

アクセスログを用いた診療文書間の関係抽出

澤野 良輔[†] 吉川 正俊^{††} 濱崎 暁洋^{†††} 岡本 和也^{††††} 黒田 知宏^{††††}

[†] 京都大学工学部情報学科 〒6068501 京都市左京区吉田本町

^{††} 京都大学大学院情報学研究科 〒6068501 京都市左京区吉田本町

^{†††} 京都大学医学部附属病院 糖尿病・内分泌・栄養内科 〒6068507 京都市左京区吉田本町 36-1

^{††††} 京都大学医学部附属病院 医療情報企画部 〒6068507 京都市左京区吉田本町 36-1

E-mail: [†]sawano@db.soc.i.kyoto-u.ac.jp, ^{††}yoshikawa@i.kyoto-u.ac.jp,

^{†††}hamasaki@metab.kuhp.kyoto-u.ac.jp, ^{††††}{kazuya,tomo}@kuhp.kyoto-u.ac.jp

あらまし 電子化された診療文書（いわゆる電子カルテ）の普及と共に、日々蓄積されるデータの二次利用への期待が高まっている。医師が毎日の業務以外に、学習や教育用に電子カルテを閲覧することがあるが、現行のシステムではIDや名前などにより患者のカルテを検索出来るものの、自分の知らない患者が一覧として表示される膨大な検索結果から、新たに診療の参考にしたい文書を見つけ出すことは困難であり、新しい手法が求められる。我々は電子カルテの閲覧を記録した既存のアクセスログを用いて、医師が患者のカルテを閲覧した事実を二部グラフで扱い、その枝の重みを入院患者の退院後に残るアクセスログに基づいて設定し、HITSを応用することにより、学習の参考になるカルテを抽出する手法を提案する。

キーワード 情報検索, アクセスログ, 診療文書, HITS

1. はじめに

日本でも医療文書が電子化され、各医療機関でもいわゆる電子カルテとして、大量のデータが扱えるようになった。一方で、未だ医療現場では電子カルテを有効に活用する仕組みは十分に整っていないのが現状である。現在は各医療機関内で患者のカルテを蓄積しているが、地域医療ネットワークであるEHR(Electric Health Record)の普及とともに、他の病院に蓄積されているカルテまでもが益々参照されることになるであろう[1]。大量の電子カルテの中から、有用な情報を見つけ出すことが求められる。

現状、医療機関に所属する職員（医師など）のニーズとして、

- 治療にとって参考になる電子カルテを参考にしたい。
- 日々の業務レポートを書くため、または研修医の教育のため、書き方の参考になるカルテを見つけ出したい。

という電子カルテを活用する上での主に二つのニーズが医師へのアンケートから存在することが分かっている[2]。

現行のシステムでは、プライバシー上の懸念もあり患者のカルテを柔軟に検索することはできず、患者のIDや現在入院している患者一覧などの定まった検索方法しか利用できていない。また、たとえ病名でカルテを検索出来るようになったとしても、数多く出てくる検索結果からどれが重要であるかをひと目で評価するのは医師でも難しく、計算機がカルテの内容を元に判断を行うのはさらに難しい。本研究ではカルテに記載されている内容ではなく、電子カルテシステム上に記録されるカルテへのアクセスログを利用することで、内容を解析することなく、今までに閲覧された事実によってカルテの重要度を評価することを目的とする。

定期的に行われるケースカンファレンス^(注1)にて紹介された患者や、以前自分が処置をした患者をもう一度見直すといったように、医学的に重要である症例を電子カルテを用いて勉強することは比較的少ないながらも存在する一方、そのような学習のためにアクセスしたログは、普段の業務によって記録されるログにより埋没してしまっているのが現状である。埋没したログの中から、有益なカルテを発掘することが本研究の課題である。

本研究の主な貢献は次の通りである。現在は二次利用目的のアクセスが少ないため、全体のアクセスログから二次利用目的のアクセスをマイニングして活用することが難しく、益々二次利用目的のアクセスが現れないという悪循環に陥っている。しかし、本研究は

- 二次利用アクセスは少ないながらも存在し、それらを注意深くマイニングすれば活用できることを示した。

また、既存研究で得られたアンケート結果を元に実験を行い、

- HITSアルゴリズムを元にした既存研究の手法を応用し、後述するアンケートから得られた4つの評価項目のうち学習すべき度合いを測る項目において改善が得られた。

- ケースカンファレンスで紹介されたカルテをマイニングする実験を行い、提案手法によりランキングした結果が、単純なHITSでランキングした結果よりも高い適合率が得られることを確認した。

本稿の構成は次の通りである。2. では、関連研究を取り上げる。3. では、実験系である電子カルテシステムの概略と、扱うアクセスログについて説明する。4., 5. にて、電子カルテシステ

(注1): 事例研究の一方法。医療現場にて起こった解決すべき内容を含む事実について、その状況・原因・対策を明らかにするために報告・共有する場である[3]。

ムが扱われる医療現場を想定してモデル化を行い、重要度を測るための手法を提案する。6. は、行った評価実験の説明である。鶴岡らが行ったアンケート結果 [2] を元に比較評価を行うため、行われたアンケートの概要説明と評価指標を説明する。また、ケースカンファレンスで紹介された患者を元に実験を行ったので、その説明もここです。実験結果を 7. にて説明し、8. で考察する。9. において本稿のまとめと巻末に謝辞、参考文献が付けられている。

