

検体検査履歴分析による検査項目組み合わせの推薦

堀埜 裕[†] Le Hieu Hanh[†] 串間 宗夫^{††} 荒木 賢二^{††} 横田 治夫[†]

[†] 東京工業大学 情報理工学院 〒152-8552 東京都目黒区大岡山 2-12-1

^{††} 宮崎大学 医学部附属病院 病院 IR 部 〒889-1692 宮崎県宮崎市清武町木原 5200

E-mail: [†]{horino,hanhhlh}@de.cs.titech.ac.jp, ^{††}{muneo_kushima,taichan}@med.miyazaki-u.ac.jp,

^{†††}yokota@cs.titech.ac.jp

あらまし 電子カルテデータの二次利用として、蓄積された医療情報の解析による有効活用が期待されている。我々は、医療行為の履歴にシーケンシャルパターンマイニングを適用することで、医療行為の典型的な流れである「クリニカルパス」抽出に向けた支援を試みてきた。本研究では、先行研究では扱ってこなかった検体検査の履歴データにパターンマイニングを適用することを試みる。医療行為の場合と違い、検体検査のパターンマイニングは数百にも及ぶ検査項目の組み合わせを考慮する必要がある。我々はそれらの組み合わせに着目しながらパターンマイニングし、次に検査すべき検体検査内容の推薦を目指す。実際の検体検査履歴データを用いた実験では、ある検査を入力して推薦した検査項目と実際に次に行われた検査項目を比較・評価し、本手法による検体検査内容推薦の精度を計測した。実験の結果から、一定の精度での推薦ができることが確認できた。またクラスター数による精度の変化を確認することにより、適切なクラスター数を検討できた。また、いくつかのパラメータと推薦精度の相関を調べた。

キーワード 電子カルテ, データマイニング, 検体検査

1 はじめに

1.1 研究背景

近年、紙のカルテの代わりに電子カルテを利用する病院が増えてきている。電子カルテを利用することによって、対象となる患者の検索やオーダーと呼ばれる医療指示の履歴閲覧の高速化などを可能とし、医療指示全体の標準化・改善に貢献している。また、電子カルテは診療記録として用いられるだけでなく、二次利用としての情報活用が期待されている。

二次利用の例としては、ある病気の患者に対して「クリニカルパス」と呼ばれる医療指示（オーダー）列の典型例を抽出することや、医療指示の流れの変形パターンである「クリニカルパスバリエーション」を検出・分析することが挙げられる。従来クリニカルパスは医療従事者の医学的経験に基づいて作成されていたが、過去の事例が膨大であるため、人力でクリニカルパスバリエーションを収集・分析して改善するのは容易ではなかった。そのため、計算機によって電子カルテをデータ工学の観点から分析し、医療行為改善の支援を行う研究が始まった。

1.2 関連研究

牧原らの研究 [1] では、電子カルテのアクセスログから、ある患者に対して行った医療指示をアイテム、医療指示の流れをシーケンス、全ての患者の医療指示の流れをデータベースと考えることで、アプリアリアルゴリズム [2] を元にしたシーケンシャルパターンマイニング（以下、SPM）を用いて頻出シーケンスの抽出を行った。牧原らは手術といった重要な医療指示を基準イベントとして、その前後の部分シーケンスでそれぞれに

マイニングを行うことで、基準とした医療指示の前後で行うべき医療指示を抽出出来ることを示した。しかし、牧原らの手法には基準イベントの後の部分のマイニング結果が膨大になってしまうこと、医療指示間の時間間隔を考慮できていないことの二つの課題が残った。

佐々木ら [3] は、この二つの課題解決のため、飽和オーダー列と呼ばれる概念 [4] を導入し、出力の情報量を損なわずに出力数を削減してタイムインターバル SPM [5] (以下、TI-SPM) を PrefixSpan [6] に用いることによって、二つのアイテム間の時間間隔を考慮した抽出を行った。しかし、TI-SPM は人為的に定めたタイムインターバル（以下、TI）に結果が依存し、最適な時間間隔を定める必要があるという課題があった。また、これらの研究では注射のような薬剤情報を含む医療指示において、投与薬剤の種類の情報が含まれていないという課題もあった。

浦垣ら [7] は、これらの課題解決のため、アイテムを（大別 Type, 詳しい説明 Explain, 薬効コード Code, 薬剤名 Name）の四つ組によって構成することで、薬剤情報を取り入れた抽出を行った。また、薬剤の効果を表す薬効コードに基づくマイニングを行うことで、薬剤名でのマイニングでは得られなかった結果が抽出可能であると示した。さらにアイテム間の時間間隔を統計処理によって算出することで、予め最適な TI セットを人為的に求めなければならないという問題に対処した。保坂ら [8] は、三つのスコアリング手法を導入することで最適な TI セットを事前に決定し、頻出シーケンスの抽出を行った。しかし、これらの先行研究には医学的に解析の価値が高いクリニカルパスバリエーションに対しての処理がなされていない、出力結果が頻出シーケンスの文字列であるため、クリニカルパスバリエーションを含む場合の視認性が悪いという二つの課題があった。

