

対話型病状判定システムのための電子カルテを用いた 病名への重み付けの効果検証

Validation of weighting diagnoses based on electronic medical records for an interactive decision support system

山内翔大^{1*} 岡本和也¹ 平木秀輔¹ 杉山治¹
山本豪志朗¹ 佐々木博史¹ 南部雅幸¹ 黒田知宏¹

Yamauchi Shota¹ Okamoto Kazuya¹ Hiragi Shusuke¹ Sugiyama Osamu¹
Yamamoto Goshiro¹ Sasaki Hiroshi¹ Nambu Masayuki¹ Kuroda Tomohiro¹

¹ 京都大学

¹ Kyoto University

Abstract: There are people who neglect their symptoms and do not go to the hospital. Therefore, we made an interactive decision support system which determines whether the symptoms of a user are dangerous. However, a series of inquiries to predict the symptom's severity could be inappropriate because the system assumes that the frequencies of all diseases constant. To solve this problem, we suggested weighting diagnose based on the frequency of each diagnose in electronic medical records. In the experiment to validate the suggested method, by giving diagnoses weights, the number of questions required to make a decision was reduced.

1 はじめに

体調不良を自覚する人々が受診を怠り、結果として症状が重篤になってしまうケースが存在する。したがって、体調不良を自覚する人々の症状が重篤になる前に受診の必要性を提示することができれば、受診の動機付けができると考えられる。

体調が悪いにも関わらず人々が受診をしない最も大きな原因は、病院に行くのが面倒だと感じることである [1]。そのため、我々はユーザに面倒さを感じさせないよう、時間的負担が小さくなるような対話型病状判定システムの構築を試みた。

構築したシステムは、いくつかの質問を繰り返すことでユーザの自覚症状を引き出し、症状に対応する病名群を抽出する。ユーザの症状に基づき疾病を判定するシステムに関する研究はいくつかされており、あらかじめ決められたルールに則り病名を抽出するシステムを構築する研究 [2] や、症状、生体情報などの多様な特徴を入力としたニューラルネットワークモデルを利用して医師の診断の補助を行う研究 [3] などが存在する。また、代表的な対話型病状判定システムとして株式会社メドレー提供の「症状チェッカー」¹⁾ や Ubie 社

提供の Dr.Ubie²⁾ が存在する。本研究とこれらの研究やシステムとの相違点は、抽出した病名群の持つ「緊急度」という独自に定義した値に基づき、受診すべきか否かを出力する点である。「緊急度」とは、候補病名群中の各病名に対し与えられた「緊急性」があるか否かの 2 値ラベルに基づき候補病名群に付与される指標である。構築したシステムは、その時々状態において候補病名群の「緊急度」に基づく情報利得が最も大きくなるような症状の有無をユーザに質問する。しかし、構築したシステムは、各病名に割り当てられた「緊急度」が各疾病の出現頻度を反映しておらず、実世界に使用する上での最適化はなされていないと考えられる。現実において、各疾病は等確率で発生するものではない。したがって、システム内のデータセット中の各病名に対し、疾病の発生頻度に基づく重み付けを行った後に「緊急度」を計算することにより、頻出病名を疾患として持つユーザに対する判定に必要な質問数を削減することが可能であると推測される。

本研究の目的は、各病名に発生頻度に基づく適切な重みを付与することにより、対話型病状判定システムのユーザへの質問数を削減することである。

*連絡先：京都大学医学部附属病院
〒606-8507 京都市左京区聖護院川原町 54
E-mail: medinfo@kuhp.kyoto-u.ac.jp

¹⁾ <https://medley.life/>

²⁾ <https://www.company.dr-ubie.com/>

2 対話型病状判定システム

本節では、我々が構築した対話型病状判定システムについて説明する。システムはユーザからの最初の入力症状及びユーザに対して行う質問へのユーザの回答を元に病名を絞り込む。そして、候補病名に「緊急性」の高い病名しかない状態になれば受診を促し、「緊急性」の低い病名しかない状態になれば問題ない旨を出力する。そして、「緊急性」が高い病名と低い病名が混在する状態であれば、情報利得に基づき新たに質問を行う。また、「緊急性」が高い病名と低い病名が混在する状態であっても、候補病名内の各疾病の症状が全て一致し、質問によってこれ以上病名を絞り込むことができないかつ、候補病名内に一つでも緊急度が高い病名が存在する場合は受診を促す。

