DSSR

Discussion Paper No.J-6

異質なデータを統合した敗血症患者の転帰予測システム

石塚治也　石垣司　小林直也　工藤大介　中川敦寛

2017 年 9 月

Data Science and Service Research
Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

異質なデータを統合した敗血症患者の転帰予測システム
Mortality prediction system based on different types of data in electric health records for septic patients

石塚 治也[1]　石垣 司[1]　小林 直也[2]　工藤　大介[2]　中川 敦寛[2]

Haruya Ishizuka Tsukasa Ishigaki Naoya Kobayashi Daisuke Kudo Atsuhiro Nakagawa

概 要

敗血症は臓器不全を伴う感染症であり、死亡率が極めて高く、重症化しやすい患者を早期に予測する事は高い臨床的意義を持つ。本研究では電子カルテから得られる重症度スコア、バイタルサイン、看護記録を組み合わせた敗血症患者の転帰予測システムを提案する。提案システムは二つの段階から成り立つ。第一段階で特徴量抽出として、重症度スコアの計算，Switching AR Model を用いたバイタルサインの状態比率抽出、潜在的ディリクレ配分法を用いた看護記録のトピック比率抽出を行う。第二段階でその三種類の変数を用いてサポートベクターマシンの学習を行う。集中治療室の実データを用いた実験により、提案システムは既存手法や看護記録を用いない手法と比較して、より正確に患者の転帰を予測できる事を示す。

1 まえがき

集中治療室（Intensive Care Unit:ICU）は病院内で密接に患者の生死に関わる部門の一つである。その為、医師、看護師は患者の状態を的確、迅速に判断し診療を行う事が求められる。しかし、時間的、人的制約により、全ての判断を人手で行う事は医療スタッフの大きな負担となる。判断を補助する為の手段として、重症化しやすい患者や、特定の症状の発症などを検知し、医療スタッフにアラートを出す意思決定支援システムが注目されている。

重症化しやすい患者を検知する為に、様々な転帰予測手法が提案されている。ICU に入室した患者の転帰予測手法には、SAPS II[1]、APACHE II[2] など重症度スコアが用いられてきた。近年では電子カルテの導入により、一人の患者から異質かつ多様なデータが取得できるようになった。これらのデータに対して、機械学習的アプローチを用いる予測手法も提案されている [3]。

ICU に入室した患者はそれぞれ異なる疾患を持つが、上述の手法の多くはその違いを考慮していない。Siontis らは ICU に入室した患者の転帰予測に関する既存研究を解析し、疾患や背景の違いにより予測精度に大きなばらつきが存在する事を報告している [4]。より的確に医療スタッフの意思決定を補助する為に、個別の疾患に焦点を当てたシステムを開発する事が望ましい。

ICU 内で観測される様々な疾患の中でも、敗血症は特に重要な疾患である。敗血症は臓器不全を伴う感染症であり、死亡率の高さ、経済コストの大きさは ICU の運営に強く影響する。2001 年に発表された米国を対象とした疫学研究では、年間 75 万人が発症し、半数以上は ICU に入室、ICU 内死亡率は 34．3% に上り、治療や対策に伴うコストは年間 160 億ドル以上であると報告されている [5]。

敗血症では、早期に患者の転帰を予測するシステムは高い臨床的意義を持つ。重症化し得る患者を早期に特定し、治療を行う事で予後を改善する可能性がある [6]。一方で、軽症な患者に対して

[1]東北大学大学院経済学研究科
[2]東北大学大学院医学系研究科

表 1: 電子カルテに記録されるデータとその異質性

種類	性質
バイタルサイン	等間隔時系列定量データ
検査データ	不等間隔時系列定量データ
デモグラフィック変数	定量データ
看護記録	不等間隔時系列定性データ

過剰な抗菌薬投与を行うと、死亡リスクを向上させるという報告もある [7]。その為、転帰予測システムに基づき、重篤な状態にある患者を検知する事は、予後を改善する上で有効である。

電子カルテ内には、表 1 に示されるようにデモグラフィック変数、バイタルサイン、検査データ、看護記録などが記録されている。[3]バイタルサイン、検査データ、看護記録は患者の入室時間に比例して記録が増加する。その為、転帰予測の際に直接特徴量として用いる事は出来ず、特徴量抽出を行いデータを構造化する必要がある。

敗血症患者の転帰予測に関する研究は、定量データからの特徴量抽出に基づくものが多く、定性データである看護記録を用いた研究は見受けられない。しかし、看護記録は、医療スタッフの観察、経験の様に定量データでは得られない情報を反映している可能性があり、有効な予測因子になり得る。

本研究では、看護記録を含む異質なデータを統合した敗血症患者の転帰予測システムを構築する。システムは 2 つの段階から構成される。第一段階で特徴量抽出として、重症度スコアの計算、Switching AR Model を用いたバイタルサインからの状態比率抽出、潜在的ディリクレ配分法を用いた看護記録からのトピック比率の抽出を行う。第二段階では、第一段階で得た特徴量を組み合わせてサポートベクターマシンの学習、予測を行う。そして、東北大学病院 ICU の実データを用いた実証実験を通じて、看護記録を予測に用いるシステムの有効性を示す。

