

とすることを表している。

C.2 問診データからの冷えの予測結果

図11に予測精度のまとめを載せた。結果として、「冷え」があるか否かの予測は91.1%の確率で正解できることが分かった。また、この予測は、233個の問診項目、および問診項目の組み合わせ（複数の問診項目の交互作用項）によるモデルにより達成された。図12と図13には、実際に「冷え」の予測に対して寄与が大きかった（ロジスティック判別モデルにおいて、問診項目にかかる係数が大きい（負の場合小さい））ことに対応する問診項目の組み合わせを示した。図12にあげたものは、係数が正の問診項目の組み合わせであり、これらの問診項目に該当した患者は西洋病名診断において「冷え」と判断される傾向にある。1位、2位をみると、問診項目において直接「冷え」に関わるものが選ばれている。これは全く自然である。しかしながら、3位、4位は直接「冷え」を尋ねた問診項目ではない項目であった。3位は、「食事の早さが早い」と「気分が憂鬱になる」の交互作用であり、そのサポートは0.1である。ここで、サポートとは、2つの問診項目「食事の早さが早い」と「気分が憂鬱になる」の両方に該当するサンプルが全体の0.1（10%）ということである。サンプル数は約10,000ということから、この2つの問診項目に同時に該当するサンプルは約1,000ということになる。4位は、3位と同様に2つの問診項目の交互作用であり、「イライラする」と「目のクマができる」であった。サポートは約13%であった。なお、9位にも冷えを直接は尋ねていない「目が疲れる」と「足がふらつく」の交互作用が現れる。このサポートは約10%である。これらの結果から、身体的に疲れており、かつ、精神的に不安定であることが「冷え」に対して正に関連していることが示唆される。

逆に、該当すると冷えがないと判定される傾向にある問診項目については、その傾向の強いものから順に図13にあげた。興味深いのは、図12と同様に「イライラする」が出現することである。しかし

ながら、図12の結果とは異なり、「イライラする」に「ものを忘れる」や「のどが渇く」といった身体的な疲れとはことなる性質を有する問診項目との交互作用が図13には見られる。また、7位には、「足の冷え」という「冷え」を直接尋ねている問診項目が出現する。一見、これは矛盾のように思えるが、「足の冷え」をだけを訴える患者というわけではなく、「足の冷え」に加えて「首のこり」や「目の疲れ」に同時に該当するという非常に特異的な患者を対象にしている。また、後の解析で詳しくデータを用いて示すが、足の冷えは、「冷え」に対する直接的な質問における該当部位「手」、「足」、「腰」、「全身」の中ではもっとも初期に現れる症状であると考えられる。したがって、比較的「冷え」の症状の軽い患者がこの交互作用に該当していると考えられる。

冷えに正に関連する問診項目のネットワークを図14にあげる。ネットワークの構築法は、研究方法で説明した方法を用いた。つまり、冷えに対して予測能力があると判定された問診項目の交互作用を構成する問診項目同士をエッジ（線）で結ぶことでネットワークは得られる。ここで、赤い色を付けた問診項目は、「冷え」に対して正の影響のある問診項目群にだけ出現する（負の影響のある問診項目群には出現しない）問診項目である。かなり複雑なネットワークを問診項目は構成していることが分かる。このなかで、例えば、「目が疲れる」のように様々な問診項目に繋がっている、いわゆるネットワーク上のハブになっているような問診項目は重要な意味があると考えられる。「冷え」に対して負の影響のある問診項目のネットワークは図15にあげた。図14と同様に、赤のノードで示した問診項目は、負の影響のある問診項目のネットワークにしか出現しないものである。

問診項目と同様に、「冷え」を漢方診断（証）から予測する数理モデルを構築した。問診項目と同様に、5つまでの証の交互作用をEclatを用いて高頻度サブセットマイニングを行い抽出し、それらをモデルに取り込み、どの証、証の組み合わせが「冷え」を予測する能力を有するかをElastic netを用いて

評価した。問診項目と同様に、冷えを予測する能力を有する証のネットワークを、正に関連する証と負に関連する証に分け、可視化したものがそれぞれ図 16、17 となる。

C.3 問診データからの虚実寒熱診断予測

判別方式の性能比較を行った結果を図 18 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。

BMI を問診データと共に用い、慶應義塾大学病院での虚実の予測を行った結果を図 19 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 37 と少なかったため、トレーニングには 20 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 91.2% を達成することに成功した。

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研究代表者である渡辺の診断と一致するの可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 20 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 21 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 21 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 21 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 22 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. 4 各施設の虚実診断状況比較結果

図23、24にある各施設の虚実診断の数であるが、まず、気がつくこととして、慶応は実証の割合が多く、亀田では虚証と診断される患者は少ない。また、富山大は中間症として診断される患者の割合が少ないことが特徴としてあげられる。

次に、慶應義塾大学病院問診データに基づき虚実予測式を構築した後、多施設において集められた問診データに適用した。図25が結果である。慶応大学病院の問診データによって構築された虚実予測式であったが、実証、やや実証については90%以上の正解率を示すことに成功した。また、虚証、やや虚証においても60%以上の正解率を得ることができた。「虚」側では「実」側よりも正解率が上がらなかったが、虚実の診断としては、慶應義塾大学病院、亀田総合病院は問診データに基づく限られた議論ではあるが確かに一致性がみられることがわかった。

富山大学病院の患者の虚実を予測した結果、亀田総合病院と同様に、「実」側において高い正答率、「虚」側においても「実」側ほどではないが、60%以上の正答率を得ることができ、やはり慶應義塾大学病院と富山大学病院の判断にはある一定の一致性があると考えられる。つまり、慶應義塾大学病院、亀田総合病院、富山大学病院は虚実判定においてその基準には問診データをキーにしてある類似性があると考えられ、何らかの補正は必要とは考えるがデータ統合の可能性は高い。特に、「実」側では高い予測能力を慶應義塾大学病院問診データで構築した予測方式は他施設のデータに対しても有しているという結果は興味深い。

麻生飯塚病院も亀田総合病院、富山大学病院と同様に「実」側の予測が高い正答率を示している。また、麻生飯塚病院は、「虚」側の予測も70%弱の正答率を達成し、他の施設に比べ高いことが見て取れる。問診データをキーにすると、より慶應義塾大学病院との類似性があるのかも知れない。

自治医科大学病院はサンプル数が少なく、虚実中間証を除くと7症例であるが、その結果の傾向はこれまで示した施設と共通である。ただし、サンプル

数が少ないため結果の分散も大きいことに注意する必要がある。この点については、自治医科大学病院においてデータ数を増やした上で再確認すべきであろう。

D. 考察

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、亀田総合病院から得られた問診データの解析、問診データから証の予測、および、多施設において集められた問診データの統合解析、および統一的予測方式構築についての可能性を探った。その結果、以下の点を見いだした。

