一湯 | ケイシニマオウイトウ |
| 厚朴三物湯 | コウボクサンモツトウ |
| 柿蒂湯 | シテイトウ |
| 赤丸料 | セキガンリョウ |
| 通脈四逆湯 | ツウミヤクシギャクトウ |

表5 東京女子医科大学による北里処方への追加・変更処方

| 処方名 | カナ |
|---------------|---------------|
| 黄解丸料 | オウゲガンリョウ |
| 枳縮二陳湯 | キシユクニチントウ |
| 桂枝加朮苓湯 | ケイシカジュツリョウトウ |
| 桂枝加朮苓附湯 | ケイシカジュツリョウブトウ |
| 甲字湯加大黄 | コウジトウカダイオウ |
| 清肺湯【万病回春 便血門】 | セイハイトウ |
| 大黄一物湯 | ダイオウイチモツトウ |
| 猪苓散料 | チョレイサンリョウ |
| 抑肝扶脾散料 | ヨクカンフヒサンリョウ |

| |
|---|
| 北里処方集の「参苓白朮散料」のよみがなを「ジンリョウ(レイ)ビャクジュツサンリョウ」に変更 |
| 北里処方集の「清湿化痰湯」のよみがなを「セイシツケ(カ)タウトウ」に変更 |
| 北里処方集の「清肺湯」を「清肺湯【万病回春 咳嗽】」に変更 |
| 北里処方集の「旋覆花代赭石湯」のよみがなを「センブクカタイ(ダイ)シャセキトウ」に変更 |
| 北里処方集の「薏苡仁湯【勿誤薬室方函】」を「薏苡仁湯【勿誤薬室方函・外科正宗】」に変更 |

漢方の特性を利用したエビデンス創出と適正使用支援システムの構築

研究分担者 宮野 悟 東京大学・医科学研究所・ヒトゲノム解析センター
共同研究者 井元清哉 東京大学・医科学研究所・ヒトゲノム解析センター
山口 類 東京大学・医科学研究所・ヒトゲノム解析センター
片山琴絵 東京大学・医科学研究所・ヒトゲノム解析センター

研究要旨

慶應義塾大学病院漢方専門外来などにおいて集められた患者個々人の自覚症状が記された貴重な診療支援材料である問診データから、漢方治療の科学的エビデンスを創出し、そのエビデンスに基づき漢方の特性を生かした診療支援システムを構築することを研究の目的とする。その成果として、問診データを用いてデータマイニングにより漢方診断である証について、虚実を90%以上、寒熱を約85%と高い精度で予測できることを示した。また、気血水においても、問診データと気血水の証との関連を数学的に解析し、予測システムの基盤を構築することができた。その中で、これまで証の決定に関連すると言われていた身体的特徴に該当する問診項目だけでは実際に診断された証を予測するのに十分ではない場合があることを示した。今後、データを増やし、漢方診断と問診データの関連をより鮮明にすることで、漢方診断を予測するシステムの高精度化は必要不可欠である。そのためには、問診データにおける項目の見直し、それらと証との非線形な関係を数理モデル化し、証と証の関係にも踏み込む必要があることが明らかとなった。また、予測結果の出力形式についても、医療現場において有用な形式を探索する必要がある。

A. 研究目的

本研究は、慶應義塾大学病院、富山大学病院、千葉大学病院、自治医科大学病院、東北大学病院、東京女子医科大学の漢方専門外来に加えて、鹿島労災病院、秋葉伝統医学クリニック、麻生飯塚病院、亀田総合病院において集められた問診データを、最先端のデータマイニング手法を駆使し解析することにより、漢方診断の科学的なエビデンスを創出し、診療を支援するシステムを構築することを目的とする。その概略図を図1に示す。

患者が自分自身の自覚症状を慶應義塾大学病院漢方専門外来など各施設に設置された問診システムに入力し、問診データは集められている。我々は、そのデータをデータマイニング手法により解析し、漢方専門医ではない医師の漢方的診断の助けとなるような診療支援に繋がる情報、処方薬を選定する支援になるような情報、そして、その治療による効果の予測を行うことを目指している。そのためには、問

診システムに蓄積されたデータからさまざまな情報を抽出するための情報解析技術が必要となる。我々は最先端のデータマイニング手法を用い、平成23年度においては、慶應義塾大学病院漢方専門外来において収集された問診データからの漢方診断「証」の予測を試みた。正確には、百を超える項目について記録された患者個々人の問診データから、漢方専門医がその患者に診断すると思われる漢方診断(「虚実」、「寒熱」、「気血水」のそれぞれ)の予測、および、処方される漢方薬の予測をコンピュータにより行う方法論、および、その計算プログラムを開発する。

B. 研究方法

慶應義塾大学病院漢方専門外来に設置されている自動問診システムにより収集された患者自身の自覚症状を表す問診データ、および漢方専門医による漢方診断の結果を用いて、問診データから漢方診断

を予測する数学的方法を構築する。問診システムには、362の項目（VAS値を入力するための項目を含む）があるが、本解析においては、昨年度の研究実績に従い、例えば手の左右、足の左右などは区別せず、単に手、足として取り扱った。その際、症状の強弱を表す Visual Analog Scale (VAS) は、左右で大きい方を採用することにした。昨年度の結果によると、例えば、手の冷え足の冷えなどで左右の差の大きい患者は見当たらなかった。その点から言うと平均を用いても構わないのであるが、もし、左右に差のある患者が将来表れた場合に、その方の主訴を見逃さないために最大値を用いることにした。また、家族構成などの項目を除き、男女を分けない解析においては、女性特有項目（月経困難症や月経痛など）は用いていない。しかしながら、女性に絞った解析においては、これら女性特有項目は使用している。

B.1 虚実

まず、漢方診断「証」のうち、虚実より予測式を構築することを始めた。虚実においては、患者は

- 「虚証」
- 「やや虚証」
- 「虚実中間証」
- 「やや実証」
- 「実証」

の5段階に分けられている。用いたデータ13640レコード中、5段階のそれぞれのサンプル数は図2にあげるように、2187、2244、7005、1098、1106となり、もっとも多い証は「虚実中間証」であり、「虚証」、「やや虚証」の方が、「実証」、「やや実証」の患者よりも多い（約2倍）であることがすぐにわかる。まずは、このデータに昨年度の研究において、問診データからの西洋病名予測において高い予測性能を示したロジスティック判別分析モデルを Elastic net を用いて推定する方法を試した。ここで、ロジスティック判別モデルとは、 i 番目のレコード（患者）の j 番目の問診項目のデータを x_{ij} と表したとき、その患者に対して2つのカ

テゴリの比較を行うときに使用し、線形モデル

を用いる。例えば、 i 番目の患者が「虚証」か「実証」かという2つのカテゴリのどちらに属するかという比較を行う際、 i 番目の患者が「虚証」に属する確率を

で計算するものである。ここで、 a_0, \dots, a_m がパラメータであり、Elastic net の推定方法は、

$$L_E(\mathbf{a}) = L(\mathbf{a}) - \sum_j \lambda |a_j| - \frac{\nu}{2} \sum_k a_k^2$$

