

原著論文

症状からの病名検索支援に基づく病院検索支援システムの提案

Hao BO*, 吉川 大弘*, 古橋 武**, 杉浦 伸一*

*名古屋大学大学院工学研究科, **名古屋大学予防早期医療創成センター

PROPOSAL OF HOSPITAL SEARCH SUPPORT SYSTEM BASED ON DISEASE SEARCH FROM SYMPTOMS

Bo HAO*, Tomohiro YOSHIKAWA*, Takeshi FURUHASHI*, ** and Shin-ichi SUGIURA*

* Graduate School of Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

** MEXT PME of Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

Abstract : Recently, public interest in health has grown, and increasing numbers of people are relying on a regular medical checkup for the early detection of disease. At the same time, the rapid development of information technology gives us much medical information, and the demands for an effective medical support system will continue to grow. The aim of this study is to develop a Hospital Search Support System (HSSS) as a part of integrated medical support system which includes electronic medical chart sharing system, specialist introduction system, medicine information system and so on. It employs Hierarchical Keyword Graph (HK Graph) for the Disease Search Support System (DSSS). DSSS can list diseases according to the strength of the relation to the symptoms inputted by a user. First, this paper shows the extraction method of symptom words for DSSS. And then this paper also shows the method to present the candidates of diseases with their rank based on the symptoms on HK Graph and the symptom severity inputted by user. This paper investigates the effectiveness of the proposed method based on the accuracy and coverage rate of symptoms and the transition of ranking of diseases with the change of symptom severity.

Keywords : HK Graph, Symptom, Disease Ranking, Hospital Search, Text Mining

1. はじめに

近年、人々の健康に対する関心が高まっている。特に、現在の高齢化社会において、疾患の早期発見が重要視されており、定期的な健康診断や人間ドックに積極的に参加する高齢者が増加している[1-3]。また、高齢者のみならず若齢者の中でも、長年の生活習慣が引き起こすとされる生活習慣病にかかる人が増加しており、疾患の早期発見および生活習慣の見直しがさけられている[4, 5]。一方、現在のユビキタス社会における情報化技術の発達とともに、インターネットを用いて医療情報を提供するシステムの開発・公開[6, 7]が急速に進められており、今後も人々の疾患の早期発見・予防を支援するシステムの実現に対する要求や必要性は高まっていくものと考えられる。現在一般に公開されている医療情報は、疾患の症状や療法に関する情報、病院・診療所の位置情報や診療科に関する情報など、多岐にわたる[8, 9]。しかしながら、これらを統合した早期発見・予防のための総合的医療支援システムはまだ少ない。例えば、医学辞書などによる病名検索では、ユーザが症状を入力することでシステムが関連する病名を表示するものの、その疾患をどの病院・診療所へ行って治療すればいいかといった情報はサポートされていない。また[8, 9]では、単に病院・診療所の位置情報や診療科情報が与えられただけで、ユーザの疾患を治療するにはどの病院・診療所を選択すればよいかは、医学知識を持たないユーザ自身の判断に委ねられている。さら

に、ユーザが具体的にどんな症状があるか明確にはわからず、ほんやりと体調が悪いと感じているとき、そもそも症状として何を入力したらよいかわからず、病名や病院・診療所を検索することが困難となる。

そこで例えば、ユーザが疾患にかかり、病名がわからず漠然と“腹痛”と“頭痛”がするという状況があったとき、それらの症状に関連する病名だけでなく、関連する他の症状もあわせて表示することで、ユーザが自分の症状や病名を把握する支援を行うことができると考えられる。さらに、ユーザの疾患や生活環境に合った病院・診療所を表示することで、ユーザ自身に専門的な知識を要求することなく、病院・診療所の選択支援が可能となる。これにより、ユーザが体調の異変を感じたとき、すばやく適切な病院・診療所で治療を受けることが可能となり、疾患の早期発見・予防に繋がると期待できる。

総合的な医療支援システムとして、Yahooヘルスケア[10]やGooヘルスケア[11]がある。[10]や[11]では、入力された症状に基づき、ユーザの病名候補を表示するシステムを提供している。これらのシステムでは、病名とその症状を説明した文章リストをデータベースとして持ち、ユーザが入力した症状とデータベースの症状文章中のキーワードとのマッチングにより、病名の検索・表示を行っている。そのため、ユーザが一般的な症状を入力したとき、検索された病名の表示数が膨大になる。また、検索結果は五十音順で表示されるため、特に表示数が多いとき、自分の疾患を判断することが困難となる。さらにこれらのシステムでは、ユーザの

症状の程度情報を考慮することや、入力した以外の症状を発想することは困難である。本論文では、ユーザが入力した症状を基に、インタラクティブに病名を推測し、ユーザの位置情報や病院・診療所の情報などを考慮して、適切な病院・診療所を表示する病院検索支援システムを提案する。