以下では 2 章で先行研究と本研究との相違点について述べ、3 章ではシステムの詳細について説明する。4 章では実証実験とその結果について説明する。5 章では実証実験の結果をもとに考察を行い、6 章を本稿のむすびとする。

2 先行研究

2.1 ICU 入室患者の転帰予測

従来の転帰予測では、SAPS II や APACHE II など重症度スコアが用いられてきた。しかし、これらのスコアは ICU 入室後 24 時間の最悪値を用いて計算されるため、しばしば患者の死亡リスクを過大評価する。

スコアリングを用いる以外の手法として、電子カルテから取得したバイタルサイン、検査データ、看護記録などに対して機械学習的アプローチを用いる予測手法の提案が行われている。Lehman らは心拍数など高頻度で観測されるバイタルサインから特徴量を抽出し、重症度スコアと併用する事で予測の精度が向上する事を示した [8]。Lehman や Ghassemi らは、ICU 入室後 24 時間以内の看護記録に対して潜在トピック推定を行っている。両者とも重症度スコアに潜在トピック推定から得た特徴量を加える事で予測の精度が向上する事を示している [9]、[10]。Cavallero らは、バイタルサイン、血液検査、看護記録を一般化状態空間モデルを用いて統合し、死亡リスクのオンライン

[3]投薬など治療に関するデータも記録されているが、本研究では使用しない為除外した。

評価を行った [11]。電子カルテに記録されたデータ、機械学習を用いた転帰予測手法の開発は多くの研究が行われている。

これらの研究の多くは ICU 入室患者全体を対象とした転帰予測を行っているが、ICU に入室した患者はそれぞれ異なる疾患を持ち、疾患や患者の背景によって予測性能が大きくばらつく事が知られている。その為、個別の疾患に焦点を当てた予測システムを構築する事が望ましい。

2.2 敗血症患者の転帰予測

敗血症患者の転帰を予測する為に、スコアリングシステムを利用する手法が用いられてきた。MEDS（Mortality Emergency Department Sepsis:MEDS）スコアを用いた研究 [12] や、SOFA（Sequential Organ Failure Assessment:SOFA）スコア [13] を用いた予測 [14] などがそれに該当する。

機械学習的アプローチを用いた研究も行われている。Taylor らは 4676 人の敗血症患者から得られた 500 の変数に対してランダムフォレストを適用する事で、決定木分析やロジスティック回帰で予測を行った場合よりも、予測精度が向上すると報告している [15]。他にも、血圧や呼吸数を特徴量として、サポートベクターマシンやナイーブベイズ分類器による学習、予測を行い、その有用性について検討した Gultepe らの研究がある [16]。しかし、これまで行われてきた敗血症患者の転帰予測に関する研究は、定量データのみを用いた予測を行っており、本研究の様に、予測の際に看護記録を利用した研究は見受けられない。

3 システム構築

本章では、システムの詳細について述べる。システムは特徴量抽出、予測という 2 つの段階から成り立つ。

第一段階の特徴量抽出では、入室後 24 時間で電子カルテに記録されたデータから、基礎特徴量、状態比率、トピック比率の 3 種類を抽出する。基礎特徴量は、医師が標準的に診断に用いる変数である。状態比率は、バイタルサインから抽出される特徴量である。バイタルサインの変動は、自律神経など体内の制御システムの変動に由来する。制御システムは治療などの外部要因と、病状の進行など内部要因との相互作用により複数の変動パターンを持ち、それらの混合で構成される。そして、特定の変動パターンが敗血症患者の転帰に関連すると指摘されている [17]。我々は、制御システム及びバイタルサインの変動が複数の自己回帰モデルの混合により表す事が出来ると仮定し、各自己回帰モデルが混合している割合を状態比率として抽出する。トピック比率は看護記録から抽出される特徴量である。看護記録の潜在トピック推定を行い、患者ごとの各トピックの比率を特徴量とする。

第二段階の予測では、第一段階で得た 3 種類の特徴量を組み合わせて、サポートベクターマシンの学習、予測を行う。

3.1 特徴量抽出

3.1.1 基礎特徴量の抽出

ICU の医療スタッフは、デモグラフィック変数や、重症度スコアを参考に患者の状態を判断する。そこで、性別、年齢、BMI の 3 つのデモグラフィック変数と、APACHE II、SAPS II、SOFA スコアの 3 つの重症度スコアを基礎特徴量として用いる。SAPS II、APACHE II は患者の ICU 内死亡のリスク評価に広く用いられている。SOFA スコアは、神経系や臓器障害を点数化し、重症度を判定する事を目的としたスコアであり、敗血症患者の転帰予測に有用であるという報告がある [14]。この 6 種類の変数を基礎特徴量として患者ごとに抽出し、抽出した特徴量を $\{BL_l^{(n)}\}_{n=1}^N$、$(l=1$、2、$\cdots$、$6)$ とおく。ここで添え字 n は患者を表す添え字であり、$n=1$、2、$\cdots$、N である。