2. 関連研究

病院の電子カルテの中から有用なカルテを抽出する研究として、鶴岡らの研究 [2] が挙げられる。鶴岡らは Web 文書のランキング手法として有名な HITS アルゴリズム [4] を応用して、医師と患者の重要度を評価した。他に病院のアクセスログを用いた研究としては、Zhang らによるもの [5] があるが、ログから役割を推定することによるアクセスコントロールや業務改善が目的であり、情報推薦が目的ではない。

数多くの電子カルテの中から一つの電子カルテの閲覧を推薦するのではなく、与えられた電子カルテの内容を元に、診断の推薦を行うシステムも研究されている。IBM の人工知能 Watson は、人のゲノム配列と電子カルテに記載されている内部情報を元に、ゆくゆくは医師に医学的根拠とともに処置の推薦を行おうとしている [6]。

また、患者の社会的繋がり (例えばどこの医療機関に通い、担当医は誰であるか等) も含めた基本情報や電子カルテ、各種検査値などを入力として監視することで、患者に起こりえる疾患のタグ付けや警鐘を出す。加えて、より柔軟な検索クエリの使用を実現し、内部情報を元に似た患者を求めることの出来るシステムも現れた [7]。

電子カルテへのアクセスログから離れば、広告を目的として、Web 検索のクエリと結果ページへのアクセスログを利用して重要度を測る研究は数多くされている。重要度を測るものでは他に PageRank [8] が有名である。

3. 電子カルテシステムに蓄積される情報の詳細

この章では電子カルテシステムの概略と、本研究で扱うアクセスログの詳細についてまとめる。

3.1 電子カルテシステムまでの病院情報システムの変遷

『医療情報サブノート』(篠原出版新社) [9] より病院情報システムの歴史に関する記述を抜粋すると、

病院でコンピュータが導入されたのは、医事会計システムである。その後すぐに、検査部で自動分析装置を制御するためにコンピュータが導入された。診察室から紙の伝票で依頼情報を伝達していたが、コンピュータネットワークにより電送するシステムに置き換えられたのが、オーダエントリシステムである。ここで初めて病院情報システムという言葉が生まれることになる。検査や処方といった一部の処理がコンピュータによって行われるようになった。その後、電子カルテシ

患者ID	12345678
職員ID	N0123456
アクセス開始年月日	2005/01/13 0:00:00
アクセス開始時間	08:29:07
アクセス端末番号	FK01
アクセス終了年月日	2005/01/13 0:00:00
アクセス終了時間	08:29:07
アクセス文書種別・操作	120R, 130R, 130U

図 1 アクセスログの項目とその内容

ステムの登場により、紙の診療録が電子媒体に置き換わるようになった。

とある。京都大学医学部附属病院における電子カルテシステムは、2004 年にそれまで稼働していたオーダエントリシステムに付け加えられる形で導入された。

3.2 電子カルテが記載する内容

電子カルテシステムで扱われる各々の電子カルテが記載する内容は、患者の名前や年齢といった基本情報から、経過記録、各種検査結果、画像など様々である。(電子カルテの規格については例えば MML [10] を参照されたい。) 本稿ではその中でも退院時サマリーという特定の文書に着目する。退院時サマリーは、ある入院患者の退院時に記載される要約であり、退院時サマリーだけを見て、医学的に学ぶべき典型例であるかどうかを判断可能であるため実験評価がしやすい利点がある。

3.3 アクセスログが記録する項目

まず、アクセスログを構成する項目とその例を図示する (図 1)。アクセスログを見れば、誰が・いつからいつまで・どの患者のカルテの・何の項目を参照したかがわかる。特に、アクセス文書種別・操作における 120R, 130U といった文字列を見ることで、どの電子カルテの項目に新規書き込み/更新/閲覧があったかを把握することが出来る。

3.4 本稿で取り扱うログの種類とその件数

京都大学医学部附属病院では、本番系と参照系という 2 つの電子カルテシステムが存在し、各々に対してログをとっている。二つの電子カルテシステムは本番系から参照系へ書き込みがあるごとに同期を行うので、参照できるデータはほぼ同じである。二つの違いは可能な操作系列であり、本番系では閲覧以外に新たに書き込みや更新を行うことができるが、参照系では閲覧しかできない。各々のシステムに対して、異なるアクセスログが蓄積されており、本番系のアクセスログ総数は参照系に比べて約 440 倍と多く記録されている。情報に大きな差があるため (図 2)、本研究では記録の多い本番系の電子カルテシステムを取り扱い、2004 年から 2012 年 12 月末迄蓄積されたアクセスログを元に解析を行う。

本番系で集積されたログ総数は 105,810,109 件であり、閲覧者は 9,445 名であった。閲覧する職員は、医師を始めとして看護師、薬剤師、大学教員、一般職員など様々な役職から構成されるが、アクセスログ上には項目として役職名は用意されておらず他のテーブルと照合する必要がある。本研究で提案する手法は、

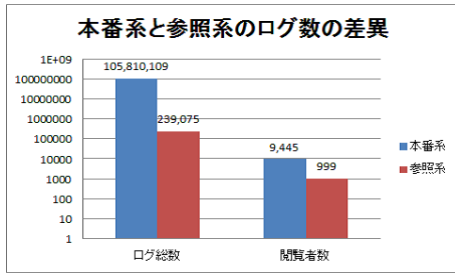


図 2 本番系と参照系のログ数 (対数目盛)