- サンプル数の多い慶應大問診データ、および漢方診断（証）のデータを用い、西洋病名診断「冷え」（「アトピー」に関しても実施）の予測を試みた。その結果、極めて高精度にこれら2つの疾患を予測することができるとわかった。
- 高頻度な問診項目に絞った解析では、各施設の傾向は高い類似性を示した。従って、高頻度の問診項目については施設をまたいだ判断を下しても良いと思われる。ただし、今回の解析を行う際に、5施設の間診票に共通の間診項目を最初に抽出し、20問診項目が共通であるという結果を得たが、これは千葉を除く4施設のおおよそ10%に過ぎなかった。もっとも、千葉の間診項目数が50と、他の施設よりもかなり少ないということもその理由としてあげられる。患者個々人の自覚症状をより反映させ、診療の強力な支援となる問診票を作成するためには各施設の間診票の見直し、標準化が今後鍵となると思われる。
- 証の予測においては、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ことが明らかとなった。

- 患者の気虚、寒熱の診断数は、施設毎にバラツキがあることが分かった。しかしながら、そのバラツキが施設における診療の違いなのか、それとも地域的な違いにより診断の数が見かけ上異なっているように見えるだけなのかを明らかにするために慶応問診データを用いて構築した虚実寒熱予測式を他施設の問診データに適用した。その結果、実証においては80～90%の高い正答率を得ることができ、虚証においても60～70%という正答率を得ることができた。また、寒熱についても2施設においては70%以上という高い正答率を達成することができ、診断の背景にはある一致性が示唆された。

E. 結論

漢方専門外来において集められた問診データは、患者個人毎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西洋病名診断を高い確率で予測できることを示した。また、診療支援を目的とし、問診データから証を予測するためにより性能の高いデータマイニング手法を探索し、実際に慶應義塾大学病院の問診データを用い虚実寒熱の予測正解率が90%、80%以上という高い性能を実証できた。この成果に基づき、多施設において集められた問診データを解析し、統合解析の可能性を探り、また、各施設における診療を問診データからの予測という観点で見た際の一致性について検討し、各施設の共通性を見出し、また、将来、多施設解析の際の補正として利用する候補となる際のある問診項目を見出すことに成功した。

F. 健康危険情報

特になし

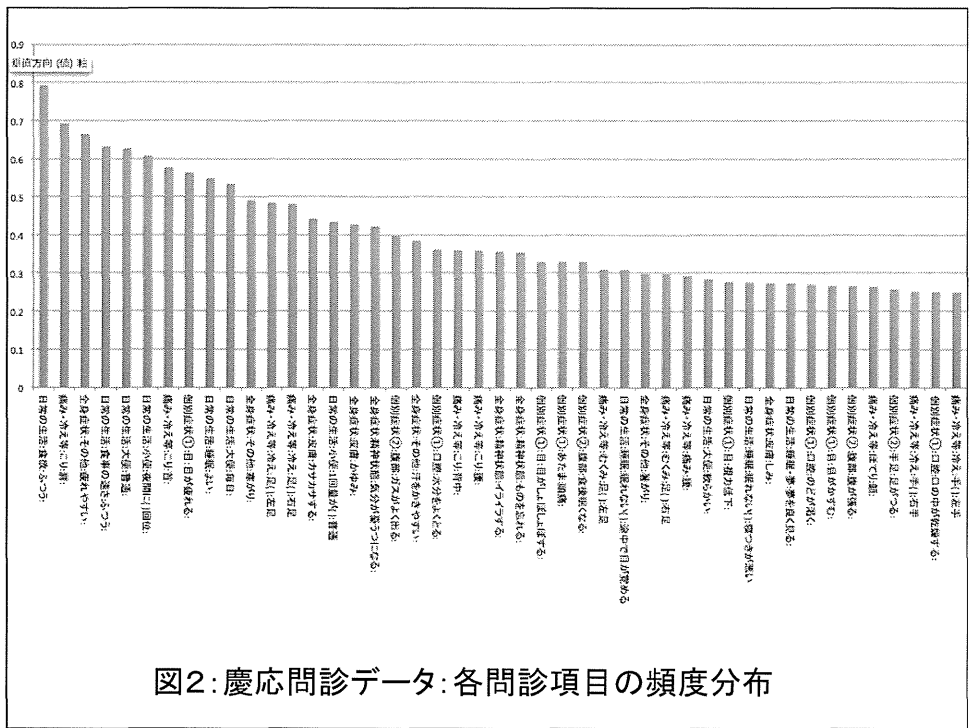
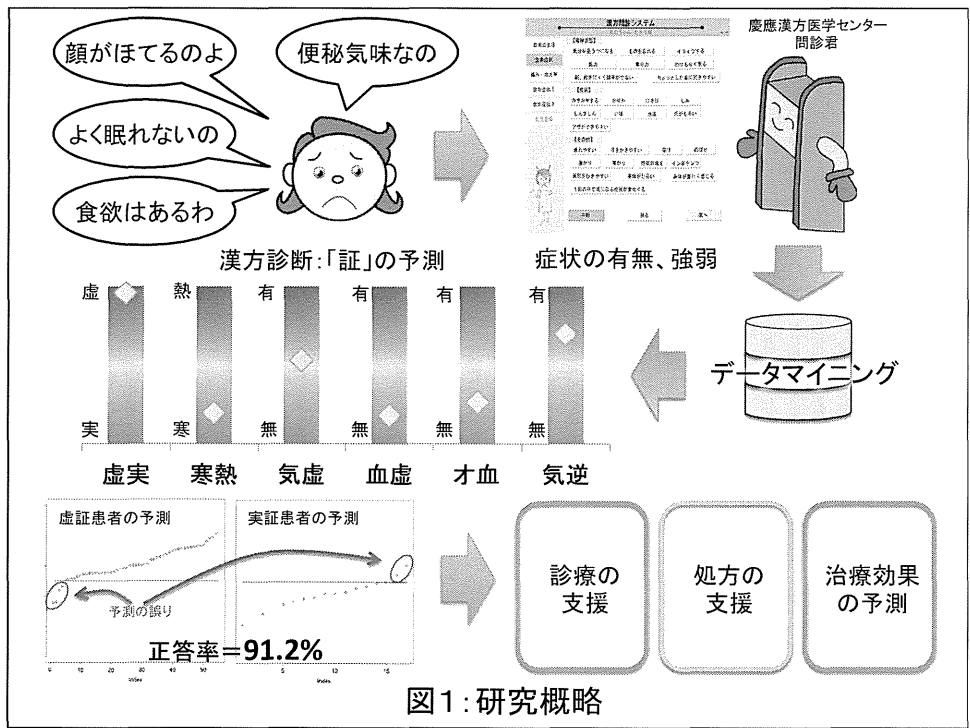
G. 研究発表

