

図45: 亀田 証から問診の対応関係

寒項目 熱項目	1 0	0 1	1 1	0 0	
期待される証	寒証	熱証	錯雑証	中間証	合計
寒証	26	0	15	4	45
寒証	57.8%	0.0%	33.3%	8.9%	1
熱証	10	7	15	14	46
熱証	21.7%	15.2%	32.6%	30.4%	1
錯雑証	2	6	6	3	17
錯雑証	11.8%	35.3%	35.3%	17.6%	1
中間証	10	7	28	2	47
中間証	21.3%	14.9%	59.6%	4.3%	1

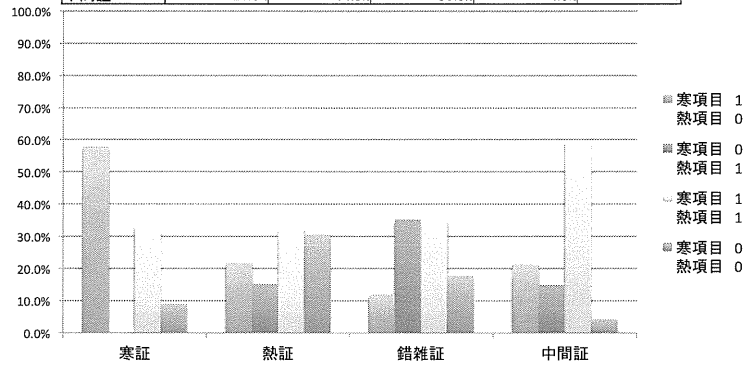


図46: 富山 証から問診の対応関係

寒項目 熱項目	1 0	0 1	1 1	0 0	
期待される証	寒証	熱証	錯雑証	中間証	合計
寒証	22	2	10	6	40
寒証	55.0%	5.0%	25.0%	15.0%	1
熱証	9	2	13	10	34
熱証	26.5%	5.9%	38.2%	29.4%	1
錯雑証	4	5	4	4	17
錯雑証	23.5%	29.4%	23.5%	23.5%	1
中間証	1	2	4	2	9
中間証	11%	22%	44%	22%	1

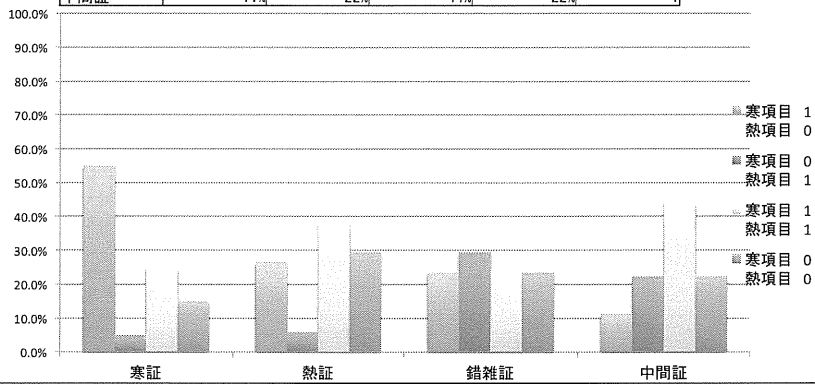
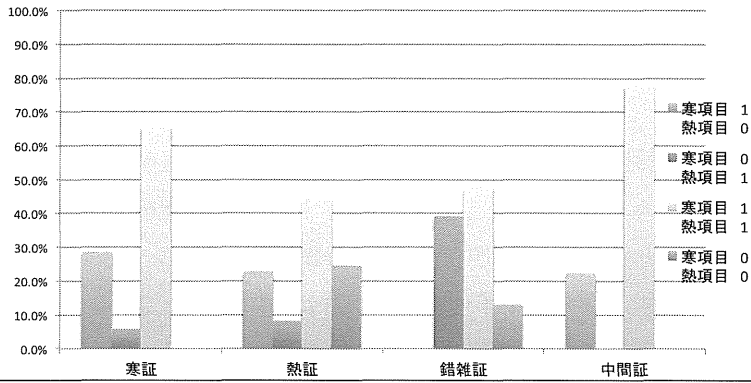


図47: 麻生飯塚 証から問診の対応関係

寒項目	1	0	1	0	
熱項目	0	1	1	0	
期待される証	寒証	熱証	錯雑証	中間証	合計
寒証	10	2	23	0	35
	28.6%	5.7%	65.7%	0.0%	100.0%
熱証	14	5	27	15	61
	23.0%	8.2%	44.3%	24.6%	100.0%
錯雑証	0	9	11	3	23
	0.0%	39.1%	47.8%	13.0%	100.0%
中間証	2	0	7	0	9
	22.2%	0.0%	77.8%	0.0%	100.0%



