

図43: 気逆 vs 血虚

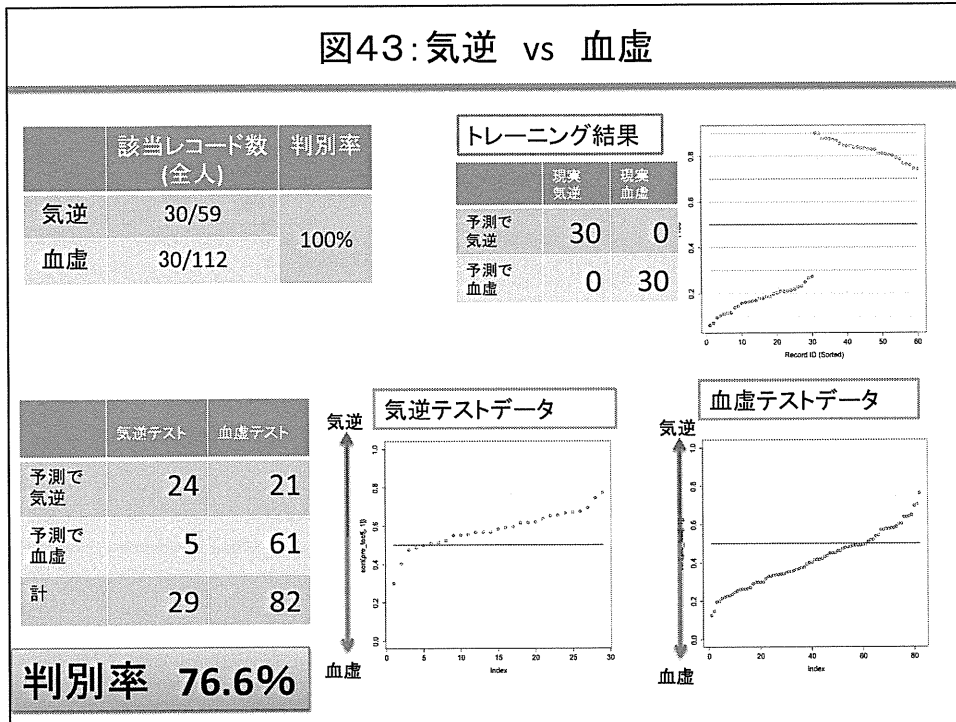


図44: 証ペアでの比較まとめ

	気虚	気うつ	気逆	血虚	お血	水毒
気虚		62.6	54.2	66.4	64.2	57.6
気うつ	62.6		53.7	75.7	66.2	68.6
気逆	54.2	53.7		76.6	63.6	62.3
血虚	66.4	75.7	76.6		72.1	73.3
お血	64.2	66.2	63.6	72.1		58.4
水毒	57.6	68.6	62.3	73.3	58.4	
合計	305	326.8	310.4	364.1	324.5	320.2
平均	61.0	65.36	62.08	72.82	64.9	64.04

気虚 : 疲れやすい, 寝汗, 食後眠くなる, 食欲がない
 気うつ: 気分が憂うつになる, のどがつかえる, 不眠
 気逆 : 動悸, いらいらする, 頭痛, ぼてり
 血虚 : 髪が抜けやすい, 爪がもろい, 皮膚がかさかさする, ふけが出やすい, 視力低下,
 目が疲れる, 目がかすむ, 目がしょぼしょぼする, 足がつる, 物忘れをする, しみ
 お血 : 月経困難症, 月経痛, にきび, 子宮筋腫, 静脈瘤
 水毒 : むくみ, めまい, 立ちくらみ, 乗り物酔い, 手足がこわばる, 頭重

図45:重要問診(気逆 vs 血虚 (判別率76.6%))

気逆	動悸, いらいらする, 頭痛, ほてり
血虚	髪が抜けやすい, 爪がもろい, 皮膚がかさかさする, ふけが出やすい, 視力低下 目が疲れる, 目がかすむ, 目がしょぼしょぼする, 足がつる, 物忘れをする, しみ

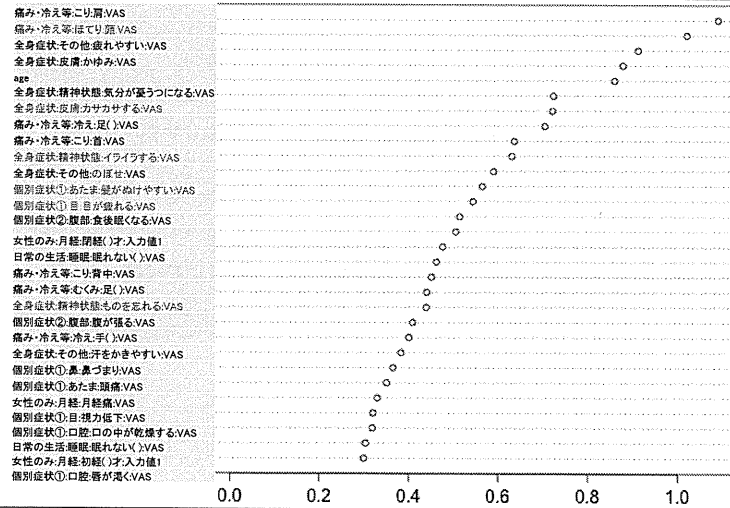


図46:重要問診(気虚 vs 気逆 (判別率54.2%))

気虚	疲れやすい, 寝汗, 食後眠くなる, 食欲がない
気逆	動悸, いらいらする, 頭痛, ほてり

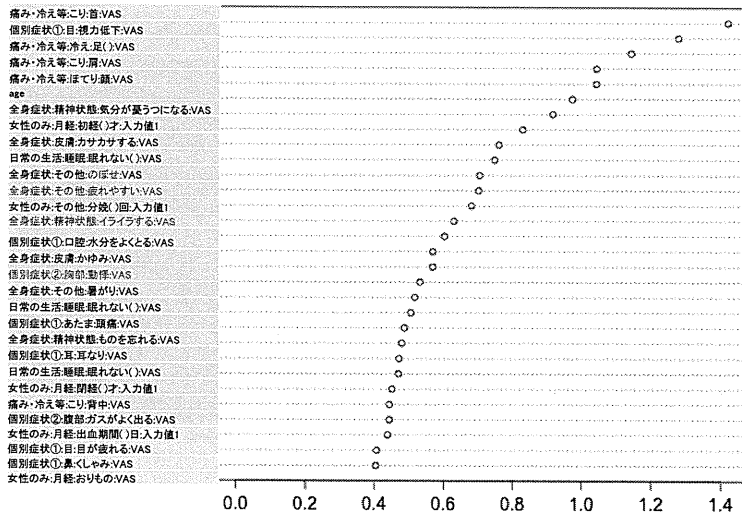


図47:エキスパートナレッジによる気血水対応問診項目

気虚	気うつ	気逆	血虚	お血	水毒
1 食欲がない	8 眠れない	54 イライラする	52 ものを忘れる	60 にきび	174 むくみ(顔)
72 疲れやすい	50 気分が憂鬱になる	186 頭痛	56 皮膚がかさかさ	348 月経痛	176 むくみ(手)
76 寝汗	255 のどがつかえる	248 動悸	62 シミ		181 むくみ(足)
270 食後眠くなる		162 ほてり(顔)	70 爪がもろい		188 頭重
		164 ほてり(手)	194 フケが出やすい		190 めまい
		169 ほてり(足)	196 髪が抜けやすい		192 立ちくらみ
			198 視力以下		262 乗り物酔い
			200 目が疲れる		278 手がこわばる
			202 目がかすむ		
			204 目がしょぼしょぼ		
			284 足がつる		