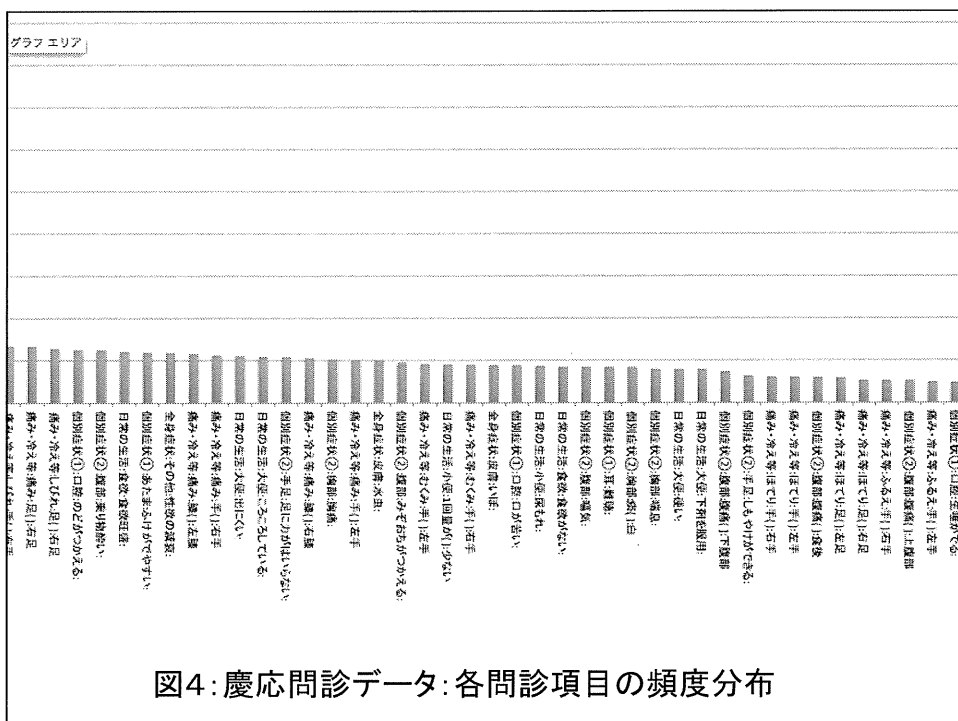
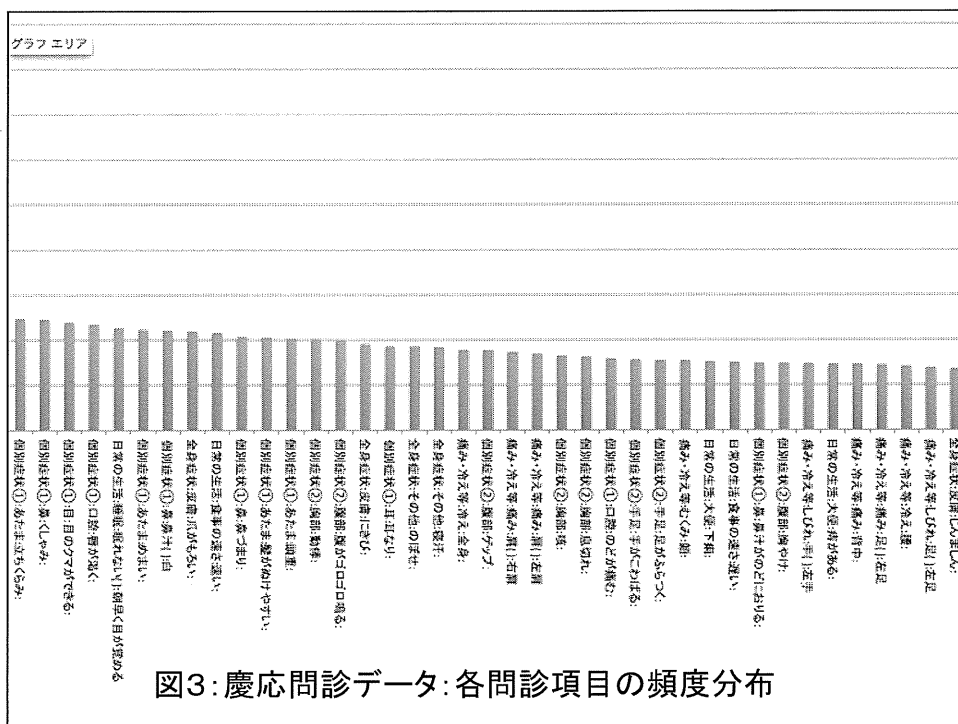
1. 論文発表

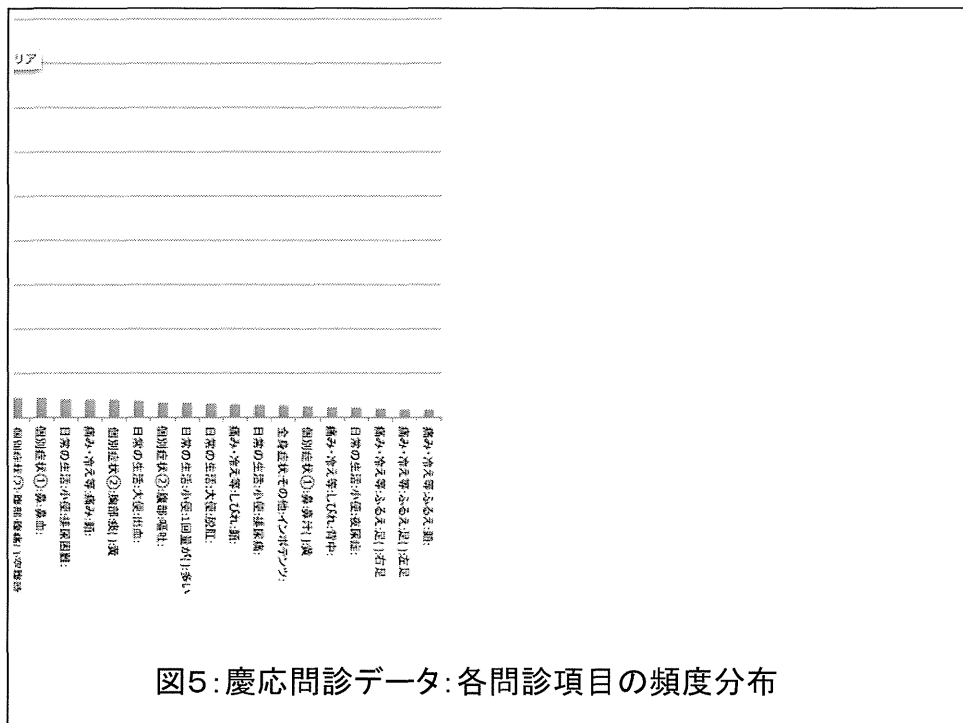
1. K. Katayama, R. Yamaguchi, S. Imoto, H. Tokunaga, Y. Imazu, K. Matuura, K. Watanabe and S. Miyano : Symbolic hierarchical clustering for visual analogue scale data, KES-Springer Smart Innovations, Systems and Technologies series, Springer-Verlag, 10, 799-805, 2011
2. K. Katayama, R. Yamaguchi, S. Imoto, K. Matsuura, K. Watanabe, S. Miyano : Transform of visual analogue scale data and their clustering, International Journal of Knowledge Engineering and Soft Data Paradigms, 3(2), 143-151, 2012
3. K. Katayama, R. Yamaguchi, S. Imoto, K. Matsuura, K. Watanabe, S. Miyano : Clustering for visual analogue scale data in symbolic data analysis, Procedia Computer Science, 6, 370-374, 2011
4. K. Katayama, R. Yamaguchi, S. Imoto, K. Matsuura, K. Watanabe, S. Miyano : Symbolic hierarchical clustering for pain vector, Intelligent Decision Technologies Smart Innovation, Systems and Technologies, 16, 117-124, 2012
5. K. Katayama, R. Yamaguchi, S. Imoto, K. Matsuura, K. Watanabe, S. Miyano : Connection between traditional medicine and disease, ACM SIGHIT Record, 2, Issue 1, 2012
6. K. Katayama, R. Yamaguchi, S. Imoto, K. Matsuura, K. Watanabe, S. Miyano : Analysis of questionnaire for traditional medical and develop decision support system, *IEEE Catalog Number: CFP1226D-US*, 762-763, (ISBN: 978-1-4673-2744-2), 2012