2.1 緊急度

本項では、本システムにおける「緊急度」について述べる。まず、はじめに各疾病に緊急を要する疾病か否かを識別する2値ラベルを付与する。例えば、「くも膜下出血」は緊急度が高いため、「緊急性」ラベルを True に、「偏頭痛」は緊急度が低いので、「緊急性」ラベルを False に設定する。そして、「緊急性」ラベルと、「緊急性」ラベルごとに分類された病名数のペア集合を「緊急度」と定義する。例えば図1においては、「緊急性」ラベルが True の病名が4つ、ラベルが False の病名が4つ存在することから「緊急度」は $([True, 4], [False, 4])$ と表現される。

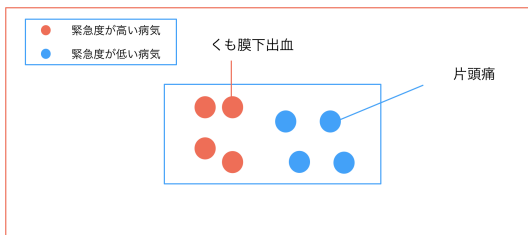


図1: 本システムにおける緊急度

2.2 情報利得

情報利得は、あるデータを分割する際、分割前データの情報エントロピーに比べて分割後データの情報エントロピーがどの程度減少するかを表現する値である。本稿における情報エントロピー I は、データ内の全病名数に対する、緊急度の True に対応する病名数の割合 d を用いて以下の式で表現される。

$$I = -d \log_2 d - (1-d) \log_2 (1-d) \quad (1)$$

また、情報利得 IG は以下の式で表現される。

$$IG = I_p - c_r * I_r - c_l * I_l \quad (2)$$

ここで、 I_p は分割前のデータにおける情報エントロピー、 I_r 、 I_l は2分割後のデータそれぞれの情報エントロピー、 c_r 、 c_l は分割前データ量に対する2分割後のそれぞれのデータ量の割合である。情報利得は、データ分割により増える情報量を意味する。したがって、有無を質問することにより情報利得が一番大きくなるような症状を、ユーザに質問することにより効率的に情報を増やし、ユーザが緊急度の高い症状を持つかどうかを素早く判断することが可能となる。例えば、図2と図3にそれぞれ「肩こり」と「意識障害」の有無を質問した場合の情報利得の計算例を記す。この2例を比較した場合、「肩こり」の有無を質問する方が「意識障害」の有無を質問するよりも情報利得が大きくなる。したがって、「肩こりがありますか？」という質問をユーザに問う事になる。

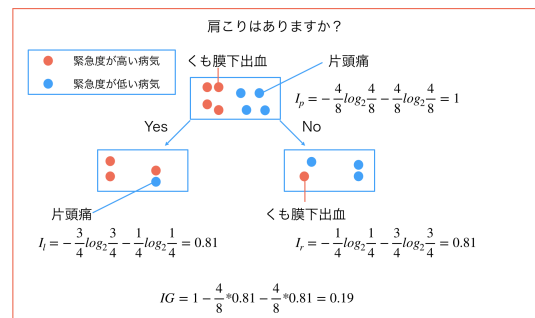


図2: 情報利得が大きくなる場合

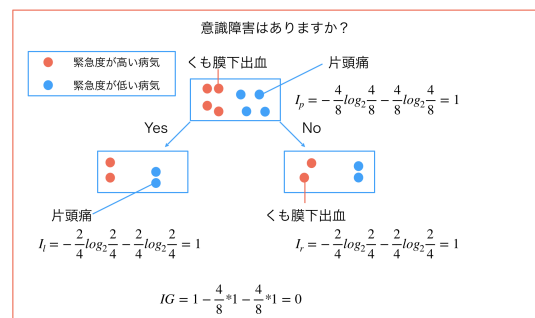


図3: 情報利得が小さくなる場合

3 病名への重み付け手法

本節では、病名に対する重み付け手法を説明する。本研究で提案する重み付けには、京都大学医学部附属病

院の電子カルテの病名オーダーにおける各病名の登場回数を使用する。重み付け手法1は各病名の出現頻度をそのまま重みとする手法である。システム内データセット中に存在する病名に対して、2005年1月～2018年11月の間に京都大学医学部附属病院に蓄えられた電子カルテの中に出現した病名オーダー回数を重みとして付与する(表1)。なお、出現回数が0の病名に対しては一律に重みを1と設定する。重み付け手法2は、手法1の重みに対し、自然対数をとった後に小数点以下を切り捨て、1を加算した値を重みとする手法である(表2)。手法1では、重みの値域が55614～1までの広い範囲をとってしまう。したがって、対数を取る事により値域の範囲を狭くすることが手法2の狙いである。

