

厚生労働科学研究費補助金

臨床応用基盤基礎研究事業

主観的個別化患者情報のデータマイニングによる
漢方・鍼灸の新規エビデンスの創出に関する研究

平成20年度 総括研究報告書

研究代表者 渡辺 賢治

平成21(2009)年 4月

目 次

I. 総括研究報告

- 主観的個別化患者情報のデータマイニングによる1
漢方・鍼灸の新規エビデンスの創出に関する研究
渡辺賢治

II. 分担研究報告

1. 自動問診システムからのデータ蓄積システムの構築4
西村 甲
2. 個別化診療情報解析システムの検討 I11
(初診患者情報からみた解析検討)
西村 甲
3. 個別化診療情報解析システムの検討 II13
(自動問診システムに蓄積された情報に基づく予後予測
解析の検討)
西村 甲
4. 個別化診療情報解析システムの検討32
(MIMA search を用いた情報可視化の検討)
美馬 秀樹
5. 鍼灸での自動問診システムの構築35
塚田 信吾
6. 東アジア伝統医学の用語集の日本語翻訳に関する研究42
石野尚吾

- II. 研究成果の刊行に関する一覧表43

主観的個別化患者情報のデータマイニングによる漢方・鍼灸の新規エビデンスの創出に関する研究

研究代表者 渡辺 賢治

慶應義塾大学 医学部 漢方医学センター センター長 准教授

研究要旨

漢方・鍼灸の新規エビデンスを創出し、漢方診療支援システムを構築することを目的とした。このために、既に開発している漢方自動問診システムの改良、問診システム開始以前のデータを用いた患者情報に関する基礎的調査、問診システムに集積されたデータによるデータマイニング手法の確立、鍼灸の自動問診システムの確立、伝統医学用語の翻訳について検討し、有意義な成績を得た。

1) これまでに使用してきた漢方自動問診システムについては、集積された患者の愁訴、漢方所見、西洋病名、漢方処方の変化を通して、より良いデータ収集、蓄積ができるよう、十分な改良がなされた。これにより、今後のデータマイニングが実施しやすくなり、診療支援プログラム作成に向けた検討が可能となった。

2) 問診システム開始以前の慶應義塾大学病院漢方クリニックの平成 17・18 年度の初診患者について、ある症状・所見に関して、随伴する症状・身体所見の特徴について検討した。経験的に把握されていた所見が、統計学的に確認することができた。

3) 漢方自動問診システムに集積されたデータによる予後予測に関する検討では、Apriori, Eclat、クラスター解析を用いて検討した。一定の症状・疾患における患者の特徴が解明された。今後、症例を増加させることで、統一されたルール作成が行えるようになることが望まれる。

4) 漢方自動問診システムに集積されたデータによるデータマイニングによる関連する症状所見の可視化の検討においては、頭痛に関する患者では、冷えや視力低下、月経異常等の患者群が示され、また冷えも同様に頭痛との関連や、月経との関連があることが確認された。

4) 鍼灸に関する問診システムの構築に関しては、鍼灸治療の臨床現場での状況を調査し、それを基に施術情報の記録を目的とした電子化システムの基本設計を行った。さらに漢方版の患者問診システムを鍼灸分野に適応するようカスタマイズを行った。臨床応用に向けて、システムの大幅な進歩が確認された。

分担研究者

西村 甲

慶應義塾大学医学部漢方医学センター講師

塚田信吾

日本伝統医療科学大学院大学・統合医療研究科
特任教授

美馬秀樹

東京大学大学院工学系研究科 准教授

石野尚吾

日本東洋医学会 会長

A. 研究目的

患者側からの主観的医療情報をベースに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、漢方・鍼灸治療の新たな臨床研究の手法を創出する。

東アジア伝統医学に対する関心の高まりは世界的に広がっている。わが国では、生薬薬物治療ならびに鍼灸治療を合わせ、漢方医学として独自の発展を遂げている。2007年8月、WHO西太平洋地区事務局では日本、中国、韓国が中心となって東アジア伝統医学の用語集を作成し出版した。このように東アジア伝統医学の標準化が進む中で、最も求められているのが臨床的エビデンスである。

西洋医学的手法である無作為割付臨床試験(RCT)は、漢方診療には馴染まない。その理由として個別化治療である、2) 主観を重視した医療である、3) 全人的治療である、ことが挙げられる。主観を重視する漢方診療では効果判定が患者サイドから為されるのが通常であり、そのような研究デザインが求められる。個人差が評価できる研究デザインを組むためには、個々のデータの集積化を行い、データマイニングの手法により解析する方法が最も適していると考えられる。また、漢方・鍼灸医学は全人的治療であることから、目標とする症状以外にも全身状態の改善度を評価する必要がある。このような観点から我々は、現在までに自動問診システムを開発し、データ集積を試みている。本システムは、患者愁訴の経時的変化を患者自身が入力することにより、症状の変化を患者側から評価できるシステムになっている。本データに加え、診療者からの「証」診断や、診察所見、薬方データを加えることで、1) 証の検証、2) 漢方・鍼灸の治療効果の評価、が可能となる。

漢方医学の治療効果の判定や、伝統医学的診断である「証」の決定の評価を、個別に、主観を軸とした評価をした研究デザインは存在しな

い。本研究は漢方医学のエビデンスを従来の手法とは全く異なる、伝統医療の特質を生かした手法にて評価しようとするものであり、国内外に例をみない。また、本システムは診療現場にて患者の情報が回収されていくため、RCTに比し、低コストで大規模の情報を集積・解析することが可能である。

米国 NIH は伝統医療を含む補完・代替医療の研究費を毎年 300 億円拠出している。こうした世界情勢を受け、中国、韓国を中心とした東アジア諸国では独自の伝統医療を世界に売り込むことに躍起である。わが国の漢方医学は西洋医学と医療システム的に統合されたユニークな存在であり、品質の安定性など優れた点を有し、科学的にも一番理解しやすい医学とされている。また、日本の鍼灸技術は世界でも最も優れているとの評価を得ている。漢方・鍼灸は、患者の主観を元に治療を構築するため、エビデンス創出には患者の主観的評価を取り入れる、という発想の転換が必要である。WHO 国際機能分類