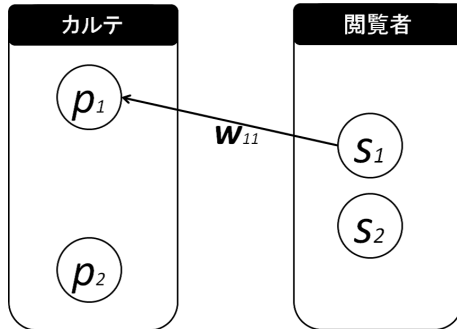


図 3 カルテと閲覧者による閲覧グラフ

図 1 で表記された項目のみを扱い、医師などの職員の職種を考慮せずにまとめて閲覧者と呼ぶことにする。

4. 患者のカルテと閲覧者による閲覧グラフ

定義 1 (閲覧グラフ)

カルテの集合 P と閲覧者の集合 S を考える。ある閲覧者 $s \in S$ があるカルテ $p \in P$ を閲覧したときにノード s と p の間に有向枝 $\langle s, p \rangle$ を張り、このような有向枝の集合 $E = \{\langle s, p \rangle \mid s \text{ が } p \text{ を閲覧した}\}$ を考える。枝 $\langle s_i, p_j \rangle$ には、 s_i が p_j を閲覧した回数や時間に応じて重み w_{ij} が付与される。各枝に重みを与える関数を W とする。このとき、重み付き二部グラフ $(P \cup S, E, W)$ を閲覧グラフと呼ぶ (図 3)。

あるノード $v \in P \cup S$ に対し、 v と有向枝で結ばれているノードの集合を $E(v)$ で表す。より厳密には、 $v \in S$ に対しては、 $E(v) = \{u \mid \langle v, u \rangle \in E\}$ 、 $v \in P$ に対しては、 $E(v) = \{u \mid \langle u, v \rangle \in E\}$ である。例えば、ある医師 s_1 が患者のカルテ p_1 を閲覧したとき、閲覧グラフ上で枝 $\langle s_1, p_1 \rangle$ が張られ、その重みは w_{11} である。我々は、閲覧グラフに対して Web 文書ランキングアルゴリズムを適用することで、ノードの重要度を測ろうと試みた。本稿では、ノードの絶対的重要度を測る上で HITS アルゴリズムを応用したものを提案する。

5. 重要度に関する提案手法の定式化

我々は、退院後に何度も見られる退院時サマリーは、やむなしに閲覧しなくてはならない普通の業務と比べると参考目的が強く有用である、という仮定のもと、入院の過去がある患者の退院時サマリーへの【退院した日付】から【2012 年 12 月 31 日】迄のアクセス履歴に注目した。一人の患者が複数回入院

を繰り返している場合、一回目の退院後から二回目の入院まで、等というように入院中でない期間に着目する。入院と退院の日時は、アクセスログに記載される項目のうちアクセス文書別・操作において入院予約または入院決定の書き込み、退院決定の書き込みがあった日付を元に算出した。

ノードの重要度を計算するのに、HITS アルゴリズムを応用する。閲覧グラフに対する HITS の計算式は次のように定義される。総閲覧者数を r 、総カルテ数を d とし、 x を各閲覧者の重要度 x_i を要素とする $r \times 1$ 行列、 y を各カルテの重要度 y_i を要素とする $d \times 1$ 行列とする。また、 e を全要素が 1 である $r \times 1$ 行列、 L を枝の重み w_{ij} を要素とする $r \times d$ の隣接行列としたとき、各閲覧者の重要度は

$$x = (\alpha(LL^T) + (1 - \alpha)\frac{1}{r}ee^T)x$$

と再帰的に計算する。

ここで、 α は式が収束するための 0 から 1 の間のパラメータであり、重要度 x はべき乗法を適用することで近似的に値を求めることができる。与えられた近似解 x を元に、 y は $y = L^T x$ と計算できる。

HITS の手法と同様に、 y を先に計算してから x を求めることもできる。その際は、 x と y 、 L と L^T 、 r と d をそれぞれ入れ替え、 e を全要素が 1 である $d \times 1$ 行列に入れ替えればよい。

5.1 日付による正規化

ここで、重要度を測るために抽出してきたアクセスログは、それぞれの患者の【退院日時】から【2012 年末】の間のデータなので、各患者の入院退院時期の違いから、対象となる期間が異なる。正規化のため、患者 j の退院日から 2012 年 12 月 31 日迄の日数を D_j とおき、各ノードの重み w_{ij} を w_{ij}/D_j と置き換える。

5.2 閲覧時間または閲覧項目数による枝の重み付け

隣接行列 L の要素 w_{ij} は、閲覧者から患者のカルテへの各々の閲覧をもとに重み付けられる。重要なカルテかどうかを判断するための重みを測るには幾つかの方法が考えられるであろう。例えば次のようなものが考えられる。

- カルテへの閲覧回数が多い
- 各々の閲覧時間が長い
- 各々の閲覧時間が長すぎず、短すぎない。
- 一回に沢山のカルテに記載されている項目を閲覧している

それぞれの重要と考えられる観測に従い定式化を行い、次章の実験で妥当性を検証する。

閲覧者 i がカルテ j を閲覧した回数を N_{ij} 、 $k (= 1, 2, \dots, N_{ij})$ 番目の閲覧時の閲覧時間を $T_{ij}(k)$ としたとき、本稿での定式化は次の通りである。上に挙げられた各々の観測に合うように重み付けが大きくなるような式であれば、他の式を導入しても構わない。