表 1: 重み付け手法1で用いる各病名の重み

病名	重み
胃潰瘍	55614
便秘症	42150
糖尿病	35580
逆流性食道炎	28279
アレルギー性鼻炎	26359
・	・
・	・
・	・
・	・
急性尿酸性腎症	1
ヤング・シンプソン症候群	1

表 2: 重み付け手法2で用いる各病名の重み

病名	重み
胃潰瘍	11
便秘症	11
糖尿病	11
逆流性食道炎	11
アレルギー性鼻炎	11
・	・
・	・
・	・
・	・
急性尿酸性腎症	1
ヤング・シンプソン症候群	1

4 評価実験

本研究の目的は、各病名に出現頻度に基づく重みを付与する事により、対話型病状判定システムのユーザへの質問数を削減する事である。したがって、以下の3種類のデータセットを用いて病状判定システムを実行し、病状判定に必要な質問数の定量的評価を行う。

1. 重みなしデータセット
2. 病名に出現頻度を重みとして付与したデータセット
3. 病名に出現頻度の自然対数をとった後、小数点以下を切り捨てた値に1を加算した値を重みとして付与したデータセット

4.1 システムデータセット

本稿では、病状判定支援システムの内部データセットとして臨床医学連結知識データベース LiLak[4]中に存在する病名症状所見関連テーブルを使用した。病名症状所見関連テーブルに格納されている病名と症状の対応関係を用いて、index を病名、column を症状とする病名症状対応行列を作成した(表3)。次に、ユーザ

表 3: 病名症状対応行列

	腹痛	頭痛	...	脱力感	意識障害
胃潰瘍	1	0	...	0	0
心不全	1	1	...	1	1
・	・	・	...	・	・
・	・	・	...	・	・
・	・	・	...	・	・
・	・	・	...	・	・
下痢症	1	0	...	0	0
脳梗塞	0	0	...	0	1

の自覚できる症状のみを使用するために、他覚症状を病名症状対応行列から除外する。LiLakに存在する症状所見テーブルでは、以下のように症状が分類されている。

1. 自覚のみ
2. 他覚のみ
3. 両方

この情報を用いて、病名症状対応行列中の症状から他覚のみで観測できる症状を除外した。さらに他覚のみ

の症状の除外により対応する症状がなくなった病名についても行列中から除外した。

次に、各病名の緊急性判定を行う。本稿では、株式会社メドレーが公開している Web サービス「医療辞典」³⁾において「緊急度高」と見なされている病名に対し、「緊急性」があると判定した。また、LiLak には存在するが医療辞典には存在しない病名においては「緊急性」はないと判定した。

4.2 評価実験プロトコル

本節の冒頭で記述した 3 つのデータセットに対し、2 つの評価実験を実施した。本項では、実験内容を説明する。

4.2.1 評価実験 1

評価実験 1 では、以下のフローに従い実験を行った。

1. 病名症状対応行列中の病名集合 I から病名 i_j ($|I| = n, 1 \leq j \leq n$) を取り出す。
2. i_j に対応する症状集合 S_j からランダムに症状 s_{jk} を取り出す。 ($s_{jk} \in S_j$)
3. 病名 i_j をもつユーザが症状 s_{jk} を入力したとしてシステムを起動する。
4. 情報利得に基づく質問を繰り返し、病状判定が終了した段階での質問数 q_j を計測する。
5. 1~5 を I 中の全病名について行う。
6. 質問数の平均である $Q_{ave} = \frac{1}{n} \sum_{j=1}^n q_j$ を求める。

なお、3 種類の実験対象のうち、1 種類でも質問数が 0 になるケースにおいては、他の 2 種類でも質問数が 0 になることから、質問数が 0 となるケースについては評価対象から除外した。この実験により、各疾病が均一確率で発生した場合における、ユーザに対する質問数の平均を比較する。

4.2.2 評価実験 2

評価実験 2 では、以下のフローに従い実験を行った。

1. 病名症状対応行列中の病名集合 I から病名 i_j ($|I| = n, 1 \leq j \leq n$) を取り出す。
2. i_j に対応する症状集合 S_j からランダムに症状 s_{jk} を取り出す。

3. 病名 i_j をもつユーザが症状 s_{jk} を入力したとしてシステムを起動する。
4. 質問を繰り返し、病状判定が終了した段階での質問数 q_j を計測する。
5. 質問数 q_j に i_j の京都大学医学部附属病院の電子カルテ内での登場回数 N_j をかけた値 $Score_j = q_j * N_j$ を求める。
6. 1~5 を I 中の全病名について行う。
7. 各疾病が電子カルテにおける病名の登場頻度で発生したと考えた場合の、ユーザに対する質問数の平均である $Q_{ave} = \frac{1}{\sum_{j=1}^n N_j} \sum_{j=1}^n Score_j$ を求める。