H. 知的財産権の出願・登録状況

該当無し







施設	問診項目数	備考
慶應義塾大学病院	362	VAS・女性特有も含む全項目
千葉大学病院	50	
麻生飯塚病院	240	
富山大学病院	231	富山と鹿島の問診用紙は同じ
鹿島労災病院	231	

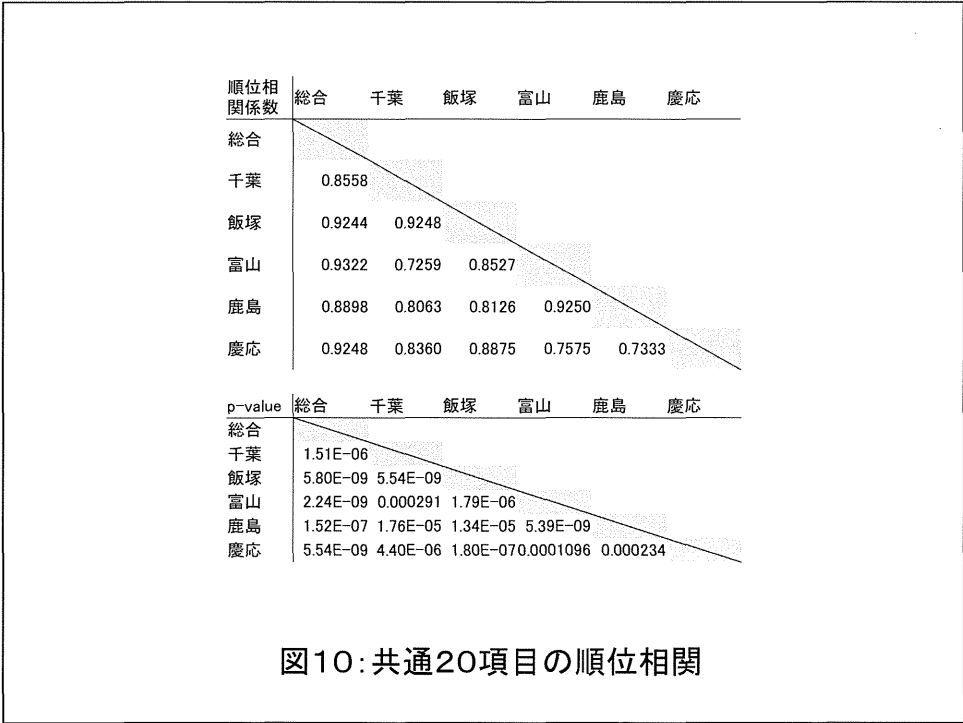
応応番 号	症状内容	千葉 番号 千葉内容	飯塚 番号 飯塚内容	富山県 島番号 富山県島内容
	23 日常生活小便(回数が)少ない	47尿が少ない	32下痢する	25下痢する
	40 日常生活大便下痢	42下痢しやすい	185痔がある	171痔の気がある
	43 日常生活大便痔がある	45痔がある	185痔がある	171痔の気がある
	72 全身症状その他疲れやすい	5疲れやすい	1疲れやすい	1疲れやすい
	125痛み・冷え等しなれ手(足)の手	18しびれや麻痺の強いところがある	217手足の気がしびれる	201手足の気がしびれる
	127痛み・冷え等しなれ手(足)の手	18しびれや麻痺の強いところがある	217手足の気がしびれる	201手足の気がしびれる
	130痛み・冷え等しなれ足(足)の足	18しびれや麻痺の強いところがある	217手足の気がしびれる	201手足の気がしびれる
	132痛み・冷え等しなれ足(足)の足	18しびれや麻痺の強いところがある	217手足の気がしびれる	201手足の気がしびれる
	135痛み・冷え等しなれ手(足)の手	19手や足が冷える	20手足が冷える	68手足が冷える
	135痛み・冷え等しなれ手(足)の手	19手や足が冷える	20手足が冷える	68手足が冷える
	135痛み・冷え等しなれ足(足)の足	19手や足が冷える	20手足が冷える	68手足が冷える
	135痛み・冷え等しなれ足(足)の足	19手や足が冷える	20手足が冷える	68手足が冷える
	190個別症状①あたまめまい	24めまいがする	139よく立ちくらみする	128めがまわることがある
	192個別症状①あたま立ちくらみ	25立ちくらみがある	140耳なりがすることがある	129よく立ちくらみする
	225個別症状①口腔のどがつかえる	30喉がつかえた感じがする	147何かがのどにつかえる感じがする	217咽喉のつかえ感
	21肩がこる(肩こり)が肩	27肩がこる	143よく肩がこる	133よく肩がこる
	212個別症状①息苦しい(呼吸)	27息苦しい	143よく肩がこる	133よく肩がこる
	239個別症状②胸部動悸	32胸こむ	152よくセキがでる	142よくセキがでる
	249個別症状②胸部動悸	33急に動悸がする	156動悸がする	145動悸がする
	264個別症状②手足足がつる	20こむらがえりがする	216よくムズムズがエリする(筋肉がひきつる)	200よくムズムズがエリする
	268個別症状②腹部ガスがよく出る	44おならがよくでる	183ガスがよく出る方だと思ふ	169ガスがよく出る方だと思ふ
	252個別症状②腹部ゲップ	34ゲップがしやすい	169ゲップがでる	156ゲップがでる
	256個別症状②腹部みぞおちがつかえる	38肋骨の下部分のお腹が重苦しい	172みぞおちの重苦しい感じがする	220みぞおちから脇あたりのつかえ感
	262個別症状②腹部乗り物酔い	26車酔いしやすい	21乗り物酔いをする	16乗物酔いをする
	266個別症状②腹部腹がゴロゴロ鳴る	41腹がゴロゴロ鳴る	182腹がゴロゴログーグーなることがある	168腹がゴロゴログーグーなることがある
	264個別症状②腹部腹が張る	40腹がはる	175腹が張ることがある	162腹のはることがある
	258個別症状②腹部嘔気	38吐き気がする	166よく嘔気(はきけ)がする	154よく嘔気(はきけ)がする
	1 日常生活食欲あまりない	35食欲がない	46食欲がない(食事を食べたくないことがある)	39食欲がない