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

本研究は集められた問診と診断・処方の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を目指す。平成24年度においては、前年度までに確立した証を軸とした問診データの可視化手法を用いて処方支援システムの構築を行った。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、実装を行った。また試験運用による使用フィードバックを集め、改善を重ねた結果条件を満たす処方支援システムの土台が構築された成果について報告する。

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24年度においては、前年度までに確立した証を軸とした問診データの可視化手法を

用いて処方支援システムの構築を行った。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、実装を行った。

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 24 年度においては、前年度までに確立した証を軸とした問診データの可視化手法を用い、処方支援システムの構築を行った。その際に

インターフェースに求められる条件にもとづいてデザインを提案し、試験運用によるフィードバックにもとづいて改善を行った。

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 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も簡潔に説明する。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について議論し、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ついても説明を行う。

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
- ・ 知識の重複、差分
- ・ 知識の抜け

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。

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、合成の要素を探るのである(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。

本研究では、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を目指している。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4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。平成22年度、23年度、24年度の初診患者約5000名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た。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。

可視化システム MIMA サーチ

我々は、上記の知識に対する可視化の方法論を実践し、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²⁾」を開発し、その実用化を行ってきた(図5)。

「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、

2

<http://ciee.t.u-tokyo.ac.jp/MimaSearch/manual/mima/overview.html>

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6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。
- ・ 選択された知識に対して特定の属性での値で集計・分類することで、選択された知識からさらに興味のある知識集合へ絞り込むための属性値候補を提示する（ファセット検索）

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、生命科学分野においては論文が加速的に増加しているが、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7(a)(b)はそれぞれ2006年、及び2007年に開催された生命科学ネットワーク・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年度に比較して、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るが、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。

可視化による診断・処方支援

一般に診断支援の方法としては予測があげられる。証診断で言えば、問診の結果などから、虚実、寒熱などのカテゴリごとに確率の高い証をそれぞれ提案する。これに対し、本研究では予測による支援を補完する形での可視化による診断支援の実現を目指した。予測による支援は診断提案が一目でわかるのに対し、可視化による支援は類似例との比較によって判断の根拠を示し、二次選択薬の候補を提示できることに利点がある。

証の診断は患者の自覚・他覚症状、腹診などの所見、生活習慣などを他の患者と比較しながら決定される。そこで患者を問診項目などによって特徴付けした上で他の患者たちとの位置づけを示すことで証の診断を支援することを目指す(図8)。この場合、「証を診断する上で似ている患者」同士を類似例として提示し、近くに表示する必要がある。つまり患者同士の近さを「診断における類似」を関連させる必要がある。このような場合、正解の証診断データを用いて患者同士の距離指標を学習することが考えられるが、23年度は簡易な学習として、患者の各特徴を証の予測への貢献度(証の予測器における貢献度)で重み付けし、距離指標を調整する手法の影響を確認した。

可視化の目的は証診断の支援であるから、最終的な評価は、専門医の目からみて可視化

によって示された患者間の関係の全体・部分像妥当かどうかとなるであろう。ただしこの評価は個別例に基づかざるをえない上に、人手によるためコストが大きい。システム開発や手法評価には自動的に評価可能な定量的指標が必要であり、可視化された患者間の距離（特徴空間上での距離）と患者の証診断データを利用するなどの方法が考えられるが、これについては検討が続けられている。

漢方データの可視化

患者間の類似関係の可視化は、既存の可視化ツール MIMA サーチを用い患者をクラスタリングした結果を各クラスタ間の距離に応じて表示させて行った。距離を診断・処方支援用に調整するための重み付けについては、患者状態を表す状態ベクトルの要素にそれぞれ重みをかけたものを用いた。重みとしては [片山琴絵, 2012] 中の random forest による証予測器(予測精度 72.4%)での重要度を重みとして採用する。

可視化結果の例を図 9、10、11 に示す。これらの例は初診時に頭痛と冷えの症状のある初診患者を可視化したものである。図 9 が重みなしの間診ベクトルを用いたもの、図 10 が重要度スコアに比例してベクトルを重み付けしたものである。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）が示され、これら円の上に表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った漢方診断を表し、

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。

重みなしではクラスタを特徴付けるものとして、肩こりなど関係の深い症状もあるものの、脈拍など虚実の診断に重要視されない項目が多い。また実際の証が異なる（色の違いで示している）患者が入り乱れたクラスタが複数ある。それに対し、重み有りの場合は証の同じ患者がクラスタを構成し、クラスタの特徴も首こり、寒がりなど虚実の診断に重要とされるものが多く見られた（図 11）。実際専門医による観察でも、重みをつけた場合にはクラスタに直観的な意味付け、たとえばあるクラスタは特定の処方に適した典型的な患者の群、ができることが多くなった。例えば、図 10 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、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提示することで処方の支援にもつながると期待される。