を最大にするパラメータ値 \mathbf{a} を推定する。ここで、 $L(\mathbf{a})$ は尤度関数、 λ は Lasso パラメータと呼ばれ、この値の大小によりどの程度の変数をモデルから除外するかが決定される。また、 ν は Ridge パラメータと呼ばれ、問診項目間のグループ構造を捉えることに寄与する。Lasso パラメータ、Ridge パラメータはともにモデルの性能に大きく影響を与えるパラメータであるが、本解析の中では全てベイズ型モデル評価基準（Bayesian Information Criterion; BIC）によって最適な値を選び用いた。また、問診項目間の交互作用を取り込むため、データマイニング手法の一つであるサブセットマイニング法 Eclat を用いて、5つの問診項目までの交互作用で、かつ、サポート（出現頻度）が0.1以上のもの17070個を線形モデルの中に取り込んで予測式を構築した。

このロジスティック判別モデルに Elastic net を組み合わせた方法は、平成22年度に行った西洋病名を問診データから予測するという研究において高い予測性能を発揮したが、今回も高い予測性能を示すことが出来るかという保証はもちろんない。従って、平成23年度においては、より高い予測精度を持つ方式の探索を目的に、ロジスティック判別モデルに Elastic net を組み合わせた方法に加え、サポートベクトルマシン、Classification and Regression Tree (CART)、Random Forest (RF) を試して、問診データから漢方診断「証」の予測においてもっとも高い性能を示した方法を利用することにした。

サポートベクトルマシンは、判別問題を解く際に、クラス間の最小距離を与えるサンプルをサポートベクトルと呼び、そのサポートベクトル間の距離をもっとも大きくする徴兵面によって判別境界面を決定する方法である。徴兵面によって分離可能なデータのことを線形分離可能と呼ぶが、線形分離可能ではないデータについては、より高次元の空間への射影を考え、その空間における線形分離を試みる。この作業を仮想的に行う方法がカーネルトリックと呼ばれる。また、線形分離可能ではない場合に、分離できないサンプルに対してスラッグ変数と呼ばれるペナルティを考慮した超平面の学習が行われる。カーネルトリックを用いることにより、非線形な境界面での判別を行うことができること特徴であり、線形モデルであるロジスティック回帰モデルに Elastic net を組み合わせた前述の方法との大きな違いとなる。

CART は、ルートノードに全てのサンプルを入れておき、分岐させていくことでサンプルのラベルが均一なノードを創っていく方法である。分岐の際には、サンプルを特徴付けている変数（ここでは問診項目に相当する）の一つの値がある値よりも大きいグループ、小さいグループに分ける。その際、ある値（閾値）は、分けられた2つのグループ（ノード）に分けられたサンプルのラベルがもっとも均一になるように求められる。また、どの変数で分離させるかも、分けられた2つのグループにおけるラベル情報をもっとも均一になる変数で分けることになる。この操作により、ルートノードは2つのノードに分けられ、また、分けられたそれぞれのノードは次の分岐においてそれぞれ2つに分けられる。つまり、計4つのノードになる。もし、どの変数で分けてもラベルの均一性がそれ以上改善しない場合は分けることはせず、そのノードに関しての分岐はこれ以上行わない。また、あるノードに含まれるサンプル数が最初に設定した数以下になった場合もそれ以上の分岐を行わないことが多い。このようにして木 (Tree) を作り、leaf ノードによって分類する方法が CART である。

Random Forest は、CART を用いてより予測能力の高い方式を構築するための方法である。その方法は、今、 n 人のサンプルがあって、それらを x_1, x_2, \dots, x_n と書いたときに、それからブートストラップ標本 $x_1^*, x_2^*, \dots, x_n^*$ を復元抽出により作成する。このブートストラップ標本を用いて CART と同様の方法によって Tree を作っていくのであるが、分岐を作る候補の変数もランダムに少数の変数を選びその中からのみ分岐に採用する変数を選ぶことにする。このようにして Tree を作るのであるが、この操作を例えば1000回繰り返して、1000本の Tree を作り、最終的な予測はこれら1000本の Tree によるそれぞれの結果のアンサンブルによって得られる。この方法の特色は、分岐を作る際、少数の変数に候補を制限して Tree を作るため、元データにおいて相関の高い変数が含まれていた際にも有用な方法（高い予測性能の期待できるという意味で有用）であると言える。

問診データから漢方診断を予測するという文脈において、もっとも予測性能の高いデータマイニング手法を決定した後は、その予測能力の向上を出来る限り行う。また、問診データと共に、患者の身体的特徴（身長、体重から計算される BMI）の利用も検討する。

B.2 寒熱

「寒熱」が「虚実」と異なる点は、虚実は5つの順序カテゴリーである「証」からなっていたが、寒熱では、「寒証」、「中間証」、「熱証」、「上下寒熱」、「手足煩熱」、「蕨冷」、「寒疝」という順序であるものと無いものが混在していることである。本厚生労働科学研究費の研究代表者を務めている渡辺賢治准教授のアドバイスを受け、このうち、「上下寒熱」と「手足煩熱」は「寒熱錯雑証」としてまとめ、「蕨冷」と「寒疝」に関しては「寒証」とした。従って、寒熱の証は

「寒証」

「中間証」

「熱証」

「寒熱錯雑証」

の4カテゴリーとなる。この中で、「寒証」、「中間証」、「熱証」は順序カテゴリーと見なせるが、「寒熱錯雑証」に関しては、「寒証」と「熱証」の性質を併せ持った「証」と見なすことが出来るため、「寒証」、「中間証」、「熱証」の順序とは別のカテゴリーとなる。このため、「寒熱」の4つの証に対する問診データからの予測は「虚実」よりも複雑なものとなることが予想される。例えば、「寒証」と「熱証」で予測式を作った際に、寒証と熱証のどちらとも言えないような場合は「中間証」か「寒熱錯雑証」のどちらかであるというような推論が成り立ち、2段階での予測式を構築することが考えられよう。また、「寒証」、「熱証」、「寒熱錯雑証」の3カテゴリーでこれらを順序カテゴリーとしては見なさず、3カテゴリーの判別ルールを構築することも考えられよう。このような「虚実」よりも複雑な状況での問診データからの予測法の構築をデータマイニング手法により行う。また、「寒熱」に関しては、手の「冷え」や「ほてり」によって比較的簡単に予測できるのではないかという直感が本当に成り立つのかということも実際に得られている問診データと漢方診断との関係を見ることで検証する。そして、それが単純なマッチングでは対応するとは言えないことを示す。

B.3 気血水

気血水においては、「虚実」、「寒熱」よりもさらに証のカテゴリーは増え

「気虚」

「気うつ」

「気逆」

「血虚」

「お血」

「水毒」

「亡津液」

「下焦の虚」

の8つのカテゴリーからなる。

このうち、「亡津液」は、患者数でカウントすると3名のみとなり、解析できるようなサンプル数が無いため削除し

た。また、「下焦の虚」については、患者数でカウントすると44名となり、他の証(例えば、気虚は650名、気うつは501名)と比べた際には明らかに少なく、サンプル数がアンバランスとなり解析結果にバイアスをもたらすことは明らかであるが、まずは入れて解析を行い、その結果を踏まえて外すべきか否かを判断することとした。残り6カテゴリーにおいて、もっともサンプル数の少ない証は「気逆」であり192名の患者が相当する。この6ないし7カテゴリーのデータに対して、様々なグループの定義を試し、予測能力の高いデータマイニング方式の確立を目指す。また、データマイニング法については、気血水については、グループの定義も様々に変化させて最適なものを探索するため、その組み合わせが膨大になってしまうため、虚実、寒熱において高い予測能力を示したものを使用することとする。また、男女を層別化しての解析や年齢での層別化しての解析などの可能性も試すこととする。この気血水は、順序カテゴリーでもなく、それぞれが独立なカテゴリーとして見なされるものであるため、データマイニングの問題としては、「虚実」、「寒熱」よりも遙かに難易度の高い問題であることは明らかである。従って、気血水に関する既存の知識(例えば、気虚という証を診断する際には、漢方専門医がどのような患者の状態をもって判断しているか)を整理して、それがデータからどのように見えてくるか、また、データがどのようにその既存知識を反映しているかについても解析を行うこととする。