本田ら [9] はこの二つの課題解決のため、シーケンス同士の共通部分を医療指示が出される日の情報を考慮して検出していくことによって、共通部分を持つシーケンスの差異「シーケンスバリエーション (以下, SV)」を検出し、グラフで表現して医療従事者への視認性の高い可視化ツールの提供を行った。しかし、SV に対しての分析を行っていない等の課題があった。この課題に対して、山田ら [10] はシーケンスの識別子である SID を保持しながらマイニングを行い、SV を評価する指標を取り入れることで、SV の分析を行った。また、シーケンスや SV の評価が円滑かつ正確に行えるような可視化手法の提案を行った。また、本田ら [11] は SV が発生する原因をロジスティック回帰分析により発見する研究を行った。

他にも、津本ら [12] は看護オーダーの実施履歴からクリニカルパスを生成する仕組みについて提案した。しかし、いずれの研究でも、検体検査が行われたことは考慮していても、その内容を考慮できていないことが課題として挙げられる。

1.3 本研究の目的

本研究では、ある検体検査の検査項目とその結果値から、次の検体検査項目組み合わせを提案することによる検体検査項目選択支援を目的とする。今まで分析されてきた医療行為の場合と違い、検体検査のパターンマイニングは数百にも及ぶ検査項目の組み合わせを考慮する必要がある。そこで、実際の電子カルテシステムに記載された検体検査の検査項目組み合わせを分類し「検査タイプ」を設定することによって、膨大な数の組み合わせを数種類の検査タイプに落とし込み、パターンマイニング・推薦を行う。

実験では、本手法の有効性を確認するとともに、パラメータの変化が推薦の精度にどう影響するかを調べる。

1.4 本稿の構成

本稿は以下の通り構成される。2 節では、本研究が提案する検体検査内容推薦の手法について説明する。続く 3 節では、実際の電子カルテシステムに対し本手法を適用し、結果に関する考察を行う。最後に 4 節でまとめと今後の課題について述べる。

2 提案手法

本節では、データをどのようにして扱うか、検体検査の分類方法、それを基にした推薦方法について説明する。おおまかな全体の流れは、

- (1) 検査データの抽出
- (2) 検査データの分類
- (3) 検査シーケンスのマイニング
- (4) 検査項目組み合わせの推薦

の 4 段階である。

また、検査項目の特殊性を考慮する手法についても説明する。

2.1 検査データの抽出

電子カルテシステムに記載された検体検査履歴のデータベースには、1 行につき 1 つの検査項目についての情報が入ってい

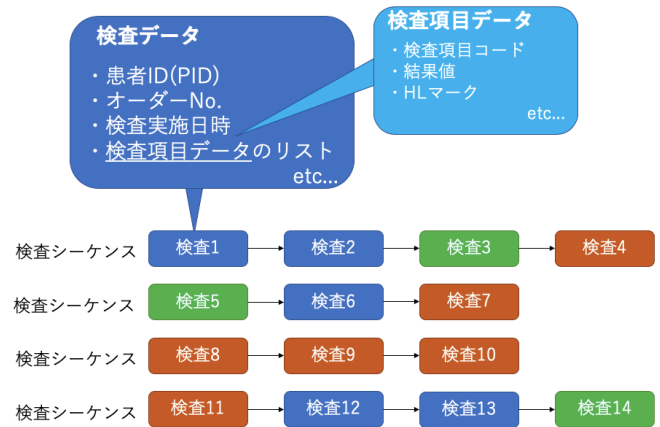


図1 抽出するデータのイメージ

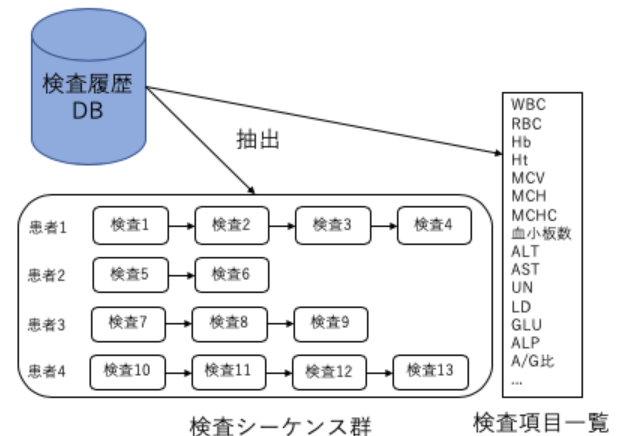


図2 検査データ抽出のイメージ

る。入っている情報は、患者を識別する患者 ID、検査を識別する検査オーダー、検査日時、検査項目コードと検査項目名、実際の値である結果値、その結果値が高いのか低いのかを示す HL マーク、どのくらい異常値かを示す異常値レベル、などである。一行に入っている、検査項目とその結果に関する情報をまとめて「検査項目データ」と呼ぶことにする。