図48:気血水対応問診項目によるクラスタリング

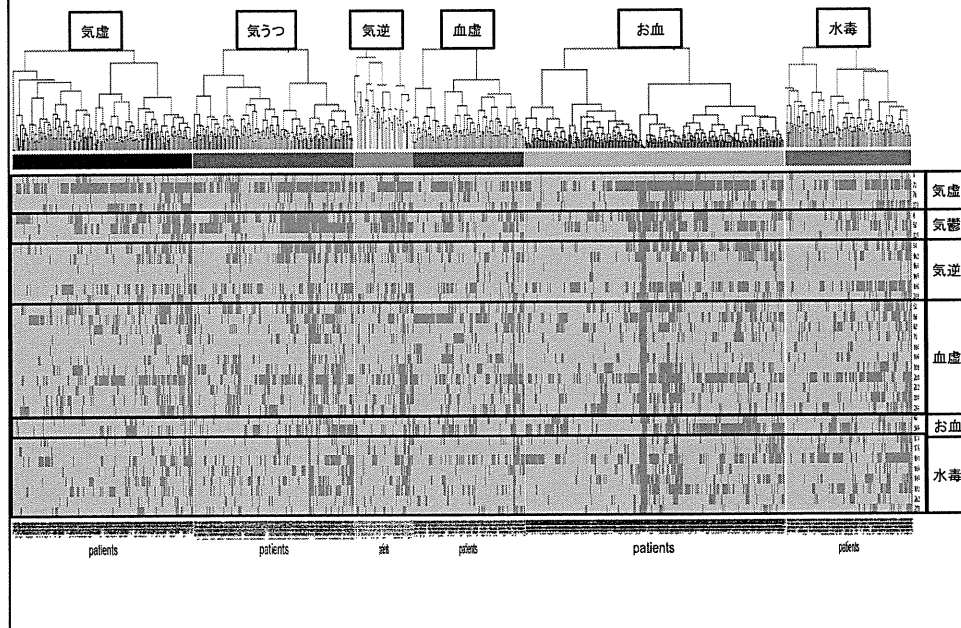


図49: 気虚 vs 気うつ

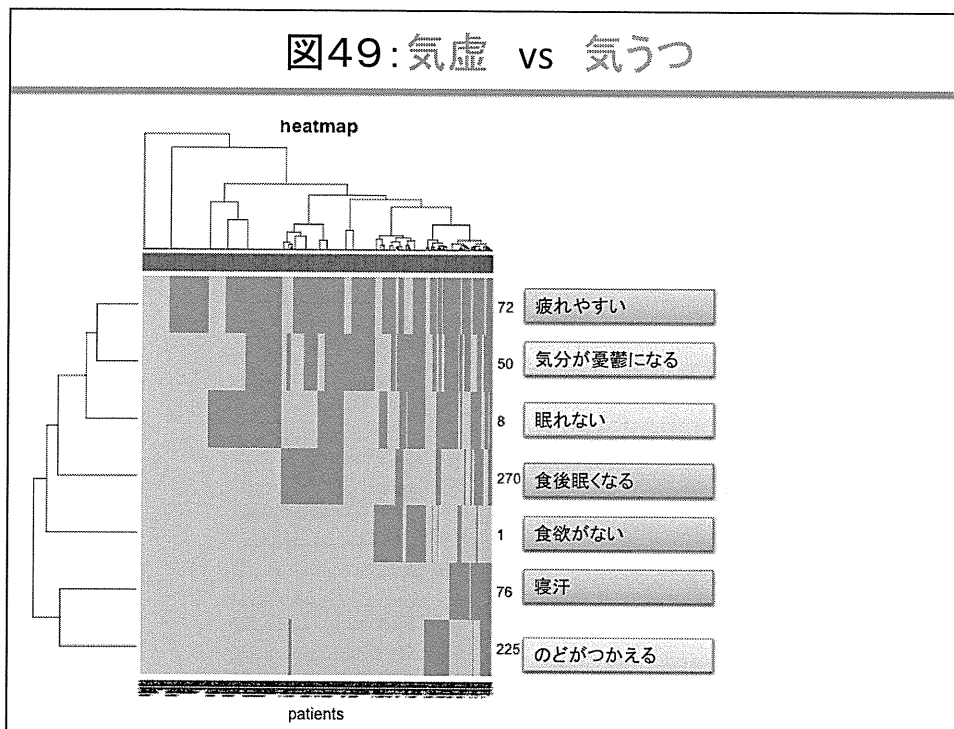


図50: 気虚 vs 気逆

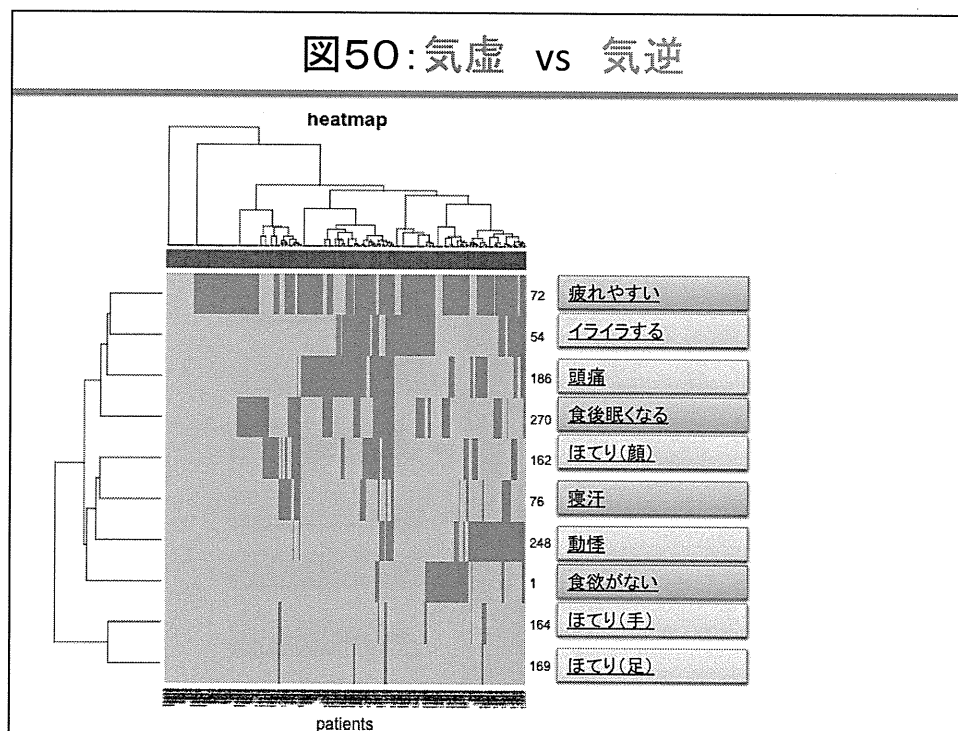


図51: 気虚 vs 血虚

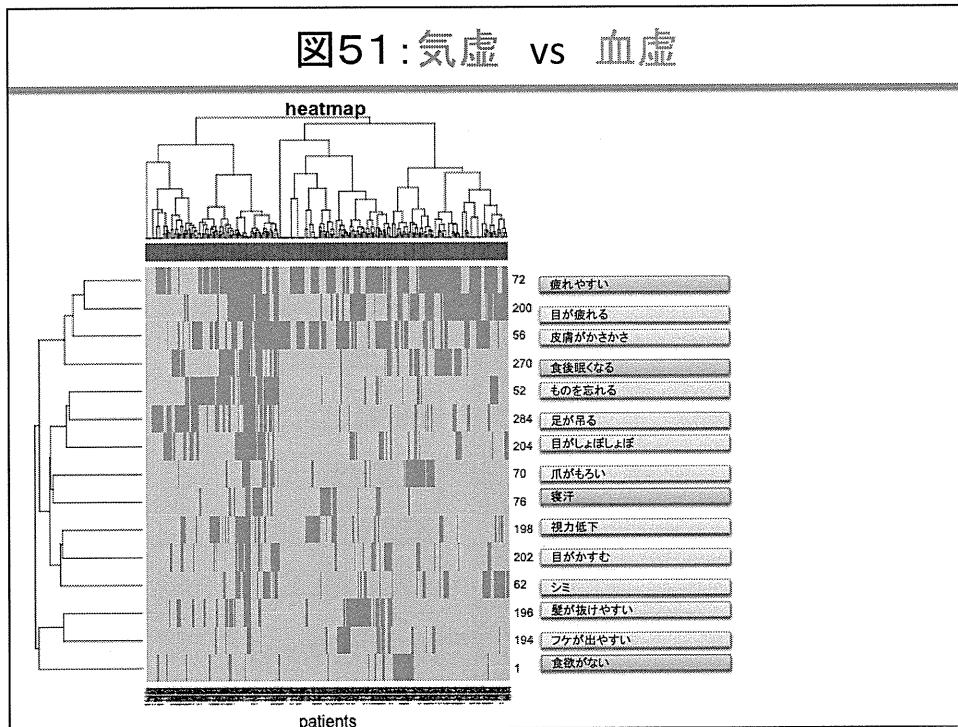


図52: 気虚 vs お血

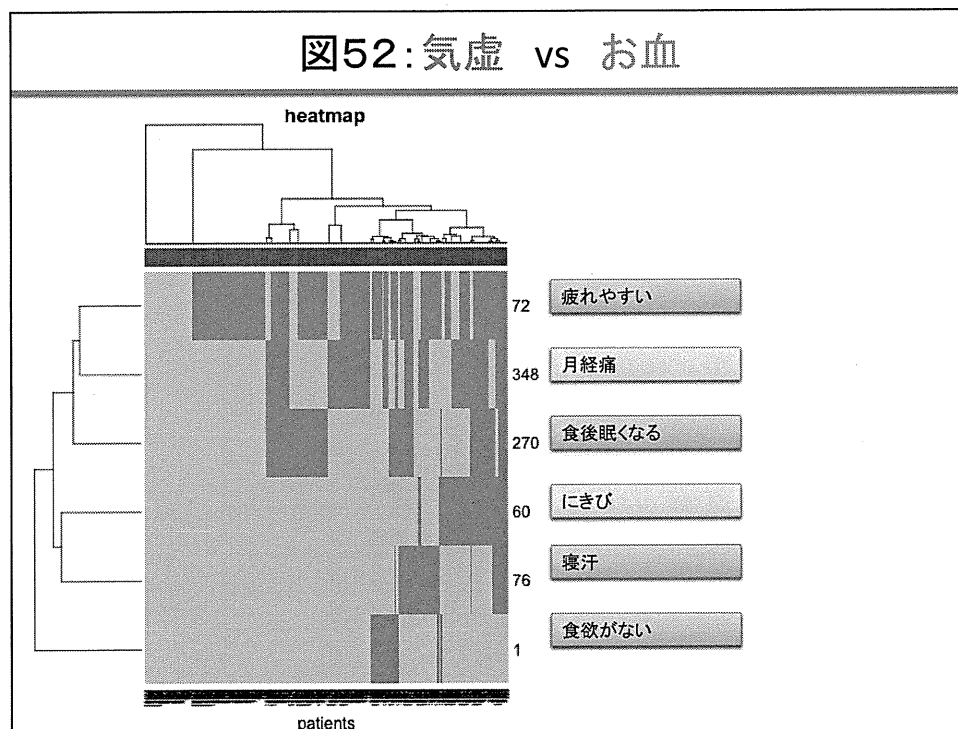


図53: 気虚 vs 水毒

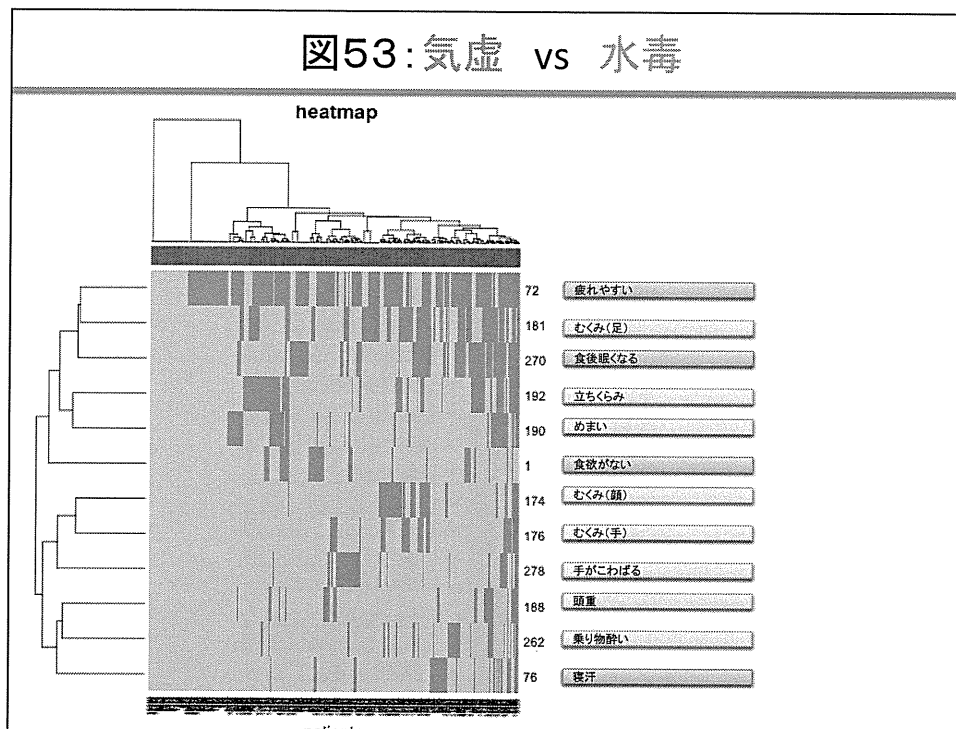


図54: 気うつ vs 気逆

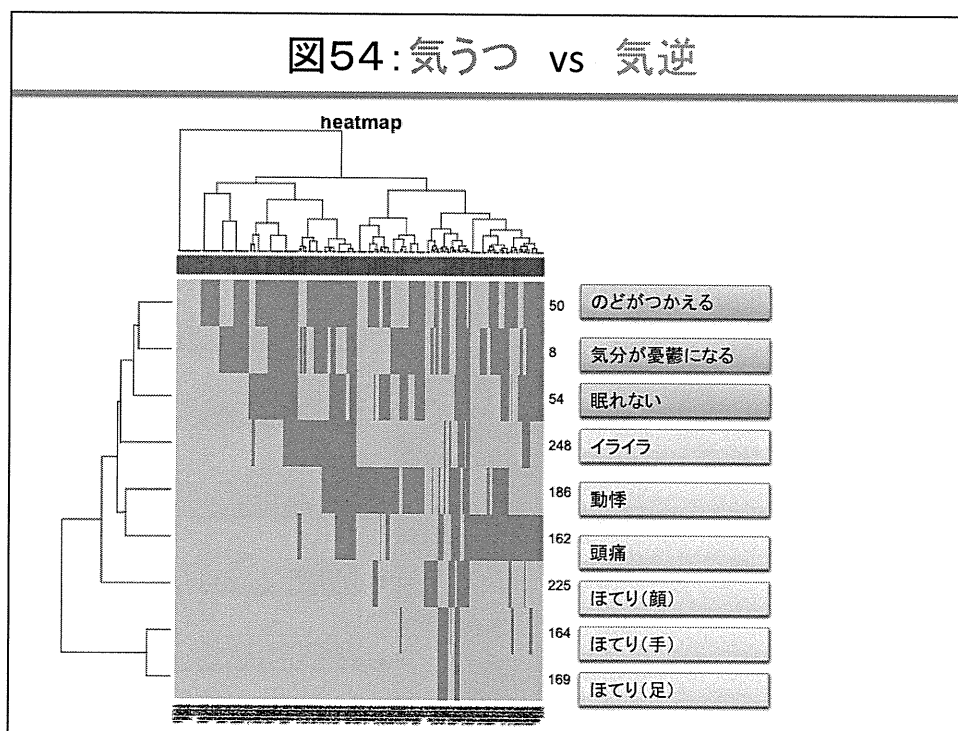


図55: 気うつ vs 血虚

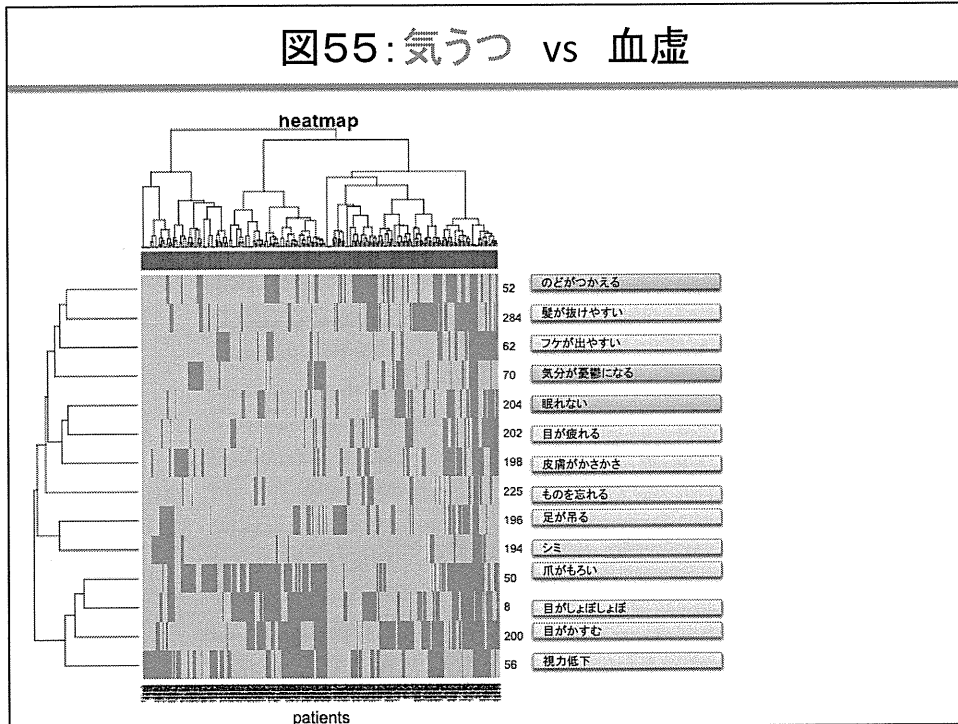