図7:5施設問診票の共通20項目

慶応		千葉		飯塚	
全身症状その他疲れやすい	66.6%	全身症状その他疲れやすい	53.9%	全身症状その他疲れやすい	64.4%
手足の冷え	51.9%	手足の冷え	51.2%	手足の冷え	56.6%
個別症状②腹部ガスがよく出る	39.5%	個別症状②腹部ガスがよく出る	29.3%	個別症状②腹部ガスがよく出る	39.3%
手足のしびれ	26.7%	手足のしびれ	28.1%	個別症状①あたま立ちくらみ	31.0%
個別症状②腹部腹が張る	26.4%	個別症状②腹部腹が張る	21.5%	個別症状②腹部腹が張る	27.8%
個別症状②手足足がつる	25.8%	日常生活大便秘がある	20.7%	個別症状②手足足がつる	26.5%
鼻水	25.8%	個別症状①あたまめまい	20.3%	個別症状②腹部腹がゴロゴロ鳴る	24.9%
個別症状①あたま立ちくらみ	24.9%	個別症状②胸部動悸	18.0%	個別症状①あたまめまい	24.3%
個別症状①あたまめまい	22.0%	鼻水	17.6%	鼻水	24.3%
個別症状②腹部腹がゴロゴロ鳴る	20.0%	日常生活大便秘下痢	16.8%	手足のしびれ	24.1%
個別症状②胸部動悸	20.0%	個別症状①あたま立ちくらみ	16.4%	個別症状②胸部動悸	20.4%
個別症状②腹部ゲップ	17.5%	個別症状②手足足がつる	15.2%	個別症状①口腔のどがつかえる	18.7%
個別症状②胸部咳	16.6%	個別症状②腹部腹がゴロゴロ鳴る	14.5%	日常生活大便秘がある	18.2%
日常生活大便秘下痢	15.0%	個別症状②腹部ゲップ	13.7%	日常生活大便秘下痢	18.0%
日常生活大便秘がある	14.6%	個別症状①口腔のどがつかえる	13.3%	個別症状②腹部乗り物酔い	17.6%
個別症状①口腔のどがつかえる	12.5%	個別症状②腹部みぞおちがつかえる	12.1%	個別症状②胸部咳	16.9%
個別症状②腹部乗り物酔い	12.3%	日常生活食欲あまりない	11.3%	日常生活食欲あまりない	15.6%
個別症状②腹部みぞおちがつかえる	9.6%	個別症状②腹部乗り物酔い	10.9%	個別症状②腹部みぞおちがつかえる	15.4%
日常生活食欲あまりない	8.7%	個別症状②胸部咳	8.6%	個別症状②腹部ゲップ	13.7%
個別症状②腹部嘔気	8.5%	個別症状②腹部嘔気	8.2%	個別症状②腹部嘔気	8.5%