この検体検査履歴のデータベースから、検査項目データを抽出していく。抽出する際に、同じ検査オーダーであるもの・検査日時が完全に一致しているもの、つまり同時に検査されたとみなせるものを 1 つのデータとしてまとめて、1 つの「検査データ」とする。

同じ患者の検査データを一連の時系列データとしてまとめ、「検査シーケンス」として、システムの中に保存していく。抽出するデータのイメージを図 1 に示す。

また、検査データを抽出することと同時に、全検査中に出てくる検査項目一覧を作成する。この一覧は検査項目コードの昇順でソートされている。検査データ抽出の一連のイメージを図 2 に示す。

2.2 検査データの分類

初めに、検査項目の組み合わせをベクトルで表現する「検査項目ベクトル」を作成する。検査項目ベクトルは、検査項目一

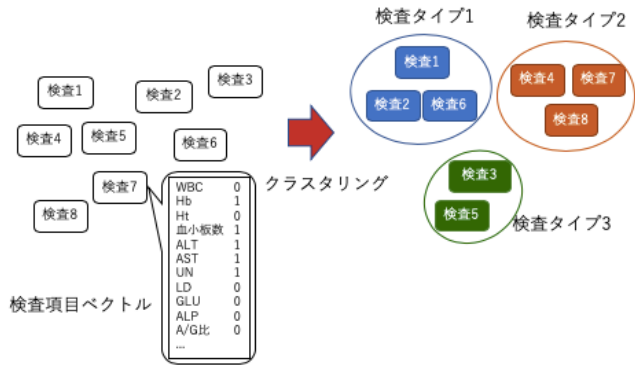


図3 データ分類のイメージ

覧に出てくる検査項目の種類と等しい次元数を持ち、各次元が各検査項目に対応している。ある検査について、検査項目一覧における i 番目の検査項目が検査されていたら検査項目ベクトルの i 次元の要素が 1 を取り、検査されていなかったら 0 を取るようにしたベクトルを、その検査の検査項目ベクトルとする。図 2.3 で例示しているものでは、Hb, 血血小板数, ALT, AST, UN に対応する値が 1 になっていて、これらの項目が検査されたことを表している検査項目ベクトルを示している。

この検査項目ベクトルを全ての検査データについて作成し、それらのベクトルを重複を省いてクラスタリングする。検査項目の組み合わせをベクトルで表すことで、一般的なクラスタリングアルゴリズムを適用できる。今回の研究では、検体検査において一般的な項目と特殊な項目の組み合わせという構造が多く存在しており、かつ血液検査と尿検査のように完全に別の項目の検査も存在しているという仮説のもと、クラスターの半径が固定になってしまう非階層クラスタリングの k-means 法は不向きだと考え、階層型クラスタリングアルゴリズムの中で分類性能が高いワード法を使用している。

検査項目ベクトルまたはそれらを併合したクラスターの併合・距離の再計算を繰り返し、全ての検査項目ベクトルを併合し、一つのデンドログラムが作成できたところで、そのデンドログラムをあらかじめ指定したクラスター数 N になるように区切り、 N 個のクラスターに分割する。このときできたクラスターの一つ一つを「検査タイプ」とする。検査データ分類の一連のイメージを図 3 に示す。

2.3 検査シーケンスのマイニング

検査タイプへの分類が完了した後、全ての検査シーケンスについて、シーケンス内の検査ごとに検査タイプがどのように推移しているかと、検査項目組み合わせ・各検査項目の結果値の高低を記録していく。

結果値の高低については、その結果値が基準値と比べて高いか低いかを表す HL マークを参照し、ある検査項目が High か Low か基準値の範囲内かを示す「HL ベクトル」を作成して記録する。

HL ベクトルは、検査項目一覧に出てくる検査項目の種類と等しい次元数を持ち、各次元が各検査項目に対応している。ある検査について、検査項目一覧における i 番目の検査項目の HL

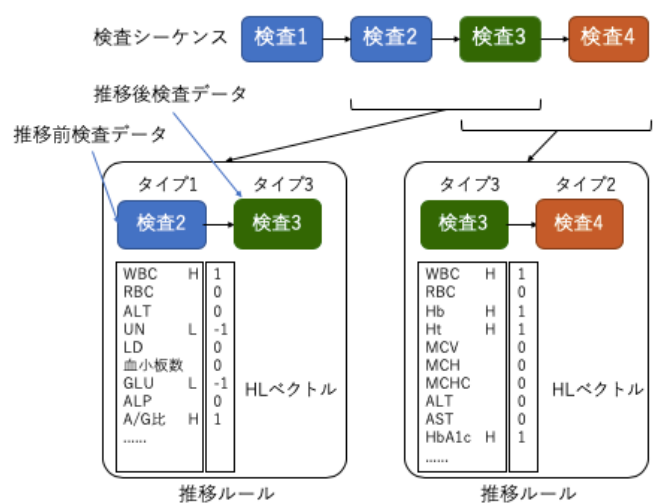


図4 マイニングのイメージ

マークが H なら HL ベクトルの i 番目の次元の要素が 1 を取り、L なら -1 を取り、基準値の範囲内なら 0 を取るようにしたベクトルを、その検査の HL ベクトルとする。HL ベクトルの具体例を図 2.4 に例示する。