図56: 気うつ vs お血

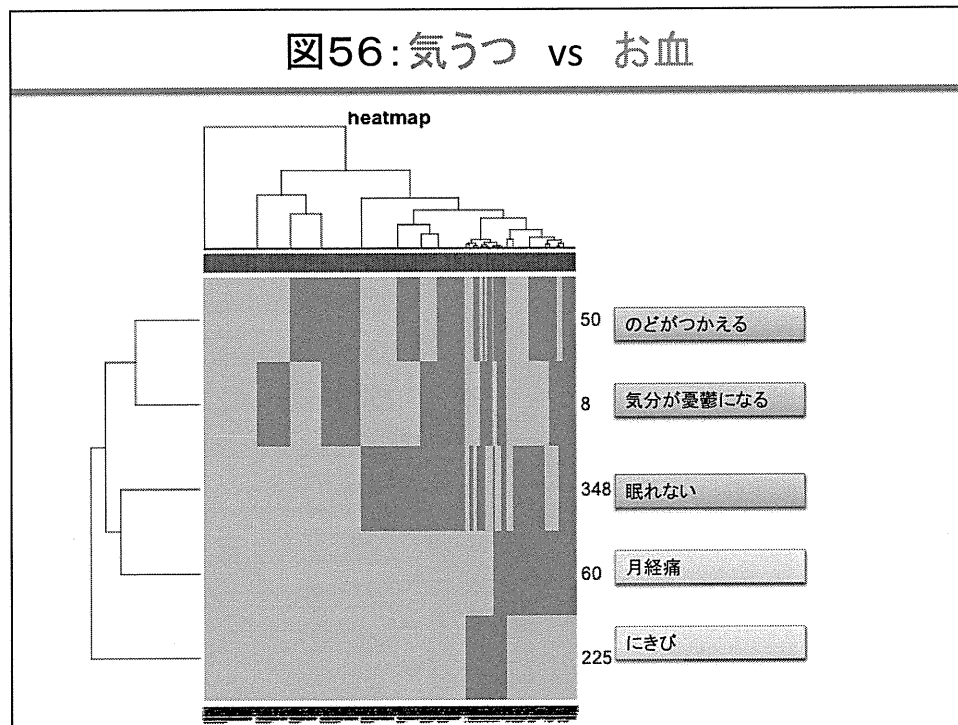


図57: 初診女性気血水のグルーピング

1576人中	気虚	気うつ	気逆	血虚	お血	水毒
気虚	176 (11.4%)	36 (2.3%)	19 (1.2%)	91 (5.8%)	59 (3.7%)	49 (3.1%)
気うつ	36 (2.3%)	155 (9.8%)	17 (1.1%)	21 (1.3%)	67 (4.2%)	61 (3.9%)
気逆	19 (1.2%)	17 (1.1%)	59 (3.7%)	10 (0.6%)	27 (1.7%)	16 (1.0%)
血虚	91 (5.8%)	21 (1.3%)	10 (0.6%)	109 (6.9%)	31 (2.0%)	39 (2.5%)
お血	59 (3.7%)	67 (4.2%)	27 (1.7%)	31 (2.0%)	268 (17.0%)	120 (7.6%)
水毒	49 (3.1%)	61 (3.9%)	16 (1.0%)	39 (2.5%)	120 (7.6%)	123 (7.8%)