(ICF) では患者中心の医療評価が求められているが、患者中心のエビデンスの創出は、単に漢方診療の世界だけではなく、幅広く医療現場に応用可能である。また、得られたデータから将来的に漢方・鍼灸治療の標準化ならびに診療支援が可能となる。

B. 研究方法

平成 20 年度においては、1. 自動問診システムからのデータ蓄積システムの構築、2. 個別化診療情報解析システムの検討 I (初診患者情報からみた解析検討)、3. 個別化診療情報解析システムの検討 II (自動問診システムに蓄積された情報に基づく予後予測解析の検討)、4. 個別化診療情報解析システムの検討 III (MIMA search を用いた情報可視化の検討)、5. 鍼灸での自動問診システムの構築、について検討した。また、6. 2007 年 8 月、WHO 西太平洋地区事務局で日本、中国、韓国が中心となって作成

された東アジア伝統医学の用語集の日本語翻訳を行い、臨床応用の利便化を図った。

C. 研究結果

1) 自動問診システムからのデータ蓄積システムの構築では、集積された患者の愁訴、漢方所見、西洋病名、漢方処方の変化を通して、問診項目、西洋医学病名、漢方病態の整備・整理、漢方処方入力のルール作成が必要であることが判明した。これらの問題点について、症状の追加、西洋病名の分類修正、処方入力のルールを作成した。

2) 個別化診療情報解析システムの検討Ⅰ（初診患者情報からみた解析検討）では、慶應義塾大学病院漢方クリニック初診患者 1691 名につき、冷えをもつ患者、漢方医学的診察、特に腹診所見に関して、特徴的所見を有する患者について、随伴する症状・身体所見の特徴について検討した。「冷えあり」と強い関係があった項目は「女性」「肩がこる」「疲れやすい」などであった。各腹候と強い関係があった項目についてみると、腹候がもつ病態とそれに対応する症状の関連性が確認された。

3) 個別化診療情報解析システムの検討Ⅱ（自動問診システムに蓄積された情報に基づく予後予測解析の検討）では、冷えがある患者は、胃腸炎、月経異常に分類されていくこと、アトピー性皮膚炎は、体力充実かつ熱性の病態、体力虚弱かつ寒性の病態などに分類されていくことが確認された。

4) 個別化診療情報解析システムの検討Ⅲ（MIMA search を用いた情報可視化の検討）では、頭痛に関する患者において、冷えや視力低下、月経異常等の患者群が示され、また冷えも同様に頭痛との関連や、月経との関連があることが確認された。

5) 鍼灸での自動問診システムの構築に関しては、鍼灸治療の臨床現場での状況を調査し、そ

れを基に施術情報の記録を目的とした電子化システムの基本設計を行い、データベースを構築した。日本の鍼灸の多彩な治療手技の特徴を生かしつつ、標準化し、資料性を保ちながら、出来る限り簡潔に記録できるよう配慮した。さらに漢方版の患者問診システムを鍼灸分野に適応するようカスタマイズを行った。

6) 東アジア伝統医学の用語集の日本語翻訳については、作業を終了し、日本語索引を加えた。

D. 考察

20 年度の検討により、自動問診システムのデータ収集、データマイニング処理に関わるデータ蓄積の方法が構築された。

データマイニングの基礎的検討により、経験的に知られてきた漢方医学の情報が、系統的に整理されることが判明した。これらから、漢方医学でいう証の検証、漢方・鍼灸の治療効果の評価の第一歩が確立された。

E. 結論

東アジア伝統医学の標準化が進む中で、最も求められているのが臨床のエビデンスである。このエビデンスの確立、証の検証、漢方・鍼灸の治療効果の評価、漢方・鍼灸治療の標準化ならびに診療支援について、基礎的な成果が確認された。これらをもとに、21 年度のより整備された環境でのデータマイニング作業を行っている。

F. 健康危険情報

該当なし

G. 研究発表

渡辺賢治：伝統医学に関する診療情報からのデータマイニングの試み，第7回医療所法研究会，2009.1.19 東京

H. 知的財産権の出願・登録状況

該当なし

自動問診システムからのデータ蓄積システム構築に関する研究

研究担当者 西村 甲
慶應義塾大学 医学部 漢方医学センター 講師

研究要旨

自動問診システムについては、集積された患者の愁訴、漢方所見、西洋病名、漢方処方の変化を通して、より良いデータ収集、蓄積ができるよう、十分な改良がなされた。これにより、今後のデータマイニングが実施しやすくなり、診療支援プログラム作成に向けた検討が可能となる。今回の検討では、問診項目の見直し、西洋医学病名の整理、漢方病態の整備、漢方処方において、頓服、試し処方、など様々な形で、通常の治療に付加される漢方薬があるが、その意図が明確になるよう処方入力のルールを作成を行った。今後、これらの修正を活かして、データ数の増加が期待される。

A. 研究目的

これまでの患者サイドにおける問診項目のタッチパネルを用いた入力に関して、下記の図(図2-1~図2-4)に提示する。これは、患者の主観的な症状・所見をvisual analogue scaleにより定量化したものととなる。これらに、診察した医師が患者の漢方医学的病態、西洋医学的病名と処方薬剤を入力する。これらが一度の受診によって蓄積されていくことになる。このような情報を解析するうえで、有効な形として蓄積されることが重要であり、その方法について検討する。

B. 研究方法

これまで、蓄積された自動問診システム上の問診項目・西洋医学的病名・漢方医学的病態・処方について、実際の症例を通して問題点を確認する。

C. 研究結果

1. 患者症状・所見の問診項目について、耳鳴りが欠落していた。
2. 皮膚症状に関して、十分な項目で満たされ

ていない。

3. 西洋医学病名が非常に多岐にわたるため、同じ病名を意図して入力しても、言葉としては、異なることがあった。例えば、アトピー性皮膚炎、アトピー性湿疹、アレルギー性皮膚炎、アレルギー性湿疹などである。
4. 漢方病態に関しては、おおまかな病態を捉えることを目標に入力内容を設定したため、処方変更が微細な漢方病態変化に基づく場合は、その処方変更意図が理解されないことがある。
5. 漢方処方において、頓服、試し処方、など様々な形で、通常の治療に付加される漢方薬があるが、その意図が理解されにくい。