漢方の証診断の支援することを目的として患者の状態を他と比べることで可視化する手法では、診断に有用な問診項目を重視した可視化のためベクトルの調整に着目し、患者の証識別器における重要度を用いて問診項目を重み付けしたベクトルを用いた可視化を行った。結果、専門医による直観的な評価では可視化結果の意味付けが改善したものの、定量

的な評価指標の必要性も確認された。よってこの可視化を用いた支援システム構築を目指すとともに、より定量的指標を用いたベクトル調整手法を模索こととした。

支援システムとしての条件とインターフェース設計

本研究では漢方の専門知識をもたない臨床医の漢方薬処方を支援するツールのひな形を構築することが目的である。そのために満たすべき条件として以下の点を考える。

機能：症状(問診項目)、西洋病名等から処方薬剤の候補を適切な数に絞り込む。またそれらの候補が実際に処方された患者群の特徴の違いを明示する。その際、漢方証を意識することなく処方候補情報までガイドできることが望ましい。

効率性：臨床現場で医師が一人の患者の問診に割ける時間は3分以下とも言われる。そのような時間的制約の中で有効な支援に結びつけるためには、上記の薬剤候補に素早く、つまり少ない操作数でたどり着けるような設計が必要である。

構築システムは、処方すべき薬剤候補 2,3 種類と、それらの薬剤から実際の処方を決定する際の判断材料となる症状や西洋病名と薬剤候補の関連を提示する目的に特化することとする。これは支援システムの目的として漢方専門知識が少ない医師の漢方薬処方を支援することがあるためである。漢方証診断と薬剤処方には密接な関連があるとされているが、このシステムにおいては専門医でない医師には難解な漢方証との関連は提示せず、西洋病名や症状といった属性と薬剤の関連に絞って可視化を行う。目的に特化したインターフェースは第 2 の条件効率性にも適合するとおも

われる。

これらの条件を踏まえ、既存の MIMA サーチを基礎に図のようなインターフェースを設計した(図 12)。左下は従来のデータ分析用インターフェースであり、右上は支援目的に特化した目標インターフェースである。特徴としては

- ・ 詳細データは初期状態では表示しない
- ・ 基礎的な属性である性別、年齢を選択式にする
- ・ 有用な属性(西洋病名、薬剤、症状)のファセット・集計を視覚的に表示
- ・ 上記ファセット・集計から検索条件の追加を直接行える
- ・ 検索条件に合致したデータの処方薬を頻度順リストとして表示

などがあげられる。

また上記であげた機能と効率性の二つの条件に関連して、問診システムと連携して動作する際には、1) 基礎属性である年齢、性別、症状(主訴が望ましい)は問診システムからのクエリによって自動的に入力されるようにし、2) 検索結果ランキングを証予測の近い順にするようにする。前者にかんしてはなるべくユーザの負担を軽くするためである。後者に関しては応答性の面から条件に合致するデータ全件を対象に可視化を行うことが現実的でないことから、どのような条件で可視化を行うデータと行わないデータを線引するかに関係している。ランキングを証予測の近い順にすることは、現在の患者と証予測の近い患者データの関係を優先して可視化を行うということである。つまり似た証を持つと思われる患者に対して、クラスタリングをしそれら患者群の違いを示すことで薬剤選択を支援することを目指している。

システムの実装モデル

可視化による支援システム自体は MIM サーチと同じくサーバクライアントモデルを採用する。サーバはクライアントからの処理要求を受けて実際の処理を行い、その結果をクライアントに返すというモデルである。一般的には、共有プリンタやデータベースなどの情報資源を集中的に管理するサーバと、サーバを利用するクライアントで構成する。³ 支援システムにおいては索引の検索、ファセットの計算などの比較的負荷の大きい処理をサーバにおいて行い、クライアントはユーザ入力からサーバへの検索クエリへの変換、検索結果から可視化クラスタリングの計算などの処理を行う。

システムの構築に際してのデータの流れを図 1 3 に示す。前処理として問診データに上記のようなデータ処理を行い、各データに可視化クラスタリングの際のベクトルデータを付与した MIMA サーチ用データを用意し、これを MIMA サーチにて索引付けする。運用時には外部（にあるクライアント）からのクエリにしたがってデータ検索を行い、合致したデータをランキングし上位のものを返す。

実際に問診システムと連携して運用する際には、新たに入力された問診とシステムによる支援後に実際に処方された薬剤のデータを元にさらにデータ蓄積が行うことも想定している（図 1 4）。

引用文献

片山琴絵. (2012). 問診データ解析による漢方医療支援システムについて. 第 16 回日本医療情報学会春季学術大会.