本論文ではこの病院検索支援システムのうち、主に症状から病名を推測する病名検索支援システムに重きを置いて検討を行う。病名検索支援システムには、共起度情報をもとに、文章データ中の語句同士の関係を階層的なキーワードグラフ構造で可視化する HK Graph (Hierarchical Keyword Graph) [12] を用いる。ユーザにより入力された症状キーワードに対して、HK Graph を用いることで関係性の高い症状を階層的キーワードグラフ構造でユーザに表示する。本システムの特徴の一つは、インタラクティブなやりとりにより病名検索を進められる点にあり、例えば、「めまいがひどくて、少し吐き気がする」といった、ユーザの症状に合わせて程度情報を与えたり、表示された症状を新たに入力したりすることで、疾患の診断結果に症状の程度を反映させることができある。本システムでは、表示された症状と入力された程度情報を用いて、それらと関係性の高い順に病名をランキング形式でユーザに表示する。そして、ユーザが気になった病名に対して、ユーザの位置情報や病院・診療所の診療科情報などを考慮して、その疾患を診断するのに適切な病院・診療所を検索し、ユーザに表示する。ただし、疾患の診断や総合所見は医療機関において医師が行うものであり、ここでは、ユーザへの早期受診を後押しする支援システムの構築を目指す。

本論文では初めに、病名検索支援システムについて述べる。病名検索支援システムでは、HK Graph を用いて、入力された症状を基に、関係性の高い症状をユーザに表示する。このとき、検索の対象となる文章は医学辞書や医療に関する Web ページである。そのため、それらの文章から語句を切り出せば、当然病名や症状とは関係のない一般的な単語が大部分を占める。そこで、Web ページや電子辞書の文章中から、症状として用いられる語句を自動で抽出する手法を提案する。次に、症状の程度情報を考慮した表示病名のランキング手法、および病院・診療所の提示方法について述べる。さらに実験として、Yahoo ヘルスケアの症状に関する Web テキストに提案手法を適用し、症状に関する語句の抽出手法、および症状の程度情報が病名ランキングに与える影響に対する検証を行う。

2. HK Graph

2.1 HK Graph の特徴

提案する病名検索支援システムでは、入力された症状と関連する症状のキーワードグラフの作成のために、テキストマイニング手法 HK Graph を用いる。HK Graph とは、文章情報に付加された属性情報と、文章中の語句の共起度を利用して語句の抽出、階層的キーワードグラフの作成を行う手法である。HK Graph では、ユーザが介在することにより、ユー

ザ自身が能動的に文章情報の解析、探索を行う事ができる。一方、文章中の語句の関係をグラフ構造に可視化する手法として、KeyGraph [14] がある。KeyGraph では、文章中の語句の出現情報および共起情報を用いて語句を抽出し、グラフ構造に可視化することで、文章の主要な語句や重要な事象とそれらの関係を把握するものである。KeyGraph では、文章中の語句の関係を表示することは可能であるが、ユーザが特定の語句を入力し、それらの語句と文章中の語句との関係を把握することは困難である。本病名検索支援システムでは、ユーザとシステムがインタラクティブに入力・表示を行うことで、病名や症状についての発想・診断を支援することを目的としているため、ユーザが能動的にグラフ構造や表示内容に関与することができる HK Graph を用いる。以下に、HK Graph のアルゴリズムを述べる。

2.2 語句への分割

初めに、文章データを語句に分割する。ここでは、対象となる文章データに対して、南瓜 [15] を用いて形態素解析および文節区切りを行い、語句への分割を行う。このとき、文節区切りのみでは、一つの文節として助詞が付加されるため、例えば“脳機能障害が”と“脳機能障害は”は別の語句とされてしまう。また形態素解析では、語句が最少単位に分割されてしまうため、複数の語句で一つの意味をなすような言葉が分割されてしまう。例として、“脳機能障害があります”という文を文節区切りおよび形態素解析を用いて分割すると、それぞれ以下のようになる。

・文節区切り

脳機能障害が、あります

・形態素解析

脳、機能、障害、が、あり、ます

そこで、HK Graph では、形態素に分割された語句に対して、その品詞情報を用いて、助詞、記号（句読点、括弧）、代名詞、接続詞、連体詞となるものを解析に必要のない語句として削除した上で、文節情報を用いて結合する。これにより、上記の例文は、以下のように語句に分割される。

・形態素解析→助詞等の削除→文節区切り

脳機能障害、あります

これにより、文節区切りでは残ってしまう助詞 “が” が削除され、形態素解析では分割されてしまう “脳” “機能” “障害” が再結合され、一つの語句として分割されていることがわかる。また、以下の条件を満たす語句についても、ノイズ語句として削除する。

(i) 数字で始まっている語句 (2週間、300ml など)

(ii) 一字のみの語句 (時、今など)