図8:共通20項目の頻度分布。慶応、千葉、飯塚

図8: 共通20項目の頻度分布。慶応、千葉、飯塚



		データ	
		冷え有	冷え無し
予測	冷え有	785	452
	冷え無し	396	7846

正しい予測
予測の誤り

予測正答率 91.10%

233個の問診項目、問診項目の組みあわせによって予測される

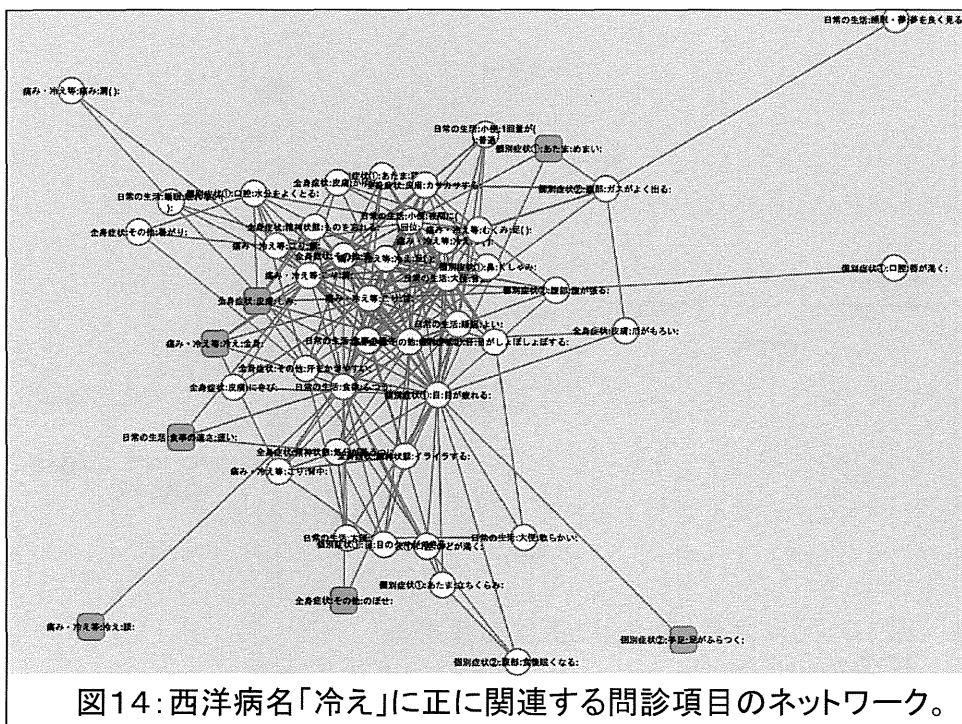
図11: 西洋病名「冷え」を問診項目から予測した際の精度

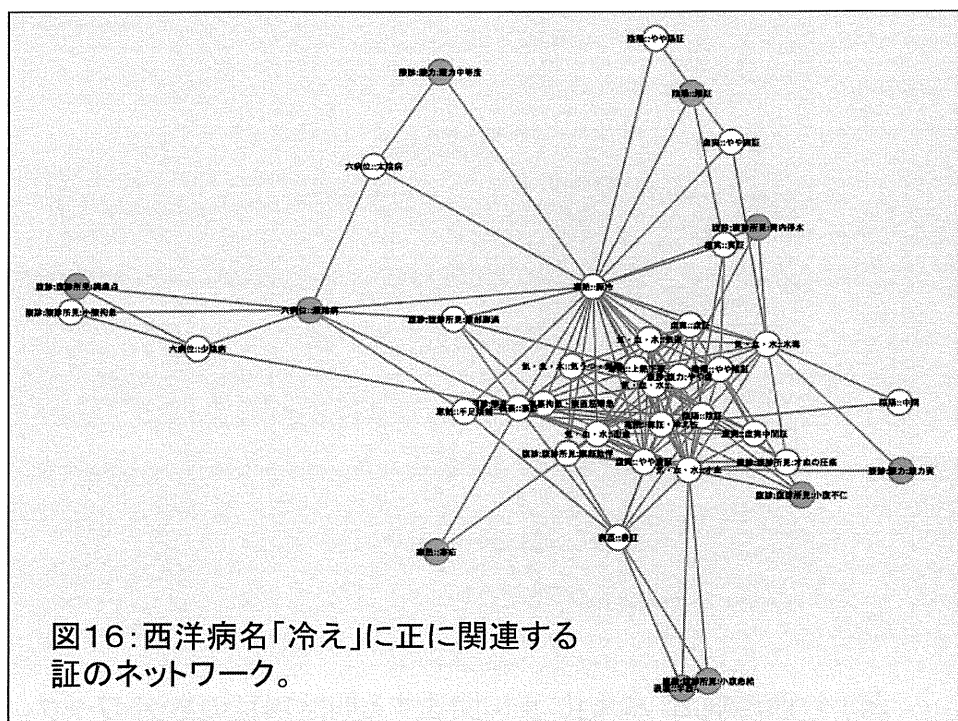
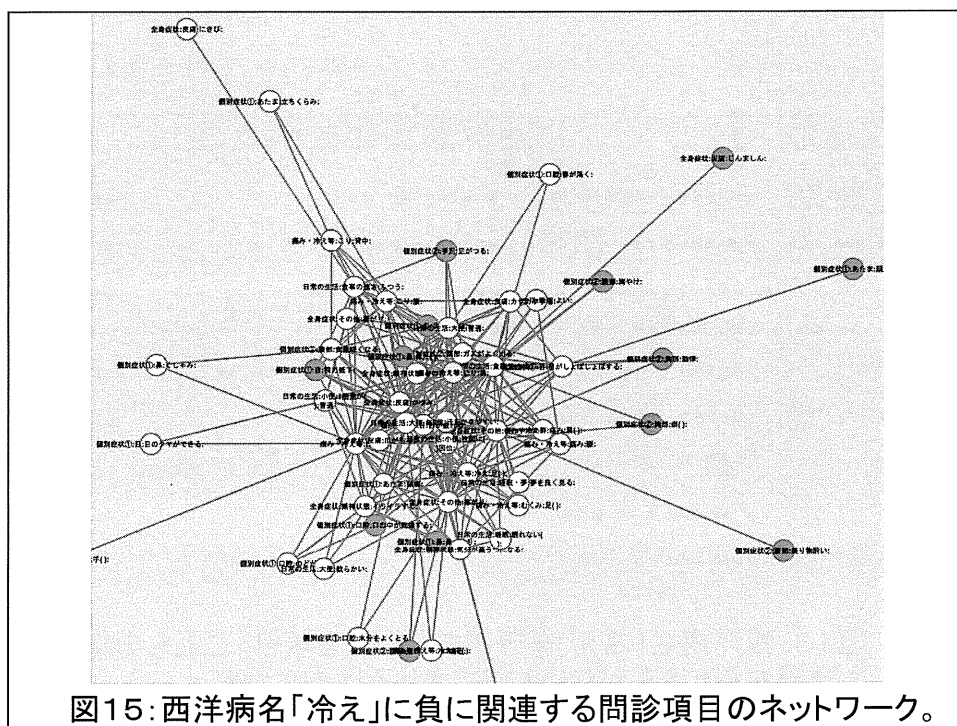
問診番号			問診番号				係数	Support		
148	痛み・冷え等	冷え	全身	0			0.897698	0.178394		
2	日常生活	食欲	ふつう	0	160	痛み・冷え等 冷え 腰	0.621715	0.108978		
6	日常生活	食事の速さ	速い	0	50	全身症状 精神状態気分が冴う 眠 つになる	0.483668	0.101171		
54	全身症状	精神状態	イライラする	0	206	個別症状① 目 目のクマが できる	0.466241	0.132292		
5	日常生活	食事の速さ	ふつう	0	155	痛み・冷え等 冷え 足()	204	個別症状① 目 目がしょぼしょぼする	0.44091	0.116996
155	痛み・冷え等	冷え	足()	0			0.430167	0.505117		
38	日常生活	大便	普通	0	74	全身症状 その他 汗をかきや すい	82	全身症状 その他 寒がり	0.388204	0.121848
22	日常生活	小便	1回量が()	普通	38	日常生活 大便 普通	82	全身症状 その他 寒がり	0.366116	0.136934
200	個別症状①	目	目が疲れる	0	282	個別症状② 手足 足がふら つく			0.324017	0.106657

図12: 西洋病名「冷え」を予測できる問診項目。これらの項目に該当すると冷えがあると判定される傾向になる。

		問 診 番 号			問 診 番 号			問 診 番 号	係数	support			
日常生活	小便	夜間に ()回位	054全身症状	精神状態	イライラする	227個別症状①	口腔	のどが渇く	-0.61414	0.104758			
日常生活	食欲	ふつう	018日常生活	小便	夜間に()回 位	70全身症状	皮膚	爪がもろい	-0.44703	0.121215			
日常生活	小便	1回量が ()	普通	52全身症状	精神状態	ものを忘れる	54全身症状	精神 状態 イライラする	-0.43602	0.100433			
日常生活	大便	毎日	038日常生活	大便	普通	186個別症状①	あたま	頭痛	-0.38	0.101382			
日常生活	食欲	ふつう	072全身症状	その他	疲れやすい	74全身症状	その他	汗をかきやす い	個別 268症状 ②	腹部	ガスがよく出 る	-0.37339	0.108134
日常生活	小便	夜間に ()回位	052全身症状	精神状態	ものを忘れる	204個別症状①	目	目がしょぼ しょぼする	-0.33181	0.125119			
痛み・冷え等	こり	首	10痛み・冷え 8等	痛み	肩()	155痛み・冷え等	冷え	足()	個別 200症状 ①	目	目が疲れる	-0.30789	0.107184
全身症状	精神 状態	イライラ する	0 ²¹ ₅ 個別症状①	鼻	鼻づまり							-0.30763	0.111404
日常生活	睡眠・ 夢	夢を良く 見る	038日常生活	大便	普通							-0.2869	0.159511