少なくともある一つの気血水に該当しているかどうか

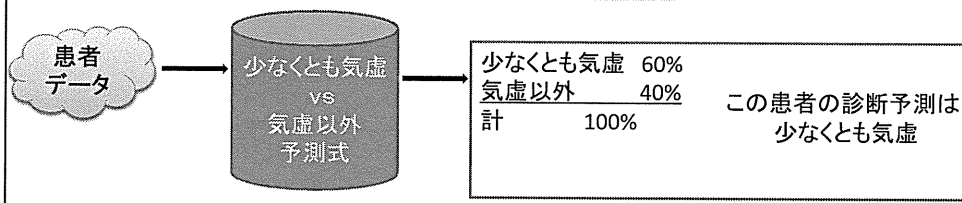


図58: 初診女性気血水の現状(トレーニング)

1393人中	気虚	気うつ	気逆	血虚	お血	水毒
気虚	168	28	11	83	51	41
気うつ	28	147	9	13	59	53
気逆	11	9	51	2	19	8
血虚	83	13	2	101	23	31
お血	51	59	19	23	260	112
水毒	41	53	8	31	112	115
合計	382	309	100	253	524	360

	気虚	気うつ	気逆	血虚	お血	水毒
〇〇	383	311	100	257	526	362
〇〇以外	1010	1082	1293	1136	867	1031
合計	1393	1393	1393	1393	1393	1393
〇〇	27.5%	22.3%	7.2%	18.4%	37.8%	26.0%
〇〇以外	72.5%	77.7%	92.8%	81.6%	62.2%	74.0%