D. 考察

患者問診内容について、追加、修正が必要である。西洋医学病名について、グルーピングを行い、大きなくくりを設定する必要がある。漢方病態については、このような問診システムを非漢方専門医が使用可能なものにするを予定しているため、変更を加えないこととする。処方内容については、主訴に対する内容のみ

について、入力する方針とすべきである。

E. 結論

自動問診システムの問題点の洗い出しができた。これらの問題点を解決することで、より精度の高い情報収集、解析が実行できると思われる。

F. 健康危険情報

該当なし

G. 研究発表

該当なし

H. 知的財産権の出願・登録状況

該当なし

『自動問診システムの機能一覧』

①患者側自動問診システム（タッチパネル）】

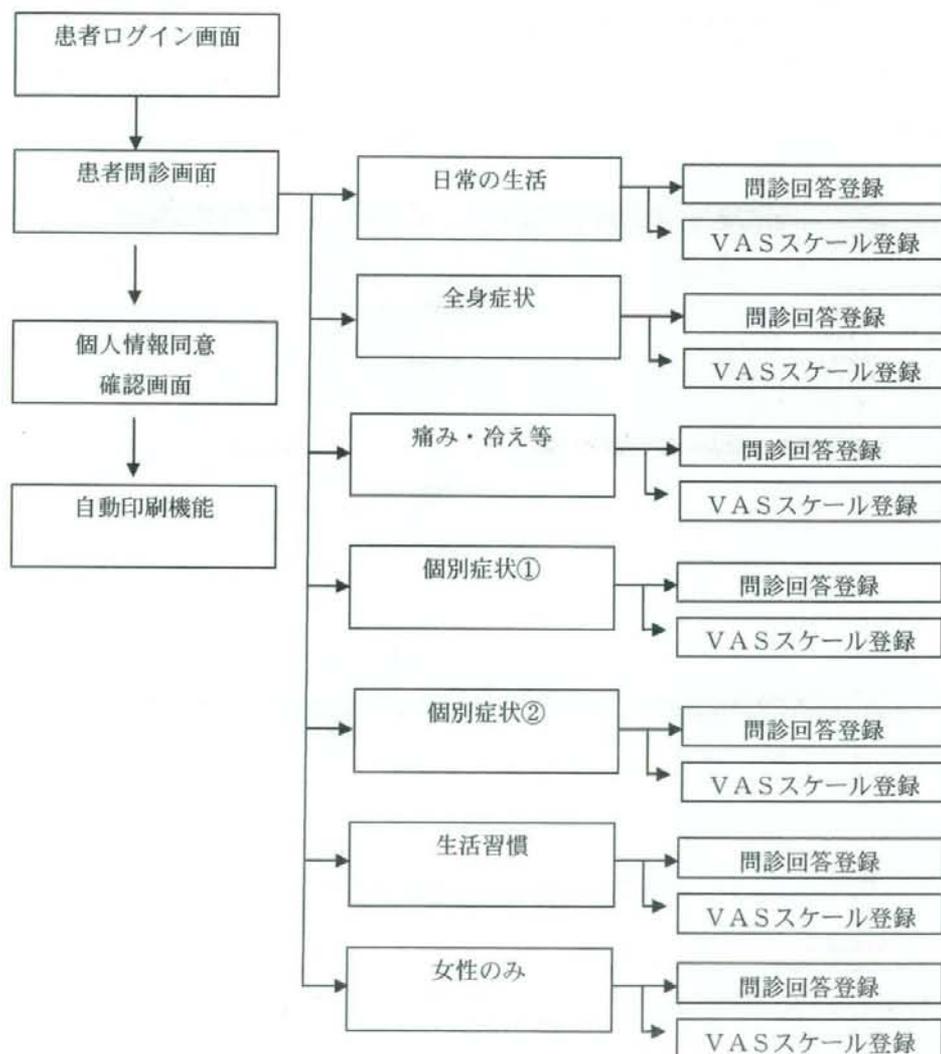


図 1-1 自動問診システムの機能一覧

【①患者側自動問診システム（サンプル画面）】

（患者ログイン）

漢方問診システム

診察券の番号を入力して確定ボタンを押してください。

診察券の番号

7	8	9
4	5	6
1	2	3
訂正	0	確定

※印刷環境がないと、自動的に最初の画面からやり直しになります。
ご注意ください。【中絶ボタンを押した時も同様です。】

（日常生活）

漢方問診システム

日常生活

【寒熱】 必必須項目
寒熱が強い ぶつう 寒熱旺盛

【寒熱の遷移】 必必須項目
寒い ぶつう 暑い

痛み・冷え等

【痛熱】 必必須項目
【寒熱】 寒くない() 暑く長く感じる

【小便】
【大便】
【小便】 1日に()回位 夜間に()回位 1回量が()
【大便】 排便困難 排便痛 屎もれ 排便度

女性のみ

【月経】
【月経】 痛月 ()日に()回位
【月経】 痛い ころころしている 普通 軟らかい
【月経】 下痛 出にくい 痛がある 量狂
【月経】 出血 下副を服用

中断 次へ

（全身症状）

（痛み・冷え等）

漢方問診システム

全身症状

【精神状態】
気分が憂うつになる ものを忘れる イライラする

【皮膚】
かさかさする かゆみ にきび しみ

痛み・冷え等

【寒熱】 じんましん いぼ 水虫 汗がもろい

【その他】
【寒熱】 寒れやすい 汗をかきやすい 喉痛 のぼせ
【その他】 腫がり 脹がり 性飲の減退 インボナンツ

【女性のみ】
※乳になる症状があればボタンを押して入力して下さい
【女性のみ】 乳房腫れ 乳が痛い 乳房硬くなる

中断 戻る 次へ

漢方問診システム

痛み・冷え等

【心身】 寒 寒 背中 熱

【痛み】
【痛み】 熱 手() 足() 肩()
【痛み】 背中 膝 膝()

【しびれ】
【しびれ】 熱 手() 足() 背中

【冷え等】
【冷え等】 熱 手() 足()

【手足】
【手足】 全身 熱() 寒() 熱
【手足】 熱 手() 足()
【手足】 熱 手() 足()
【手足】 熱 手() 足()