³ 「ネットワーク大事典」日経 BP 社

C. 研究結果

データ処理

対象データは慶應大学漢方医学センターにて記録された問診と漢方診断、処方のデータを用い、2011 年 12 月までの初診患者データ約 3100 名の内データ整理済み、かつ身長体重のデータのある 493 人分を用いた。身長体重のあるデータに限ったのは、証予測において身長体重のないデータの分類性能が著しく低下したためである。患者は自覚・他覚症状や身長体重、脈拍などの数値（以下まとめて問診項目と呼ぶ）によって特徴付けし、全ての項目を平均値を使うなどして[0,1] の実数値に変換したものをを用いた。

これに上記のようなベクトル調整処理を施し、可視化のためのクラスタリングに用いる特徴ベクトルとした。

ツール構築

実際のシステム構築においては、サーバ側では Apache フトウェア財団のオープンソース全文検索エンジン Solr 3.4.0⁴ をベースとした MIMA サーチサーバプログラムを用いた。クライアントからサーバに検索条件が送られた際はそれがサーバ側で solr に対する検索要求に変換されて solr がデータ検索を行う。クライアントは HTML5 仕様と Ajax と呼ばれる技術群を用いて実装された MIMA サーチクライアントを用いた。

これらの組み合わせを用いたことで、

- ・ ファセットとする属性の変更、検索対象となる属性の変更など検索仕様の変更が容易

・ インターフェースとクライアントーサーバ間

⁴ <http://lucene.apache.org/solr/>

への検索要求が独立であるためにクライアントのインターフェースの変更が容易となり、ツール開発において避けられない試行錯誤のサイクルの循環が速くなることを意図している。

またこの形式のクライアントはブラウザと Javascript さえ動作する環境ならすぐに(インストール無しで)動作が可能であることも選択理由の一つである。

実際のツール

実際に構築された支援ツールのひな形を図 15 に示す。想定される使い方にそって説明すると、まず画面左上で基本的な検索条件、つまり性別、年齢層、特徴的な症状(問診項目)を設定する。図では上から、症状、性別、年齢層を入力して可視化するデータを絞り込むようになっている。その条件でデータが検索され、上で述べたように虚証予測スコアの近い 50 件(変更可能)がクラスタリング、可視化されて表示される。現在は可視化ツールを起動してからこれらの項目を入力するが、実際に問診システムと連携して起動される際には問診データからこれらの項目が設定されるものとする。

さらにグループ分けと表示されている箇所では現在の検索条件を満たすデータ群を属性値(西洋病名や処方薬剤)でグループ分けし(図 16, 17)、さらにグループを選択することでその西洋病名を持つデータに絞って可視化を行うことができる(図 18)。これは属性による検索条件付を直感的に行うためである。

画面左下では薬剤のファセットと問診項目のファセットを表示する。ファセットとは上で述べたように、検索条件にあったデータ群に対してある属性値の頻度順リストを表示し

てデータ絞り込みを補助するものである。今回は薬剤と問診項目に対するファセットであるから、薬剤に関しては検索条件に一致した初診患者に処方された薬剤が多い順(処方された人数が多い順)に表示されることになる(図 19)。またファセットリスト上の薬剤名をクリックすることで、「その薬剤を処方されたデータ」という条件を検索条件に追加したり逆に除いたりという操作が直感的にできるようになっている(図 20)。問診項目ファセットに関しても同様である(図 21、22、23)。

試験運用によるフィードバックと改善

構築したシステムを本研究事業の分担研究機関の医師を対象に試験運用し試用フィードバックを集めた。本研究では漢方の専門知識をもたない臨床医の漢方薬処方を支援するツールのひな形を構築することが目的であった。よって本来の対象ユーザは漢方専門医以外の臨床医であるが、ひな形開発段階では専門知識をもった医師らによる支援の質の評価が必要である。よって今回の試用は漢方専門医によって行われた。また試験運用は問診システムとは連携せずに行われた。

フィードバックの分類と抜粋を表 1 に示す。フィードバック 5、6 に関して想定したモデルケースでは、症状で検索、さらに他の症状や西洋病名で絞り込み、薬剤リストから処方判断へという流れを想定していたが、それ以外の検索要求が見られた。これは対象ユーザでない専門医ならではの意見なのか、一般的に処方の判断に有用なものであるのか検証が必要である。7 は実質クロス集計への要求であり、想定使用モデルでは必要性が小さいと思われていたが、このようにクロス集計を視覚的に表示できれば、直観的なデータ理解と

いう可視化ために役立つとも考えられる（図24）。8,9 に関しては手法とデータの本質に関わるためD.考察にて述べる。1,2,3,4 に関しては本来のユーザの検索要求を考慮しつつ改善していくべきである。1 は現状年齢層を10代、20代などで洗濯しているがそれを自由に設定した検索が望ましいという指摘、また2 に関してはファセットリストのリスト項目数を増減する機能の要求であり、より細かな設定による検索要求であると思われる。4 に関してはすでに改善を行ったためその詳細を述べる。