図59: 気虚 vs それ以外

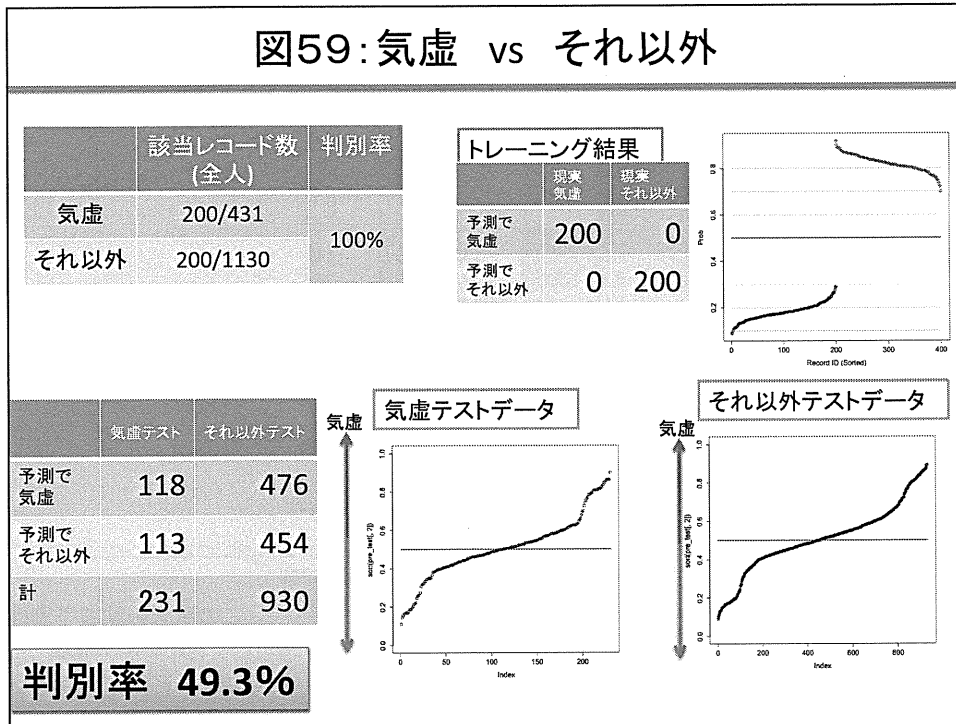


図60: 気うつ vs それ以外

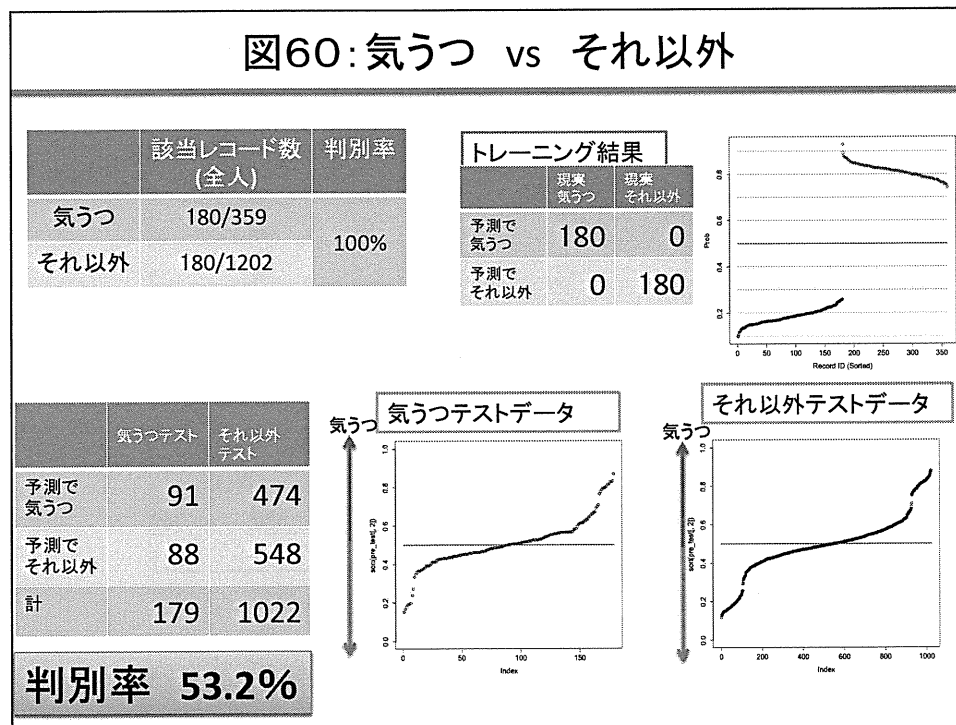


図61: 気逆 vs それ以外

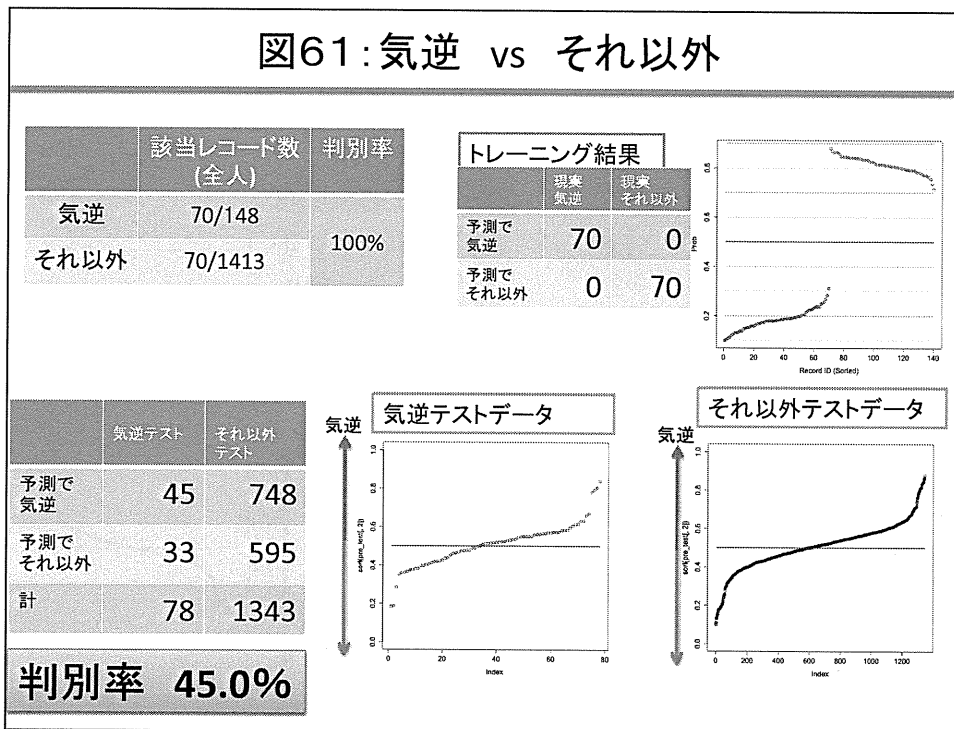


図62: 血虚 vs それ以外

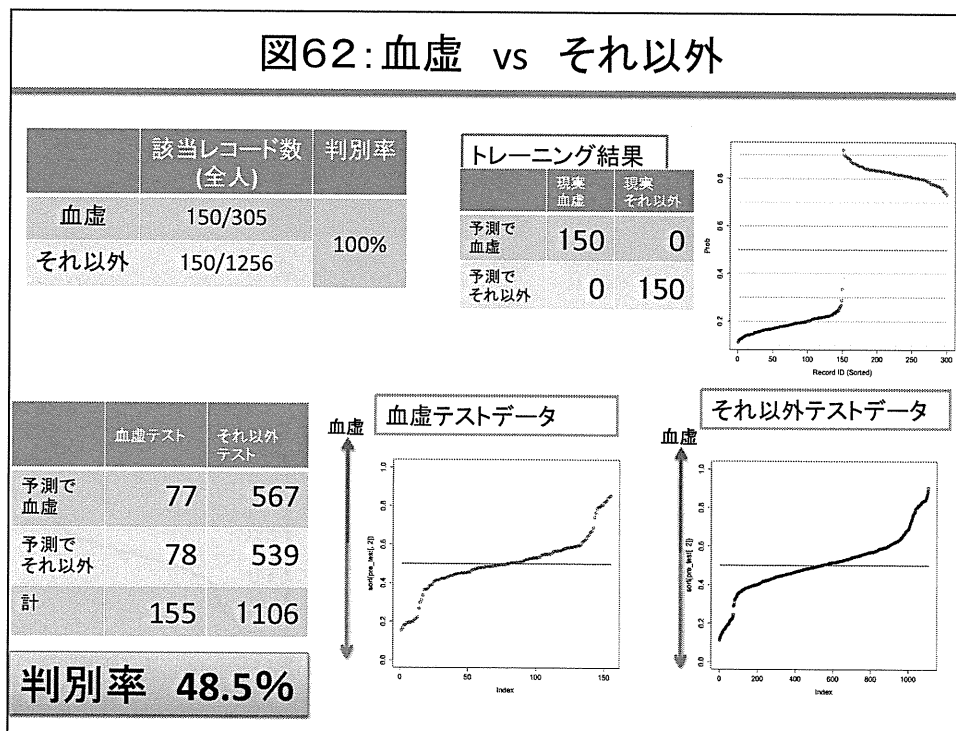


図63:お血 vs それ以外

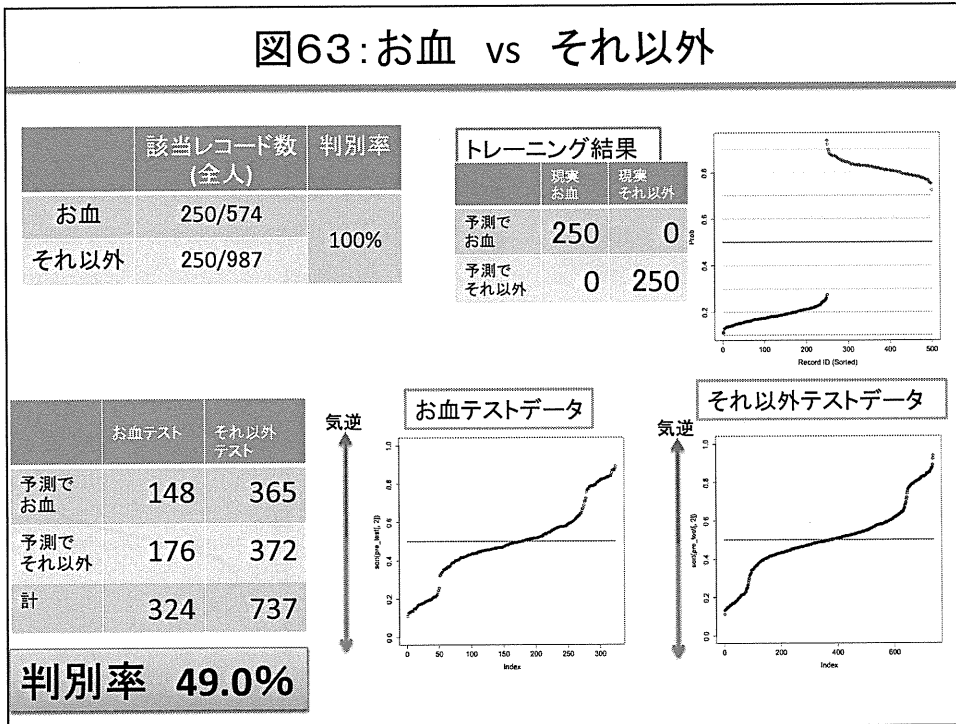


図64:水毒 vs それ以外

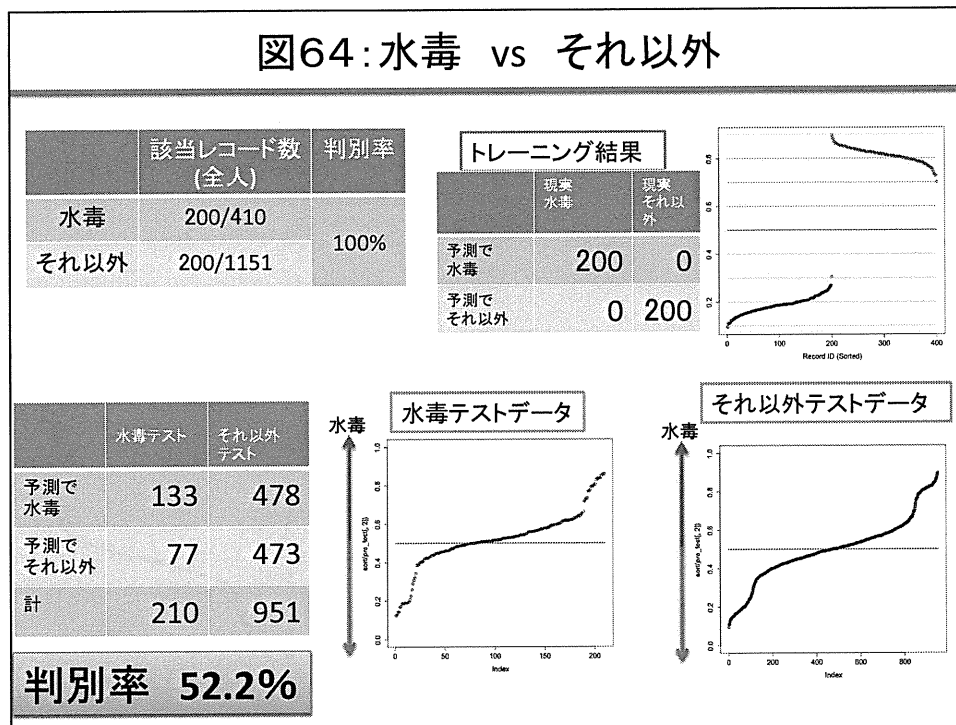
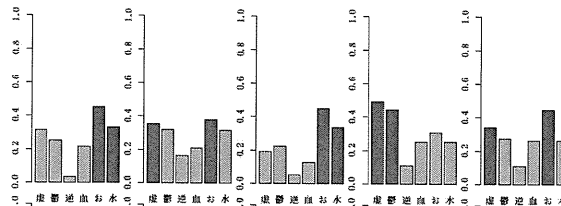


図65:各患者での予測例

2つともに正解



1つのみ正解

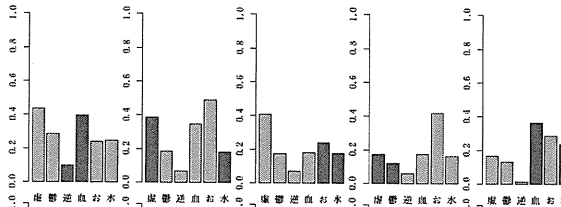


図66:結果の解析(1)

テストデータで
予測の上位2位までに、実際の診断が入っていた割合

実際の診断	人数	正解数	正解率
1つ	48	22	45.8%
2つ	120	84	70.0%
合計	168	106	63.1%

気血水 1つ持ち	人数	割合	気虚	気うつ	気逆	血虚	お血	水毒
全外し	27	54.2%	3	5	8	5	1	5
1位で 正解	14	29.2%	3	1	0	3	6	1
2位で 正解	7	14.5%	2	2	0	0	1	2
合計	48	100%	8	8	8	8	8	8

図67:結果の解析(2)

テストデータで
予測の上位2位までに、実際の診断が入っていた割合