中断 戻る 次へ

（個別症状①）

（個別症状②）

漢方問診システム

全身症状

【おたまり】
【おたまり】 積痛 積寒 必まい 立ちくらみ
【おたまり】 ぶけがやすい 腫がぬけやすい

痛み・冷え等

【目】
【目】 視力低下 目が腫れる 目がかすむ

痛熱

【目】 目がしょぼしょぼする 目のクマができる

【鼻】
【鼻】 くしゃみ 鼻汁() 鼻汁がのどにおりる
【鼻】 鼻づまり 鼻血

【口咽】
【口咽】 口が強い 生唾ができる のどが痛む のどがつかえる
【口咽】 のどが乾く 口の中が乾燥する 舌が高く 水分をふよとる

【舌】
【舌】 舌なり 数割

中断 戻る 次へ

漢方問診システム

痛熱

【動脈】
【動脈】 痛 痛() 痛重 痛切れ

全身症状

【動脈】
【動脈】 動悸 胸痛

痛み・冷え等

【動脈】
【動脈】 ゴック 胸やけ みぞおちがつかえる 嘔気

痛熱

【動脈】 嘔吐 走り物酔い 腹が痛む 腹がゴロゴロ鳴る

【動脈】
【動脈】 ガスがよく出る 食欲減くなる 腹痛()

【手足】
【手足】 手がこわばる 足に力がはいらない 足がふらつく
【手足】 足がつる しもやけができる

中断 戻る 次へ

(生活習慣)

遠方問診システム
大塚 太郎様

日常生活	【好きな食べ物】
食生活	甘いもの 塩辛いもの 辛いもの すっぱいもの
痛み・冷え等	ほっこりもの 冷たいもの 温かいもの 肉
個別症状①	魚介類 煮物 刺身 漬物
個別症状②	【嗜好品】
タバコ	アルコール 飲まない 飲む、週に()日
女性のみ	タバコ 吸わない 吸っている()才()才
	吸っている()才()本/日
	コーヒー 紅茶 日本茶 その他 ()日に()杯
	【家族構成】
	未婚 既婚
同居人	配偶者 父 母 祖父 祖母
	兄弟 姉妹 子供 その他
	中断 戻る 次へ

(女性のみ)

遠方問診システム
大塚 太郎様

日常生活	【妊娠】
食生活	妊娠()才 産前()才 産後()才()年()月
痛み・冷え等	月経周期()日 出血期間()日 出血量()
個別症状①	おりもの 月経痛
個別症状②	【妊娠の可能性】
個別症状③	なし あり
生活習慣	【避妊手段】
女性のみ	なし あり
	【その他】
	ピル使用 分()回 自然授乳()回
	人工授乳()回 妊娠中薬歴 不正出血
	中断 戻る 次へ

図 1-2 患者側自動問診システム (サンプル画面)

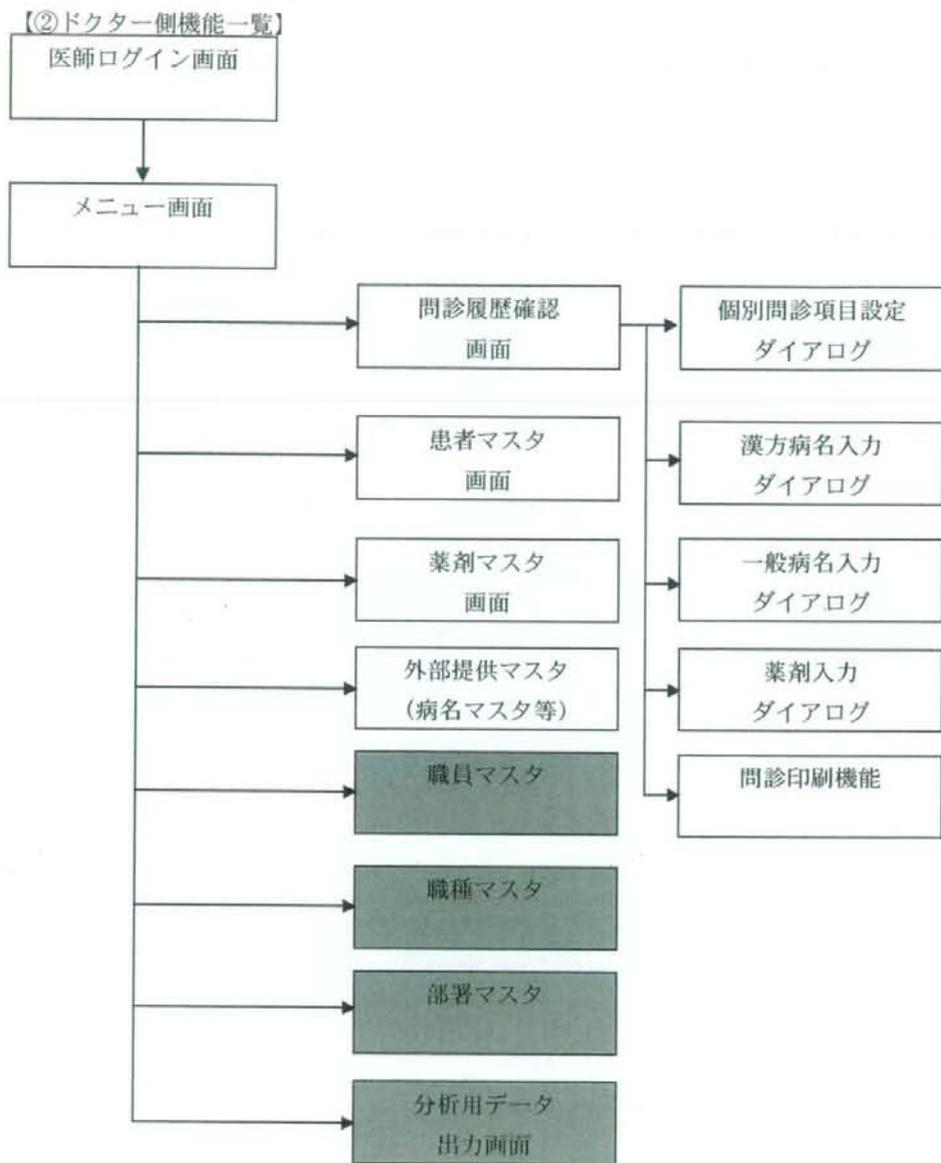
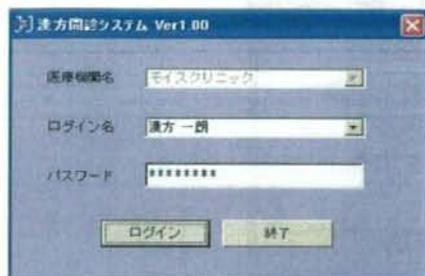