4の「薬剤の名前が長くてグループ表示が見えにくい」という指摘は、表示の問題だけではない。この時点での薬剤名としては、データ内の薬剤名、例えば「ツムラ葛根湯エキス顆粒（医療用）」といった名称を用いており、製剤元や形状の違いなどで同じ葛根湯でも複数の薬剤として表示・集計されていた（図25）。これは表示が見づらいだけでなく、薬剤処方支援にとって本質的ではない特徴で集計が分かれていることになり、支援の質にも影響があったと考えられる。そのため、データ上の薬剤名に対して薬剤グループとも言える属性を新たに定義し、例えば「ツムラ葛根湯エキス顆粒（医療用）」は「葛根湯」グループとして集計・表示するように処理を行った。データ処理は問診システムに登録された薬剤種からまとめるべき薬剤グループへの変換表を用意するだけである（表2）。またこの変換の利点は処方支援の対象からは除きたい薬剤を指定する（例えば、よく苡仁湯等にはグループを割り当てないなど）ことで、より処方支援に特化した集計を実現できることである。今回は専門知識を持つ医師によって人手でつくられた変換表を用いたが、薬剤名称はある

程度パターンがあり、除外リスト以外は自動に処理することも可能であると考えられる。

また3 に関しては現在ファセットは頻度順、つまり条件に合致したデータ内で多く出現した問診項目が上位に表示されている。ただし、問診項目は日常生活に関する項目などがあり、これらは処方候補の絞り込みにはあまり貢献しないとかんがえられている。よって4と同様の手法で3も改善が可能である。

D. 考察

前述したように、フィードバック8,9はより手法とデータの本質に関わる問題である。

8は例えば、問診項目が「頭痛があるか」である場合に、「頭がいたい」で検索しても合致したとは見なされない問題である。現在のシステムは検索したい症状を文字列で入力する方式である。これは問診項目を示して多数の項目から選択するより簡便である。しかし検索は入力文字列と問診項目表現の文字列比較によって行われることになるため、表面的な文字の違い（「頭痛」と「頭がいたい」）によって、同じ症状と見なされない問題が生じる。現状ではユーザが入力を始めると問診項目表現から前方一致する表現を候補として表示してすることによって、問診項目に一致するよう誘導しているが（図26）入力途中のキーワードと問診項目の前方が一致する必要があるためこれのみで対応するには限界がある。

これに対し現状では2種類の対応が可能であると思われる。一つは情報検索の分野でよく行われる query expansion のように、入力キーワードと同値、もしくは似た症状を表す表現をキーワードに追加して検索を行うことである。これには問診項目の症状と似た症状を表す表現リストを人手または自動で用意す

る必要がある。またより正確な検索を求める人へ向けて、問診項目のリストを示し選択された項目群とデータ上の項目を直接比較するというアプローチも考えられる。

またフィードバック 9 はデータ上の問診に対する薬剤処方が有効だったかどうかを示す、もしくは有効だったデータのみを支援に用いるほうがより質の高い支援になるという指摘と思われる。現在のデータでは有効処方だったかどうかは蓄積されていないためすぐに実現するのは困難であるが、今後の漢方データ解析によって実現する可能性はある。つまりその後の経過、つまり処方後の問診の変化から有効であったかの判定を自動で行い、それを用いるものである。この実現には時系列データの解析手法の漢方データへの適用 (図 27, 28) など課題が多いが、有効性判定はさらなるエビデンスの確立にも貢献することから重要な技術であるといえる。

E. 結論

本報告では、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、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目的のため、まず可視化による処方支援の可能性と、可視化による処方支援システムが満たすべき条件について議論した。また、条件を満たすシステムを提案し、実際に既存の可視化システム MIMASearch を用いたシステムの実装について述べた。実際にシステムを試験運用し得られたフィードバックからの改善により、条件を満たす処方支援の土台となるべきシステム構築の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、可視化を用いた診断・処方支援の実用化が期待される。

F. 健康危険情報

該当なし

G. 研究発表

1. 論文発表

Hideki Mima : An Issue-oriented Syllabus Retrieval System based on Terminology-based Syllabus Structuring and Visualization, In Proceedings of The First Workshop on Speech and Language Processing Tools in Education*(SLP-TED) COLING-2012, 15th December, 2012. Mumbai, India

2. 学会発表

なし

H. 知的財産権の出願・登録情報

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

図表

表 1 フィードバック抜粋

インターフェースに関するもの

1	年齢の区切りを自由に設定しての検索を行いたい
2	ファセットリスト（薬剤、問診）により多くの項目数を
3	問診ファセットには表示する必要の薄い項目(日常生活に関する項目等)がある
4	薬剤の名前が長くてグループ表示が見えにくい