C 研究結果

C.1 虚実

まず、全13640レコードを用いて、Elastic netによってロジスティック判別モデルを推定する方法を用いた。その結果を図2にまとめた。

行った解析は

(1) 「虚証」 vs 「やや虚証」 + 「虚実中間証」 + 「やや実証」 + 「実証」

(2) 「やや虚証」 vs 「虚証」 + 「虚実中間証」 + 「やや実証」 + 「実証」

(3) 「虚実中間証」 vs 「虚証」 + 「やや虚証」 + 「やや実証」 + 「実証」

(4) 「やや実証」 vs 「虚証」 + 「やや虚証」 + 「虚実中間証」 + 「実証」

(5) 「実証」 vs 「虚証」 + 「やや虚証」 + 「虚実中間証」 + 「やや実証」
の5パターンである。

結果、「虚実中間証」を除いて判別の正解率は80%を超えるものであったが、その内容に問題点を見出した。図2下の5つのパネルが(1)～(5)のそれぞれの結果であるが、左の黒で表した患者が、vsの右に書いた証を診断された患者(x軸がサンプルの番号を並べ替えたものに対応)とその証であると計算された確率(y軸)である。パネル内の赤で書かれた患者はx軸にサンプルの番号(y軸の値に応じて並び替えている)とy軸が証の確率である。具体例で説明しよう。一番左のパネルは「虚証」と「虚証以外の証」の判別である。黒のサンプルは「虚証以外の証」に総投資、赤のサンプルが「虚証」のサンプルである。y軸は「虚証」である確率を表すので、赤のサンプルでy軸の値が0.5以上のサンプルは虚証であると判断されるので正解となる。また、逆に、虚証以外のサンプル(黒のサンプル)においては、0.5以下になったときに正解なので、0.5のラインより下のサンプルが正解となる。従って、黒で表した「虚証以外の証」のサンプルは大半が正解となり、赤で表した「虚証」のサンプルは半分強が誤りとなっている。

このように、2つのカテゴリにおいて正答率が大きく異なってしまうことが分かった。その理由についてもさまざまな解析を行い突き止めることが出来た。もっとも大きな問題は、サンプル数のバランスが大きく偏っていることである。同じ例で説明すると、「虚証」vs「虚証以外の証」での比較においては、大半のサンプルは「虚証以外の証」を診断されたサンプルであり、「虚証」でるサンプルはマイノリティとなってしまっている。もし、極端に片方のカテゴリに属するサンプルが全体の99%を占めてしまうような場合が生じたら、データをなにも見ることなく、全てのサンプルをマジョリティであるカテゴリに割り付けても99%は正解となる。

予測性能を単に正答率だけで見ると、このような不合理が生じるため、その内容にまで配慮する必要がある。平成23年度に行った我々の解析においては、単に高い正答率ということだけではなく、このようにその内容(どのように間違っているか)についても十分に配慮し、データ数が各カテゴリ(証)でアンバランスな場合には、単に高い正答率だけを求めるのではなく、そのバランス(間違え方)にも配慮して証の予測式を構築した。

このサンプルのアンバランスは、よりサンプル数の少ない実証の予測においてさらにクリティカルなる。正解率は94%以上と高いものの、それはほとんどが「実証ではない」患者に対する正解となっていることが分かる。従って、図2の正答率はあまり過信することはできないことが分かる。

それでは、各患者について、このような(1)～(5)の比較による確率値が5つ求まることになったので、患者内の5つの確率の大小でもってその患者がどの証なのかを判断することにする。すなわち、ある患者に対して、(1)～(5)の比較における確率値((1)の比較における確率値とは、その患者が「虚証」である確率を表し、(3)における確率値とは「虚実中間証」でる確率を表す)が、例えば、(1)0.7、(2)0.5、(3)0.4、(4)0.3、(5)0.3であったならば、その患者は確率のもっとも高い(1)の「虚証」と判断するルールである。この方式と確率の絶対値を見ることにより、その患者をどの証と判断するかについてそのエビデンスの強さも表現することができる。

この各患者内での虚実の5つの証に対応する確率値のコントラストによって患者の証を予測した結果を図3にまとめた。5つの確率値の中でもっとも高い確率値に対応する「証」が漢方専門医の診断した「証」と一致した正答率は64.4%であった。これは、要するに3択の問題なので、もし、ランダムに「証」を選択すると、正答率は33.3%となる。すなわち、ランダムの倍近くは正解できていることになる。また、5つの確率値のうち、1位と2位に対応する「証」が漢方専門医によって診断された「証」

を含んでいる割合は88.1%となった。これも、もし、ランダムに1位と2位に相当する「素証」を選んだとすると66.7%が正答率になるので、ランダムより20%強の正答率が上乘せされたことになる。また、男性、女性に分けて正答率を評価してみたが、1位との一致、1位、2位との一致の両方において、男女での大きな違いは見られなかった(若干、男性の正答率が低い傾向にあったが、これもサンプル数のアンバランスが原因かと推測される。実際、女性患者の方が男性患者よりも数が多い)。

予測したい虚実の「証」は5つあり、それは、統計学の用語では順序カテゴリーと呼ばれるものとなる。すなわち、カテゴリーに順序が付いている。そのような場合、同じ間違いでも漢方専門医が「虚証」と診断したときに「やや虚証」と予測するのと、「実証」と予測するのでは、誤りの程度が異なる。前者は後者よりも正解に明らかに近く、このような予測が出来るのであればたとえ実際の診断にぴったりにあたらずとも漢方専門医以外のいしにとっては有用な情報である可能性もある。それでは、5つの証のうちで確率ももっとも高かった証とその隣の証が漢方専門医の診断した「証」と一致するかを調べた結果が図4である。正解率は82.3%であり、かなりの正解率になることが分かる。ここでは、正解率に男女の差があるかどうか、年齢による差があるかどうかをみるための情報を載せている。男女差については、それほど大きな違いはないが、先ほどと同様に若干男性に不正解の割合が高い傾向が出ている。また、年齢については、正解患者、不正解患者の年齢分布を計算した。分布としての明らかな差としては、不正解の患者は年配の方の割合が高いことがわかる。逆に、正解した患者は若い患者が多いことがわかる。

いま紹介した結果は、全レコードを使用した結果であった。このデータは、ある患者が繰り返し受診した際、その都度問診システムを用い自覚症状を記録していくため、同一患者のデータが(記録された時期はもちろん異なるが)繰り返し計測されていることとなる。つまり、前述したサンプル数のアンバ

ランスが引き起こす結果のバイアスを考えると、繰り返し受診する患者の状態は徐々に変化していくと考えることが出来るため、それらはあるまとまった構造をデータの空間において形成すると想像される。そのような場合に、繰り返し受診する患者を重点的に正しく判別する方式の方がより高い正解率に繋がるため、受診回数の多い患者をより重点的に正しく判別するような判別方式が構築されるであろう。