3.1.2 Switching AR model と状態比率

SARM（Switching Autoregressive Model:SARM）によるバイタルサインのモデリングと、それに基づく状態比率の抽出について説明する。状態比率の抽出は、Lehman らの方法に沿って行う [8]。バイタルサインは 1 分に 1 回の頻度で観測される一次元時系列データであると仮定し、患者 n における t 期の観測値を $x_t^{(n)}$ と表す。t については、$t=1$、2、$\cdots$、$T^{(n)}$ である。ここで、患者 n について観測されたバイタルサイン系列を $\mathbf{x}^{(n)} \equiv \{x_1^{(n)}$、$x_2^{(n)}$、$\cdots$、$x_{T^{(n)}}^{(n)}\}$ と表す。また、$\{\mathbf{x}^{(n)}\}_{n=1}^N$ をコホートバイタルサインと呼ぶ。

SARM は隠れマルコフモデルの一種であり、K 個の隠れ状態の遷移に伴う自己回帰（Autoregressive:AR）モデルの切り替わりによって、コホートバイタルサイン $\{\mathbf{x}^{(n)}\}_{n=1}^N$ をモデリングする。K 個の AR モデルに関するパラメータは、係数行列と分散共分散行列であるが、係数行列を $\{A_p^{(k)}\}_{k=1}^K$、$(p=1$、2、$\cdots$、$P)$、分散共分散行列を $\{Q^{(k)}\}_{k=1}^K$ とする。添え字 p は AR モデルの次数を表す。隠れ状態の遷移については。$s_t^{(n)}$ を患者 n の t 期における隠れ状態とし、患者 n の隠れ状態の系列を $\mathbf{s}^{(n)} \equiv \{s_1^{(n)}$、$s_2^{(n)}$、$\cdots$、$s_{T^{(n)}}^{(n)}\}$ と表す。また、$s_t^{(n)}$ はマルコフ過程に基づいて遷移すると仮定する。状態遷移に関するパラメータは、状態の初期分布と状態遷移行列であるが、状態の初期分布を $\{\pi^{(n)}\}_{n=1}^N$、状態遷移行列を Z とする。ここで、状態の初期分布 $\{\pi^{(n)}\}_{n=1}^N$ は患者ごとに固有だが、状態遷移行列 Z、AR モデルの係数行列 $\{A_p^{(k)}\}_{k=1}^K$、$(p=1$、2、$\cdots$、$P)$、分散共分散行列 $\{Q^{(k)}\}_{k=1}^K$ は患者間で共通である事を注記しておく。これは、各患者のバイタルサインは K 個の患者間に共通の AR モデルの混合で表現されるが、患者毎に隠れ状態の混合の比率が異なるため、得られる観測値の系列に変化が生じる事を意味する。

$\{\mathbf{s}^{(n)}\}_{n=1}^N$ が既知ならば、$x_t^{(n)}$ は定義したパラメータを用いて次の様に表す事が出来る。

$$x_t^{(n)} = \sum_{p=1}^{P} A_p^{(s_t^{(n)})} x_{t-p}^{(n)} + w_t^{(s_t^{(n)})}$$

$w_t^{(s_t^{(n)})}$ は分散共分散行列に $Q^{(s_t^{(n)})}$ を持つ正規ノイズ項である。しかし、実際には $\{\mathbf{s}^{(n)}\}_{n=1}^N$ は潜在変数であり、観測する事が出来ない。その為、EM（Expectation-Maximization:EM）アルゴリズムを用いて、尤度を最大化するパラメータと、$s_t^{(n)}$ が状態 k を取る事後確率 $Pr(s_t^{(n)}=k)$ をすべての n、t、k に対して求める。SARM における EM アルゴリズムの実装については、Murphy により紹介されている方法を用いる [18]。EM アルゴリズムは 2 つのステップからなり、その 2 つのステップを反復する事で、尤度を最大化するパラメータと隠れ状態の事後確率を推定する。第一ステップにあたる E ステップでは、forward-backward アルゴリズムを用いて、隠れ状態の事後確

率 $Pr(s_t^{(n)}=k)$ を計算する。第二ステップにあたる M ステップでは、直前の E ステップで得た $Pr(s_t^{(n)}=k)$ を所与として、尤度を最大化するように SARM のパラメータ Z、$\{A_p^{(k)}\}_{k=1}^{K}$、$(p=1、2、\cdots、P)$、$\{Q^{(k)}\}_{k=1}^{K}$、$\{\pi^{(n)}\}_{n=1}^{N}$ を求める。EM アルゴリズムを行う際に隠れ状態の個数 K と、AR モデルの最大次数 P は分析者が指定する必要がある為、AIC（Akaike's Information Criterion:AIC）[19] が最小となったモデルを選択する。