実際の診断	人数	正解数	正解率
1つ	48	22	45.8%
2つ	120	84	70.8%
合計	168	106	63.1%

気血水 2つ持ち	人数	割合	気虚	気うつ	気逆	血虚	お血	水毒
全外し	35	30%	10	12	17	16	0	15
1つ正解	72	58.3%	25	25	23	22	30	19
2つ正解	13	10.8%	5	3	0	2	10	6
合計	120	100%	40	40	40	40	40	40

どの気血水で
正解したのか

データ解析、及び診療支援システムの可視化に関する研究

研究分担者 美馬 秀樹 東京大学大学院工学系研究科

研究要旨

本研究では、漢方医学のエビデンスの確立を目指し、患者の症状を軸とした臨床評価の手法を研究するとともに、エビデンスに基づき漢方医学における診断を支援するシステムの構築を目指す。そのため漢方薬及び鍼灸治療に伴う患者の自覚症状の推移を外来に設置した入力端末を活用して系統的に収集する。電子カルテの医療情報に対してオントロジー工学技術、可視化の技術を利用したデータマイニングを行って、治療効果の判定や漢方・鍼灸の診断「証」と症状との関連性を解析し、漢方の診断と治療の科学的検証をすると共に、解析結果を利用して有効な漢方処方を促進するための支援システムの構築を目指す。本年度は、データマイニングと可視化分析技術を統合したシステムに apache 財団による全文検索システム Solr を統合し、ユーザーによる分析を支援する機能を組み込んだ。本システムに患者の症状とくだされた漢方診断に関して実際のデータを入力し分析を行うと共に、診断を軸とした可視化を行うことで、本システムによる診断支援の予備的評価を行い、当該分野において新たなデータ分析手法創成の可能性を確認した。