受診回数の多い患者が正しく判別されると言うことは、そのような患者にとってはメリットがあると思われるが、一般に、学習データにおいて、そのような患者にバイアスのかかった学習をしてしまうと、テストデータに対する予測性能の向上はあまり望めない。従って、次に、患者初診時の問診データのみを用いて虚実を判定する方式の構築を試みる。初診時での問診データと言うことで、主訴の見極めにも関連するかと思われ、非常に大切な判断を行う時期における診療支援を狙っている。

各患者初診時の問診データに絞ったことで、漢方診断を予測する方式構築に用いることの出来るサンプル数は、図5のようになる。大きく数が減ることがわかる。この初診時の問診データに対して、前述した

(1) 「虚証」 vs 「やや虚証」 + 「虚実中間証」 + 「やや実証」 + 「実証」

など(1)～(5)の判別方式を構築して、5つの虚実における証の確率を計算したものが図5の下の5つのパネルとなる。また、図5の表の一番右のカラムには「係数の個数」とあるが、これは、Elastic netによって、不要と思われる変数(問診項目)にかかる係数は0(ゼロ)となり、モデルからその問診項目は除かれるが、この個数とは、係数がゼロではない問診項目の個数である。つまり、証の予測に使用した問診項目の個数に等しい。ここで注目すべきは、係数の個数の少なさである。ほとんど問診項目を使用することが出来ていない。これは、それぞれの項目の証に対する関連が弱いとElastic netにより判断され、係数ゼロとなっている。このような場合の原因は(1)本当に関連がない、(2)線形の関係

ではない非線形な関係がある、(3) 問診項目間に相関が高いものが多数有り、それらの変数をモデルに取り込むために Ridge パラメータを大きくせざるを得ず、すると係数の値が小さく見積もられるため、Lasso 項によりゼロと判断されてしまう。という3つの原因が考えられる。ここでは、(2)、もしくは(3)がその主たる原因であると考え、Elastic net だけではなく、他のデータマイニング手法、統計的判別分析の手法を試すこととした。そして、その中で最も性能の高かった方式を採用することとした。試した手法は、

- ロジスティック判別モデル (Elastic net)
- ロジスティック判別モデル(最尤法による推定、AICによる変数選択)
- サポートベクトルマシン (SVM)
- Penalized SVM
- Classification and Regression Tree (CART)
- Random Forest (RF)

また、5つの証をそれぞれ予測するのは、例えば、「虚証」と「やや虚証」は非常に近い状態であると思われるし、「実証」と「やや実証」に関しても同様である。すなわち、これら5つの証は段階的な意味がある順序カテゴリーであるため、それぞれを独立に予測するのは効率的ではない。従って、「虚証」vs「実証」の比較を行い、その確率値によって、どの程度、虚証としての性質を強く持つか、逆に実証としての性質を強く持つかを表すようにした。従って、この方式により、図1の概略図によって示したような段階的な情報提供を可能とする。

判別方式の性能比較を行った結果を図6にまとめた、Random Forest は VAS 値を使わずに、単にその項目が選ばれたかどうかという情報を使った結果 (without VAS) と、VAS 値を各個人ないで normalize したデータを用いた結果の2種類の結果をあげた。また、その他の方法は全て normalized VAS を用いた結果である。サポートベクトルマシンと Random Forest がこの中ではかなり高い予測性能を示すことが出来た。逆に線形モデルであるロジスティック判別モデルや Classification and Regression Tree

(CART) は予測性能があまり上がらなかった。すなわち、患者の自覚症状が記録された問診データから漢方診断虚実の証への写像は、線形ではない医師の判断がなされていると言うことがこの解析から示唆され、また、高い予測性能を達成しようとしたときには、問診項目間の高い相関の中で有効に働くような方法が必要となると言うことがわかる。実際、平成22年度の研究において、問診項目間の相関構造をデータマイニング手法 apriori や eclat を用いて探索したが、非常に高い相関を示す問診項目がいくつも抽出され、また、apriori のルールマイニングにおいてもそれらの高相関な項目間に統計的に意味のあるルールを見いだすことが出来ている。その意味では、Random Forest が最も高い予測性能を示すことが出来たということは理にかなっていると言える。

Random Forest は、一般に、高次元データでかつ変数間に強い相関構造があるときにでも高い予測性能を示すことが出来るという特徴があるが、その学習過程における各変数の重要度が計算されているという副産物がある。その情報を図7にあげた。もっとも重要度が高いと判定された問診項目は、「日常生活：食事の早さ：早い」であった。この重要度が計算されている結果からは、各項目が選ばれた、または VAS の値が大きかったということが、虚証に関する特徴として選ばれているのか、それとも実証に関する特徴として選ばれているのかということは分からない。しかしながら、もちろんデータに戻って確認したり、Random Forest が作る各 Tree で注目している項目が含まれる Tree を抽出することで、重要とされた項目がどちらの特徴を表しているのかを知ることが出来る。この「食事が早い」という項目は、これが選ばれていると実証的な傾向があると判断される項目であった。これは理にかなっているであろう。2位は「全身症状：その他：寒がり：VAS」である。これは VAS 値の項目であるため、自分自身を寒がりであると感じていればいるほど虚証の傾向があるという結果になっていた。3位は「全身症状：その他：汗をかきやすい：VAS」である。これは、汗をかきやすいと思っていればいるほど実証である

傾向があるという結論であった。1、2、3位に上がった項目以外にも、虚実を判定する際に妥当な項目が多く上がっているという結論を得ることができた。

トレーニングデータとして、虚証、実証それぞれから120名ずつのデータをランダムに抽出しRandom Forestを学習し、残りのデータを用いてテストしたわけだが、スタートとなる120名をさまざまに変えて学習し、テストするという手順を繰り返した。その結果を図8にあげる。この3つは正答率の低かったものであるが、それでも71%以上と、他の手法よりも高い予測性能を達成できていることが分かった。また、この成績があまり振るわなかった結果を解析することで問診データによる漢方診断の予測に関する問題点も見えてきた。

図8にあげた3回において、テストで予測が誤ったケースを解析した。その結果、この3回共に予測が誤ったケースが一例見つかった。それは67歳女性のデータで漢方診断は「虚証」となっていた。しかしながら、問診データに戻ってみると、この患者は問診システムにほとんど情報が入力されていないことが分かった。例えば、先ほどの「全身症状：その他：寒がり：VAS」という項目は、全く症状がなければ0、ひどい寒がりであると感じている方は100を付けるであろう。この0と100という値は、取り得る値の最小値、最大値となり、数の上では対象であると言えよう。もし、0が実証であるということに対応し、100が虚証であるということに対応するならば、数字の対称性と、その意味の対称性がマッチし、0にも強い意味が出てくる。しかしながら、そうではないケースが多いと予測される。すなわち、100であるならば強く虚証であると判断されるが0は実証であるという判断材料にはならないというケースである。このような場合、問診システムに入力が少ない患者の予測は、予測をするための情報が不足しているわけであり、極めて困難、あまりにも少ない時には不可能となる。そこで、この結果を踏まえ、答えている問診項目がある一定以上の患者に絞って解析を進めることとした。