検査タイプの推移については、シーケンスの i 番目の検査と $i+1$ 番目の検査を見たときに、 i 番目の検査タイプを「推移前検査タイプ」、 $i+1$ 番目の検査タイプを「推移後検査タイプ」として、推移前・推移後検査タイプの組を記録していく。

推移前検査タイプ、推移後検査タイプ、推移前検査の HL ベクトルの組を「推移ルール」とし、推移前・推移後検査タイプの組が同じものごとに推移ルールの集合を作成する。検査シーケンスのマイニングの一連のイメージを図 4 に示す。

2.4 検査項目組み合わせの推薦

ある検査結果を入れると次にすべき検査の検査タイプを推薦する方法について述べる。まず初めに、入力された検査結果から、その検査データの検査項目ベクトルを作成する。次に、作成した検査項目ベクトルと検査データの分類時に作成したクラスターとの距離を計算し、一番距離が小さいクラスターを求め、そのクラスターを入力された検査の検査タイプとする。次に、入力された検査の検査タイプを推移前検査タイプとして持つ推移ルール集合の中から、入力された検査データの検査項目ベクトル・HL ベクトルと一番類似度が高い推移ルールを探し、その推移ルールの推移後検査タイプを算出する。その後、その検査タイプの中で、条件を満たす検査項目の組み合わせを推薦結果として出力する。

検査項目組み合わせの推薦の一連のイメージを図 5 に示す。また、提案手法の全体の流れのイメージを図 6 に示す。

2.5 検査項目の特殊性の考慮

検体検査の検査項目には、多くの検査で検査されるような一般的なものと、特定の用途でしか検査されない特殊なものに大別できると考え、特殊な検査項目に着目するために「uniqueness」と「分離度」という指標を導入する。これらの定義を以下に

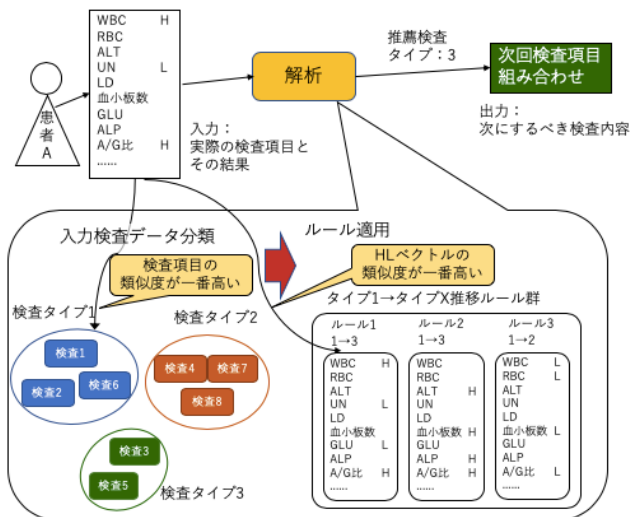


図 5 推薦のイメージ

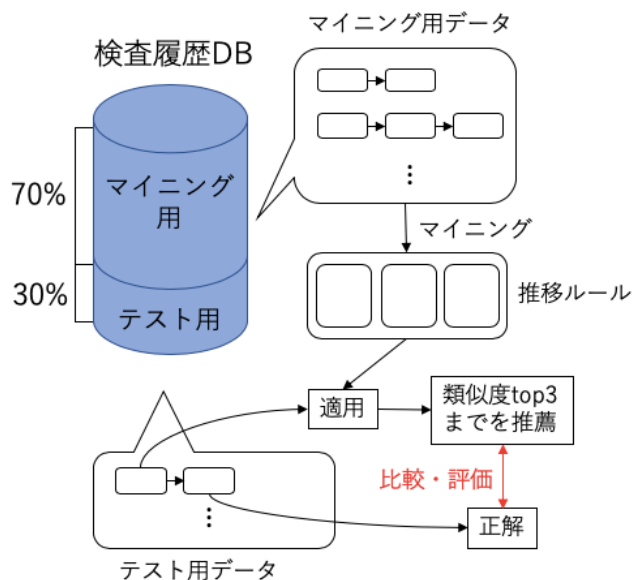


図 7 実験のイメージ

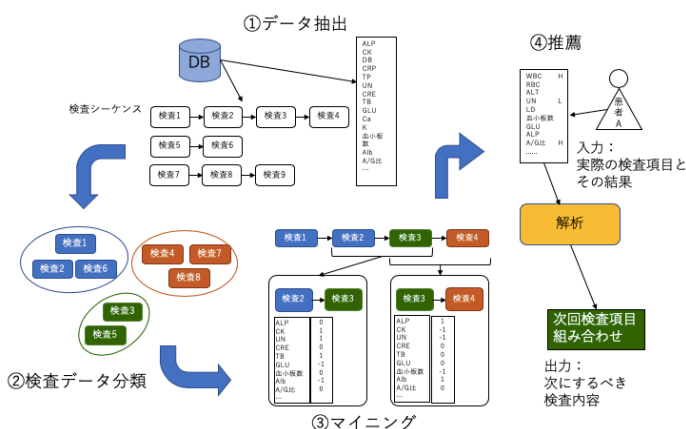


図 6 全体の流れのイメージ

示す。

定義 1.