想定した使用モデルの変更に関するもの

5	入力による検索を症状だけでなく薬剤名等も対象にして検索を行いたい
6	グループ分けを複数の属性で多重に行いたい

新たな可視化手法に関するもの

7	西洋病名でグループ分けした際に各病名（を持つ人）に対する薬剤の集計グラフを見たい
---	--

手法・データに対するもの

8	検索できない症状がある
9	有効症例かどうかを示せばよりよい支援になるのではないか

表2 薬剤名からグループへの変換表（抜粋）

データ表記	グループ名
クラシエ柴苓湯	柴苓湯
クラシエ当帰芍薬散	当帰芍薬散
ウチダの八味丸M	八味（地黄）丸
ヨクイニンエキス錠「コタロー」	（集計対象外）
オースギ八味地黄丸料エキスT錠	八味（地黄）丸
コタロー黄連解毒湯エキスカプセル	黄連解毒湯
コタロー桔梗石膏エキス細粒	桔梗石膏
コタロー九味檳榔湯エキス細粒	九味檳榔湯
コタロー梔子柏皮湯エキス細粒	梔子柏皮湯
コタロー大柴胡湯去大黄エキス細粒	大柴胡湯去大黄
コタロー八味丸料エキス細粒	八味（地黄）丸
コタロー竜胆瀉肝湯エキス細粒	竜胆瀉肝湯
三和附子理中湯エキス細粒	附子理中湯
太虎堂のきゅう帰調血飲エキス顆粒	芎帰調血飲
ツムラ安中散エキス顆粒（医療用）	安中散
ツムラ胃苓湯エキス顆粒（医療用）	胃苓湯
ツムラ茵陳蒿湯エキス顆粒（医療用）	茵陳蒿湯
ツムラ茵陳五苓散エキス顆粒（医療用）	茵陳五苓散
ツムラ温経湯エキス顆粒（医療用）	温経湯
ツムラ温清飲エキス顆粒（医療用）	温清飲
ツムラ越婢加朮湯エキス顆粒（医療用）	越婢加朮湯
ツムラ黄連解毒湯エキス顆粒（医療用）	黄連解毒湯
ツムラ黄連湯エキス顆粒（医療用）	黄連湯
ツムラ乙字湯エキス顆粒（医療用）	乙字湯
ツムラ葛根湯エキス顆粒（医療用）	葛根湯