2.3 ベースの選択

次に、ユーザが解析したいと考える内容、すなわちここではユーザの症状に対して適切なキーワードを、2.2 で抽出された語句から選択する。本来、ユーザが症状を自由に「入力」することが望ましいが、“頭痛” と “頭が痛い” など、記述

のあいまい性や同義性に対する検討は今後の課題とし、現段階では症状を「選択」する方法を用いる。ここで選択された語句は解析の基となるため、「ベース」と呼ぶ。2.4で述べる主ノードの抽出では、ベースとの共起度が計算され、呈示する語句が抽出されるため、ベースが異なればキーワードグラフの呈示結果も異なる。この点は、ユーザが能動的にグラフ構造に関与するというHK Graphの特徴の一つである。

2.4 主ノードの抽出

選択されたベースを利用して、文章データ中の語句から、ベースとの共起度が高い語句の抽出を行う。共起度の算出にはJaccard係数を用いる。ベースとして選択された語句を B_i 、抽出対象となる文章データ中のある語句を W_j としたとき、 B_i と W_j の共起度は次式で表される。

$$\text{Jaccard}(B_i, W_j) = \frac{N_s(B_i \cap W_j)}{N_s(B_i \cup W_j)} \quad (1)$$

ここで $N_s(X)$ は語句 X を含む文章の数である。ベースが2つ以上選択されている場合、(1)式を用いて得られた各ベースとの共起度を掛け合わせた値を指標とし、ベース全体との共起度が高い語句（全結合語句）、特定の組み合わせのベースに対して共起度が高い語句（複結合語句）、各ベースに対して共起度が高い語句（単結合語句）の抽出をそれぞれ行う。ここで抽出された語句を“主ノード”と呼ぶ。

2.5 副ノードの抽出

“副ノード”には、ユーザが選択した主ノードに対して関係性の高い語句が抽出される。つまり(1)式において、ベースを主ノードに置き換えたとき抽出される語句となる。ただしシステムの画面上ではベースと主ノードのみが表示され、ユーザが、主ノードを選択（クリック）したときにその主ノードに対する副ノードが表示される。また、それぞれの副ノードについて、対象となる主ノードの他に、最も共起度が高い他の主ノードとのリンクを表示する。

2.6 階層的キーワードグラフの呈示

図1に、HK Graphにより作成された階層的キーワードグラフのイメージ図を示す。図において、 $B_1 \sim B_3$ がベース、 A_1 が全結合語句、 $M_{12} \sim M_{23}$ が複結合語句、 $S_{11} \sim S_{32}$ が単結合語句、 Sub_1 、 Sub_2 が S_{32} に対して共起度の高い副ノードである。このように、ベースと主ノードはリンクの接続により関係が表され、語句どうしの共起度の値はリンクの太さにより表されている。

3. 病院検索支援システム

3.1 症状語句の抽出

2.2において、対象となる文章を医学辞書や医療に関するWebテキストとすることで、一般的な文章と比較して、抽出される語句は症状に関するものが多くなる。しかしこのよう

に対象文章を限定しても、症状とはあまり関係のない語句も多く含まれる。HK Graphでは、ベースと共に起度の高い語句を抽出して呈示するため、症状との関連を示すにはあまり適切ではないがベースとの共起度は高い語句も抽出されてしまう。図2は、“発熱”と“鼻水”をベースとして、HK Graphを適用した結果である。この図において、“落ちてきて”、“重いなど”、“先立って”、“場合”などの語句は、“発熱”や“鼻水”と共に起してはいるものの、関連症状の発想支援や病名検索支援に用いる語句としては適切であるとはいえない。そこでここでは、Web情報を用いて、2.2で得られた語句から、症状に関連した語句（以後、“症状語句”と呼ぶ）の選別を自動で行う手法を提案する。

初めに、2.2で述べた通り、南瓜を用いて語句を分割し、ノイズ語句の削除を行う。次に、得られたすべての語句に対し、“症状”とのアンド検索を行い、そのヒット件数を調べる。具体的には、Yahoo Web API [16] を用いて、以下のように検索を行う。

- (i) 得られた各語句を単独で検索し、そのヒット件数を調べる。
- (ii) “症状語句”（症状（半角スペース）各語句）で検索し、そのヒット件数を調べる。
- (iii) 得られた各ヒット件数を用いて、(2)式により共起度を計算する。