図13: 西洋病名「冷え」を予測できる問診項目。これらの項目に該当すると冷えがないと判定される傾向になる。





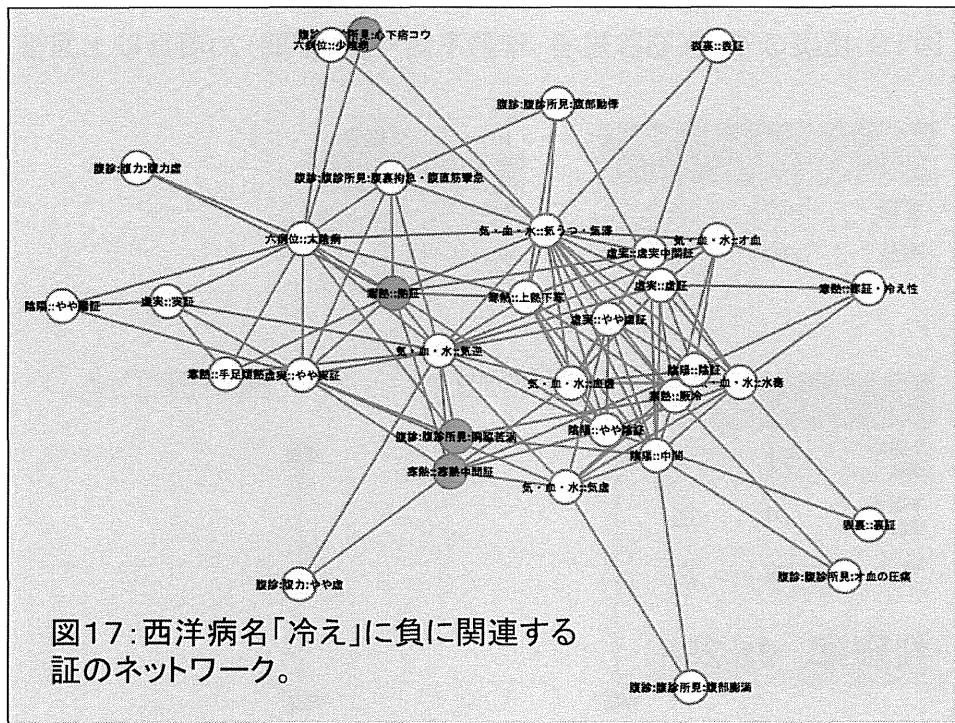


図18: 判別方式の性能比較

ロジスティック判別モデル

55.70%	検証テスト	実証テスト
予測で 虚証	102	19
予測で 実証	74	15
計	176	34

SVM

72.85%	虚証テスト	実証テスト
予測で 虚証	136	17
予測で 実証	40	17
計	176	34

Penalized SVM

67.60%	虚証テスト	実証テスト
予測で 虚証	123	15
予測で 実証	53	19
計	176	34

CART

60.00%	虚証テスト	実証テスト
予測で 虚証	105	14
予測で 実証	71	20
計	176	34

Elastic net without VAS

68.57%	虚証テスト	実証テスト
予測で 虚証	126	16
予測で 実証	50	18
計	176	34

Elastic net normalized VAS

75.71%	虚証テスト	実証テスト
予測で 虚証	135	10
予測で 実証	41	24
計	176	34

図19: 虚実の予測 (初診患者・年齢不問・性別不問・20項目以上回答)

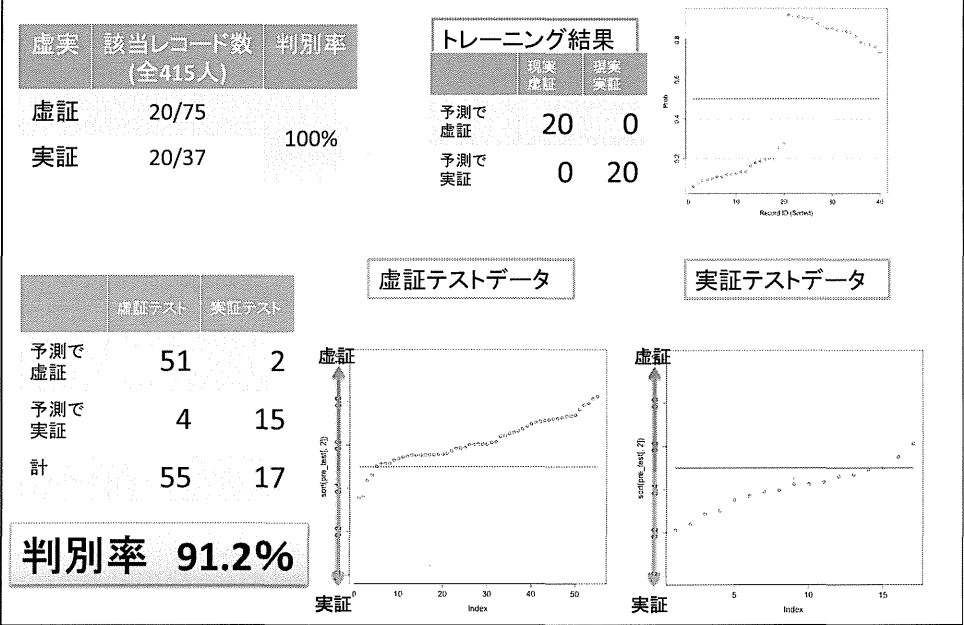
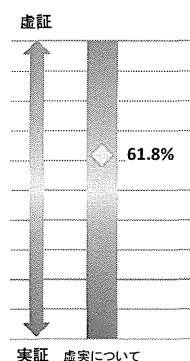


図20: 患者番号3368 問診データからの証(虚実、寒熱)予測 (NHK名医にQより)

6 日常生活:食事の速さ:速い:	1.00	
10 日常生活:睡眠:眠れない():VAS	0.26	
12 日常生活:睡眠:眠れない():VAS	0.71	
14 日常生活:睡眠:眠れない():VAS	0.38	
38 日常生活:大便:普通:	1.00	
53 全身症状:精神状態:ものを忘れる:VAS	0.73	
69 全身症状:皮膚:水虫:VAS	0.64	
75 全身症状:その他:汗をかきやすい:VAS	0.32	
79 全身症状:その他:のぼせ:VAS	1.00	
83 全身症状:その他:寒がり:VAS	0.55	
91 痛み・冷え等:こり:肩:VAS	0.88	
154 痛み・冷え等:冷え:手():VAS	0.58	
159 痛み・冷え等:冷え:足():VAS	0.96	
161 痛み・冷え等:冷え:腰:VAS	0.78	
163 痛み・冷え等:ぼてり:顔:VAS	0.97	
203 個別症状①:目:目がかすむ:VAS	0.82	
205 個別症状①:目:目がしょぼしょぼする:VAS	0.55	
230 個別症状①:口腔:口の中が乾燥する:VAS	0.96	
232 個別症状①:口腔:唇が渇く:VAS	0.80	
269 個別症状②:腹部:ガスがよく出る:VAS	0.90	
271 個別症状②:腹部:食後眠くなる:VAS	0.66	
279 個別症状②:手足:手がこわばる:VAS	0.34	
287 個別症状②:手足:しもやけができる:VAS	0.33	
BMI	BMI	20.08

図21: 患者番号3368、虚実の判定



虚証患者20、実証患者20のデータにより構築した
虚実予測方式により番号3368の患者の虚実を
予測する。

結果
虚証度61.8%

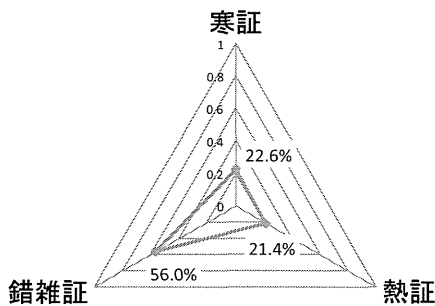
結論
中間証からやや虚証の傾向があると予測される。

先生の診断

虚 実 中 間

NHK 名医にQより

図22: 患者番号3368、寒熱の判定



寒証患者70人、熱証患者70人、
錯雑証患者40人のデータにより
構築した寒熱予測方式により
番号3368の患者の寒熱を予測する。

結果
寒証度22.6%
熱証度21.4%
錯雑証度56.0%

結論
錯雑証の傾向があると予測される。

先生の診断

寒 熱 (上熱) 下寒

NHK 名医にQより

図23:施設別虚実判定数

	亀田	自治医科	富山	飯塚	慶応
虚証	4 (2.8%)	1 (6.7%)	17 (17%)	18 (13.8%)	75 (18.7%)
やや虚証	64 (45.0%)	3 (20%)	45 (45%)	46 (35.4%)	28 (7.0%)
中間証	61 (43.0%)	8 (56.3%)	28 (28%)	51 (39.2%)	223 (55.5%)
やや実証	12 (8.5%)	3 (20%)	9 (9%)	13 (10%)	39 (9.7%)
実証	1 (0.7%)	0 (0%)	1 (1%)	2 (1.5%)	37 (9.2%)
合計	142	15	100	130	402