あるクラスター c における検査項目 i の $uniqueness =$ クラスター c での検査項目 i の出現率を a , 全クラスターでの検査項目 i の出現率の標準偏差を σ として, 検査項目 i の出現率が a より 2σ 以上低いクラスターの個数

定義 2. 分離度 = $uniqueness$ が 1 以上の検査項目での $uniqueness$ のクラスター内平均値 / $(N - 1)$

単にクラスター内での出現率が高い項目だけを推薦すると, 一般的な検査項目の推薦が多くなるが, 医学的には特殊な検査項目が重要だと考えた. そのため, 推薦時にクラスター内での出現率が一定以上の項目を推薦する方法と, クラスター内で $uniqueness$ が一定以上の項目を推薦する方法をそれぞれ試す.

3 実験

これまでに, 本研究で使用する検体検査項目の推薦手法について説明した. 本節では, 宮崎大学医学部附属病院から提供された実際の電子カルテデータに対して, 提案手法を適用してどのような結果が得られるのかを観察する.

3.1 実験概要

3.1.1 実験目的

実際の検体検査履歴データに対して提案手法を適用し, 推薦した検査内容と実際にその次に行われた検査内容がどのくらい一致しているのかを確認する. また, クラスター数 N が推薦の精度に与える影響と, クラスタリング精度と推薦の精度の関連性を確認する.

3.1.2 実験内容

実際の検体検査履歴データから本手法と同じ方法でデータを抽出する. 抽出したデータのうち一定の割合のものをランダムに選んでテスト用データとして区別しておき, 残りのデータでマイニングを行う. マイニング後, テスト用の検査データを入力として検査項目の推薦を行い, 実際にその次に行った検査項目とどのくらい一致しているのかを確認する. このとき, 推移ルールは類似度が高い順にトップ 3 まで選ばれており, トップ 1 のみ・トップ 2 までのうち F 値が一番高かったもの・トップ 3 までのうち F 値が一番高かったものの 3 段階で評価をしている. この条件で, クラスター数 N を $N=6$ から $N=30$ まで 2 ずつ変化させていき, 結果がどのように変わるのかを調べる. この時, 全てのクラスターで集計した精度と, 質の悪いクラスターを除いて集計した精度をそれぞれ考える.

また, クラスター数を変化させていき, クラスターの分離度がどうなるのかを調べ, 分離度と推薦精度がどのように関連しているかを調べる. 実験のイメージを図 7 に示す.

3.1.3 実験対象のデータ

本研究では宮崎大学医学部附属病院の電子カルテシステムに記録されている, 2015 年 1 月 1 日から 2018 年 4 月 20 日の範囲の実際に使用された検体検査の履歴データを対象とする. この検査履歴データは宮崎大学医学部附属病院で使用されている電子カルテシステム WATATUMI [13] によって取得されており, 個人情報保護の観点より患者を一意に特定できるような情報を含まない. 対象データセットを以下の表 1 に示す.

表 1 対象データセット

マイニング用シーケンス数	5,425
マイニング用検査データ数	15,227
テスト用シーケンス数	904
テスト用検査データ数	4,820
検査項目の種類	1,338

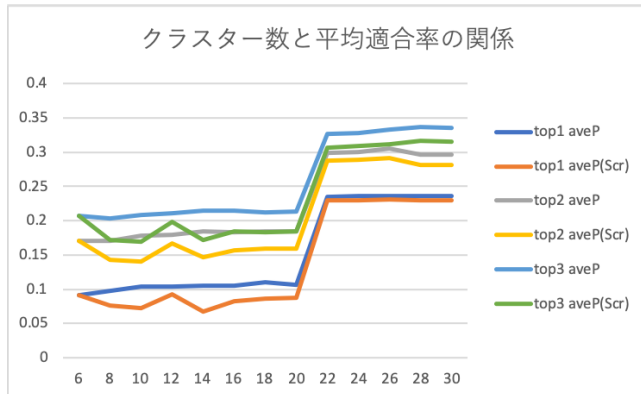


図 8 クラスター数と平均適合率の関係

3.1.4 評価指標

推薦精度の評価指標として適合率 (precision), 再現率 (Recall), F 値の 3 種類の指標を使用している。それぞれの計算式を以下に示す。

定義 3.

$$Precision = \frac{\text{実際の次の検査の検査項目と推薦項目の一致数}}{\text{推薦した項目数}}$$

定義 4.

$$Recall = \frac{\text{実際の次の検査の検査項目と推薦項目の一致数}}{\text{実際の次の検査の項目数}}$$

定義 5.