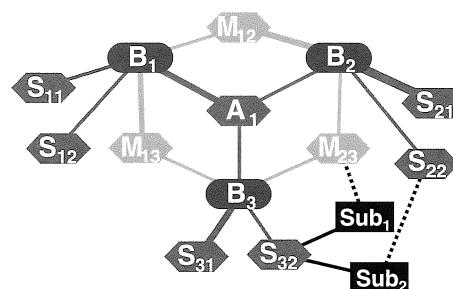


図1 HK Graph のイメージ

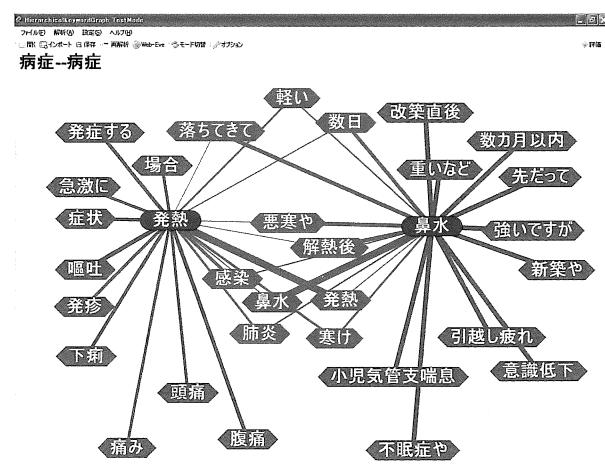


図2 “発熱”，“鼻水”をベースとしたキーワードグラフ

ここで、共起度には、(2)式に示す *Dice* 係数を用いる。*Jaccard* 係数ではなく、*Dice* 係数を用いるのは、*Jaccard* 係数を使った場合、“症状”という語句が一般的な語句であることでヒット件数が膨大になってしまうためであり、語句のヒット件数そのものは少なくとも、症状としてよく用いられる語句を見落とすのを防ぐためである。*Dice* 係数の式は以下のように表される。

$$Dice(Symp, W_i) = \frac{N_h(Symp \cap W_i)}{N_h(W_i)} \quad (2)$$

ここで、*Symp* は“症状”であり、*W_i* は 2.2 で得られた各語句、*N_h(X)* は語句 *X* のヒット件数である。本手法では、算出された共起度に閾値を設け、その閾値以上の語句を症状を表す語句、もしくは症状と関連した語句であるとみなす。

3.2 HK Graph による症状キーワードグラフの呈示

3.1 で抽出された症状語句を用いて、階層的キーワードグラフを構築することで、入力した症状と関連のある症状や、それらの関係性が把握できる。これによりユーザが、呈示されることで初めて気づく自覚症状や、例えば時間的なずれにより入力しなかった他の症状を発想できることが期待される。ユーザは、呈示された階層的キーワードグラフに対して、以下の操作を行うことにより、新たな症状を反映することが可能である。

- (i) 呈示された関連症状を新たにベースに加え、再度症状キーワードグラフを構築する。
- (ii) 呈示された関連症状に対する重みを与える。
- (i) の方法では、2.3 で述べたように、抽出される主ノード、すなわち呈示される症状そのものが変わるために、更なる関連症状の発想支援につながると考えられる。また本システムでは、(ii) のように、入力した症状や呈示された症状の程度情報を重みの形で与えることで、3.3 で述べる病名のランキングに症状の程度を反映させる。

3.3 病名ランキング

本システムは、HK Graph により呈示された症状およびそれらの程度情報を用いて、関係性の高い病名の呈示を行う。病名と症状との関係性は、医学辞書など、病名とその症状に関する記述文章を用いて以下のように定義し、定量的に算出する。これにより、関係性の高い順に、ランキング形式でユーザへの病名の呈示を行う。初めに、ユーザにより選択された症状を少なくとも一つ含む病名を抽出し、これを呈示する病名の候補とする。次に、この病名の候補群と、ユーザにより重みを与えられた症状との関係性を、以下の式で定義する。

$$R(D_i) = \sum_{j=1}^{n_n} \sum_{k=1}^{n_{si}} \frac{Weight(N_j) \times Jaccard(N_j, W_{ik})}{n_{si}} \quad (3)$$

(3)式において、*D_i* は *i* 番目の病名、*n_n* は階層的キーワードグラフ中のノードの総数、*n_{si}* は *i* 番目の病名の症状に関する文章中の症状語句の総数、*N_j* は *j* 番目のノード、*Weight(N_j)*

はノード *N_j* の重み、*W_{ik}* は *i* 番目の病名の症状文章中の *k* 番目の症状語句である。

病名に関する症状文章中の症状語句と、HK Graph により呈示された症状キーワードとの共起度の和を症状文章中の語句数で割ることで、階層的キーワードグラフと網羅的に共起する症状が多い病名をより上位にランキングしている。また、ユーザにより与えられた症状の程度を、重みとして共起度に直接掛け合わせることで、大きな重みを与えた症状に関係する病名ほど、より高くランキングされる。ただし本システムでは、病名ランキングはユーザが病名候補を把握するための支援として行うものであり、ランキングの上位の病名によって“診断”を行うものではない。関連する病名がより上位に呈示されるほど、ユーザの検索が容易になることが期待できる。