AIC によるモデル選択を行った後に、状態比率 $MP_k^{(n)}$ を特徴量として抽出する。状態比率は、患者 n のバイタルサイン系列 $\mathbf{x}^{(n)}$ のうち、k 個目の AR モデルが混合している比率を示す量である。EM アルゴリズムの E ステップで得られる $Pr(s_t^{(n)}=k)$ を用いて次の様に計算される。

$$MP_k^{(n)} = \frac{1}{T^{(n)}} \sum_{t=1}^{T^{(n)}} Pr(s_t^{(n)}=k)$$

この $MP_k^{(n)}$ をすべての n、k に対して計算し、予測の際の特徴量とする。

3.1.3 トピック比率

看護記録からのトピック比率抽出について説明する。対象は、ICU 入室後 24 時間以内に記録された全ての看護記録である。通常、ICU のスタッフは一日の間に一人の患者につき複数回看護記録を記入する。その複数の看護記録をまとめて一つの文書とする。患者ごとに文書を生成した後に、形態素解析を行う。形態素解析とは、単語の品詞情報等に基づき文書を最小単位である形態素に分解し、形態素の品詞を判別する為の解析である。形態素解析の際に、名詞、形容詞と判別された形態素を抽出し、文書を再構成する。再構成の結果得られた患者 n に関する文書を $d^{(n)}$ とする。

その後に、文書集合 $\{d^{(n)}\}_{n=1}^{N}$ に対して潜在的ディリクレ配分法（Latent Dirichlet Allocation:LDA）を用いて潜在トピック推定を行う [20]。LDA は、文書が M 個の潜在的なトピックを持ち、それらのトピックを媒介して単語が生成される事を仮定したトピックモデルである。LDA では、文書におけるトピックの出現分布と、各トピックにおける単語の出現分布に多項分布を仮定し、事前分布にディリクレ分布を仮定する事で潜在トピック推定を可能にしている。潜在トピック推定には、周辺化ギブス・サンプリングを用いる。トピックの出現分布に関する事前分布のハイパーパラメータ α は $\alpha=\frac{50}{numberTopics}$、単語の出現分布に関する事前分布のハイパーパラメータ β は $\beta=\frac{50}{numberVocab}$ とする。ここで、$numberTopics$ は潜在トピックの個数、$numberVocab$ は文書集合内のユニークな単語の数である。周辺化ギブス・サンプリングの反復回数は 5000 回とする。トピック推定で得られた $d^{(n)}$ における潜在トピック m の比率を $TP_m^{(n)}$、$(m=1、2、\cdots、M)$ とする。潜在トピックの個数 M は、第二段階で行う予測の精度に基づき決定する。

3.2 予測

第二段階では、第一段階で抽出した特徴量を組み合わせて計画行列を構成し、サポートベクターマシン（Support Vector Machine:SVM）[21] の転帰に対する学習と、それに基づく予測を行う。SVM は二値判別の為のアルゴリズムであり、カーネルトリックを用いて非線形の識別関数を構成出来る。SVM のカーネル関数としてはガウシアンカーネルを採用する。

図 1 にシステムの概観を、表 2 に構築した 4 種類のシステムとその特徴量の組み合わせを示す。電子カルテからデモグラフィック変数、コホートバイタルサイン $\{\mathbf{x}^{(n)}\}_{n=1}^{N}$、検査データ、文書集合 $\{d^{(n)}\}_{n=1}^{N}$ を取得し、特徴量抽出を行う。抽出した基礎特徴量 $BL_l^{(n)}$ を $(n、l)$ 要素に持つ配列

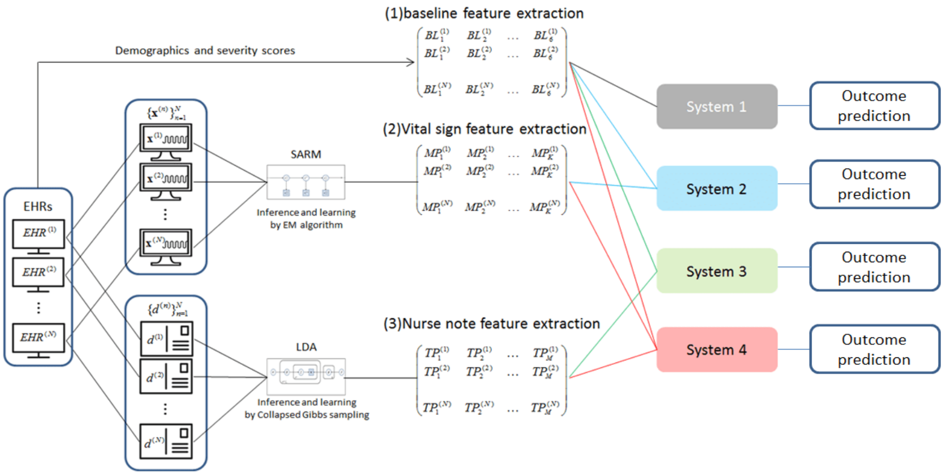

図 1: システムの概観

表 2: 実験に用いたシステム一覧

候補	BL	MP	TP
システム 1	○		
システム 2	○	○	
システム 3	○		○
システム 4	○	○	○

を BL、状態比率 $MP_k^{(n)}$ を $(n、k)$ 要素に持つ配列を MP、トピック比率 $TP_m^{(n)}$ を $(n、m)$ 要素に持つ配列を TP とする。システム 1 は BL のみを用いて、システム 2 は BL、MP を用いて、システム 3 は BL、TP を用いて、システム 4 は 3 種類の特徴量すべてを用いて転帰に対する学習、予測を行う。システム 1、2 は定量データのみを用いる従来型のシステムである。対して、システム 3、4 は従来の手法とは異なり、看護記録からの特徴量を用いる予測システムである。