図 1-3 ドクター側機能一覧

【②ドクター側機能一覧（サンプル画面）】
（医師ログイン）



（履歴確認画面：問診）

（履歴確認画面：漢方病名・証）



（履歴確認画面：一般病名）

（履歴確認画面：薬剤情報）

(分析用データ出力)

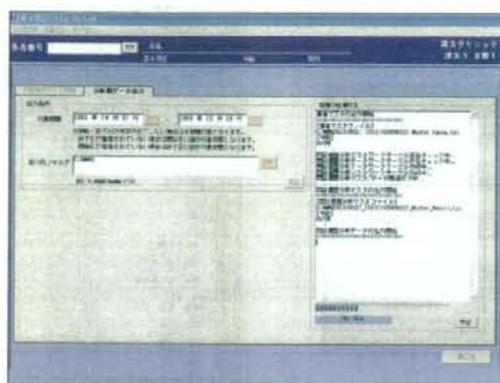


図 1-4 ドクター側機能一覧 (サンプル画面)

個別化診療情報解析システムに関する研究Ⅰ：初診患者情報からみた解析検討

研究担当者 西村 甲
慶應義塾大学 医学部 漢方医学センター 講師

研究要旨

慶應義塾大学病院漢方クリニック初診患者 1691 名につき、冷えをもつ患者、漢方医学的診察、特に腹診所見に関して、特徴的所見を有する患者について、随伴する症状・身体所見の特徴について検討した。「冷えあり」と強い関係があった項目は「女性」「肩がこる」「疲れやすい」「首がこる」「不眠」「脈沈」などであった。各腹候と強い関係があった項目についてみると、虚証でみられる腹証については、「肩こり」「冷え」「疲れやすい」「不眠」と関連性が高かった。実証でみられる腹証については、「肥満」「いぼ」「ほてる」「物忘れ」「皮膚がカサカサする」「鼻血」「目がかすむ」と関連性が高かった。虚証、実証どちらでも比較的良好にみられる腹証については、「肩がこる」「疲れやすい」「脈沈」と関連性が高かった。経験的に把握されていた所見が、統計学的に確認することができた。

A. 研究目的

自動問診システムにおける問診データの集積、解析方法を確立する際に、漢方診療の特徴を踏まえて検討することが重要になる。漢方外来を受診する患者は血液データ、画像検査所見、など数値化あるいは、それに近似できるデータよりも主観的な症状に関心があり、その症状の改善を希望することが多い。まず、初診患者から得られた情報を基に各種症状・所見がどのように関連しているかを把握することは、解析をすすめる上で重要な検討である。

B. 研究方法

慶應義塾大学病院漢方クリニックを平成 17 年度、18 年度初診患者 1691 名につき、データベース化しているが、漢方医学で比較的重要な症状である冷えと頭痛について解析方法を検討する。SPSS Japan Inc のデータマイニング・ソフトである Clementine を用いて、冷えをもつ

患者、漢方医学的診察、特に腹診所見に関して、特徴的所見を有する患者について、随伴する症状・身体所見の特徴について検討した。

C. 研究結果

「冷えあり」と強い関係があった項目の上位大部分は異常なし所見であったが、陽性所見の上位 10 項目は「女性」「肩がこる」「疲れやすい」「首がこる」「不眠」「脈沈」「目が疲れる」「歯痕あり」「舌下静脈怒張あり」「寒がり」であった。また痛みに関しては肩・首がこる、目が疲れるに加えて「背部痛」「頭痛」「月経痛」「腰痛」といった項目に比較的高い関係性が認められた。また男女別に解析を行うと女性では 20 歳代をピークとするのに対して男性では年齢を追うごとに増加する傾向があった。

各腹候と強い関係があった項目の上位大部分は異常なし所見であったが、虚証、実証、どちらもみられる腹候という観点から各々に共通な陽性

所見及び症状を挙げる。虚証で比較的良好にみられる腹候として、「腹直筋攣急」、「振水音」、「小腹拘急」、「腹力虚」、「小腹不仁」については、「肩こり」「冷え」「疲れやすい」「不眠」と関連性が高かった。実証で比較的良好にみられる腹候として、「腹力実」については、「肥満」「いぼ」「ほてる」「物忘れ」「皮膚がかさかさする」「鼻血」「目がかすむ」と関連性が高かった。虚証、実証どちらでも比較的良好にみられる腹候として、「心下痞硬・心下堅」、「胸脇苦満」、「腹部動悸」、「腹満」を考えると、「肩がこる」「疲れやすい」「脈沈」と関連性が高かった。

腹候と他の所見との関連性をみると、「脈沈」と「舌下静脈怒張」が関連性が高かった。

D. 考察

いくつかの症状・所見と他の症状・所見の関連性が確認された。このような情報を逐次蓄積していくことで、データマイニングによる診療における患者状態の予測に役立っていくものと思われる。

E. 結論

これまででも、漢方所見と症状を関連性が指摘されていたが、これらの報告と本研究成果とを結びつけて、解析に活かしていく事が重要となる。

F. 健康危険情報

該当なし

G. 研究発表

該当なし

H. 知的財産権の出願・登録状況

該当なし

個別化診療情報解析システムに関する研究Ⅱ：自動問診システムに蓄積された情報に基づく予後予測解析の検討

研究担当者 西村 甲
慶應義塾大学 医学部 漢方医学センター 講師
研究協力者 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 准教授