3.4 病院・診療所の呈示

名古屋大学予防早期医療創成センターでは、独自に日本国内の病院・診療所データベースを構築してきた。この病院・診療所データベースには、各病院・診療所の郵便番号、所在地、電話番号、診療科、施設基準、病床数などの属性データが存在している。3.3 で示された病名から、ユーザの気になった病名に対して、このデータベースの診療科情報および所在地情報を用いて、その疾患を診断するのに適切な病院・診療所の呈示を行う。また、図 3 に示す Yahoo ヘルスケアの医療データの例のように、本論文で用いた病名データでは、各病名に対する診療科カテゴリーが分類されている。そこで本システムでは、3.3 で示された病名から、ユーザの気になった病名に対する診療科情報を用いて、病院・診療所データベース中の診療科情報との比較を行い、治療に適切な病院の絞り込みを行う。さらに、携帯電話の GPS 機能を用いてユーザの位置情報を取得し、病院・診療所データベース中の所在地情報に基づいて、ユーザの位置から近い病院および診療所を地図上に表示する。このように、本論文では、その疾患を診断するのに適切な病院・診療所であることの基準として、ユーザが選択した病名の診療科を持っており、ユーザの位置から近い病院・診療所であることを用いる。また、ユーザの現在の位置ではなく、例えば自宅から近い病院や診療所を呈示することも可能である。

3.5 システムの実装

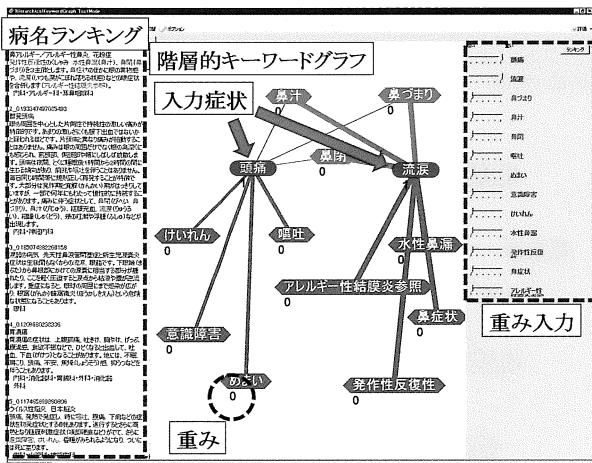
3.1 から 3.4 で述べた方法を、システムとして実装した。

疾患名	かぜ症候群
診療科	小児科、内科
症状	鼻汁、咳、くしゃみ、痰(たん)、発熱などが主な症状です。吐いたり下痢をしたりといった症状がみられることもあります。

図3 Yahoo ヘルスケアの医療データ例

図4にシステムのインターフェースを示す。図4(a)に示す病名検索支援システムでは、画面左に病名ランキングの表示ウインドウ、中央にHK Graphによる症状キーワードグラフ、右に重み入力バーの3つの部分からなる。

症状キーワードグラフ画面では、ユーザにより入力された自覚症状（青）、およびそれに関連する症状（緑（単結合）、橙（複結合）、赤（全結合））が表示される。重み入力バーでは、ユーザの入力した症状を含め、HK Graphにより抽出された関連症状すべての重みが入力できる。本システムでは、0.0（全く感じない）から1.0（ひどく感じる）まで、0.1刻み10段階で入力可能であり、初期状態では、入力した症状の重みは1.0、関連症状の重みは0.0となっている。病名ランキングでは、病名とその症状文章および対応する診療科が表示される。また、キーワードグラフで表示された症状語句には色づけされており、ユーザが病名を把握しやすい工夫されている。さらに病院・診療所の表示では、登録されたユーザの住所、もしくはGPS等で得られた位置情報に基づき、病名ランキングから選択された疾患を扱う病院・診療所の位置情報とその行き方を、図4(b)のように表示する。



(a) 病名検索支援システム



(b) 病院検索支援システム

図4 実装したシステム

4. 提案手法の評価

4.1 症状語句の抽出

3.1で述べた通り、HK Graphにより構築されたキーワードグラフには、症状とはあまり関係のない語句も多く含まれている。提案するシステムでは、3.2で述べたユーザに対する症状の発想支援を行うことにより、病名への関係づけにおける信頼性の向上を期待している。ただし表示される不要語句の多さが発想支援に与える影響を定量的に評価することは困難であるため、本論文では、症状以外の語句が多く表示されることで、症状の発想の妨げになると仮定し、3.1で提案した症状語句の抽出手法に対する定量的な評価を行う。ここでは、Yahooヘルスケアの医療データに3.1の手法を適用し、症状語句の抽出性能について検証する。

4.1.1 症状語句の定義

使用したデータは、インターネット上に公開されている[10]、502病名およびそれらの症状の文章である。これらの文章に対し、2.2で述べた語句の分割・削除・結合、および重複した語句の削除を行うことで、11,018語句を抽出した。本実験では、提案手法の定量的評価のため、この11,018語句の中から、症状を表わす語句および症状に関連した語句を人手により選別し、これを抽出すべき症状語句の正解とした。その結果、11,018語句の中の4,670語句を症状語句と定義した。定義した症状語句は、以下のような語句である。