4　実証実験

本章では、東北大学病院 ICU の実データを用いたシステムの実証実験について述べる。前半部では、使用したデータセットや予測の評価方法など実験条件について、後半部では、実証実験の結果について述べる。本文中における解析は全て R（ver．3．3．1）を用いて行った。

4.1　実験条件

4.1.1　データセット

データセットには、東北大学病院 ICU から収集したデータを用いた。対象は、2008 年 1 月から 2015 年 1 月の間に ICU に入室し、かつ、敗血症の診断基準 Sepsis-3[22] を満たす 20 歳以上の患

者 137 名である。ただし、我々が構築したシステムは入室後 24 時間以内の全てのデータを用いて予測を行うため、診断基準を満たしていても、ICU 滞在時間が 24 時間未満である患者については除外している。予測の対象となる転帰については ICU 内死亡、入室後 28 日死亡の 2 種類とした。転帰が ICU 内死亡となった患者は 21 人 (15.3%)、入室後 28 日死亡となった患者は 25 人 (18.2%) であった。第一段階の特徴量抽出に必要なデモグラフィック変数、検査データ、バイタルサイン、看護記録については各患者の電子カルテから取得した。

4.1.2 データの前処理

取得したデータから、状態比率、トピック比率を抽出するために、バイタルサイン、看護記録に対して行った前処理について説明する。

まず、バイタルサインに対して行った処理について説明する。東北大学病院 ICU では、20 種類のバイタルサインが 3 秒に 1 回の頻度で記録されている。状態比率の抽出対象となるバイタルサインを決定する必要があるが、平均血圧（Mean Arterial Pressure:MAP）を対象とした。電子カルテから入室後 24 時間以内の MAP を取得するが、その際に、MAP が 20mmHg 未満、または、200mmHg より大きい観測値は欠損値として処理した。次に取得した MAP の系列を、1 分間の中央値の系列へと変換した。[4]ICU 内で観測されるバイタルサインは多くの欠損を含む為、MAP の中央値の系列にも欠損が生じる。その為、MAP の中央値の系列から、欠損を含まない最も長い部分時系列を切り出し、SARM を適用する為、線形トレンドを取り除いた。トレンドを除去した部分時系列の中で、$|x_t^{(n)} - x_{t-1}^{(n)}| > 30$ となった場合、$x_t^{(n)}$ を欠損値と見なし、状態空間モデルにより補完した。状態空間モデルによる欠損値の補完については、Petris[23] らの方法に従った。

次に、看護記録の処理について述べる。まず、入室後 24 時間以内のすべての看護記録を電子カルテから取得し、患者ごとに一つの文書に統合する。その後、オープンソースの形態素解析環境 *MeCab*、*RMeCab*[24] を用いて形態素解析を行い、名詞、形容詞を抽出した。結果として、3927 個の看護記録から、137 個の文書、3616 個のユニークな単語を抽出した。

4.1.3 評価指標

予測性能の評価指標には、AUC（Area Under the ROC Curve）、感度、特異度を用いた。AUC は、学習したシステムを用いて転帰予測を行った時に正解数が不正解数よりも大きくなる確率を表しており、分類器の性能の良さを表す指標として用いられる。完全にランダムに判別を行った時 AUC の値は 0．5 となり、値が 1 に近いほど高い判別能力を持っている。感度、特異度はそれぞれ次の様に定義される。

$$感度 = \frac{TruePositive}{TruePostive + FalseNegative}$$

$$特異度 = \frac{TrueNegative}{TrueNegative + FalsePositive}$$

ここで、$TruePositive$ はシステムが死亡と予測し、実際の転帰が死亡だった患者の合計、$FalsePostive$ は死亡と予測し、実際の転帰が生存だった患者の合計、$TrueNegative$ は生存と予測し、実際の転帰が生存だった患者の合計、$FalseNegative$ は生存と予測し、実際の転帰が死亡だった患者の合計を示す。感度が高い事は、転帰が死亡である患者を誤って生存と予測する可能性が低い事を意味しており、特異度が高い事は、転帰が生存である患者を誤って死亡であると予測する可能性が低い

[4]本研究における SARM は、バイタルサインが 1 分間に 1 回記録される事を仮定している。しかし、東北大学病院 ICU ではバイタルサインは 3 秒に 1 回の頻度で記録される為このような処理を行った。