A. 研究目的

医療情報と共にオントロジー工学技術、可視化の技術を利用したデータマイニングによって、治療効果の判定や漢方・鍼灸の診断「証」と症状との関連性を解析し、伝統医学の新たな臨床研究の手法を開発し、漢方の診断と治療の科学的検証を目指す。

B. 研究方法

科学の拡大、深化、それに伴う分野の分化を背景に、自律分散的に創造、管理される

知識¹の活用に際し、問題点として以下を考える。

- ・ 情報過多、知識過多

¹ 「情報」及び「知識」という言葉はドメインや文脈、また状況により様々な意味を持つ。例えば、Web を対象とした情報抽出は、HTML 文書から特定の部分のテキストを抽出することを指す場合が多いが、自然言語処理では、さらにテキストから固有名詞等の特定の情報を抽出することを指す。よって、本稿においてもそれらを厳密に定義しないが、構造化処理の対象をテキストとした際の「知」及び「知識」の対象として、用語及びその重要度等の属性により特徴付けられたパーティクル（文、段落、節、文書等の単位、またはそれらと関連付けられたコンテンツ）を扱うものとする。

- ・ 過度の細分化
- ・ 縦割り型（階層型）知識管理

結果として、

- ・ 知識の相互の繋がり
- ・ 知識の重複、差分
- ・ 知識の抜け

が不明瞭となっているのが現状である。

例えば、コンピュータ2000年問題、大銀行の統合のように、誰一人、システムの全体像を把握していないという状況が生じる原因となる。分野や組織、時勢を超えて知識を理解し、活用するためには、知識の全体像を明らかにすることが先決であり、総じて、「他を知る、他を分かる」ことが非常に重要となる。

逆に多様な知を関連づけることにより、新しい価値を創出することが可能になる。例えば医療において、近年、医学と工学の連携により発明された技術として、『3次元血管造影診断技術』がある。血管造影は心筋梗塞の重症度診断等、様々な診断で利用されているが、従来は腕や大腿部の動脈から細い管（カテーテル）を入れて造影剤を流し込み映画撮影するものであった。これは、治療法の選択等にも欠くことの出来ない有用な検査であるが、患者の時間的、体力的な負担が大きく、簡単に繰り返して行えるものではなかった。これに対し、『3次元血管造影診断』では、ITによる高速センシングと3次元CG（Computer Graphics）を利用した可視化技術により、短時間に検査を行うことができ、患者の負担軽減の観点からもその価値は計り知れない。この発明は、図1に示すように、『医療』と『情報工学』に係る知の構造化、さらには『造影』

と『可視化』という知の合成なしには、なし得なかったものである。

本研究におけるオントロジー工学、及びそれを活用したデータマイニングでは、この考えの基にモデル化、及び設計を行い、システム構築を進めた。

C. D. 研究結果および考察

<漢方データマイニングのモデル>

漢方データマイニングの目的は、膨大な問診等のデータを対象とし、

- a) 知識間の隙間を埋める知識の発見
- b) 知識間をまたぐ知識の発見

を行うことにある。本研究での知識とは、漢方問診における診断や患者、「証」、漢方薬等のそれ自体、及びそれらの関連や、それらを決定するためのノウハウを指すものとする。一般に、全ての知識の関連は絶対的に定義できるものではないため、全ての場合において上記を厳密に区別し処理を進めることは困難であるが、既にある例として、a)に対してはバイオインフォマティクス等、b)に対しては環境科学等がそれに当たると言える。これを実現するためのマイニングモデルとして以下を考える（図2）。

1) 全体像の把握

知識の既存の関連や属性に基づく関連を抽出し、知識間の個々の関連から全体の関連を明らかにする。細分化や縦割りの弊害等により、失われがちな関連をも見つけ出すことが重要であり、オントロジー、可視化、見える化等の技術が重要な要素となる。

2) 抽象化と詳細化

膨大な量の知識の全体像を把握するためには、抽象化は必須である。抽象化された領域より必要とする知識を選択した後、その領域の詳細化へと進めることで、必要な知識の絞り込みが容易になる。言わば、「森を見て、木を見る」操作である。

3) 合成

様々な知識から新たな知識を創造するためには、既存の知識を如何に再利用するかが重要である。異なる分野の知識を上記、抽象化等の操作により選択し、合成することで、より新しい知識の創出が期待される。また、創出された新たな知識を次の合成の種へとリサイクル、リファインメントを繰り返すことで、知識はより成熟する。

これらの分析、可視化、及び操作が、個人、及び任意の視点によりリアルタイムに行えることが重要である。つまり、任意の視点で詳細化、抽象化の階層を上下しつつ、関連のある、もしくは関連が必要な知識を選択し、合成の要素を探すのである(図3)。さらには、次の瞬間にこれら新たに創出された知識が次の合成や抽象化の対象となる。このように、知識の連続的創出と活用を促し、さらに高度な知識の再活用へと昇華させるためには、知識創出、活用の「螺旋」を形成できることが重要である。以下では、これらのモデル化に基づき設計、開発した技術に関し、詳述する。

<漢方データマイニングシステムと「MIMAサーチ」>

我々は、上記の方法論を実践し、テキスト情

報を対象とした知の構造化を支援する機構として「MIMAサーチ¹⁾」を開発し、その実用化を行ってきた。「MIMAサーチ」においては、自然言語書処理や多変量解析等の統計処理を活用することで、膨大なデータや情報より瞬時に必要とする知識を抽出し、さらに抽出した知識間の関連性を自動で計算する。一般に、自然言語処理とは、形態素解析、構文解析、意味解析等により計算機を用いて言語の理解を行うことを指す。従来、これらを用いた仮名漢字変換、機械翻訳システム、用語(概念)抽出システム、全文検索システム等のアプリケーションが開発されており、現在では、計算機の発展により大量の言語情報を高速に処理することが可能となっている。「MIMAサーチ」の特長は、図4に示すように、論文や、報告、アンケート等に記述されている自由文(テキスト)を自然言語処理により解析し、その統計情報に基づき、オントロジーとして重要な用語(概念)を自動的に認識・抽出することにある。さらに、抽出したオントロジーと問診データ等のデータを統合し、シームレスに扱うことで、知識を総合的に処理し、知識間の意味的関連とその関連の強さを定量的に計算する。さらに、それらを視覚的に捉えることができるよう、関連、及びその強さをグラフモデルにより可視化する。つまり、単なる個々の診断等の内容を捉えるだけでなく、患者や診断結果間の意味的な関連に基づいて全体を俯瞰し、知識を抽象化して捉えることができることを意味する。

より具体的には、「MIMAサーチ」は以下のような特徴を持つ。

- ・ 必要とする分野全体の知識、日々創出されるリアルタイムな情報、共創的に創出される知識を含む様々な形態の知識群を統合し、データベースとして蓄積する。
- ・ 上記データベースより、ある分野や領域、または分野横断的に任意の知識を抽出し、抽出された知識全体の関連を可視化する。
- ・ 知識間の関連として、あらかじめ定義された情報、もしくは手続きにより導出される類似、包含（差分）、部分全体、因果、を含むオントロジー的関連が参照できる。
- ・ 上記はキーワード等により指定される任意の視点を反映できる。
- ・ 上記により指定、もしくは計算された関連を基に、関連の強い知識同士をまとめ上げる（クラスタリング/クラシフィケーション機能）。
- ・ 上記のまとめ上げを任意の抽象度で可視化する（階層的クラスタリング機能）。
- ・ 任意の知識を選択し、また必要な知識を加えることで新たな知識を創出し、データベースに追加できる。

例えば、これらにより、複数のソース、分野から横断的に知識を検索、抽出し、関連度指定、抽象度指定により表れたクラスターより任意の知識を選択、さらにこれらを合成するという流れが実現可能である。例えば、生命科学分野における論文の加速的な増加は冒頭でも述べたが、爆発的に増加する論文の全てに一人の人間が目を通すのは既に不可能な状況になっていることは容易に想像できる。にもかかわらず、論文の査読や発明特許の申請等においては、既知の事項との重複がない