- (i) 症状一般（発熱、嘔吐など）
 - (ii) 身体の一部（腕、皮膚など）
 - (iii) 属性（女性、幼児など）
 - (iv) ~症（後遺症、認知症など）
 - (v) ~病（うつ病、歯周病など）
 - (vi) 擬音語（カサカサ、ズキズキなど）
- (i) 以外にも、“頭がズキズキする”、“女性特有の症状”などの記述は、症状の状態や性質を表していると考えられるため、(ii) (iii) (vi) の語句も症状に関する語句とした。また症状と病名は、医学分野において明確に分類されている[17]が、本システムは医学知識を持たない一般のユーザを対象としているため、これらの明確な分類はせず、(iv) (v) の語句も症状に関する語句と定義した。

4.1.2 抽出正答率および網羅率

提案手法において、(2)式のDice係数による共起度に対する閾値を変化させて語句を抽出した。抽出した語句の中で、4.1.1で定義した症状語句をどの程度含んでいるかを検証するために、正答率を以下のように定義した。

$$\text{正答率} (\%) = \frac{N(S \cap E)}{N(E)} \times 100 \quad (4)$$

$N(X)$ は集合 X の要素数であり、 S は症状語句の集合（4,670語句）、 E は抽出された語句の集合である。また、それぞれの閾値による語句抽出の際、定義した症状語句（4,670語句）

の中の語句をどの程度抽出しているかの指標として、網羅率を以下のように定義した。

$$\text{網羅率} (\%) = \frac{N(S \cap E)}{N(S)} \times 100 \quad (5)$$

図5に、閾値を変化させたときの正答率・網羅率の結果を示す。横軸が閾値であり、縦軸が正答率および網羅率である。図では表されていないが、閾値を用いないときの正答率・網羅率はそれぞれ42.4% ($4,670 / 11,018 \times 100$), 100%である。図から、閾値を高くすると、抽出語句に対する症状語句の正答率が高くなっていることがわかる。これは、指標として用いた(2)式による共起度が、症状を表す語句の抽出に対応していたことを示している。図において、閾値を変化させたとき、正答率は最大で約80%まで高くなっているが、同時に網羅率は低くなり、約10~20%となってしまっている。

本論文では、正答率および網羅率の双方を考慮した適切な閾値の決定を行うために、相互情報量を用いる。相互情報量が最大となるとき、(4), (5)式における $S \cap E$ と $\bar{S} \cap \bar{E}$ の双方が最大化されることを意味する。相互情報量の式を以下に示す。

$$I(S; E) = H(S) - H(S|E) \quad (6)$$

$$H(S) = -\sum_{i=1}^n p(s_i) \log_2 p(s_i) \quad (7)$$

$$H(S|E) = -\sum_{i=1}^n \sum_{j=1}^m p(s_i, e_j) \log_2 p(s_i|e_j) \quad (8)$$

これらの式において、 $p(x)$ は事象 x が起こる確率、 s_i は語句が症状語句である／ないという事象、 e_j は語句が抽出される／されないという事象、 n, m はそれぞれ s, e の事象数、すなわち $n = 2, m = 2$ である。

この実験より、相互情報量が最大となった閾値は 0.2 であり、このときの正答率は 70.7%，網羅率は 72.3% であった（図5）。経験上、入力した以外の症状の発想支援および病名検索支援を行う上では、表示された語句の 6~8 割程度が症状との関係の強い語句となることが求められる水準であると考える。

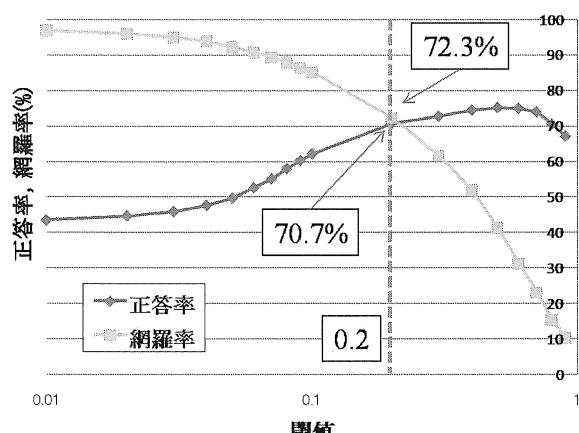


図5 閾値の変化による正答率および網羅率の推移

えられるため、十分とはいえないが、実用上は有効な結果であるといえる。

4.1.3 階層的キーワードグラフの作成

提案手法により抽出された語句を用いて、階層的キーワードグラフを作成した。得られた階層的キーワードグラフを図6に示す。図6を図2と比較すると、症状に関連のない語句がほぼなくなり、適切なキーワードグラフが生成されていることがわかる。