【閲覧回数】

$$w_{ij} = N_{ij}$$

【閲覧時間】

$$w_{ij} = \sum_{k=1}^{N_{ij}} \beta^{T_{ij}(k)}$$

【閲覧時間：中央値】

$$w_{ij} = \sum_{k=1}^{N_{ij}} \gamma \frac{|T_{ij}(k) - MED|}{\lambda}$$

MED は収集したアクセスログの閲覧時間 (秒) の中央値である。

【アクセス文書種別・操作数】

$$w_{ij} = \sum_{k=1}^{N_{ij}} \omega_k$$

where

$$\omega_k = \sum_{l=1}^{a_k} \frac{1}{2^l}$$

a_k は k 番目のアクセスにおいて、アクセス文書種別・操作で記録された個数である。例えば図 1 の場合 $a_k = 3$ である。

5.3 入院中の書き込み時間による重み付け

患者の入院中に多くの時間をかけて書き込み・更新をした閲覧者は、その患者のことをよく知っていると考えられる。患者の入院期間に全く関わりがなかった閲覧者がたまたま退院後にカルテを閲覧したものよりも、関わりの深い閲覧者が再度そのカルテを閲覧した方が作務的であり、より閲覧者と患者のカルテを結ぶ枝に意味があると考えられる。ある患者の入院中に書き込み・更新を行った閲覧者が、退院後にまたその患者の退院時サマリーを閲覧したときはその枝の重みを大きくし、逆に入院中に全く書き込んだことのない閲覧者による閲覧の重みは小さくするよう、次のように定式化する。

$r \times d$ 行列の L の各要素 w_{ij} に以下の w'_{ij} を掛け合わせる。 w'_{ij} は、患者 j の入院中に患者 j のカルテを見た閲覧者集合を $D(j)$ ($E(j)$) として

$$w'_{ij} = \begin{cases} \frac{1}{2} + \frac{1}{2^k} & (i \in D(j)) \\ 0 & (otherwise) \end{cases}$$

ここで式中の k は、患者 j の入院期間中に行われた書き込み・更新の累積時間を閲覧者ごとに求め、累積時間が多かった順に並べた時の、閲覧者 i の順位である。例えば、一番患者 j のカルテに書き込み時間をかけた閲覧者については重みが $\frac{1}{2} + \frac{1}{2^1} = 1$ 、二番目に書き込み時間が長かった閲覧者については重みが $\frac{1}{2} + \frac{1}{2^2} = 0.75$ となる。

6. 評価実験

6.1 評価指標としての医師によるアンケート

鶴岡ら [2] の行ったアンケート結果と、京都大学医学部附属病院 糖尿病・内分泌・栄養内科で開かれたケースカンファレンスで紹介された患者のカルテを元に評価実験を行い、結果を比較考察する。以下にその評価指標を説明する。

6.1.1 指標 1：アンケート

京都大学医学部附属病院に蓄積された電子カルテのうち、直腸がんを患った 52 人の患者の退院時サマリーを抽出し、事前調査により良いとされるカルテに必要な以下の 3 つの項目について、医師 1 名に 1 点～5 点の 5 段階評価を依頼した。

- (1) 必要な情報が過不足なく記載されている
- (2) 読みやすく記載されている
- (3) 学習すべき典型例、または希少例

上記の 3 項目に加え、3 項目の合計点数 (3 ~ 15 点) を 4 つ目の項目として後の評価指標とする。

- (4) 1, 2, 3 の合計点数

アンケートは、それぞれの電子カルテが学習にとって有用であるかどうかを確かめ、1. に記述した二つの電子カルテの利用目的のうち、治療にとって参考になるかどうかを評価する上での指標となる。

6.1.2 指標 2：ケースカンファレンス紹介カルテ

2 つ目の実験として、京都大学医学部附属病院の糖尿病・内分泌・栄養内科で 2009 年から 2014 年までにケースカンファレンスで紹介された患者のカルテを元に、本提案手法で得られた重要度評価の結果と一致するかを検証する。ケースカンファレンスは一週間から二週間に一回定期的に開催され、学習にとって有用であるなど、共有すべきと考えられる症例が紹介される。2009 年から 2014 年までの対象期間中、紹介された患者の件数は 178 件であった。この 178 件の患者カルテを、糖尿病薬の処方歴をもつ患者のカルテ 7493 件からマイニング出来るかを実験する。ケースカンファレンスで取り上げられた患者のカルテの中から 25 件を正解集合とし、ケースカンファレンスで取り上げられてはいないが、糖尿病の薬を一度でも処方された患者のカルテ 100 件とを合わせた 125 件のデータセットを作成した。

6.2 評価方法

6.2.1 医師によるアンケートとの相関

アンケートで得られた各項目ごとの評価値と提案手法の重要度との関連を、スピアマンの順位相関係数で評価する。スピアマンの順位相関係数 ρ は次の式で計算する。

$$\rho = \frac{T_x + T_y - \Sigma D^2}{2\sqrt{T_x T_y}}$$

$$T_x = \frac{N^3 - N - \Sigma_i (t_i^3 - t_i)}{12}, T_y = \frac{N^3 - N - \Sigma_i (u_i^3 - u_i)}{12},$$

N はランキングを行う要素数、 D は比較する二つのランキングにおける同一要素の順位の差であり、同順位が複数存在する場合には、それらの順位の平均値を順位とする。カルテの場合、それぞれのカルテはアンケートによって得られた順位と提案手法で得られた順位を持ち、それらの差が D である。得られた順位として、例えば 5 位となったカルテが 3 件あるとすれば、それら 3 件の順位を $\frac{(5+6+7)}{3} = 6$ 位とする。 T_x や T_y は二つのランキングに対応し、 t_i や u_i は各々のランキングにある同じ順位の要素数の個数である。2 位に同じ順位が 3 件あれば、3. 3 位に同じ順位が 2 件あれば 2, のようになる。