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

適合率と再現率はトレードオフの関係になっているため、一方が上がれば一方が下がる。この二つをまとめて総合的に評価するのが F 値である。F 値は $0 \leq F \leq 1$ の値を取り、1 に近いほど推薦性能が高いことを示す。

また、特殊な検査項目に着目するため、全てのテストの精度を集計したもの (All) と、クラスターの質が悪いもののテストを除いて集計したもの (Screening) で考える。ここで、質が悪いクラスターとは、分離度が uniqueness の全体平均より σ 以上低いものとしている。この処理で、一般的な項目での検査を除いている。

3.2 実験 1：出現率による推薦

実験 1 は、クラスター内出現率が高い検査項目を推薦し、その精度を調べた。今回は、推薦検査タイプのうちクラスター内出現率が 0.3 より大きい検査項目を推薦した。この実験により、項目の特殊性にかかわらず、検査全般の推薦精度を測る。

3.3 実験 1：実験結果

クラスター数と平均適合率・平均再現率・平均 F 値・最大 F 値の関係それぞれ図 8, 図 9, 図 10, 図 11 に示す。なお、

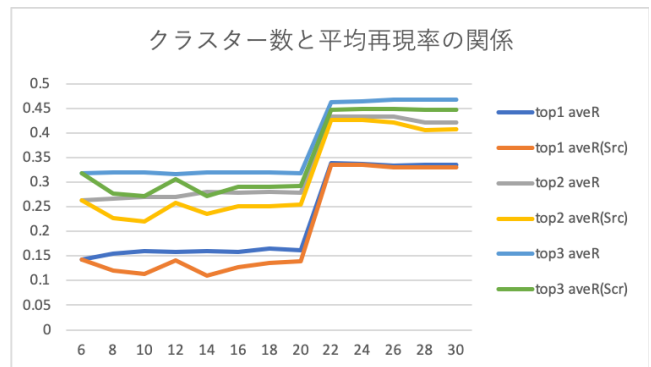


図 9 クラスター数と平均再現率の関係

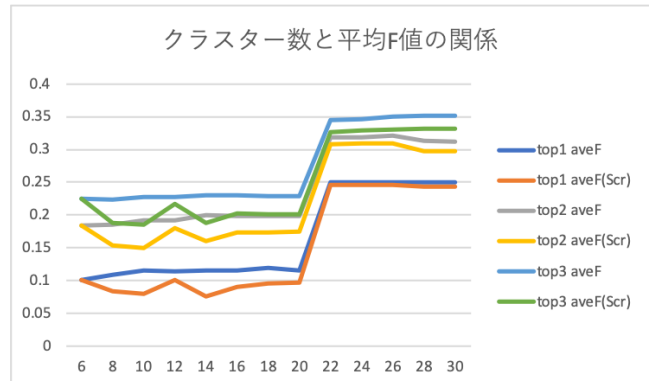


図 10 クラスター数と平均 F 値の関係

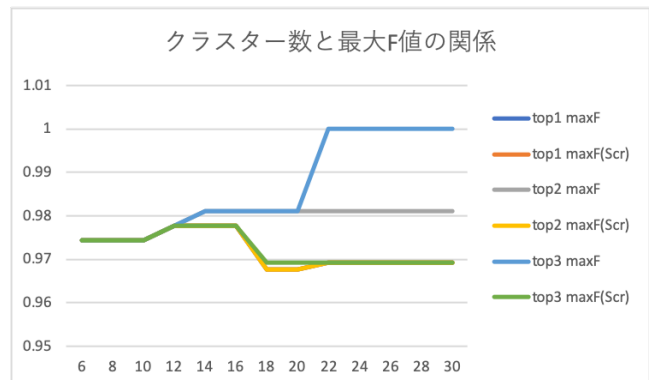


図 11 クラスター数と最大 F 値の関係

(Scr) と書いてあるものが Screening, 無表記のものが All の結果を示している。また、top1・top2・top3 となっているのは、それぞれ推移ルールの類似度トップ 1 のみ・トップ 2 までのうち F 値が一番高かったもの・トップ 3 までのうち F 値が一番高かったものを表している。

全体的に、All より Screening の方が推薦精度が低いことが読み取れる。このことから、一般的な検査項目の推薦精度が推薦精度に大きく寄与していることが分かる。さらに、適合率と再現率を比較した時に再現率の方がやや高くなっていることから、推薦すべき検査項目を広めに推薦できている傾向があることが分かる。また、類似度が高い上位 1 つのルールのみを見るよりも、上位 2 つ・上位 3 つまでを見た方が、着実に精度が上がっていることが分かる。さらに、N=22 以上で All・Screening ともに精度が大きく上昇していることが読み取れる。