4.2 病名ランキングの推移

最後に、表示された症状の重みを変化させたときの病名ランキングに与える影響を調査した。数十、数百件の病名が表示されたとき、ユーザがそれらすべてに目を通すことは困難であり、上位数件しか見ないユーザも少なくないと考えられる。そのため、何らかの意味を持つ病名が上位に表示されることは、病名検索支援として有益であると考えられる。本論文では、3.3で示した症状との関係性をランキングの基準としているため、ここではユーザが着目した症状（重みを変化させた症状）に対して関係が強い病名が上位にランキングされているかどうかを調査した。ユーザが自覚症状として、“発

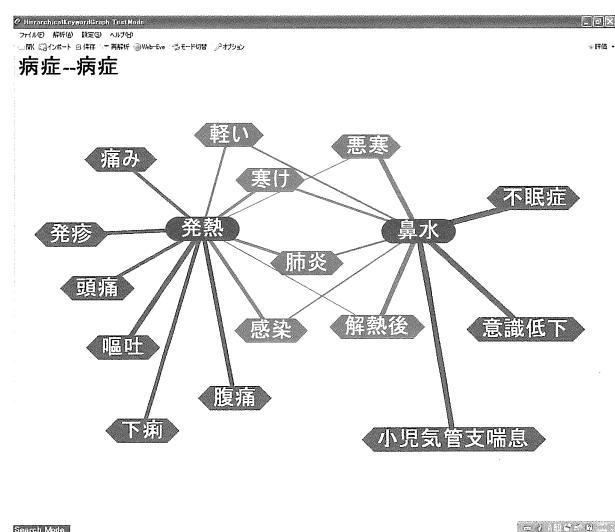


図6 症状語句抽出後の階層的キーワードグラフ

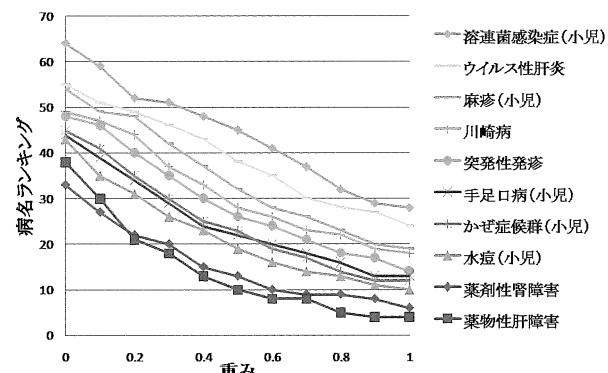


図7 病名ランキングの推移

熱”と“鼻水”を入力したとき、関連症状が図6のように呈示されたとする。この結果により、ユーザが“発疹”に着目し、その重みを変化させたとき、病名ランキングの推移を図7に示す。図の右側にある病名は、“発疹”を症状にもつ病名群である。呈示された症状のキーワードグラフにより、症状の入力時には意識しなかった症状が発想され、その重みを増加させることで、それらの症状と関係性が高い病名が上位にランキングされやすくなることが確認された。ランキングの上位のものから順に病名を検索していくことを仮定すれば、関係性の高い病名をより早く発見することが可能となることが期待される。

本論文で提案するシステムでは、3.3で示した症状と病名との関係性のみに基づいて病名のランキングを行っているため、図7のように、“薬剤性腎障害”や“川崎病”などの本来発生数が少ない病名も上位にランキングされている。本システムの実用化を目指す上では、各病名の発生頻度もランキング算出に反映することで、ユーザの疾患として可能性の高い病名をより上位に呈示するよう、改良を行っていく必要があると考えられる。また、症状に関する文章中には“軽度の疲労”，“軽い頭痛”など、症状の程度が低いことが、その疾患の可能性が高いことを表している場合もあるため、これらの病名ランキング基準についても今後検討していく必要がある。

5. おわりに

本論文では、ユーザが入力した症状を基に、関連する症状を呈示することで自覚症状の発想支援を行い、複数の症状およびそれらの程度情報を考慮して病名の呈示を行う病名検索支援システムを提案した。また、診療科情報を用いて病院・診療所の絞り込みを行い、ユーザの位置情報と病院の所在地情報に基づき近くの病院・診療所を地図上に表示する病院検索支援システムを提案した。初めに、HK Graphを用いて、症状から病名を推測する病名検索支援システムについて述べ、Webページや電子辞書の文章中から、症状として用いられる語句を自動で抽出する手法を提案した。本システムの特徴の一つは、インターラクティブなやりとりにより病名検索を進められる点にあり、ユーザの症状に合わせて程度情報を与えたり、呈示された症状を新たに入力したりすることで、病名ランキングに症状の程度を反映させることが可能である。次に、症状の程度情報を考慮した呈示病名のランキング手法、およびユーザの位置情報や病院・診療所の診療科情報などを考慮した病院・診療所の呈示方法について述べた。実験として、Yahooヘルスケアの症状に関するWebテキストに提案手法を適用し、症状に関する語句の抽出手法、および症状の程度情報が病名ランキングに与える影響に対する検証を行った。今後は、症状の呈示方法が発想支援に与える影響の調査、疾患発生頻度および症状文章における症状の程度情報の病名ランキングへの反映、診療科以外の情報を用いた病名と病院との関連づけなどを行っていく予定である。また、