図 1 生命科学における知の構造化の例

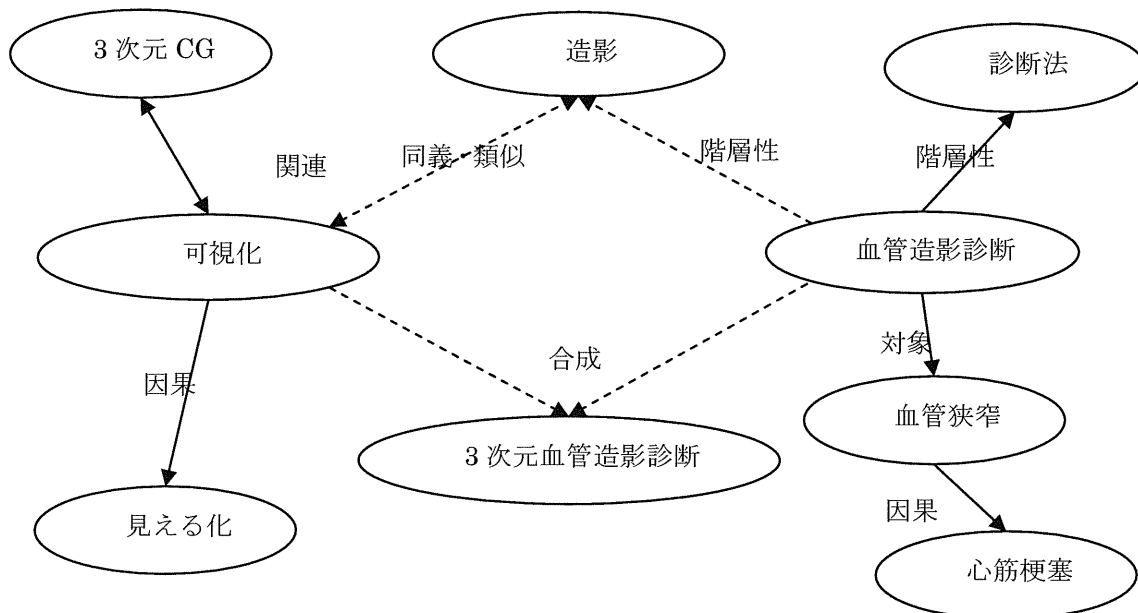


図 2 データマイニングと可視化のモデル

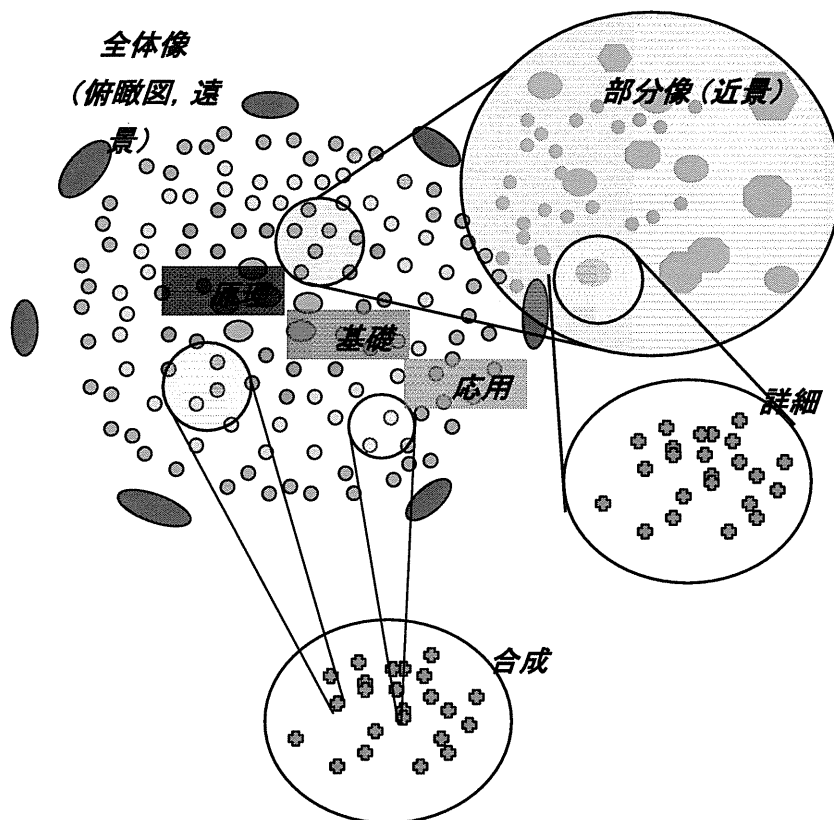


図3 知識の継承と合成