か等の、関連する分野の知識を網羅的に把握する必要がある。このような目的においても、知の構造化技術を利用することで、まずは分野全体の知識を俯瞰し、全体の中での位置づけをつかんだ上で、さらにその位置の詳細を確認するといった、「全体像」から「詳細像」、さらにはまた「全体像」へといったズームインとズームアウトを繰り返すことで、関連する知識をより効率的に探すことが可能である。より具体的には、まずMIMAサーチで全体を俯瞰し、意味的なまとまりのある部分に絞り込んで検索を進め、主として関連している可能性の高い論文を把握した上で個々の関連を取り出し、検証するという詳細化のアプローチにより検索や比較の対象を絞ることが考えられる。例えば、図5(a)(b)はそれぞれ2006年、及び2007年に開催された生命科学研究ネットワーク・シンポジウムで発表された論文（それぞれ304件、324件）をMIMAサーチにより可視化したものである。先にも述べたように、図では、内容が関連する論文がより近くなるように配置されており、よりまとまりのある論文群（クラスター）にはその内容に応じて『分子メカニズム』のような重要な用語を基に計算したトピックラベルが自動で振られている。また、さらに大きなまとまりを円で囲み、『臨床医学』のような分野名のラベルを割り当てている。年度をまたいだ恒常的なテーマが存在するのと同時に、『分子機構』や『メタボリックシンドローム』のような、それぞれの時勢に応じてテーマとなる研究が変遷していく様子が見て取れるのが興味深い。また、イノベーション支援や知識創造

支援の観点では、境界領域の設定のような分野を横断した関連が増加することが望ましいと言えるが、図において、2006年度から2007年度への分野の変遷を見ると、2006年度と比較して、2007年度には『新学術領域』と『臨床医学』や『工学』との関連が増加し、全体の繋がりがより明確になり、全体像が凝縮されていることが見て取れる。これは主に、シンポジウム等により人的交流が増加したためと見ることができ、このように、ある種の仮説検証のプロセスの一部として、このような知の構造化と可視化を活用することも可能である。

本年度は、ユーザビリティのさらなる向上のため、apache 財団による全文検索システム Solr への対応を行い、MIMA サーチの可視化結果の最適化支援を強化した（図6）。MIMA サーチの索引付けに Solr を統合することにより、検索結果をさらに絞り込むために有用な属性値を提示するファセット機能、データに基づき検索クエリの候補を提示するサジェスト機能などのユーザー支援機能が実現された。これらの機能は、MIMA サーチによるマイニング作業、全体像での位置づけ把握から、絞り込んだ知識の詳細像を確認する、などのユーザー作業をよりスムーズにするための機能である。またユーザインターフェースの面においては、Solr はモジュール化された拡張性の高い構造をしているため、昨年度 HTML5 化した MIMA サーチと組み合わせることで、利用目的ごとにこれらの豊富な機能を柔軟に付け替えることが可能になり、より目的志向のインターフェースを提供できるようになった。

本研究の目標は、漢方医学を対象とした診療のエビデンス創出、及び、創出されたエビデンスを基にした診断支援システムの構築である。今日では、医師の7割以上が日常診療で漢方を用いているにも関わらず、漢方専門医以外は漢方薬の使用処方限定されているのが現状である。これは、漢方診療が、同病異治・異病同治に示されるような、個別化診療であることが主な原因である。また漢方医学的診断の特徴である「証」の特定は Know-How のような経験知から導かれるものであり、未だ十分に科学的・統計学的な解明が為されているとは言い難い。つまり、臨床的エビデンス（診断・治療）が得られるようになれば、一般臨床医もある一定のレベルで効果的な漢方治療ができると期待される。これを目的とし、本研究では、まず、漢方薬及び鍼灸治療に伴う患者の自覚症状の推移を、外来に設置した入力端末を活用して系統的に収集し、電子カルテの医療情報と共に知の構造化の技術を利用することによって、治療効果の判定や漢方・鍼灸の診断「証」と症状との関連性を解析し、伝統医学の新たな臨床研究の手法を開発することで、漢方の診断と治療の科学的検証を行うことを目指している。例えば、図7には「証」（虚実）における「やや虚証」と「虚証」の診断に対する微妙な差異を統計と可視化により明確化した例であるが、本研究により、このような Know-How に係る経験的、暗黙的知識の「見える化」が行えると期待される。実際のデータ収集においては、問診システムにて評価された患者の状態に対し、投与された薬剤や処置を問診終了後に医師が入

力することとし、再診時以降は、患者の視点から評価された症状の連続的な変化が、治療経過と共に時系列で記録されている。現在、平成22年度、23年度初診患者3100名余りの診療がデータベース化されており、このデータを基にMIMAサーチによる解析を進めている。従来のこのような研究は、収集したデータの統計情報を基に、定量的分析を行うのみであったが、さらに、MIMAサーチを利用することで可視化技術を活用した定性的分析手法との統合的解析を行っていることも本研究の大きな特徴である。

図8は初診時に冷えと頭痛があった患者に関してMIMAサーチによる分析を行い、「証」（虚実）を軸とした可視化を行った結果を示している。図では、グラフモデルによって患者間の関連の可視化が行われている。それぞれの点が患者を表し、患者間に引かれている線の長さや太さがそれら患者間の関連の強さを示す。つまり、症状が似ている患者がより近くに配置されるようになる。また、図中の大きな円により関連の強い患者群（クラスター）が示され、これら円の上に示される内容が問診に表れる特徴を示している。点の色は患者にくだされた漢方診断を表し、同じ色の患者は同じ漢方証に診断されたことを示している。図9は図8のうち「証」（虚実）における「やや虚証」か「虚証」の診断を受けた患者に関してさらに分析可視化を行ったものである。図8から図9へはMIMAサーチ上で属性を絞り込むことにより容易に移行可能である。図9から読み取れるように、これらの患者は、冷えと頭痛の他に首こりのある患者群A、

寒がりかつ目の疲労がある患者群Bなどに分かれていることが示されている。また患者群Aは全員が「虚証」と診断されているのに対し、患者群Bには「やや虚証」とされた患者も含まれていることがわかる。このように新たな患者がどの群に最も近いかわかる、どのような位置付けにあるか等の観察により、証の診断を支援することが可能になると考えられる。さらに新たな患者と近い患者群の処方をも提示することで処方の支援にもつながると期待される。

E. 結論

本報告では、分野を超えた知識の活用を目的に、オントロジー工学、及び可視化技術を統合した漢方データマイニングのモデル化、また、モデル実現のためのシステム設計とその一実装である「MIMAサーチ」に関して述べた。また、実際に「MIMAサーチ」を活用した漢方データ分析とその可視化による診断支援に向けた研究成果について報告した。

問診結果等のリアルタイムに流れてくる情報や知識を取捨選択しつつ、価値のある知識のみを抜き取り、そこから仮説を構築し、既に蓄積された知識を活用しながら、仮説を検証するサイクルにおいては、必要な知識を必要なタイミングで獲得できることが不可欠である。その意味でも、オントロジー工学や可視化への期待は非常に大きい。

1) Mima, H., Ananiadou, S., Matsushima, K.: Terminology-Based Knowledge Mining for New Knowledge Discovery, ACM Transactions on Asian Language Information Processing,

Vol. 5, (2006) pp.74-88.

F. 健康危険情報

該当なし

G. 研究発表

- (1) 美馬秀樹, 「生命科学における知の構造化」, Biophilia, Vol.7, No.3, pp.47-53, (2011).
- (2) Hideki Mima, Katsuya Masuda, Susumu Ota, and Shunya Yoshimi, “Revealing Modern History of Japanese Philosophy Using Natural Language Processing and Visualization”, In Proceedings of Osaka Symposium on Digital Humanities (OSDH) 2011, pp. 47-50, (2011).

H. 知的財産権の出願・登録状況

1. 特許取得

該当なし

2. 実用新案登録

該当なし

3. その他

なし