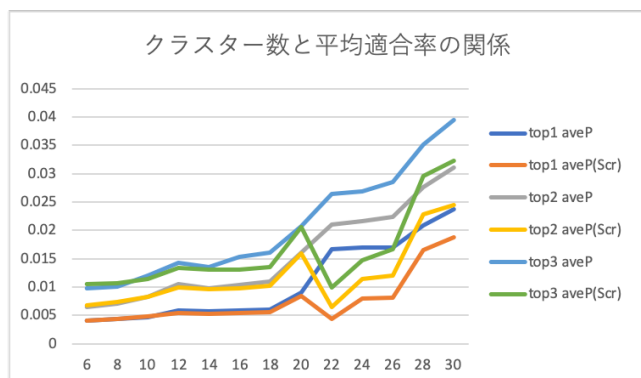


図 12 クラスター数と平均適合率の関係

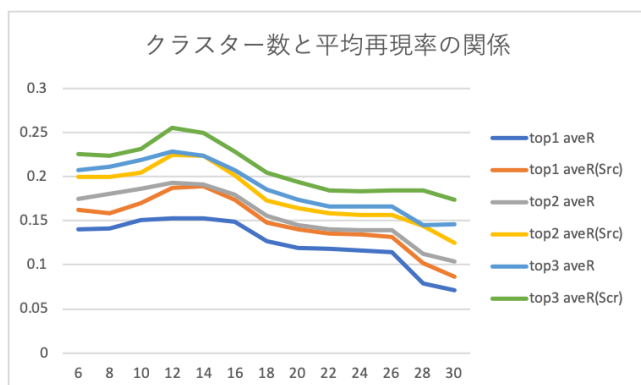


図 13 クラスター数と平均再現率の関係

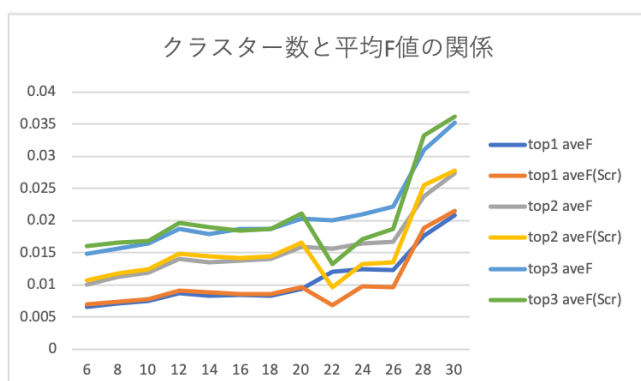


図 14 クラスター数と平均 F 値の関係

このことから、 $N > 20$ のクラスターになんらかの精度に関する要素があると推測できる。

この実験での平均 F 値は、All の場合で最大 0.352、Screening の場合で最大 0.332 を記録した。

3.4 実験 2：uniqueness による推薦

続く実験 2 では、特殊な検査項目の推薦を行うことを考えて、uniqueness が一定以上のものを推薦した時の精度を調べる。今回は $\text{uniqueness} > N/2$ のものを推薦する。この実験により、あまり一般的でない検査項目を推薦する際の推薦精度を測る。

3.5 実験 2：実験結果

クラスター数と平均適合率・平均再現率・平均 F 値・最大 F 値の関係をそれぞれ図 12、図 13、図 14 に示す。平均適合率は

クラスター数を増やすごとに上昇していき、反対に平均再現率はクラスター数を増やすごとに減少していくことが読み取れる。これは、クラスター数を増やすことで特殊な検査項目をよく分離できるためだと考えられる。この適合率の上昇がどこまで続くのかについては、データを増やして検証することが今後の課題として挙げられる。

また、適合率は Screening の方が数値が低く、再現率は Screening の方が数値が高いことが読み取れる。このことから、uniqueness が高い項目を推薦すると、推薦する項目が多すぎる傾向があることが分かる。今回は $\text{uniqueness} > N/2$ という条件で推薦を行ったが、この条件を変えて適切な閾値を設定することが今後の課題として挙げられる。

また、適合率は上昇していき、再現率は減少していったが、この 2 つを総合的に見る F 値を見ると、クラスター数を増やすごとに上昇していつていることが分かる。

実験 2 でも実験 1 と同様に、Screening での再現率・F 値で、 $N=22$ で大きく精度が変わっていることが分かる。特殊な項目のみに着目して大きく精度が落ちていることから、特殊な項目に関するクラスターについてなんらかの精度を左右する要因があると推測できる。この要因の解明は今後の課題として挙げられる。

4 おわりに

4.1 ま と め

本研究では、検体検査履歴データにパターンマイニングを適用することによって、ある検体検査項目と検査結果から次に検査すべき検査内容を推薦することを目的とした。医療指示をパターンマイニングする場合と違い、検体検査のパターンマイニングは数百にも及ぶ検査項目の組み合わせを考慮する必要があった。我々はそれらの組み合わせを、検査項目組み合わせをクラスタリングを用いて分類し、膨大な検査項目の組み合わせをいくつかのまとまりに落とし込んだ。それらの分類を用いてパターンマイニングし、次に検査すべき検体検査内容を推薦する方法を提案した。実際の検体検査履歴データを用いた実験では、ある検査を入力して推薦した検査項目と実際に次に行われた検査項目を比較・評価し、本手法による検体検査内容推薦の精度を計測した。実験の結果から、一定の精度での推薦ができることが確認できた。また、クラスター内出現率とクラスター内 uniqueness のそれぞれに基づいて検査項目を推薦する基盤を作ることができた。