表 3: 基礎特徴量の群間差:ICU 内死亡

特徴量	生存群	死亡群
年齢	66.8 ± 12.8	66.2 ± 11.8
性別 (人)	75 (64.1%)	19 (90.4%)
BMI	21.9 ± 4.22	21.8 ± 3.89
SAPS II	40.1 ± 12.1	47.0 ± 11.6
APACHE II	23.1 ± 5.63	24.9 ± 6.57
SOFA	7.33 ± 3.10	9.76 ± 3.71

表 4: 基礎特徴量の群間差:入室後 28 日死亡

特徴量	生存群	死亡群
年齢	67.1 ± 12.8	64.6 ± 11.7
性別	75 (66.3%)	19 (76.0%)
BMI	22.1 ± 4.19	21.7 ± 4.09
SAPS II	40.7 ± 12.3	43.2 ± 12.3
APACHE II	22.8 ± 5.42	25.9 ± 6.78
SOFA	7.35 ± 3.11	9.28 ± 3.73

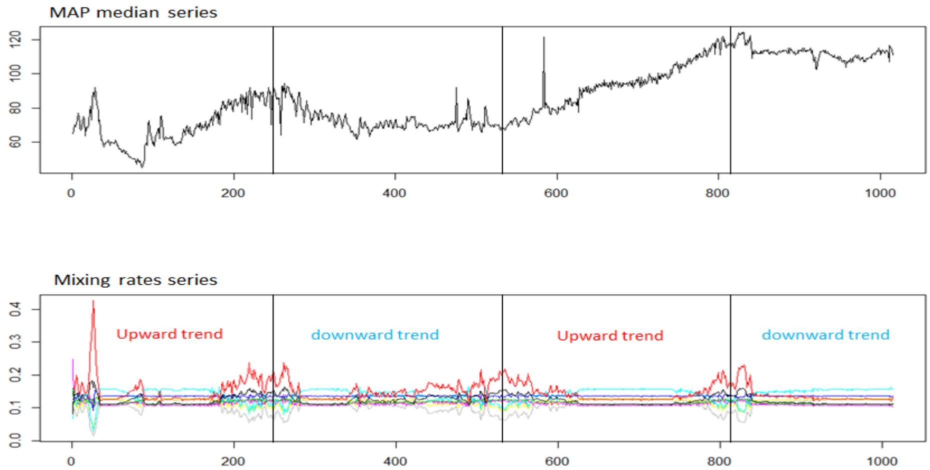

図 2: 隠れ状態の混合比率と MAP 中央値系列の推移

事を意味する。

4.2 結果

4.2.1 特徴量抽出結果

第一段階の特徴量抽出では、3 種類の特徴量抽出をデータに対して行った。基礎特徴量の抽出結果を表 3、4 に示す。性別についてはカイ二乗検定、その他の変数については Welch の t 検定を行った。転帰が ICU 内死亡の時、性別、SAPS II、SOFA について、転帰が入室後 28 日死亡の時、APACHE II、SOFA について生存群、死亡群の間に有意差が存在していた。[5]

状態比率は、AR モデルの最大次数の候補を $P = [3, 4, 5]$、隠れ状態数の候補を $K = [2, 3, \cdots, 30]$ とし、その組み合わせをモデル候補とした。EM アルゴリズムにより事後確率とパラメータの推定を行い、AIC を算出した。AIC によるモデル選択の結果、最適なモデルとして、$P = 3,\ K = 8$ のモデルが選択された。ある患者の (1)MAP 中央値の推移（MAP median series）、(2) 各時点にお

[5]性別は男性ならば 1、女性ならば 0 をとるダミー変数である。

表 5: 評価指標の比較：ICU 内死亡

候補	AUC	感度	特異度
システム 1	0.581	0.639	0.656
システム 2	0.546	0.523	0.738
システム 3	**0.732**	0.634	**0.808**
システム 4	0.725	**0.667**	0.795

表 6: 評価指標の比較：入室後 28 日死亡

候補	AUC	感度	特異度
システム 1	0.587	0.630	0.658
システム 2	0.556	0.683	0.635
システム 3	**0.695**	0.709	**0.769**
システム 4	0.670	**0.852**	0.590

ける 8 つの AR モデルの混合する割合 $\{Pr(s_t^{(n)} = k)\}_{k=1}^8$ の推移（Mixing rates series）を図 2 に示す。血圧の推移傾向が変化する前後で、AR モデルの混合する割合が変動している事が分かる。

トピック比率の抽出に関して、潜在トピック数を $M = [10, 20, \cdots, 50]$ とし、LDA を用いて潜在トピック推定を行った。その際に、R パッケージ *lda* を用いて潜在トピック推定を行い、トピック比率を抽出した。第二段階で得た AUC に応じて、潜在トピック数を 20 とした。

4.2.2　予測結果

第二段階の予測では、2 種類の転帰それぞれに対して、SVM の学習、予測を行った。それぞれの転帰に関する評価指標の値を表 5、6 に示す。表には 3-fold cross validation の元で計算した評価指標の値を記載している。