慶応漢方専門外来に設置されている問診システムに登録されている問診項目が200程度であると言うことを鑑みて、その1割の20項目以上を答えている患者のデータのみを解析に用いることにした。もちろん20という数字に特別な意味はない。しかしながら、あまりこの数を大きくしすぎると解析できる患者数が極端に少なくなってしまい、解析した結果に意味が無くなってしまふ恐れがある。また、この数字を小さくしすぎると、データに解釈不可能なサンプルが存在することになり、予測能力の低下に繋がるであろう。それでは、20項目以上というスクリーニングを行った結果、どの程度のデータがフィルタアウトされて除かれ、どの程度が残ったのかを図9にまとめた。なお、このデータから2011年12月に集計された問診データにアップデートされている。このデータを用いて、20項目以上に回答しているデータのみを用いてRandom Forestにより虚実を予測した結果を図10である。テストデータの判別率は72.4%と他の判別方式よりも良い結果を示している。また、同様の方法により、データを20項目以上の回答などと絞り込みを行うことなく学習しテストした結果を図11にあげる。トレーニングデータに関しては、20項目以上と絞り込みを行った結果と同様に100%の分類を達成しているが、テストデータを用いて評価においては67.0%の正答率にとどまった。このことから、回答している問診項目数による解析対象の絞り込みは有効に働いていると言える。

これまでに、虚実の予測としてRandom Forestは他のデータマイニング手法(Elastic net、サポートベクトルマシン、CARTやそれらのバリエーション)よりも高い正答率を達成することを示し、また、誤りの原因であるようなサンプルの特徴を明らかにして、解析対象と出来る条件を検討し、実際にその絞り込みが漢方診断の予測結果を改善することを示すことが出来た。しかしながら、初診患者に絞った際の虚実の予測は正答率80%にはまだ到達していない。しかも、先ほどの図7で見たように、Random Forestが判断する虚実の予測に重要な問診項目には、

これまでの知見からは妥当な問診項目が多数あがっているにも関わらずである。これだけ妥当な問診項目がきちんと同定されていて、データの非線形性、高相関構造化においても有効な方法を多数の方法から様々な評価を行って選んだにも関わらずである。もちろん、漢方専門医は虚実の診断をするのに問診データだけを診ているわけではない。実際に患者を見、会話し、雰囲気を感じ、そのような総合的な判断で証が判定されているならば、問診データに全ての情報が含まれているわけではない。従って、問診データには含まれていないが、必ず門神事には医師の手元にある情報で可能性のある情報を追加情報として利用することを試みた。ここでは、患者の体格に関わる情報である BMI を用いることにする。慶応漢方専門外来の問診データにおいて BMI がすでに問診システムに入力されているデータの件数を図 1 2 にまとめる。カルテ上にはほとんど全ての患者に BMI は計算することが出来るよう身長、体重の記載があるが、問診システムにまだその情報が移行されていないため、現状では 6 2 0 人の患者に対してのデータが利用可能となっている。その中で 4 0 2 人の患者が 2 0 項目以上の回答があり、これらのデータを BMI と共に用いることにする

図 1 3 に各患者の BMI 値と虚実の証の関係を示す。明らかに実証になるほど BMI 値は大きくなる傾向にあるが、BMI だけが虚実の決め手となるとはいえない難しいと言うことも分かる。実際、実証患者の BMI 最小値と虚証患者の BMI 最大値は後者の方が 7 程度大きく、この範囲に相当数の患者が実証、虚証共にいることが分かる。しかしながら、問診データと共に用いると有効かもしれないと思わせるデータであることは間違いあるまい。

この BMI を問診データと共に用い、凶日の予測を行った結果を図 1 4 にあげる。BMI が問診システムに反映されている実証患者の数が 3 7 と少なかったため、トレーニングには 2 0 サンプルずつの利用となった。この辺りは、今後問診システムへ BMI 情報をうつしていき、かつ、今後のデータを BMI 付きで記録していくことで改善することが出来よう。BMI

を併用することで虚実の予測正答率は飛躍的に向上した。判別正答率は 9 1. 2 % を達成することに成功した。

図 1 5 は、これまで用いていなかった「やや虚証」、「中間証」、「やや実証」を構築した虚実予測方式により予測した結果を示している。この結果から、「やや虚証」はきちんと「虚証」として、また、「やや実証」もきちんと「実証」として予測されていることが分かる。「中間証」は「虚証」に予測される人数と「実証」に予測される人数がちょうど半々くらいとなり、中間であるという特徴が良く捉えられていると言えよう。

この BMI を入れた際に構築した Random Forest により判定した変数の重要度を図 1 6 にあげる。やはり BMI がもっとも重要であるという結論を得たが、2、3、4 位にあがっている「全身症状：その他：汗をかきやすい：VAS」、「全身症状：その他：寒がり：VAS」、「全身症状：精神状態：気分が憂うつになる：VAS」もこれらの重要度を合わせると BMI と拮抗するような状況である。また、図 1 3 で見たように BMI だけではこのような高い正答率は望むことは出来ない。問診データとの組み合わせによりこの 9 1. 2 % という高い正答率を達成できたと言うことに高い意義がある。

それでは、この重要度が高い上位 3 0 項目のみを用い虚実の予測を行ったら、やはり高い予測性能を示すことが出来るであろうか？その結果を図 1 7 に示す。結果的には、テストデータによる予測正答率は 8 4. 7 % と BMI を用いない Random Forest での予測よりは高い正答率を示すことが出来た。また、図 1 8 や先ほどと同様に「やや虚証」、「中間証」、「やや実証」を予測した結果である。8 3. 6 % と比較的高い予測性能を示すことが出来た。しかしながら、「虚証」、「実証」のテストデータにおける予測性能から考えると、先ほどの 9 1. 2 % からはかなりの性能低下が見られる。3 0 項目では不十分であり、もう少し下の重要度の項目まで拾っていく必要があると思われる。この問診項目の絞り込みは問診システムのクオリティを上げていくためにも必

要なステップであると考えている。後で示す寒熱、気血水の予測や処方との関連を考慮に入れて、より本質的な問診項目とは何かという問いに答えていくべきであろう。

C.2 寒熱

次に寒熱の証の予測を行う方式の構築を試みる。前述したとおり、寒熱では、「寒証」、「中間証」、「熱証」、「上下寒熱」、「手足煩熱」、「厥冷」、「寒疝」という順序であるものと無いものが混在している。その集計結果は、図19が、「上下寒熱」と「手足煩熱」は「寒熱錯雑証」としてまとめ、「厥冷」と「寒疝」に関しては「寒証」とした。従って、寒熱の証は

「寒証」

「中間証」

「熱証」

「寒熱錯雑証」

の4つとなる。先ほどの虚実は一択の選択であったが、この寒熱に関しては複数の選択があり、その内訳についても図19に載せている。もっとも数の多い証は「寒証」であり、次に「中間」、「厥冷」、「上下寒熱」と続く。「厥冷」は「寒証」として取り扱われるため、「寒証」はさらに多くなることになる。

寒熱は、一般的な概念として先ほどの虚実よりも分かりやすいのではないかというのが我々の直感であった。虚実においては、さまざまなデータマイニング手法、VAS値の取り扱い方、データのアンバランスを解消するサンプリングなどを駆使してやっと70%後半の予測正解率を達成した。この寒熱においては、問診項目自体に「寒がり」や「暑がり」に関連する項目があるため、その項目に対するチェック状況を見るだけで分かるのではないかという予測がある。実際、「寒がり」に関しては、全身、手、足、腰に関して「冷え」を問うているし、「暑がり」に関しては、同様に全身、手、足、腰に関して「ほてり」を問うている。前者を寒項目、後者を熱項目とよぶと、寒項目にチェックが入っているか否か（該当する症状を自覚しているか否か）、熱項目にチェックが入っているか否かという情報が「寒証」、「熱証」と