また、得られた相関係数は 52 要素から成る標本によって得ら

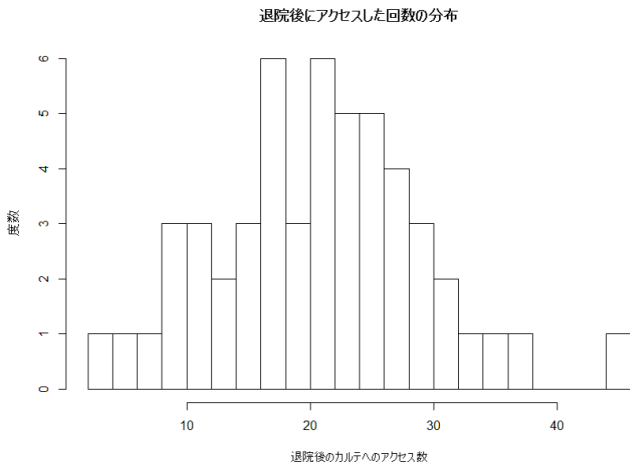


図 4 患者 52 人の退院後、カルテにアクセスされた回数の分布

れた値なので、母集団においても相関があるかを統計的に確かめる必要がある。次章では、有意水準 5%と有意水準 1%で母相関係数が有意に 0 と異なるかどうかを検定した結果を説明する。

6.2.2 ケースカンファレンス紹介カルテとの適合率

ケースカンファレンスで紹介された患者のカルテ集合を正解集合とし、提案手法で得られた重要度のランキングとの適合率、MAP (Mean Average Precision) を計算する。上位 i 件の適合率 $P(i)$ は上位 i 件中の正解文書数 R_i によって

$$P(i) = \frac{R_i}{i}$$

MAP は平均適合率 AP の平均値であり、平均適合率 AP は i 番目の正解が得られるたびに計算された適合率の平均値である。得られた正解文書数を n とし、

$$AP = \frac{1}{n} \sum_n P(i)$$

と表せる。

7. 実験結果

7.1 医師によるアンケートとの相関

対象とする 52 件の患者のそれぞれの退院後に、退院時サマリーが閲覧されたログを全体のアクセスログから抽出した。その結果、それぞれの患者に対して平均 21 件ほどのログが存在した。図 4 にその分布図を示す。

作成された四つの重要度評価項目による各カルテの重要度ランキングと提案手法による重要度ランキングを元にスピアマンの順位相関係数を求めた。提案手法では重み付けの選択肢として、

- 枝の重みを退院後から現在までの時間で正規化するかどうか
- 枝の重み付けとして、閲覧回数、閲覧時間、アクセス文書種別・操作数のどれを用いるか
- 入院中の書き込み・更新時間の長い閲覧者による枝の加重

各重み付けと日数による正規化の有無

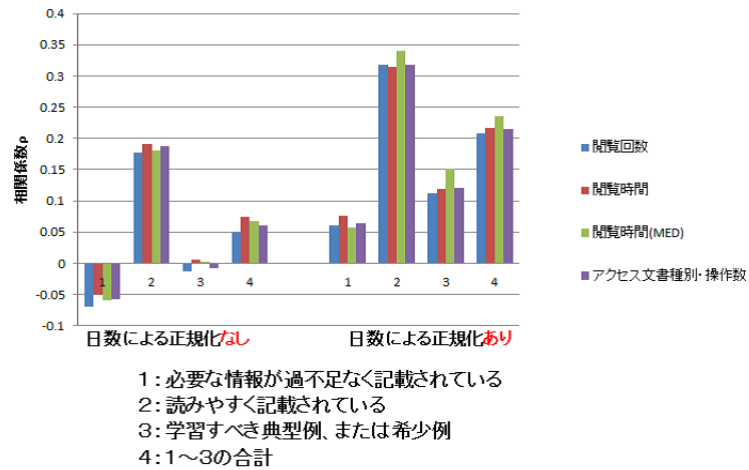


図 5 各重み付け手法ごとに退院後の日数による正規化の有無を反映した相関係数

を考えた。次からの小節ではそれぞれに対して実験を行い、どの重み付けが妥当であったかを評価する。

7.1.1 日付による正規化と枝の重み付け

まず退院後の期間による枝の重み付けの正規化の是非と、閲覧回数、閲覧時間、閲覧時間 (中央値)、アクセス文書種別・操作数のどれがよいかを比較する。ここで、重み w_{ij} は 1 より小さい正の値の足しあわせであり、各カルテへの閲覧が記録されたログの数だけ計上する。ロガー一つ一つの閲覧の重みは、閲覧にとって重要と思われる観測の影響が大きいほど徐々に 1 に近づく増加関数になるよう設定したい。(例えば、閲覧時間が長いほど 1 に近い値をとるようにする等) 本稿では各々のパラメータについて、閲覧時間の β は 0.0001、中央値については γ は 0.99、 λ については 10 とし計算した。各々のパラメータをいくらに設定すると適切かの吟味は今回行っていない。

各重み付けごとに日数で正規化を行ったものと行わなかったものとのスピアマンの順位相関係数は図 5 のようになった。日数による正規化は正規化ありの方が 4 項目全てにおいて、高い相関係数の値を出すこととなった。一方、枝の重み付けは多少の差はあるものの、目に見えて大きな違いは存在しない。