研究要旨

予後予測に関する検討では、自動問診システムにより収集・蓄積されたデータを、Apriori, Eclat、クラスター解析を用いて検討した。冷えがある患者は、胃腸炎、月経異常に分類されていくこと、アトピー性皮膚炎は、体力充実かつ熱性の病態、体力虚弱かつ寒性の病態などに分類されていくこと、が判明した。予後予測に関するルール作成は、データをもとにその特徴をつかみ、データとその特徴に応じてルール作成を行うため、個々のデータの特徴を把握することが非常に重要になる。さらに、症例を増加させることで、統一されたルール作成が行えるようになることが望まれる。

A. 研究目的

自動問診システムにおける問診データの集積、解析方法を確立する際に、漢方診療の特徴を踏まえて検討することが重要になる。漢方外来を受診する患者は血液データ、画像検査所見、など数値化あるいは、それに近似できるデータよりも主観的な症状に関心があり、その症状の改善を希望することが多い。このため、患者の症状の特徴を把握することは、症状変化を捉えること、治療選択の妥当性、治療方法の見通しなどを行ううえで、重要と考えた。

B. 研究方法

平成20年5月から運用を開始した患者自動問診システムにより、集積された患者データをもとに、受診総延べ数について、「冷え」症状を有する患者群の特徴、アトピー性皮膚炎と診断された患者群の特徴について、検討した。

C. 研究結果

まず、問診データなしで行える解析について、行った。

病名、漢方病名、処方された薬の3つ組の情報が各患者に関して得られるので、高頻度に出現する(病名、漢方病名、薬)の3つ組みはデータマイニングの方法である Apriori を用いれば解析できることになる。当たり前の情報しか得られないかもしれないが、常識の確認、意外な非常識が得られる可能性もある。「漢方のエキスパートではない医師が漢方薬を処方する時に助けになる情報」を作る第一歩になるともいえる。

まず、Apriori と Eclat を行うことにした。

バスケット分析として、 $A \rightarrow B$ というルールを設定した場合、例えば、漢方病名 A と診断された人は漢方病名 B と診断される、というルールとすることができる。support ($A \rightarrow B$) は、ルール「 $A \rightarrow B$ 」のサポートを意味して、support (A

$A \rightarrow B = s(A \cap B) / M$, $s(A \cap B)$ は A と B を同時に診断された患者数、M は患者の総数として定義される。 $\text{confidence}(A \rightarrow B)$ は、ルール「 $A \rightarrow B$ 」の信頼度を意味して、 $\text{confidence}(A \rightarrow B) = \text{support}(A \rightarrow B) / \text{support}(A) = \text{Pr}(B|A)$ となる。 $\text{lift}(A \rightarrow B)$ は、ルール「 $A \rightarrow B$ 」のリフトを意味して、 $\text{lift}(A \rightarrow B) = \text{confidence}(A \rightarrow B) / \text{support}(B) = \text{Pr}(B|A) / \text{Pr}(B) = M \cdot s(A \cap B) / s(A) \cdot s(B)$ となる。これは、条件部を空とした場合との信頼度の比率をあらわすことになり、リフト値が小さければ、結論の B は何らかの理由により単独で診断され、条件 A との関連は薄い。リフト値が 1 よりも大きければ何らかの意味のあるルールであることが期待される。

Eclat はルールではなく、高頻度 (support が大きい) に現れる病名の集合を抽出するものである。

問診システムに蓄積された西洋病名では、1029 レコードに対して 277 西洋病名、漢方病名に関しては、1001 レコードに対して 46 漢方病名、薬剤に関しては、1036 レコードに対して 154 薬剤+処方量であったが、処方量については、今回は除外して検討した。このうち、3 つのデータセットに共通するレコードを抽出して、953 レコードを解析対象にした。Apriori を用いて解析するにはレコード数があまり多くなかったため、同じ患者が複数回来院しているレコードは区別して用いた。同じ患者のデータをまとめると、3 つのデータセットに共通の患者の数は 387 例であった。患者往診歴は図 4-1 の通りであった。漢方病名が入力されていないものは、来院数にカウントされないように設定したため、漢方病名を持たない患者は来院数が 0 になっている。

漢方病名診断頻度

漢方病名診断頻度は図 4-2 の通りであった。

漢方病名の Apriori

漢方病名だけに Apriori を適用し、ルールを抽出した。support > 0.05 で 67 ルールが抽出された。複数の漢方病名を診断されることが多かった。図 4-3 に示す。トップ 15 は、図 4-4 の通りであった。

西洋病名の診断頻度

西洋病名の診断頻度は図 4-5 の通りであった。

西洋病名の Apriori

西洋病名だけに Apriori を適用しルールを抽出した。support > 0.05 ではルールを抽出できなかった。複数の西洋病名を診断されているケースが少ないと考えられた。

support > 0.01 に緩めて一つだけルールが抽出された。(腹痛症→下痢症) (support 0.0126, conf 0.923, lift 43.98) であった。support > 0.005 で 82 ルールが抽出された。ルールのトップ 30 は図 4-6 の通りであった。

漢方病名は複数で診断されることが多く、反対に西洋病名は単独で診断されることが多いことは、大きな違いであると考えられた。

薬剤の処方頻度

薬剤の処方頻度については、図 4-7 の通りであった。

薬剤の Apriori

薬剤だけに Apriori を適用し、ルールを抽出した。support > 0.05 ではルールを抽出できなかった。support > 0.01 に緩めてみると、二つだけルールが抽出できた。(ウチダのブクリョウ M→ウチダのシャクヤク M) (support 0.0105, confidence 1.000, lift 63.53), (ツムラ柴朴湯エキス→ツムラ桂枝茯苓丸エキス) (support 0.0105, confidence 0.833, lift 13.46) であ

った。西洋病名と同じく複数処方が少ないことが推測された。support > 0.005 で 203 ルールが抽出され、トップ 25 は図 4-8 の通りであった。

漢方病名&西洋病名

漢方病名と西洋病名の二つのデータ行列をつなげたデータ行列に対して、Apriori を適用した。support > 0.05 で、漢方病名だけから Apriori で抽出された (図 2-4) 67 ルールに加え、1 ルールのみ増えただけであった。support > 0.005 で 13,597 ルールとなり、多すぎて評価不能であった。有効な可視化法により情報抽出は可能かもしれない。そこで、西洋病名において、最も多くの患者に診断されていた「アトピー性皮膚炎」(117 レコード) にまずは着目した。結論部が「アトピー性皮膚炎」であるルールを上記の 13,597 のルールから抽出した。これにより、29 ルールを抽出された。これらについて、クラスター解析を行うと図 4-9 の通りであった。