評価実験 1 と同様に、3 種類の実験対象のうち、1 種類でも質問数が 0 になるケースにおいては、他の 2 種類でも質問数が 0 になることから、質問数が 0 となるケースについては評価対象から除外した。この実験により、各疾病が電子カルテにおける病名の登場頻度で発生したと考えた場合の、ユーザに対する質問数の平均を比較する。例えば、登場頻度から考えて、ヤング・シンプソン症候群のユーザが 1 人出現する間に胃潰瘍のユーザが 55614 人出現すると考えられる。したがって上記フロー 5 において、胃潰瘍のユーザがシステムを起動したと仮定した場合には 55614 人が使用したとみなして質問数に 55614 を掛ける。

4.3 実験結果

本項では本節の冒頭で記述した 3 種類のデータセットに対する 2 種類の実験の結果を記述する。

4.3.1 評価実験 1 の結果

実験結果を表 4 に示す。

表 4: 評価実験 1 の結果

重み	なし	登場回数	$\log(\text{登場回数})$
平均質問数	2.92	2.98	2.90

一番平均質問数が小さくなったのは電子カルテにおける出現回数の自然対数を重みとして各病名に付与したデータセットであった。また、出現回数を重みとして付与したデータセットを用いた場合には、重みなしのデータセットを用いた場合よりも質問数が増加した。

³⁾<https://www.medley.jp/service/medley.html>

4.3.2 評価実験 2 の結果

実験結果を表 5 に示す。

表 5: 評価実験 2 の結果

重み	なし	登場回数	log(登場回数)
平均質問数	2.71	2.59	2.84

一番平均質問数が小さくなったのは電子カルテにおける出現回数を重みとして各病名に付与したデータセットであった。また、出現回数の自然対数を重みとして付与したデータセットを用いた場合には、重みなしのデータセットを用いた場合よりも質問数が増加した。

5 考察

提案した手法が本研究の目的である病状判定システムの質問数削減に貢献したかどうかを実験結果から確認する。

評価実験 1 の結果から、全ての疾病が等しい確率で発生する環境においては、各病名の出現回数の自然対数を重みとして付与したデータセットが最も適していることがわかった。しかし、平均質問数において重みなしデータセットとの差は僅かしかなく、誤差であると考えられる。さらに、最小値が 0、最大値が 55614 という値域の大きい登場回数を重みとした場合の平均質問数が増加してしまっていることから、この環境下では値域の大きい重み付けによって質問数が増加してしまうことが示唆された。

次に、評価実験 2 の結果から、実世界における各疾病の発生確率を考慮した環境においては、各病名の出現回数を重みとして付与したデータセットが最も適していることがわかった。また、重みなしデータセットと、出現回数を重みとしたデータセットとの間では平均質問数に約 4.5 % の差が生まれていることから、実世界環境においては、実世界環境における病名の頻度を重みとして付与する事で実世界環境に適したデータセットを作ることが可能であると示唆された。

また、評価実験 1 よりも評価実験 2 の方が 3 種類全てのデータセットにおいて平均質問数が減少した。このことから出現回数が多い疾病の方が平均質問数が少なくなる傾向があることが示唆された。特に、各病名に出現回数をを重みとして付与したデータセットにおいては検証実験 1 と比較して検証実験 2 の方が平均質問数において約 13 % の減少が見られた。この結果から、各病名に出現回数をを重みとして付与したデータセットを用いることにより実世界において登場頻度が

高い病名をもつユーザに対してシステムが行う質問数が削減できることが示唆された。

次に、実際にシステムを利用した際のいくつかの質問フローに対する考察を行う。表 6、表 7 に質問フローの例を示す。

表 6、表 7 の中の左の列は、カルテ中での各病名の出現回数をそのまま各病名に重みとして付与したデータセットを用い、右の列は、重み付けを行わなかったデータセットを用いた場合の質問フローである。

表 6: 「骨粗鬆症」のユーザが「腰痛」を入力した場合の質問フロー

重み	出現回数	なし
患者の病名	骨粗鬆症	骨粗鬆症
患者の入力症状	腰痛	腰痛
質問 1	腰部圧痛	脊椎肋骨角叩打痛
回答 1	no	no
質問 2	脊椎肋骨角叩打痛	無症状
回答 2	no	no
質問 3	嚥下困難	
回答 3	no	
	終了	終了