ICU 内死亡に関する予測の比較

システム 3 において、AUC が 0.732、特異度が 0.808 で最大となった。感度はシステム 4 において最大となり、0.667 であった。一方で、定量データのみを用いるシステム 1、2 は、トピック比率を用いるシステムと比較すると AUC、特異度が低い値を取っている。AUC が最大となるシステム 3 と、最も低い値をとったシステム 2 では約 0.18 の差があった。感度については、システム 2、4 の間に約 0.14、特異度についてもシステム 1、3 の間に約 0.15 の差がある。

入室後 28 日死亡に関する予測の比較

転帰を入室後 28 日死亡とした時の予測についても、システム 3 において AUC が 0.695、特異度が 0．769 となり、4 つのシステムの中で最大となった。感度はシステム 4 の 0.852 が最大であった。システム 4 については、AUC についても 0.670 とシステム 3 に次ぐ値であるが、ICU 内死亡と比較して差が大きくなっている。一方で，システム 1、2 については、ICU 内死亡と同様に、トピック比率を用いるシステムと比較して評価指標は低い値を取った。AUC が最大となるシステム 3 と、最も低い値をとったシステム 2 では 0.14 程度の差があった。感度については、システム 1、4 の間に約 0.22、特異度については、システム 3、4 の間に約 0.17 の差があった。

5　考察

本章では、システムの有効性について考察する。敗血症患者の予後を改善する上で、転帰予測の正確性、予測を行うタイミングが重要になる。5．1 節では、実証実験の結果をもとに正確性に関

する議論を行い、5．2 節では予測を行うタイミングについて議論する。最後に、看護記録を用いるシステムの臨床的意義を述べる。

5.1 正確性

AUC の面で、トピック比率を用いるシステム 3、4 は、どちらの転帰を予測する場合でも、定量データのみを用いるシステム 1、2 よりも高い値を示した。すなわち、トピック比率を用いる予測は、従来手法よりも正確に患者の転帰を予測できる。

感度については、2 種類の転帰両方で、3 種類すべての特徴量を用いるシステム 4 が最大となった。感度の高いシステムは、死亡リスクの高い患者を見落とす可能性が低くなる。すなわち、システム 4 は、他のシステム候補と比べて、重症化し得る患者をより鋭敏に検知出来る。

特異度については、2 種類の転帰両方で、システム 3 が最大となった。特異度が高いシステムは、死亡リスクの低い患者に対して誤って死亡アラートを出す可能性が低くなる。敗血症の治療において、軽症な患者に対する過剰な抗菌薬投与は、死亡リスクを増大させる可能性が指摘されている。[7]。過剰な治療を導かないという点で、システム 3 は他のシステム候補に勝る。

以上より、AUC、感度、特異度の面で、トピック比率を用いるシステム 3、4 は、定量データのみを用いるシステムと比較して、より正確に患者の転帰を予測出来る。

5.2 予測のタイミング

ICU 内の医師は、患者の重症度を評価し、介入に関する意思決定を行う。患者の重症度を評価する為に、SAPS II や APACHE II などが一般的に用いられている。これらのスコアは、患者が ICU に入室してから 24 時間が経過した時点で計算される。また、敗血症患者の転帰予測に関する研究の多くは、入室後 24 時間のデータを用いて行われる。我々が構築した 4 つのシステムも、重症度スコアの計算や、従来の研究と同じタイミングで予測を行っている。その為、医師は、従来手法と比較して、遅れる事なく患者への介入内容を決定できる。

5.3 臨床的意義

ICU の医療スタッフは一人につき複数人の患者を担当する。また、電子カルテ内のデータは手動で入力が行われている物も多い。その為、意思決定支援システムは、医療スタッフのデータ入力に関する負担が小さい物が望ましい。システム 3、4 で用いた看護記録は、通常の業務ルーティンの中で記入される。その為、転帰予測に看護記録を利用する事は、医療スタッフに新たなデータの入力を要求しない。

以上より、看護記録から抽出したトピック比率を用いるシステム 3、4 は、従来の手法と同じタイミングで、かつ、医療スタッフの負担を大きくする事なく、より正確に敗血症患者の転帰を予測出来る。

6 むすび

本研究では、電子カルテ内の看護記録を含む異質なデータを統合した敗血症患者の転帰予測システムを構築した。そして、看護記録から抽出したトピック比率を用いるシステムが従来手法と同じ

タイミングで、医療スタッフの負担を大きくする事なく、かつ、より正確に転帰を予測出来るという意思決定支援システムとして有効性を持つ事を示した。

今後の研究の発展として、SARM における患者毎の異質性を導入したモデルの開発が挙げられる。本研究では、SARM における状態遷移行列は患者間で共通である事を仮定したが、各患者は年齢や合併症などに応じて、それぞれ異なる治療を受ける。その為、状態遷移行列も個人ごとに異なる可能性を持つ。各患者の異質性を反映したモデルへと SARM を拡張する事で、バイタルサインをより精緻にモデル化し、予測精度の更なる向上を図る。

参考文献

[1] L.G. JR, L.S, and S.F. ”A new simplified acute physiology score (saps ii) based on a europian/north american multicenter study,” Journal of American Medical Association, vol.270, no.24, (1993)

[2] J.J Escarce and M.A. Kelly, ”Admission source to the medical intensive care unit predicts hospital death independent of APACHE II score,” Journal of American Medical Association, col.264, no.18,(1985)

[3] M. Ghassemi, L.A. Celi, and D.J. Stone, “ State of the art review: the data revolution in critical care, ” Critical Care, vol.19, (2015)

[4] G.C. Siontis, I. Tzoulaki, and J.P. Ioannidis, “ Predicting death: an empirical evaluation of predictive tools for mortality, ” Archives of Internal Medicine, vol.171, no.19, (2011).