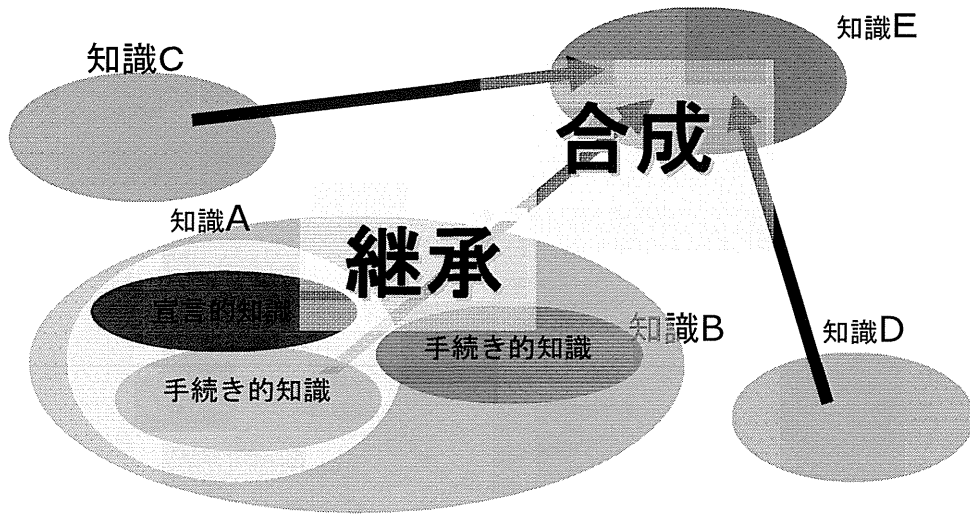


図4 「証」と問診との関連

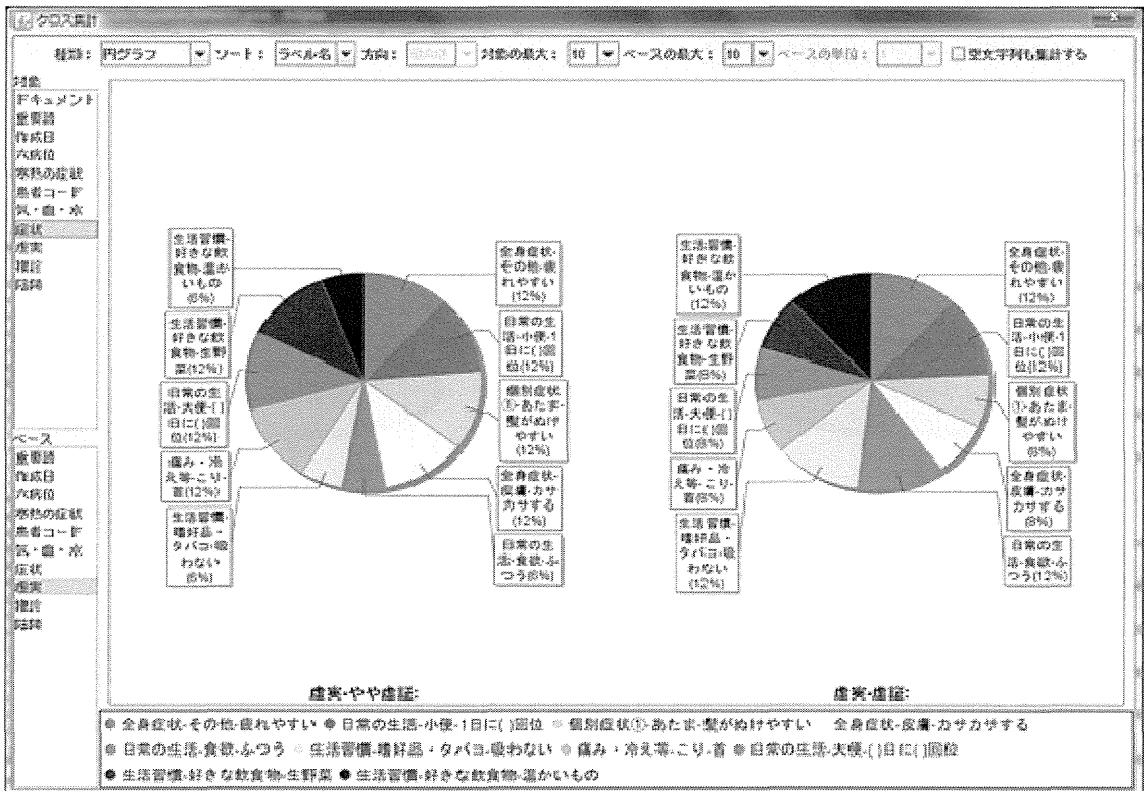


図 5 MIMA サーチ

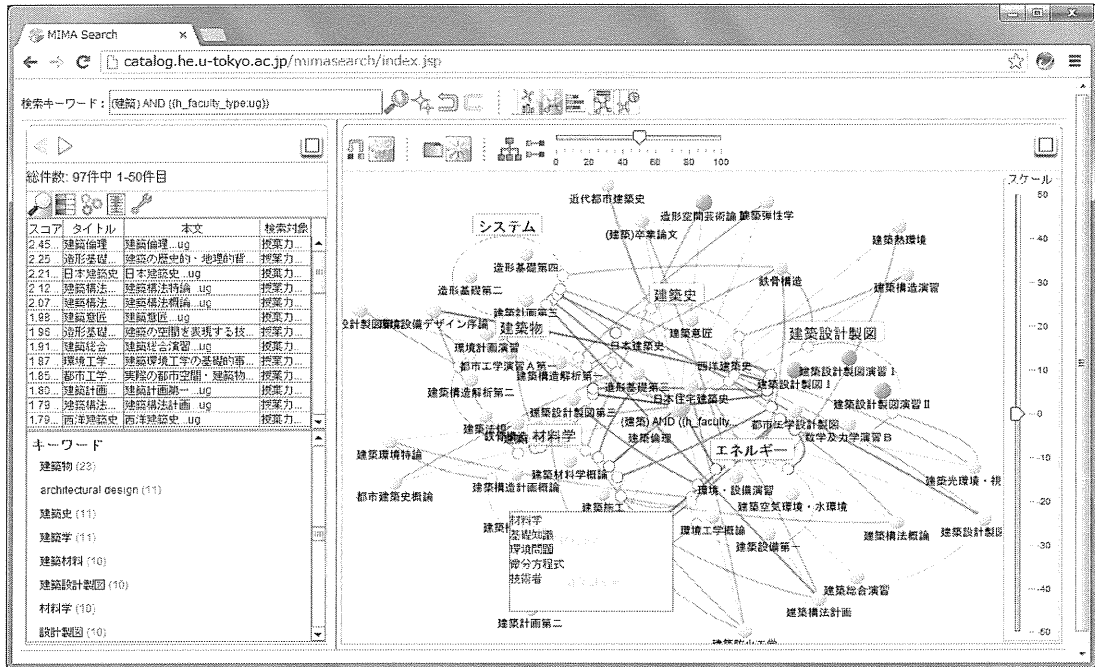


図 6 特徴の抽出と関連性の認識