どのような関係にあるか問うことをまず調べた。

ルールはシンプルに以上の考えが反映されるように構築した。寒証と熱証のみ結果を示すが、もし、ある患者が上記の寒項目の中の少なくともどれか一つに該当するという問診システムへの入力があり、熱項目のどれにも該当しないということであれば、その患者は寒証とした。また、熱証はその逆で、寒項目に関しては一つも該当が無く、熱項目において少なくとも一つ以上の該当があるというものである。そのた、中間証、上熱下寒なども同様の考えによりコーディングした結果を図20にまとめている。このルールに従い、患者を分類し、寒証に分類された患者が実際に漢方専門医から診断された寒熱の証と一致率（ルールによる振り分けが正しいか否か）を調べたところ60.8%しか正解しなかった。また、熱証に関しても、熱証と図20のルールで判定された患者のうち、実際に熱証と診断されていた患者は50.0%しかなかった。

この結果から分かることは、単に寒項目、熱項目にチェックがあるだけでは寒熱の証は決まらなく、その度合いや他の問診項目によって説明される患者の体の状態も同時に考慮しなければならないということである。すなわち、寒熱に関しても、問診データからその証を予測するという問題は決して易しい問題ではなく、多数の問診データの多変量的解析が必要になるということがわかる。もし、慶応漢方専門外来問診システムにおいてVASを記録していなければ、寒熱の予測は困難であるか、場合によっては高い正答率は望めないかもしれないということがこの結果から推測される。

まず、Random Forestが虚証と同様に寒熱についても高い予測性能を発揮することが出来るかどうかをチェックした。まず、症例を「寒証」と「熱証」に絞り、Elastic netをNormalized VASのデータに適用し、その予測性能を評価した。その結果は図21にまとめた。予測正答率は59.3%であり、図20で示した寒熱の難しさがそのまま反映された結果となった。問診項目も3個しかモデルに取り込むことが出来ていないし、トレーニングの段階ですでに6

8%前後しか患者に正しく寒証であるか熱証であるか判定できていない。注意いただきたいのは、ここでは、虚実で行った問診項目20個以上に回答というサンプルのセレクションは行っていない。ベースラインとなる予測正答率を得ることが目的である。

次に、Random Forest を用い、Elastic net と同様に Normalized VAS のデータに対してモデルの学習を行い、予測性能を評価した。このときに正答率は75.0%であった。Elastic net よりも、虚証の場合と同様に Random Forest は高い予測性能を発揮できるということが分かった。また、この解析において、大きく予測を間違った（漢方専門医の診断は寒証なのにもかかわらず、寒証であるという確率が20%以下と見積もってしまった）2名の患者の問診データを精査したとこと、虚実と同様にやはり問診データにほとんど回答がない症例であった。従って、寒熱に関しても、虚実と同様にあまりにも問診システムに情報のないサンプルは解析から除くこととした。

図23が20項目以上回答している患者のデータに絞った場合の予測性能である。ここで、データは2011年12月に集計された最新版に変更した。20項目以上に絞ったことにより、虚実と同様に予測性能は向上し、80.4%の正答率を達成することができた。

図24に寒熱の予測に重要と判断された問診項目をあげた。1位は「全身症状:その他:寒がり:VAS」、2位は「痛み・冷え等:冷え:足:VAS」、3位は「痛み・冷え等:ほてり:顔:VAS」となっている。寒熱の最初に見ていた寒項目、熱項目に相当する問診項目がやほりのきなみ高い重要度であると判断されている。Random Forest では、これらの項目の関連性を考慮し、さらに、他の問診項目の状態や関連性を考慮することにより高い予測性能を達成していると思われる。

さらに予測正答率の向上を目指し、さまざまなデータの絞り込みを行った。ポイントは、解析できる患者の数をあまり減らすことなく、予測正答率を大きく向上させることの出来る絞り込みを行うことである。そのなかで、患者の年齢を18~65歳に絞

ったところ予測正答率が84.7%を達成し、この絞り込みは有用であるという結論を得た。その結果を図25にあげている。

この患者データの絞り込みと Random Forest による寒熱の予測方式により、今まで解析に用いていなかった「上熱下寒」、「蕨冷」、「中間証」の患者の証を予測した。結果を図26に示す。上熱下寒に関しては、寒証、熱証どちらが正解というものではないが、若干寒証が多く予測されるという結果を得た。また、「蕨冷」は「寒証」であるが、これはきちんと寒証が多く、正答率85%を達成している。また、中間証においては、「寒証」、「熱証」のどちらの特徴も有さないというものであるが、熱証の方に多く予測されているという結果になった。

ここまでで、虚実、寒熱に関して問診データから漢方診断「証」を予測する方式があらかたかたまってきた。ちょうどそのとき、平成24年3月放送のNHK 名医にQという番組において、この問診システムを紹介したいという打診を渡辺賢治准教授が受けたと報告があり、その番組の中で実際にこのシステムを用いて患者の証が渡辺准教授の診断と一致するのか否かをテストしたいという申し入れがあった。NHK 側が手配した方が問診システムに入力したデータを図27に示した。この方は問診システムにおいて23個の項目に答えている。今まで構築した予測システムで解析可能な20項目以上の回答という条件にマッチし、閾値ぎりぎりながら解析対象とすることが出来た。問診データだけを頂いて、完全に我々にはブラインドで解析結果だけを返すというテストを行った。

まず、虚実の判定を行った。虚実を予測する Random Forest に上記の問診データを入力し、アウトプットを得る。アウトプットは、この方が「虚証」である確率である。対比のさせ方に従い、確率0はこの方が「実証」であることを示す。結果は、この方は61.8%の確率で虚証であるということであった。結果を図28に示す。これは、強い確信をもって虚証であるとは到底言えない結果である。したがって、判定は中間証からやや虚証の傾向があると

回答した。渡辺准教授の診断結果を図28下に載せたが、診断は中間と言うことであった。

次に、寒熱の判定を行った。予備的に、「寒証」であるか「熱証」であるかを2カテゴリーの Random Forest により判定したところ、ほとんど中間に確率は得られた。そこで、「寒証」、「熱証」に「錯雑証（上熱下寒に相当）」を加えて3カテゴリーの Random Forest によって判断することにした。その結果を図29に示した。結果、「錯雑証」が56.0%となり、「寒証」、「熱証」それぞれの確率22.6%、21.4%を合わせた確率よりも高くなり、我々はこの方は錯雑証であると結論づけた。渡辺准教授の診断結果を図29下に載せた。この寒熱に関しても虚実と同様に診断をデータマイニングによる方法によって正しく予測することができると実証できた。

C.3 気血水

最後に、問診データから気血水の予測を行うデータマイニング法を構築する。虚実において多数のデータマイニング法の中から漢方診断を予測するのにもっとも適したデータマイニング手法として選んだ Random Forest は、寒熱においても高い予測性能を発揮することが出来た。この結果を踏まえ、気血水においても Random Forest に基づき、虚実、寒熱における患者の絞り込み（回答問診項目20以上）を用いて予測方式を構築する。

図30が気血水の状況を表す。気血水は8つの証からなり「気虚」、「気うつ」、「気逆」、「血虚」、「お血」、「水毒」、「亡津液」、「下焦の虚」である。これらの一つだけが診断されるのではなく、複数を診断することが多い。図30のテーブルは、気血水の証を1つ、もしくは2つ診断された患者に対するクロス票になっている。縦に「気虚」の列を見たとき、「気虚-気虚」のセルが293というのは、気虚一つだけ診断された患者が293名いたと言うことである。「気虚-気うつ」のセルの69は「気虚」と「気うつ」の二つを診断されたのが69名であったと言うことである。また、気血水を3つ