[5] D.C. Angus, W.T. Linde-Zwruble, J. Lidicker, G. Clermont, J. Carcillo and M.R. Pinsky, ”Epidemiology of severe sepsis in the United State: analysis of incidence, outcome, and associated costs of care, ” Critical Care Medicine, vol.29, no.7, (2001)

[6] A. Kumar, D. Roberts, K.E. Wood, B. Light, J.E¿ Parrillo, S. Shama, R. Suppes, D. Feinstein, S. Zanotti, L. Taiberg, D. gurka, M. Cheang, ”Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock,” Critical Care Medicine, vol.34, no.6, (2006)

[7] A. Kumar, N. Safdar, S. Kethireddy and D.Chateau ”A survival benefit of combination antibiotic therapy for serious infections associated with sepsis and septic shock is contingent on the risk of death: A meta-analytic/meta-regression study,” Critical Care Medicine, vol.38, no.8,(2010)

[8] L.weiLehman, R.P. Adams, G.B. Moody, A.Malhotra, R.G. Mark,and S.Nemati, ”A physiological time series dynamics-based approach to patient monitoring and outcome prediction,” IEEE Jounal of Biomedical and Health Informatics, vol.19, no.3, (2015)

[9] L. weiLehman, M. Mohammed Saeed, W. Long, J.Lee, and R. Mark, “ Risk stratification of icu patients using topic models inferred from unstructured progress notes, ” AMIA Annual Symposium Proc., vol.2012, (2012)

[10] M. Ghassemi, T. Naumann, F. Doshi-Velez, N.Brimmer, R. Joshi, A. Rumshisky, and P. Szolovits, "Unfolding physiological state: Mortality modelling in intensive care units," Proc. 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (2014)

[11] K.L. Caballero and R. Akella, "Dynamic modelling patient's health state from electric medical records: A time series approach," Proc. 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (2015)

[12] N.I. Shapiro, R.E. Wolfe, R.B Moore, E.Smith, E. Burdick and D.W. Bates, "Mortality in Emergency Department Sepsis (MEDS) score: a prospectively derived and validated clinical prediction rule," Critical Care Medicine, vol.31, no.3, (2009)

[13] J.I. Vincent, R. Moreno, J.Takala, S.Willarts, A. De Mendonca, H. Bruining, C.K. Reinhart, P¿M.Suter, L.G. Thijis, "The SOFA(Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure," Intensive Care Medicine, vol.22, (1996)

[14] A.E. Jones, S. Trzeciak and J.A. Kine, "The sequential organ failure assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation," Critical Care Medicine, vol.37, no.5, (2009)

[15] R.A. Taylor, J.R. Pare, A.K. Venkatesh, H. Mowafi, E.R. Melnick, W. Fleischman, and M.K. Hall, "Prediction of IN-hospital moratality in emergency department patients with sepsis: a local big data-driven, machine learning approach," Academic emergency medicine vol.23, no.3, (2016)

[16] E. Gultepe, J.P. Green, H. Nguyen, J. Adams, T. Albertson, and I. Tagkopoulos, "From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system," Jounal of the American Medical Informatics Association vol.21, (2014)

[17] L. weiLehman, A. Johnson, C. Sudduth, R. Mark, S. Nemati, "Dynamics of multivariate vital sign time series and severe sepsis among patients in critical care," Journal of Critical Care, vol.38, (2017)

[18] K.P. Murphy, "Switching Kalman filters," Dept. of Computer Science, University of California, Berkeley, Tech. Rep, (1998)

[19] H. Akaike, "A new look at the statistical modelidentification," IEEE Trans. Autom. Control, vol.16, no.6, (1974)

[20] D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent Dirichlet allocation," J. Machine Learning Research, vol.3, (2003)

[21] V. Vapnik, "Statistical Learning Theory," John Wiley & Sons,Inc., (1998)

[22] M. Singer, C.S. Deutschman, C.W. Seymour,etal, "The third international consensus definitions for sepsis and septic shock(Sepsis-3)," Journal of American Medical Association, vol.315, no.8, (2016)

[23] G. Petris, S. Petrone, and P. Campagnoli. Dynamic Linear Models with R. use R! Springer-Verlag, (2009)

[24] 石田基弘, Rmecab, http://rmecab.jp/wiki/ index.php?RMeCabFunctions, (2016)