7.1.2 入院中の書き込み時間による重み付け

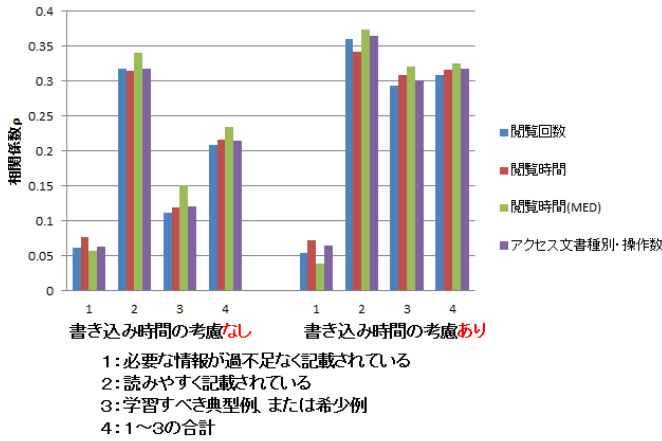
次に、効果のあった日付による正規化を行った上で、入院中の書き込み・更新時間の長い閲覧者による枝の加重を追加したのが図 6 である。この図の左半分は、図 5 の右半分と同じである。

項目 1 においては、アクセス文書種別・操作数による重み付け以外は各々の相関係数の値が下がっているものの、他の項目 2 から 4 においては書き込み時間の考慮ありの方が高い相関係数を得られた。

ここで、退院後日数による正規化と、書き込み・更新時間の重み付けを考慮した上で、閲覧回数や閲覧時間といったそれぞれの枝の重み付けの違いによる相関係数が母集団における母相関係数においても有意な違いを与えるかどうかを検定した (図 7)。

表の中で、色が塗られた部分は有意水準 1%, 5%で母相関係

書き込み時間による重み付け有無



HITSを応用した既存手法(左)と提案手法(右)の相関係数

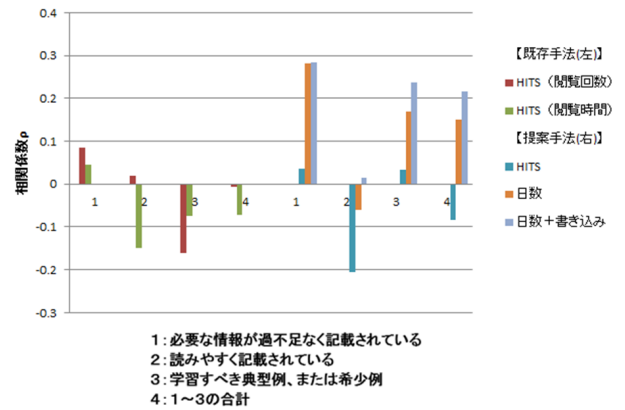


図 6 各重み付け手法ごとに入院時の書き込み・更新時間による重み付けの有無を反映した相関係数

図 9 心不全の標本に対して提案手法により求めた重要度との相関係数

	1	2	3	4
閲覧回数	0.0543	0.3603	0.2933	0.3095
閲覧時間	0.0721	0.3423	0.3096	0.3164
閲覧時間(MED)	0.0390	0.3739	0.3211	0.3249
アクセス文書種別・操作数	0.0655	0.3655	0.3017	0.3179

有意水準1%:
 有意水準5%:

図 7 枝の重み付け手法ごによる母相関係数が有意に 0 と異なるかどうかを検定

1 に対して高い相関を示していることがわかる。これらは本提案手法が得ることの出来なかったものである。

数が 0 であることが棄却された相関係数を表している。どの手法も項目 1 に対しては母集団において有意に相関があるとは言えず、項目 2 から 4 に対しては、有意水準の違いはあるものの有意な相関が認められた。4 つの枝の重み付けの違いは、閲覧時間が長いほど枝に重み付けをしたものを除いては統計的には同じ保証が得られた。

7.1.4 心不全の標本に対する追加実験
 先ほどまでに取り扱ったアンケートは直腸がんを患った 52 名の患者のカルテに対して、各々の重要度を医師からのアンケートで測ったものである。同様に、循環器内科で作成された心不全を病名に含む患者 42 名のカルテに対して、提案手法により相関係数を求めたのが図 9 である。提案手法は、退院後のアクセスログを元に単純に HITS アルゴリズムを適用させたもの (HITS)、日数による正規化を行ったもの (日数)、加えて入院中の書き込み時間により重み付けをしたもの (日数 + 書き込み) の三つにより評価した。図には鶴岡らの先行研究 [2] により得られた相関係数を比較の参考として入れてあるが、鶴岡らのものは本稿よりも少し数の多い 49 件のデータで計算しているため直接の比較はできない。結果、項目 1, 3, 4 に関してはアンケートにより得られた重要度と弱い正の相関が得られたが、項目 2 については弱い負の相関が現れた。こちらの標本においても日数による正規化と書き込み時間の考慮により、単純な HITS アルゴリズムによる結果よりも高い結果を得ることが出来た。一方、同じく母相関係数において相関があるかどうかを 5% の有意水準で検定を行ったところ、こちらは 4 項目ともに母集団での相関を統計的に保証することが出来なかった。

7.1.3 既存手法との比較
 既存手法は 4 つの手法を評価している。対象となるカルテが閲覧された記録を独自の手法で抽出してきた上で、

- HITS アルゴリズムにおいて枝を閲覧回数で重み付けしたもの
- HITS アルゴリズムにおいて枝を閲覧時間で重み付けしたもの
- HITS を使わない単純な閲覧回数の総計
- HITS を使わない単純な累積閲覧時間の総計