以上診断された患者数はテーブル下の欄にまとめた。今後は、気血水は最大2つまで診断するというルールになったと聞いているため、この欄外のケースはイレギュラーなものとなると考えられる。

通常、統計的判別分析においては、複数のグループにおいてサンプルの属する真のグループは一つである。しかしながら、この気血水においては、真のグループが2つを許容する状況にある。このため、工夫した解析方法が必要となる。真のグループが複数であるときの確立されたデータ解析手法はまだ一般に無い。

まず、この2つまで診断されてテーブルにまとめられる患者に絞り（今後はそうなるため）、男女別に何らかの傾向が見られるかを確認した。そのテーブルを図31に示す。元々の男女比は、テーブルの左に示したとおり、男性が25%、女性が75%と女性の割合が高い。このバランスが崩れている気血水の組み合わせを見ていくと、例えば、「気虚-気うつ」は男性が48%、女性が52%となって男性に多いことが分かる。また、お血を含む診断では女性が多いことが分かる。特に、「お血-血虚」は94%が女性である。また、血虚についても女性が多く、「血虚-水毒」の組み合わせは実に98%が女性となる。このことから、男性と女性で気血水の診断のされ方が異なることが示唆される。

そこで、男性と女性に分けて、図30と同様のクロス表を作成した。そのテーブルを男性に関しては図32、女性に関しては図33に示した。

次に、気血水のクロス表の情報と虚実との間に何らかの相関がないか調べた。例えば、「気虚」のみを診断された患者は「虚証」、「中間」が多く、「実証」はほとんどいない。なお、ここで「虚証」とは「虚証」と「やや虚証」を合わせたものである。また、「実証」とは「実証」と「やや実証」を合わせたものである。もともとの虚実の比率は、「虚証」30%、「中間」53%、「実証」17%であるので、例えば、実証が多い組み合わせとしては、「気うつ-水毒」の46%、「気うつ-気逆」の41%などがある。「気うつ」の診断を受けた患者は、「気

虚」、「血虚」を除き「実証」が多い。このような虚実と気血水の証の関連は、各官報診断において証の予測を行った後にその確からしさを判定するのに用いることが出来ると考えている。

「下焦の虚」はサンプル数が極端に少ないため、今後の解析から外すと言うことにした。残りの7つの証を対象に7カテゴリーの Random Forest を構築して気血水の最初の当たりを付けるための解析とした。前述したように、気血水予測のデータマイニングとしての問題の難しさは、真のグループが複数あると言うことである。その困難をまずは回避するために、気血水の一つだけ診断された患者に絞って解析を上記の解析を行うことにした。その結果を図35、36にまとめた。図35には、使用した各証のサンプル数、およびトレーニングエラーを、図36には、トレーニングには用いなかったデータを用いたテストの結果を載せている。

これまでに Random Forest を用いた場合と同様にトレーニングは100%の分類を達成している。テストデータでの評価であるが、正答率は24.5%となった。これは、今までの結果からは非常に低い正答率であるかと思われるかもしれないが、今までは2つのうちどちらかという比較を行って、その正答率を計算していたのに対して、今回は、7つのカテゴリーの中でどれかという判断をしている点が異なる。7つのカテゴリーからランダムに選択するとすると正答率は1/7で14.3%となる。したがって、ランダムな証の選択よりは遙かに高い予測性能であることはわかる。しかしながら、もちろん満足のいくような正答率ではないことは明らかである。虚実、寒熱と同様により高い正答率を目指す必要がある。

まず、正答率が低い原因の一つはトレーニングに用いたサンプル数が7カテゴリーにもかかわらず各カテゴリー30と少ないことがあげられる。これは「下焦の虚」が全部で38名の患者しかいないことに起因する。データ数のアンバランスによるバイアスを回避するためにはこの「下焦の虚」に合わせてトレーニングサンプルを抽出する必要があるからで

ある。そこで、元々のサンプル数が少ないことを考えて、この段階としては「下焦の虚」を除いて他の証に対しての予測能力を向上させる方がより現実的であろうと判断し、「下焦の虚」を除くことにした。したがって、残りの6カテゴリーでの予測を行う。

「下焦の虚」を除いた恩恵を受けて、6カテゴリーそれぞれで54のトレーニングサンプルを準備することが出来た。気逆が84であるため、そこで制限がかかるが、7カテゴリーの場合に比べると倍近くのサンプルをトレーニングに用いることが出来るのは大きな進歩を生む可能性がある。しかしながら、結果はそれほどの改善はされず、図38にあるように30.0%までの向上となった。6つのカテゴリーそれぞれに計算される確率で1位と2位のどちらかには正解が含まれるという数をカウントしたところ、テーブル最下段の示す正答率となった。こちらでは50%以上と、ランダムでは1/3で33%であることを考えると、かなりよく予測できるようになったと言えよう。

この6カテゴリー一括の判定は、渡辺准教授とのディスカッションにより現実の診断と少々乖離があることが明らかになった。その乖離とは、この予測方式が6つのカテゴリーそれぞれにおいて各患者がその証を診断される確率を計算している。しかしながら、実際の診療においては、証ごとにこの患者には当てはまるか否かという判断をされることが多いということであった。そこで、他のグループを定義する方法を試してみることにした。

次に試したのは、気血水において証の一つだけ診断されている患者を対象に「気虚」と「気虚以外」に患者を2グループに分け、各患者がどちらに属するかを予測するというものである。結果を図39にまとめた。2グループの予測になるので正答率がかなり向上することが期待されたが、思いのほか性能は向上せず56.9%の正答率に留まった。他の証（「気うつ」と「気うつ以外」や「気逆」と「気逆以外」など）についてもすべての組み合わせにおいて試したがそれほど改善はしなかった。

次に、女性だけに絞り、6カテゴリーの Random

Forest を試みた。各カテゴリーからのトレーニングサンプルは30であり、結果のみ図40にまとめた。男女を分けずに解析をしたときとさほど大きな改善はなかったが、Random Forest による問診項目の重要度に注目すると、女性だけに絞るメリットも見えてきた。図41に Random Forest が選ぶ重要問診項目をまとめたが、女性に絞ったことで使用することが出来るようになった女性のみを対象としている問診項目が多く重要変数に上がってくるようになった。男性を混ぜてしまうとこれらの問診項目は使用することが出来ないということを合わせて考えると、女性と男性を分けて解析することでこれらの項目が使用できるようになり、これの項目を上手く使用することで予測正答率の向上が達成できれば非常に有用であろう。

それでは、より問題をクリアにするために、よりシンプルな問題を通して気血水のデータの特徴を検討した。まずは、気血水の一つだけ診断されている患者に絞り、「気虚」と指弾された患者と「気うつ」と診断された患者で2グループの予測を行った。結果を図42に示す。トレーニングは100%の分類を達成したが、テストデータにおいては62.6%の正答率に留まった。また、「気逆」と「血虚」の比較を図43にあげる。この比較においては、予測正答率76.6%と比較的高い水準のものとなった。このような各証のペアでの比較を全て行い、結果をまとめた。図44に示す。証ペアでの比較からは、血虚との比較では気虚以外においていずれも70%以上の正答率を達成している。しかしながら、気虚はどの証との比較においても70%以上の正答率には届かず低い水準となっている。平均正答率で見ても気虚は低い。一方、血虚は高い平均正答率であることが分かる。テーブルの欄外には、エキスパートの知識として、各証がどのような問診に関連すると思われるかという情報をまとめた。