表 7: 「狭心症」のユーザが「冷汗」を入力した場合の質問フロー

重み	出現回数	なし
患者の病名	狭心症	狭心症
患者の入力症状	冷汗	冷汗
質問 1	上腹部痛	心窩部痛
回答 1	yes	no
質問 2	片頭痛	胸痛
回答 2	no	yes
質問 3		チアノーゼ
回答 3		no
質問 4		顔面蒼白
回答 4		yes
質問 5		片頭痛
回答 5		no
質問 6		心原性ショック
回答 6		no
	終了	終了

表 6、表 7 から、データセット中の病名に重みを付与することにより質問数だけではなく、質問内容が変化していることがわかる。また、1 名の医師に表 6、表 7 を見てもらったところ、以下のような感想をもらった。表 6 について

1. どちらかという重みありの質問遷移のほうが自然である。
2. 重みありのデータセットを用いた方では骨粗鬆症を疑って質問をしているように思える。
3. 重みなしのデータセットを用いた場合には、腎盂腎炎を疑って質問をしているように思える。

表7について

1. 疑われる病名の頻度を考えると狭心症の方が心筋梗塞よりも頻度が高いので、重みありの方が自然だと思われる。
2. 冷汗と心窩部痛は心筋梗塞の典型的な症状である。重みなしの方は心筋梗塞を疑っているのがよくわかる。医師の思考と近いのは重みなしのデータセットを用いた場合かもしれない。

医師の感想に対する考察を行う。まず、表6の例についてであるが、電子カルテにおける出現回数は、骨粗鬆症が21764回、腎盂腎炎が2277回である。病状判定システムが、重みありのデータセットを用いた場合には骨粗鬆症を疑い、重みなしのデータセットを用いた場合には腎盂腎炎を疑う質問を行っているという医師の考察から、データセット中の病名に対して重みをつけることにより、システムは疾病の発生頻度に基づいて疑うべき病名を定め、その病名にたどり着くような質問を選定していることが示唆された。次に、表7の例についてであるが、電子カルテ中における出現回数を比較すると、狭心症が23315回、心筋梗塞が1707回であり、狭心症の方が頻度が高い。病状判定システムが、重みありのデータセットを用いた場合には狭心症を疑い、重みなしのデータセットを用いた場合には心筋梗塞を疑う質問をしていることから、表6の例と同様に、データセット中の病名に対して重みをつけることにより、システムは疾病の発生頻度に基づいて疑うべき病名を定め、その病名にたどり着くような質問を選定していることが示唆された。しかし、表7の例における医師の感想にあるように、重みをデータセットに付与することにより医師の思考とやや離れた質問フローをたどる場合があることが示唆された。この差異は、病名への重み付けを行ったデータセットを用いた病状判定システムにおいては出現頻度だけに特化して質問を選定している一方で、医師が疾病の頻度だけでなくその危険性も考慮して患者に問診を行うことから生まれる差異だと考えられる。

6 おわりに

本研究では、我々が実装した病状判定システムにおいて、出現頻度が高い疾病をもつユーザが使用した場

合の、病状判定に必要な質問数の削減を試みた。課題解決にあたり、システム内のデータセット中の病名に出現頻度に基づく重み付けを行うことを提案した。実際にデータセット中の病名に出現頻度に基づく重み付けを実施し、重み付けの前後で質問数に差異が生まれるかを検証した。結果、出現頻度に比例する重みを付与することにより出現頻度が高い疾病をもつユーザの使用時には、システムがユーザに行う質問数が削減されていることが確認できた。また、質問内容についても、出現頻度が高い疾病を疑うような内容を質問していることが示唆された。

参考文献

- [1] メディカルライフ研究所: 「生活者の“受療行動”※に関する調査」を実施, http://www.medicallifelab.jp/cont/pdf/release_20130328.pdf, (2013年3月28日:最終閲覧日2019年2月22日)
- [2] Jimmy Singla: The Diagnosis of Some Lung Diseases in a Prolog Expert System, *International Journal of Computer Applications*, vol. 78, no. 15, pp. 37-40, September 2013.
- [3] Filippo Amato, Alberto Lopez, Eladia Mara Peamendez, Petr Vahara AleS HampF, Josef Havel: Artificial neural networks in medical diagnosis, *J. Appl. Biomed.* 2013, 11, 4758.
- [4] 大江和彦: 臨床医学連結知識データベースを用いた問診と診断想起システムの考察, *Consideration of medical interview and a system for evoking diagnoses with the use of The LiLak :Linked Clinical Knowledge Database*, 医療情報学会・人工知能学会 AIM 合同研究会資料, SIG-AIMED-002-08,2016