の四つの手法を用いてランキング、評価してある (図 8)。

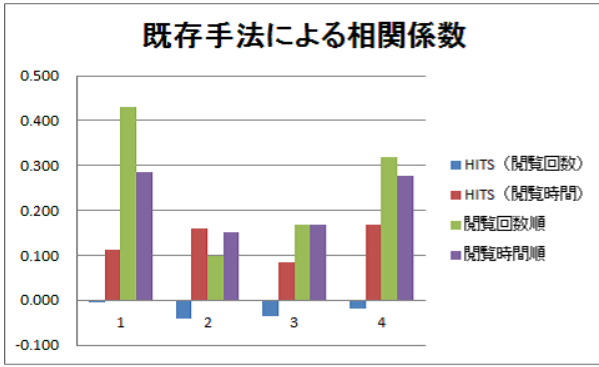
7.2 ケースカンファレンス紹介カルテとの適合率
 単純な HITS アルゴリズムに加え、日数で正規化したもの、入院期間中の書き込み更新時間を考慮した提案手法によるカルテの上位 15 件の重要度と、ケースカンファレンスで紹介された患者のカルテとの適合率は図 10、上位 15 件に対する MAP は表 1 のようになった。

既存研究のグラフを元に、母相関係数が有意に 0 と異なるかどうかを見てみると、既存研究においては HITS アルゴリズムを応用した手法で統計的に相関があることを有意水準 5% で保証できなかった。また、項目 2 や 3 については、既存手法で用いられた 4 つの手法ともに母集団についても相関を保証することが出来なかった。これらの点については、本提案手法がより良い結果を与えたことを示している。一方、HITS を用いず単純に閲覧回数や累積閲覧時間でソートしたもののランキングが項目

HITS	提案手法 (日付)	提案手法 (日付 + 書き込み)
0.422	0.821	0.577

表 1 各手法ごとのランキング上位 15 件に対する MAP

提案手法による枝の重み付けは、どちらも閲覧時間の中央値



	1	2	3	4
HITS (閲覧回数)	-0.004	-0.042	-0.037	-0.017
HITS (閲覧時間)	0.114	0.161	0.085	0.169
閲覧回数順	0.432	0.099	0.168	0.320
閲覧時間順	0.285	0.151	0.170	0.278

有意水準1%:
 有意水準5%:

図 8 既存研究での相関係数 (左) と母相関係数が有意に 0 と異なるか否かの検定 (右)

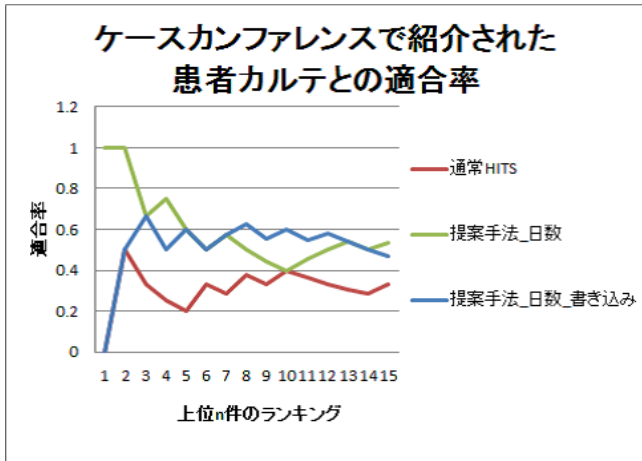


図 10 提案手法により求めた上位 15 件のカルテの重要度と、ケースカンファレンスで紹介されたカルテとの適合率

を元に計算してある。提案手法による適合率と、全てのアクセスログに単純な HITS を適用させたときの適合率を比べて見ると、提案手法による適合率の方が上位 15 件に対して高い適合率を出す結果となった。アンケートによる評価実験では、書き込み時間の考慮はより良い結果を出したが、ケースカンファレンス紹介カルテとの適合率では、書き込み・更新時間を考慮した提案手法で得られたランキング 1 位のカルテが正解データではなかったため、MAP も日付の正規化のみを行ったものの方が良い結果となった。

8. 考 察

8.1 医師によるアンケートとの相関

重要度を測る実験では、既存の HITS を応用した手法に対して直腸がんを患った 52 件の患者カルテについては高い相関係数を得ることが出来た。評価指標の中でも項目 2, 3, 4 において統計的に保証された良い結果が得られた。一方、追加実験における心不全を患った 42 件の患者カルテに対しては、項目 1, 3, 4 に対して比較的良好な値が得られたが、統計的に相関を保証することは出来なかった。提案手法によって得られた相関のある項目は、二つの標本で異なったことになる。異なった結果を生んだ項目 1, 2 は退院時サマリーに記載されている内容の簡潔

さを測るための指標であり、項目 3 は学習に最適な典型例・希少例を測るための指標である。提案手法は項目 3 のような症例研究のための項目を評価出来ているが、いまだ研修医に書き方の参考になるような電子カルテを見つけ出すことが出来ていないといえる。項目 1, 2 の結果が悪かった理由の一つとして、他の人の書き方を評価する機会があまり表立って医療機関で行われていないことが考えられる。ケースカンファレンスのように、多くの医師にとって有益かと選ばれたカルテ (症例) の研究会は定期的に行われているが、集まって他の人の書き方を参考にしたり、書き方の分かりやすいカルテを共有する機会がないため、どの医師の書いたカルテが分かりやすいかなどを把握できず、電子カルテシステムを使って参照出来ないため、結果アクセスログに残らないことが考えられる。また、Wiki のように他の人が書いた記述を訂正することがないため、校正を重ねて記述が洗練されていくような過程が存在しないことも、分かりやすいカルテを上手く検出出来なかった理由の一つであろう。