4.2 今後の課題

実際に現場で使うためには、さらに推薦の精度を上げる必要があると考える。推薦の精度を上げるために、検査ライブの分類精度と推薦精度の関連を調べるとともに、検査タイプの分類精度を上げること、推薦する検査項目の精度を上げることが課題として挙げられる。前者の課題解決のために、クラスタリングの評価を行うことが改善策として考えられる。後者の課題解決のために、各推薦の結果を分析し、推薦項目の正解・不正解

の要因推定をすることが改善策として考えられる。推薦結果を分析するにあたり、データを増やし実験の範囲をさらに広げることも検討する必要がある。また、今回は検査項目と結果値の高低しか考慮していないが、患者の年齢、性別などといった因子を考慮できていないことも課題として挙げられる。これらの因子として何が挙げられるのかを特定し、揃えた上で分析を行う必要があると考えられる。

謝 辞

本研究で宮崎大学医学部附属病院の電子カルテオーダログを医療行為改善の研究に用いることは、宮崎大学の HP [14] に記載されており、宮崎大学の倫理審査委員会及び東京工業大学の人を対象とする研究倫理委員会の承認を得ている。関係者各位の協力に感謝する。

文 献

- [1] 牧原健太郎, 荒堀喜貴, 渡辺陽介, 串間宗夫, 荒木賢二, 横田治夫. 電子カルテシステムの操作ログデータの時系列分析による頻出シーケンスの抽出. DEIM Forum 2014, F6-2, 2014.
- [2] R. Agrawal, R. Srikant. Fast algorithms for mining association rules in large databases. Proceeding of the 20th International Conference on Very Large Data Bases, pp. 487-499, 1994.
- [3] 佐々木夢, 荒堀喜貴, 串間宗夫, 荒木賢二, 横田治夫. 電子カルテシステムのオーダログデータ解析による医療行為の支援. DEIM Forum 2015, G5-1, 2015.
- [4] X. Yan, J. Han, R. Afshar. CloSpan: Mining closed sequential patterns in large databases. Proc.SIAM Int'l Conf. Data Mining (SDM '03), pp. 166-177, May 2003.
- [5] Y. Chen, M. Chiang, M. Ko. Discovering time-interval sequential patterns in sequence databases. Expert Systems with Applications 25, pp. 343-354, 2003.
- [6] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, M. Hsu. PrefixSpan: Mining sequential patterns efficiently by prefix-projected pattern growth. Proceeding of 2001 international conference on data engineering, pp. 215-224, 2001.
- [7] K. Uragaki, T. Hosaka, Y. Arahori, M. Kushima, T. Yamazaki, K. Araki, H. Yokota. Sequential pattern mining on electronic medical records with handling time intervals and the efficacy of medicines. First IEEE workshop on ICT solutions for health, proc. 21st IEEE international symposium on computers and communications, pp. 20-25, 2016.
- [8] 保坂智之, 浦垣啓志郎, 荒堀喜貴, 串間宗夫, 山崎友義, 荒木賢二, 横田治夫. 医療履歴の時系列解析におけるシーケンス間類似度評価による時間間隔調整の導入 DEIM Forum 2016, G7-5 2016.
- [9] Y. Honda, M. Kushima, T. Yamazaki, K. Araki, H. Yokota. Detection and visualization of variants in typical medical treatment sequences. Proceeding of the 3rd VLDB workshop on data management and analytics for medicine and healthcare. Springer, pp. 88-101, 2017.
- [10] 山田達大, 本田祐一, 萱原正彬, Le Hieu Hanh, 串間宗夫, 小川泰右, 松尾亮輔, 山崎友義, 荒木賢二, 横田治夫. SID を保持するシーケンシャルパターンマイニングによるクリニカルパスバリエーション分析. DEIM Forum 2019, 2019.
- [11] 本田祐一, 山田達大, 萱原正彬, Le Hieu Hanh, 串間宗夫, 小川泰右, 松尾亮輔, 山崎友義, 荒木賢二, 横田治夫. 患者の固有情報及び動的状況を考慮したクリニカルパス分岐要因推定. DEIM Forum 2019.
- [12] Shusaku Tsumoto, Shoji Hirano, Haruko Iwata. Construction of clinical pathway from histories of clinical actions in

hospital information system. 2016 IEEE International Conference on Big Data (Big Data), pp. 1972-1981, 2016.

- [13] 電子カルテシステム WATATUMI.
http://www.corecreate.com/02_01_izanami.html
- [14] 宮崎大学医学部附属病院医療情報部.
<http://www.med.miyazakiu.ac.jp/home/jyoho/>
- [15] 佐藤義治 (2009). 多変量データの分類: 判別分析・クラスター分析 朝倉書店
- [16] 東京大学医学部附属病院 検査部 患者さんへのご案内.
<http://lab-tky.umin.jp/patient/kensa.html>