漢方病名、西洋病名、薬剤それぞれの解析から

漢方病名は、基本的に複数の病名である患者の状態を記述しているように見られる。一方、西洋病名、薬剤は、患者一人につき一種で対応している場合が多い。従って、漢方病名、西洋病名、薬剤を単純に並べて、Apriori を適用しても、漢方病名の中だけから高頻度のルールが抽出される可能性が高い。実際、3 つ並べて support > 0.05 では、ルールは 69 種、漢方病名だけでも 67 ルールなので、二つしか増えない。目的の「漢方病名、西洋病名、薬剤」の組み合わせをえるためには support を低くせざるを得ない。

西洋病名から対応する漢方病名、薬剤へのマッピング

(西洋病名、漢方病名、薬剤)の三つ組みのルールを直接つなげたデータセットから単純な Apriori の適用で抽出することはできなかった。そこで、まず、ある西洋病名に着目し、その西洋病名を診断された患者に絞り、その患者に診断された漢方病名、処方された薬剤の情報を抽出することにした。例えば、西洋病名から対応する漢方病名のマッピングは図 4-10 のようになり、矢印→の部分の情報を抽出する。

アトピー性皮膚炎 (117 レコード in 西洋病名)

アトピー性皮膚炎 117 レコードにおける漢方病名の診断頻度をみてもみると、40 レコード以上に診断されている漢方病名は、図 4-11 の通りであった。前述の西洋病名の Apriori (図 3-6) による 82 ルールには、アトピー性皮膚炎は含まれていなかった。つまり、アトピー性皮膚炎は、単独に診断されることが多い。

「アトピー性皮膚炎」特異的漢方病名マイニング

アトピー性皮膚炎と診断されたレコードに高頻度に診断される漢方病名で、アトピー性皮膚炎と診断されなかったレコードにはあまり診断されなかった、もしくはその逆のものをマイニングした。結果は図 4-12 の通りであった。

Odds Ratio > 1 の場合、アトピー性皮膚炎と診断されたレコードがアトピー性皮膚炎と診断されなかったレコードより、有意にある漢方病名を診断されていることを示し、< 1 の場合は逆に、アトピー性皮膚炎と診断されたレコードがアトピー性皮膚炎と診断されなかったレコードより、有意にある漢方病名を診断されなかったことを示している。

冷え症 (97 レコード in 西洋病名)

97 レコード における漢方病名の診断頻度をみても、40 レコード 以上に診断されている漢方病名は図4-13に示す5つであった。

西洋病名において冷え症を含む高頻度病名サブセットのクラスタリング

ルールではなく、西洋病名データに含まれる冷え症を含む高頻度西洋病名のサブセットをEclat アルゴリズムにより抽出した。冷え症を含むサブセットで support > 0.005 のものは22個であった。クラスタリングの結果は図4-14の通りであった。

アトピー性皮膚炎特異的漢方病名集合マイニング

117 レコードが西洋病名データにおいて「アトピー性皮膚炎」と診断されていた。まず、漢方病名の高頻度集合について、Eclat を用いて抽出した。高頻度集合は、全レコードのデータから抽出した。最終的にはアトピー性皮膚炎レコードにおいて高頻度に出現し、アトピー性皮膚炎ではないレコードには現れないような漢方病名集合を抽出したいので、そのためには117レコードから候補の漢方病名集合を定義した方が合理的ではあるが、レコード数が117と少ないため、今回は全レコードから候補の漢方病名集合を定義した。4637個の漢方病名集合 (support > 0.005) を抽出した。

4637個の高頻度漢方病名集合からアトピー性皮膚炎レコードに特異的な高頻度漢方病名集合を抽出した。ある漢方病名集合に含まれる漢方病名全てを同時に診断されたレコードがアトピー性皮膚炎と診断されたレコードに有意に集まっているか否かを Fisher の正確検定を用いて評価した。H0: odds ratio = 1, H1: odds ratio > 1とした。H1の意図は、アトピー性皮膚炎レコードに有意に多く診断されている漢方病名集合を抽出することにある。検定の多重性を考慮

し、各高頻度漢方病名集合の p-value は q-value に変換した。

82個の高頻度漢方病名集合が q-value < 0.1であった。結果は図4-15の通りであった。

アトピー性皮膚炎特異的高頻度漢方病名集合とレコードのクラスタリング

アトピー性皮膚炎特異的高頻度漢方病名集合とレコードのクラスタリングを行うと、図4-16のようになった。

アトピー性皮膚炎特異的高頻度漢方病名集合のクラスタリング

アトピー性皮膚炎特異的高頻度漢方病名集合のクラスタリングを示すと図4-17の通りであった。

アトピー性皮膚炎特異的高頻度漢方病名集合にマッチしなかったレコードの解析

各アトピー性皮膚炎特異的高頻度漢方病名集合にマッチするレコードにどのような薬剤が処方されているかを解析した。アトピー性皮膚炎と診断された117レコード中46レコードは82個の特異的高頻度漢方病名集合には1つもマッチしなかった。これら46レコードに処方されている薬剤を調べてみた。これらは図4-18の通りであった。

また、46レコードと薬剤のクラスタリングを行うと図4-19のようであった。さらに処方されていない薬剤を除いた場合の46レコードと処方された薬剤のクラスタリングは、図4-20の通りであった。

D. 考察

いくつかの症状・西洋病名と他の症状・所見の関連性が確認された。このような情報を逐次蓄積していくことで、データマイニングによる診療における患者状態の予測に役立っていくも

のと思われる。まだまだ、症例数が少ないため、十分な解析が行えていないことが問題点といえる。今後、症例数の蓄積が重要である。

E. 結論

これまでも、漢方所見と症状に関連性が指摘されていたが、これらの報告と本研究成果とを結びつけて、解析に活かしていく事が重要となる。症例数を十分蓄積していく上で、関連施設