どちらの標本についても、日付による正規化は高い効果を出していることがわかる。業務データが多く混在する中で、退院後に絞ってデータを選出し正規化・重み付けをすることに一定の効果があると考えられる。入院期間中の書き込み時間の考慮については、先の直腸がんに対する 52 件と異なる標本に対する追加実験の書き込み・更新時間を考慮したものがより良い結果を出したことから、入院期間の間に書き込みを行った閲覧者が退院後にもう一度閲覧した場合、そのカルテの重要度は高いことが期待できる。

8.2 ケースカンファレンス紹介カルテとの適合率

単純な HITS アルゴリズムと比べて、日付の正規化を行った提案手法が高い適合率を出したことから、退院後アクセスに着目することの有効性が示せたと言える。一方、アンケートの実験結果と反して、書き込み更新時間を考慮した手法は、ランキング一位を外してしまう結果となり、日付による正規化のみを行ったものに対して優位性を示すことは出来なかった。これは、提案手法の定式化では、入院中に全く書き込みを行わなかった閲覧者が患者の退院後にカルテを閲覧したとしても、そのカルテは重要と判断されないように定式化したためであると考えられる。実際にケースカンファレンスで取り上げられると、大学

院生など普段は閲覧しない人による閲覧が増えると考えられる。そのため、患者の治療は担当していないがカルテを閲覧した閲覧者の考慮や、その閲覧数の増加の割合を考慮する必要があるだろう。

また、本実験では糖尿病の薬の処方歴がある患者を対象として実験を行ったため、125件のデータセットの中には糖尿病で入院されたのではなく、例えば整形外科など、他の手術のために入院された患者のカルテも含んで実験を行っている。実験データを糖尿病の患者だけで作成することで、より良い精度を期待できる。

9. おわりに

我々は現在の病院システムで既に扱えるアクセスログを利用して、カルテの重要度予測を行う方法を提案した。HITSを元に入力データを退院後に絞って正規化し、各々の閲覧を重み付けすることで、既存研究のアンケートで測られた4つの項目のうち、学習すべきカルテを評価する項目3において精度をあげることが出来た。また、ケースカンファレンスで紹介された患者のカルテは重要であると考えられるため、その患者のカルテをマイニングする実験を行い、単純なHITSよりも提案手法の方が良い適合率をあげることを確認した。

現在は業務上やむなく残るアクセスログが精度を落とす大きな原因となっている。今後電子カルテシステムの向上とともに、医療従事者がより電子カルテの上で学習を行うようになると、有益な電子カルテに対するアクセスが相対的に増大し、より精度を上げやすい環境になるであろう。

我々は引き続き、他のログ情報に基づき業務記録の中に埋没された貴重なカルテを抽出してくる手法を研究しつつも、新たな実験データで評価実験を行う予定である。現在は電子カルテの一部である退院時サマリーしか対象としていないので、他の部分も含めて重要度を測ることが考えられる。提案手法は、重要度を測る手法であったが、そのまま提案手法を適用したところ、五割強の精度でケースカンファレンスで紹介された患者のカルテを的中させることが出来た。ケースカンファレンス紹介カルテに特化した手法を考案することも今後の研究として考えられる。ケースカンファレンスで紹介された患者のカルテをマイニングする場合、症例報告を受けて有名になったカルテは閲覧が一時的に急増すると考えられるため、閲覧数の増加の割合などを用いて評価する方がより自然な手法であろう。また、今回はグラフアルゴリズムを用いて手法を提案したが、機械学習などを用いて実験を行うことも考えられる。また他の課題として、退院後のアクセスログと期間を絞ることにより得られるデータ数が減ってしまうため、より大きなデータ数でも同じ結果が出るかを試す必要があるだろう。

謝 辞

本論文の執筆にあたり、京都大学大学院 情報学研究科 EHR 共同研究講座の吉原博幸京大名誉教授、糸直人准教授から沢山の有益な助言を頂きました。心より感謝申し上げます。

文 献

- [1] 吉原博幸: ドルフィンプロジェクト: 地域医療連携システムの現状と今後, *Heart View*, Vol. 13, pp. 38–45 (2009).
- [2] 鶴岡誠子: 閲覧履歴等の外部情報を用いた診療文書の重要度評価, 修士論文, 京都大学大学院情報学研究科 (2013).
- [3] 岩間伸之: 援助を深める事例研究の方法 対人援助のためのケースカンファレンス(第2版), ミネルヴァ書房, chapter 1 (2005).
- [4] J.Kleinberg: Authoritative sources in a hyperlinked environment, *Proc. of SODA1998*, pp. 668–677 (1998).
- [5] W. Zhang, C. A. Gunter, D. L. J. T. and Malin, B.: Role prediction using electronic medical record system audits, *AMIA 2011 Annual Symposium*, American Medical Informatics Association, pp. 858–867 (2011).
- [6] Allain, J. S.: From Jeopardy! to Jaundice: The Medical Liability Implications of Dr. Watson and Other Artificial Intelligence Systems, *La. L. Rev.*, Vol. 73 (2013).
- [7] Haggai Roitman, Sivan Yogev, Y. T. D. W. K. and Messika, Y.: Exploratory search over social-medical data, *In Proceedings of CIKM* (2011).
- [8] Page L, Brin S, M. R. W. T.: The pagerank citation ranking: Bringing order to the web, *Technical report, Stanford Digital Library Technologies Project* (1998).
- [9] 日本医療情報学会医療情報技術師育成部会(編): 第2版 医療情報サポート, 篠原出版新社, p. 248 (2011).
- [10] 特定非営利活動法人 MedXML コンソーシアム: Medical Markup Language (MML) Version 3.0 規格書 (2003). http://www.medxml.net/mm130/MMLV3Spec_050817.pdf.