図24:施設別 虚実判定

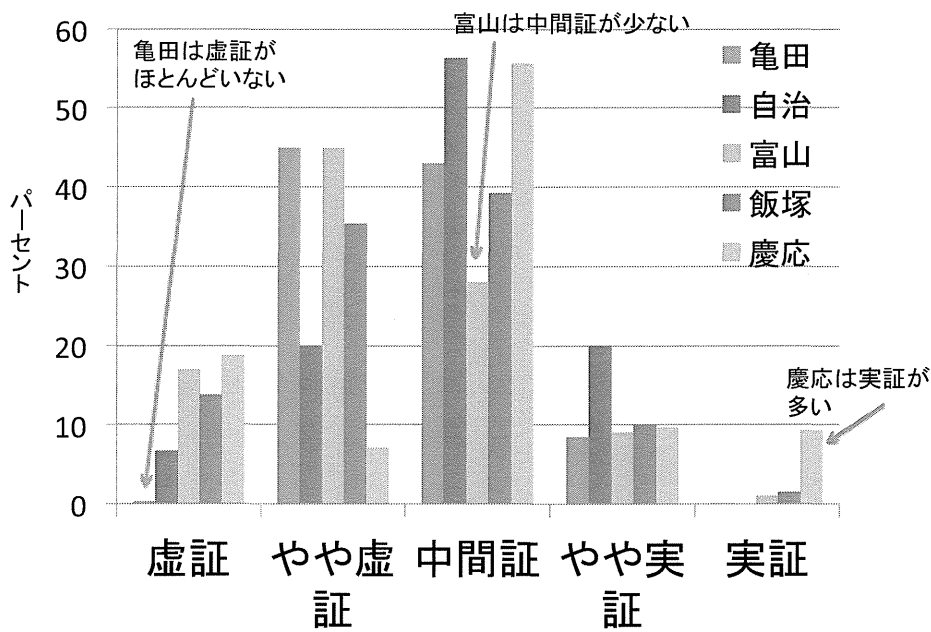


図25:虚実予測結果 正解率

亀田総合病院

虚実	実際の人数	正しく判定された人数	正解率
虚証	4	3	75
やや虚証	64	38	59.3
中間証	61		
やや実証	12	11	91.7
実証	1	1	100

富山大

虚実	実際の人数	正しく判定された人数	正解率
虚証	17	13	76.2%
やや虚証	45	25	55.6%
中間証	28		
やや実証	9	7	77.8%
実証	1	1	100%

麻生飯塚

虚実	実際の人数	正しく判定された人数	正解率
虚証	18	13	72.2%
やや虚証	46	30	65.2%
中間証	51		
やや実証	13	11	84.6%
実証	2	2	100%

自治医

虚実	実際の人数	正しく判定された人数	正解率
虚証	1	1	100%
やや虚証	3	2	66.7%
中間証	8		
やや実証	3	3	100%
実証	0		

問診データ解析、及び可視化による効率的な処方支援に関する研究

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

本研究は漢方医療における問診と診断・処方データをデータマイニング手法によって解析し、得られた問診項目と証診断・処方との関連性を可視化して、医師による漢方薬処方を効率的に支援するシステムを構築することを目的とする。本研究の可視化による処方支援は、予測対象のデータと類似した参照データを分類・提示することで診断・処方の判断を補助することに特徴がある。平成 22 年度、平成 23 年度においては蓄積されつつあった問診データの解析と、証を軸としたデータの可視化手法を開発し、平成 24 年度においてその可視化技術を用いた処方支援システムの構築を行った。支援ツール実現においては、データ分析のための可視化インターフェースから処方候補の提示という目的に特化したインターフェースを設計、実装を行った。試験運用によるフィードバックから改善を重ね、目的特化型処方支援システムを実現した。

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、東京女子医科大学の漢方専門外来に加えて、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データマイニング手法によって解析し、得られた問診項目と証診断・処方との関連性を可視化して、医師による漢方薬処方を効率的に支援するシステムを構築することを目的とする。

本研究の可視化による処方の支援は、予測・推薦といった形での支援とは異なり、対象のデータと類似した参照データを分類して提示することに特徴がある。その利点としては診断・処方の判断根拠となるデータや二次

選択薬などの選択肢を示すことができる点があり、これらは診断・処方を直接推薦する予測方式と補完しあう関係にある。可視化による支援を実現するためにはまず、単に処方によって問診データを分けるのではなく、似た状態の患者、つまり似た問診内容を持つ問診データを自動的に判別、分類する手法が重要となる。また利用者として想定している医師が一人の患者に避ける時間といった制約を考慮すると、効率的に、つまり少ないクリック数で、判断に十分なデータを提示するというシステムのインターフェース面も重要である。

平成 22 年度、23 年度においては、蓄積されつつあった問診データの解析を行うとともに、証を軸とした問診データの類似関係の可視化手法の開発を目指した。平成 24 年度において

は、それらの可視化手法を用いた処方支援システムの構築を目指した。つまり、既存の可視化システム MIMA サーチ上で証を軸として問診データを可視化し、さらにデータ分析のためのインターフェースから処方候補の提示という目的に特化したインターフェースを提案するとともに、実装を行うこととした。

B. 研究方法

本研究では、集められた問診データをデータマイニング手法によって解析し、得られた問診項目と証診断・処方との関連性を可視化することにより、医師による漢方薬処方を効率的に支援するシステムの構築を目的としている。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、漢方専門医以外による漢方薬処方限定された形で行われているのが現状である。これは、漢方診療が本来同病異治・異病同治に示されるような個別化診療であり、漢方医学的診断の特徴である「証」の特定が Know-How のような経験知から導かれるものであることが主な原因である。証と紐付いた効果的な処方が広く行われるには、証の科学的・統計学的な解明や、Know-How に係る経験的、暗黙的知識の「見える化」が貢献できるであろう。本研究では後者の経験的、暗黙的知識の「見える化」による処方支援システムの構築を目指す。このようなシステムの実現には、1) 薬剤の処方だけでなく漢方証診断と密接に結びついた可視化手法と 2) 目的の情報（本件の場合、適切な数の処方候補とその特徴比較）へ効率的に到達できるインターフェースが必要である。平成 22 年度、23 年度においては、蓄積されつつあった問診データの解析を行い 1) に関して証を軸とした問診データの類似関係の可視

化手法を開発した。平成 24 年度においては、1)の可視化手法を用いた処方支援システムを実装、2) の目的特化型インターフェースを実現し、さらに試験運用によるフィードバックにもとづいて改善を行い、システムを構築した。

以下ではまず、可視化による診断・処方支援の意義と特徴について可視化の概念と漢方データの特徴から述べる。また本研究の支援システム構築に利用した既存の可視化システム MIMA サーチについても説明する。次に 22, 23 年度に行った診断・処方支援に有用な可視化を行うためのデータ処理について述べ、続いて本年度の支援システムの構築の詳細として、インターフェースに求められる条件と、それらに基づいたデザインを提示する。またシステム実装のモデルに関しても説明を行う。

可視化とは

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の活用の際し、問題点として以下を考える。

- ・ 情報過多、知識過多
- ・ 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

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。

- ・ 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。

漢方問診データと可視化

データマイニングの目的は、膨大なデータを対象とし、

- 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、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、生命科学に関連して既にある例として、a)に対してはバイオインフォマティックス等、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