の協力が不可欠といえる。

F. 健康危険情報

該当なし

G. 研究発表

該当なし

H. 知的財産権の出願・登録状況

該当なし

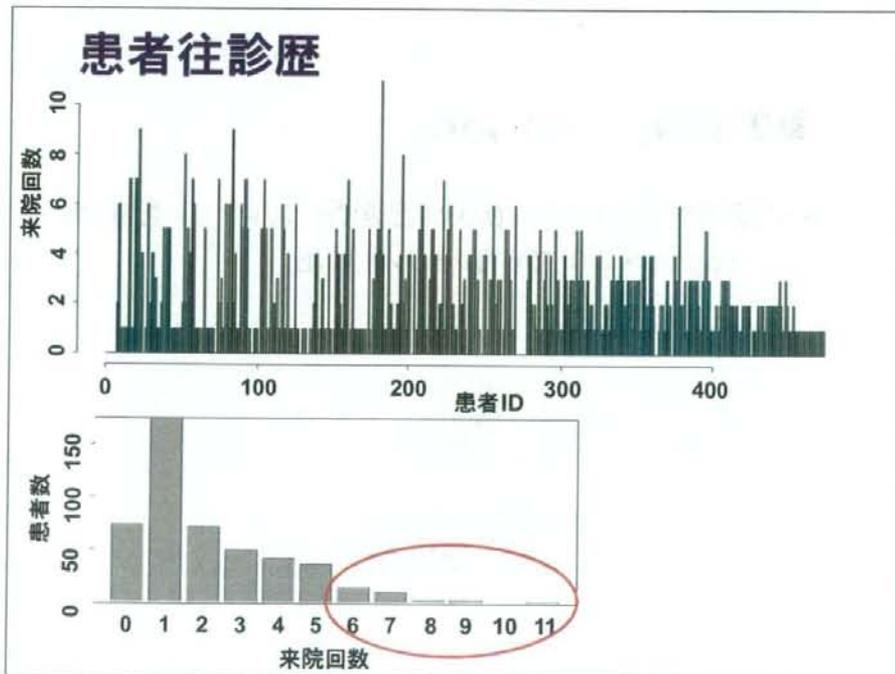


図 3-1 患者往診歴

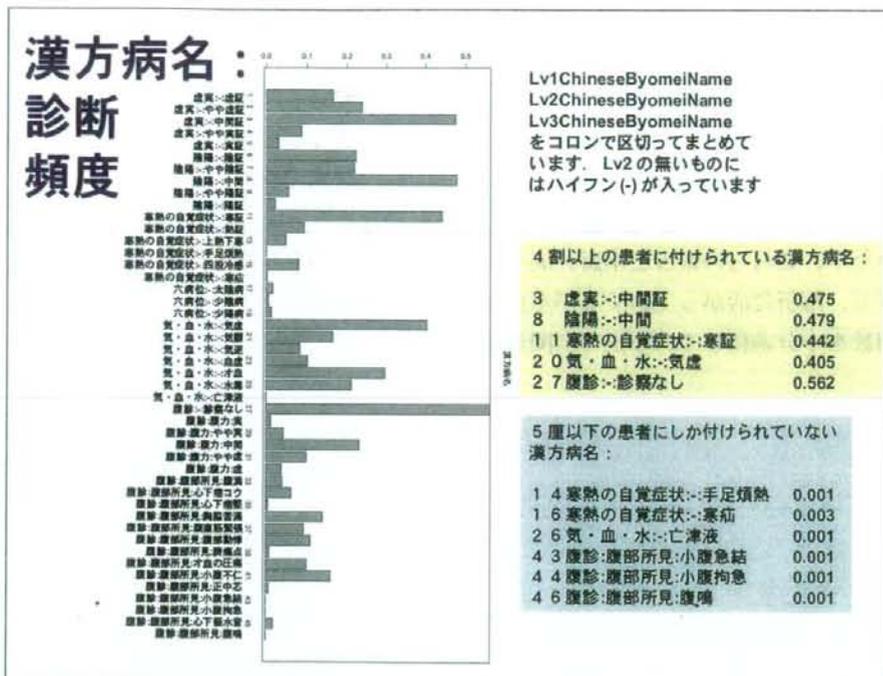


図 3-2 漢方病名診断頻度

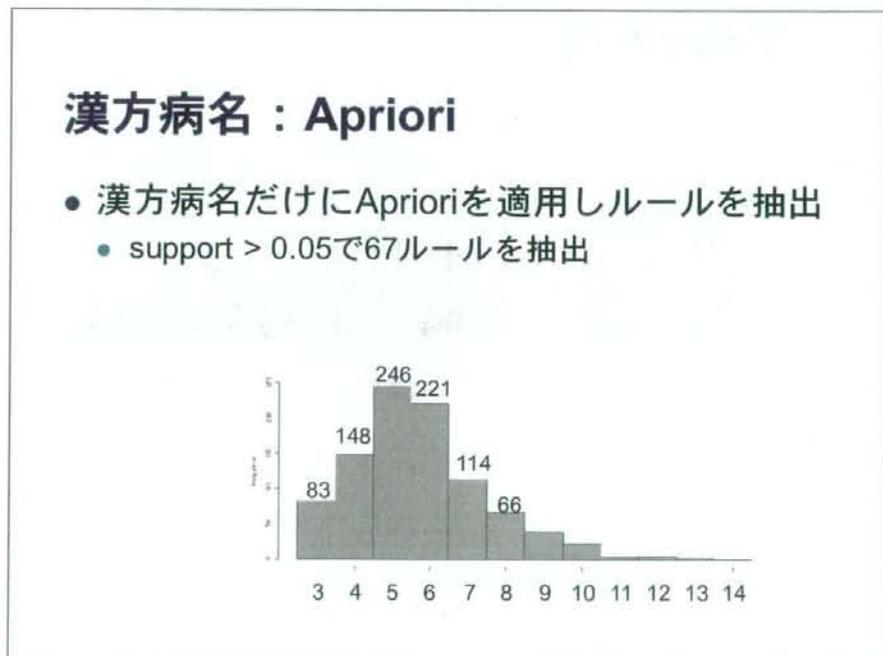


図 3-3 漢方病名：Apriori