これらの改良を行った上で、実際の病院のカルテ情報を用いて病名ランキングの妥当性を検証することや、実際にユーザに病院検索支援システムを利用してもらい、本システムの実用性を検証するとともに、実用化に向けたインターフェースの改良等を行っていきたい。

参考文献

- [1] 内田健夫：“メタボリックシンドローム健診に対する日本医師会の考え方”，医学書院（2007）
- [2] 河内卓ら：“ガンを防ぎ治す本－ガンの早期発見と最新治療法・生活の中の予防法・告知・心の問題・医療機関”，婦人生活社（1993）
- [3] 尾久征三、永田頌史：“職場のメンタルヘルスをどうすすめるか（10）二次予防－早期発見、早期対応”，中央労働災害防止協会，Vol.58, No.10, pp.73-75 (2007)
- [4] 松井宏夫：“医療シリーズ特集生活習慣病と病院さがし患者に役立つ実力ある専門医（1）生活習慣病予防・早期発見を目指す内科専門医”，毎日新聞社，Vol.82, No.46, pp.69-71 (2003)
- [5] 岡田了三：“成人病（生活習慣病）とその早期発見法”，Vol.2, No.2, pp.47-49 (2000)
- [6] 平松治彦、宮本正喜、上原邦昭：“部門間の関係に基づくインシデントレポート共有システムに関する研究（インターネット応用）”，Vol.106, No.150, pp.1-6 (2006)
- [7] 廣田雄大、山下智也、周藤安造：“Web アプリケーション方式による画像診断連携システムの開発（病院情報システム/画像表示）”，Vol.106, No.343, pp.59-63 (2006)
- [8] クチコミ病院情報：<http://www.kuchikomi.tv/>
- [9] お医者さんガイド：<http://www.10man-doc.co.jp/index.html>
- [10] Yahoo ヘルスケア：<http://health.yahoo.co.jp/index.html>
- [11] Goo ヘルスケア：<http://health.goo.ne.jp/>
- [12] 岡部貴博、吉川大弘、古橋武：“インシデントレポート解析のための多重接続型階層的テキストマイニング手法の提案”，第22回ファジィシステムシンポジウム，pp.211-214, 2006
- [13] 松尾比呂志、内野一：“意味属性に基づくテキストベース検索方式”，情報処理学会論文誌，Vol.32, No.9, pp.1172-1179, 1991
- [14] 大澤幸生、ネルスE.ベンソン、谷内田正彦：“語の共起グラフの分割・統合によるキーワード抽出”，電子情報通信学会誌，Vol.J82-D-I, No.2, pp.391-400, 1999
- [15] 工藤拓、松本裕治：“チャンキングの段階適用による係り受け解析”，情報処理学会論文誌，Vol.43, No.6, pp.1834-1842, 2002
- [16] Yahoo Web API：<http://developer.yahoo.co.jp/>
- [17] 渡辺純夫：“症状から診る内科疾患”，メジカルビュー社（2005）

Hao BO (非会員)



2007年名古屋大学工学部電気電子情報工学科卒業、同年、名古屋大学大学院工学研究科計算理工学専攻博士前期課程に入学、現在に至る。テキストマイニングに関する研究に従事。

古橋 武 (非会員)



昭和60年名古屋大学大学院工学研究科博士後期課程電気系専攻修了。工学博士。平成16年名古屋大学大学院工学研究科計算理工学専攻教授。現在に至る。ソフトコンピューティング、感性工学に関する研究に従事。平成8

年日本ファジィ学会論文賞受賞。IEEE、日本知能情報ファジィ学会、電気学会等の会員。

吉川 大弘 (正会員)



1997年名古屋大学大学院博士課程修了。同年カリフォルニア大学バークレー校ソフトコンピューティング研究所客員研究員。1998年三重大学工学部助手。2005年名古屋大学大学院工学研究科COE特任准教授。2006年10月同研究科准教授。現在に至る。主としてソフトコンピューティングとその応用に関する研究に従事。博士（工学）。

杉浦 伸一 (非会員)



昭和60年名城大学医療薬学専攻科修了。名古屋大学医学部にて医学博士取得。平成15年名古屋大学医学部附属病院医療経営管理部講師、平成18年予防早期医療創成センター講師兼任。現在に至る。Geographic Information SystemおよびMobile Infra Structureと医療情報の統合に関する研究に従事。平成12年JSPENリサーチフェローシップ、平成16年ファイザーヘルスリサーチ財団海外派遣。