比較においてもっとも正答率の高かった「気逆」vs「血虚」と正答率の低かった「気虚」vs「気逆」において重要と判断された問診項目をそれぞれ図45、46にまとめた。この2つの図から分かること

は、成績の良かった「気逆」vs「血虚」は、医師の選んだ知識ベースの項目がきちんと上位に入っているが、一方、成績の悪かった「気虚」vs「気逆」では、重要項目には出てくるがトップクラスではないということがあげられる。これは、医師側の判断の難しさもあるということを示唆しているのかもしれない。

各証について、エキスパートナレッジと言うべき図44にまとめられた特徴に対応する問診項目をまとめたものが図47である。これらの問診項目と気血水6つの証がどのように関連しているかを可視化した。その結果が図48である。本来であれば、各証に特徴的な問診項目はその証の患者に多くマッチ（紫色）し、その他の証を診断された患者ではあまりマッチしない（水色）というのがデータ解析としては易しい状況であろう。しかしながら、この図48が示しているのはそのような楽観視が出来る状況ではなく、この問題の難しさである。ほとんどモザイク模様であるという良い状況からの予測を行わなければならない。この状況では、ある問診項目の値を見てと言ったような単変数の解析は全く役に立たず、やはりこれまでにやってきたような多変量解析が必須となることが分かる。各証ペアにおけるこれら気血水対応問診項目の状況がよりはっきりと理解可能なように、証ペアそれぞれでの解析を行った。結果を図49～56にまとめた。ヒートマップの上に赤と青で分けているカラーバーが各患者がどちらの証を診断されたかを表している。ほとんどの結果において、この赤と青の情報がこの知識ベースの問診項目の情報ではセパレートできていないことが分かる。また、図49が顕著であるが、「疲れやすい」という問診項目は、気虚に特徴的項目であるとなっているが、ほとんどの患者が疲れやすいにチェックを入れるような状況である。実際、平成22年度の成果において、この「疲れやすい」は高頻度な問診項目としてあがっている。また、同じ図49を用いて説明するが、他の結果においても同様のことが見られるが、これら特徴的問診項目に何もチェックが入っていないにもかかわらず気血水の証が診断されている患者が決して少なくない人数いるとい

う事実である。これらの患者は、この特徴的問診項目以外の情報により証を診断されているものと考えられる。虚実の時に用いたBMIに相当する情報が必要なのかもしれない。

次に、少々グループの定義を変えた解析を試みる。これまで、気血水の証を一つだけ診断されていた、テーブルの対角成分の患者のみを対象としていたが、今回は2つの証を診断された患者も用いることが可能な枠組みを構築した。その原理を図57にまとめた。ここでは気虚を例に説明している。このテーブルにある患者を「気虚」という診断を含む患者のグループと「気虚」という診断は含まないグループの2つのグループに分割する。すなわち、テーブルでは、気虚の列にいる患者は前者のグループ、テーブルの下側三角(情報は対象なので下側三角で全員となる)で気虚の列以外の患者が後者のグループとするとこのグルーピングが達成できる。このグループ分けにより、各患者について、少なくとも「気虚」の診断をすべきか否かが判定でき、これを他の証に関しても同様にグルーピングを行うことにより、各証を少なくとも診断に含むか否かを判断することが出来ることになる。

各比較でのサンプル数をまとめたものが図58になる。この比較は、医師から尋ね聞いた実際の診断の手順とかなり類似した思想である。しかしながら、図59～64に各証の予測結果を示すが、現状は決して高い予測性能は示すことが出来ていない。しかしながら、各患者に戻り、その患者の中での6つの証の確率のコントラストを考え、どの証がもっとも確率が高いかという判断をしたとき、1位、もしくは2位までに正解が含まれている割合は63.1%となった。これはランダムに選択したときには40%弱としかならないものである。従って、各患者での予測と言うことを考えるとかなり予測能力は高くなっていることを示している。図65に各患者における確率値と正解を示した。赤のバーグラフが正解の証に対応する。

図66、67において、結果の解析を行った。その結果、「お血」に関しては、かなり高い予測能力を有していることが明らかとなった。気血水の証を一つだけ有しているテスト8例中6例はきちんと1位で「お血」であると判断され、1例は2位での判定であった。しかしながら、気血水

を一つだけ診断されたテストの中では「気逆」は8例全て1位にも2位にも入っていなかった。この傾向は、気血水の証を2つ診断されていた症例についても同様であり、「お血」は極めて高い予測能力を示す一方、気逆はほとんどあたらないような状況である。これは、気逆のそもそものサンプル数が少ないことも原因の一つであると考えている。来年度は、この気血水の予測能力を向上させ、虚実、寒熱、気血水の3つともに実用に耐えることの出来る方式の確立を目指す。

D. 考察

患者の自覚症状が記録された問診データから漢方診断である「証」の予測を行うデータマイニングに基づく方式の構築を行った。さまざまなデータマイニング手法を網羅的にテストし、その結果、以下の点を見いだした。

- 患者が各問診項目に「当てはまる」、「当てはまらない」という2値の情報では証の予測は困難であり、VASの情報を上手く使用しないと高い予測性能は達成できない
- ロジスティック判別モデルのような線形性は、問診データからの漢方証の予測ではほとんどの場合成り立たない。情報の非対称性・非線形性を考慮できる手法が必要となる。
- 使用する患者のフィルタが極めて重要となる。もともと患者はヘテロな集団である。出来る限り特殊な例はトレーニングサンプルから外し、より汎化能力が上がるようなデータセットを作る必要がある。
- 問診項目間には高い相関構造がある。その密な相関構造は、Elastic net のような線形モデルの有効性を著しく低下させる。この高相関な変数群においても有効な方法を用いる必要がある。
- データ数のアンバランスは避けては通れない。単純な手法の適用ではアンバランスにより生じるバイアスにより全く意味が無いにも関わらず高い予測性能を示し、意味の無い漢方証の予測になってしまう。

E. 結論

漢方専門外来において集められた問診データは、患者個々人の自覚症状が記された貴重な診療支援材料である。実際、この問診データを用いてデータマイニングの手法を用いることで漢方診断「証」を高い確率で予測できることを示した。特に、虚実、寒熱に関しては、前者は90%以上、後者は80%半ばの予測正答率を達成した。しかしながら、気血水に関しては、そのデータの詳しい特性を明らかにし、今後そのデータ背景を活かした判別方式の構築を行いより予測能力の向上を目指す必要がある。今後よりデータを増やし、漢方診断と問診データの関連をよりクリアにすることで、漢方診断のエビデンスを創り出し、診療支援システムをより高精度に磨いていくことが必要不可欠である。

F. 健康危険情報

特になし

G. 研究発表

1. 論文発表

1. K. Katayama, R. Yamaguchi, S. Imoto, K. Matsuura, K. Watanabe, S. Miyano (2012) Transform of visual analogue scale data and their clustering, *International Journal of Knowledge Engineering and Soft Data Paradigms*, 3(2), 143-151.
2. K. Katayama, R. Yamaguchi, S. Imoto, K. Matsuura, K. Watanabe, S. Miyano (2011) Clustering for visual analogue scale data in symbolic data analysis, *Procedia Computer Science*, 6, 370-374.

H. 知的所有権の取得状況

該当無し

