

疾病間での転移学習を用いたICU入室患者のリスク予測

Risk Prediction for ICU Patients via Disease-Based Transfer Learning

則のぞみ *1 鹿島久嗣 *1 山下和人 *2 猪飼宏 *3 今中雄一 *2
 Nozomi Nori Hisashi Kashima Kazuto Yamashita Hiroshi Ikai Yuichi Imanaka

*1 京都大学大学院情報学研究科知能情報学専攻
 Graduate School of Informatics, Kyoto University

*2 京都大学大学院医学研究科医療経済学分野
 Graduate School of Medicine, Kyoto University

*3 山口大学医学部附属病院医療情報部
 Department of Medical Informatics and Decision Science, Yamaguchi University Hospital

In the intensive care unit (ICU), clinical staffs are faced by diverse patients with a wide variety of diseases, which calls for disease-specific risk stratification for ICU patients. Yet, few studies have investigated data-driven risk modeling for “infrequent diseases”, namely, diseases whose number of patients is comparatively small in the entire population. In this study, we explore the feasibility of transfer learning for the risk modeling of infrequent diseases in the context of mortality prediction in the ICU. Conducting a pilot experiment, we show that by transferring knowledge from other diseases, the predictive performance for each disease could become comparable to or better than that of one of the most standard methods, L2-regularized logistic regression, and the predictive performance for infrequent diseases was more improved than for frequent diseases compared with a multi-task learning method that makes use of clinical domain knowledge; for some specific infrequent diseases, the predictive performance was improved significantly from these existing methods.

1. はじめに

集中治療室 (ICU) 入室患者の死亡リスク予測は、死亡リスクが高い患者に事前にアラートを出すといった医療支援への応用や、医療経済学分野で行われる ICU 間パフォーマンス比較等の医療研究支援に繋がる重要な問題である。これまでも死亡リスク予測のための様々なモデルが探求されてきた [Ghassemi 15] が、従来の研究では疾病ごとではなく ICU 入室患者全体を対象に共通の一つのモデルが構築されることが一般的であり、“疾病によって死亡リスクを説明するルールが異なる”というような“疾病コンテキスト”は十分に考慮されてこなかった。近年は、疾病コンテキストを考慮するために疾病を単位としたマルチタスク学習として定式化する試みも行われている [Nori 15] が、全体の疾病中比較的患者数の少ない疾病である“低頻度疾病”に対する予測モデル構築のための研究はほとんど行われていないのが現状である。

一般的な傾向として、一つの低頻度疾病の患者数は少ないものの全ての低頻度疾病の合計患者数は無視できないものとなる。例として希少疾患の分布が挙げられる。米国では約 7,000 の疾病が希少疾患として定義されており 2009 年時点で合計 2,500 万人から 3,000 万人程度が希少疾患患者であると推定されている [Griggs 09] が、これは、米国人口の約 8% から 12%, 8 人から 12 人に 1 人が希少疾患患者であることを意味する。このような傾向は QIP プロジェクト *1 で収集された日本国内約 170 病院、約 20 万人の患者から成る ICU データセット中에서도確認された。図 1 は疾病毎の患者数のヒストグラムであり、各疾病は ICD10 3 桁コードで定義されている。多くの疾病、具体的には約半数の疾病が患者数 10 以下の最も低頻度な疾病に

分類されており、約 80% の疾病が最も低頻度な疾病と、患者数 100 以下の二番目に低頻度な疾病のいずれかに分類されている。希少疾患の定義は国によって異なるが、日本と米国における、患者数の割合に基づく定義に似た定義として、2,000 人以上に 1 人という基準を上記のデータセットに当てはめると、そのような疾病の合計患者数は全体の患者の約 6% を占めることになり、この場合も無視できない数となる。

低頻度疾病に対する個別化された予測モデルの構築は重要であると考えられるが、データの希少性への対処が大きな課題となる。本論文では、この課題への解決策として転移学習を用いることを提案し、実際の病院のデータに適用した初期分析の結果を報告する。

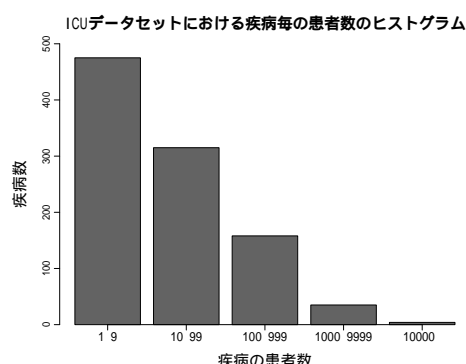


図 1: ICU データセットにおける疾病毎の患者数のヒストグラム。多くの疾病が低頻度の疾病である。

連絡先: 則のぞみ, 京都大学情報学研究科知能情報学専攻,
 nozomi@ml.ist.i.kyoto-u.ac.jp

*1 <http://med-econ.umin.ac.jp/QIP/>

2. 転移学習を用いた疾病に個別化された予測モデルの構築

本論文では、低頻度な疾病に対する個別化されたリスク予測モデル構築のために、Bickel らによる転移学習手法 [Bickel 08] を用いる。

2.1 問題設定

疾病 z は未知の同時分布 $p(\mathbf{x}, y|z)$ に紐付けられると仮定する。ここで、 \mathbf{x} は患者を表す特徴ベクトルであり、 y は患者のアウトカムを表すラベルである。訓練データのサンプル $D = \{(\mathbf{x}_1, y_1, z_1), \dots, (\mathbf{x}_n, y_n, z_n)\}$ は全疾病から構成され、患者 i について、属性 \mathbf{x}_i : 患者を表す M 次元の特徴ベクトル、クラスラベル $y_i \in \{0, 1\}$: 患者のアウトカム (0 と 1 はそれぞれ院内生存, 死亡を表す), 元のタスク $z_i \equiv \{1, \dots, T\}$: 患者の有する疾病が既知であるとする。全サンプル D は同時分布 $p(z)p(\mathbf{x}, y|z)$ に従う。我々のゴールは、各疾病 z について $p(\mathbf{x}|z)$ に従う未知の事例 \mathbf{x} に対する正しいラベルを予測する仮説 $f_z: \mathbf{x} \rightarrow y$ を学習することである。つまり、各疾病 z について下記の期待損失が最小化されなければならない:

$$\mathbf{E}_{(\mathbf{x}, y) \sim p(\mathbf{x}, y|z)}[\ell(f_z(\mathbf{x}), y)]. \quad (1)$$

2.2 アプローチ

全疾病に対して入手可能な全データを蓄積し、 $\sum_z p(z)p(\mathbf{x}, y|z)$ に従うサンプルのプールを得る。ターゲット疾病 t について、蓄積された事例集合 (プール) 中の各事例に対して疾病固有のリサンプリング係数 $r_t(\mathbf{x}, y)$ を仮定する。このリサンプリング係数はプールをターゲットの分布 $p(\mathbf{x}, y|t)$ に整合させるために導入される係数であり、下記のように定義できる:

$$\begin{aligned} \mathbf{E}_{(\mathbf{x}, y) \sim p(\mathbf{x}, y|t)}[\ell(f(\mathbf{x}, t), y)] \\ = \mathbf{E}_{(\mathbf{x}, y) \sim \sum_z p(z)p(\mathbf{x}, y|z)}[r_t(\mathbf{x}, y)\ell(f(\mathbf{x}, t), y)]. \end{aligned} \quad (2)$$

リサンプリング係数は下記のように表現できることが示されている [Bickel 08]:

$$r_t(\mathbf{x}, y) = \frac{p(t|\mathbf{x}, y)}{p(t)}. \quad (3)$$

$p(t|\mathbf{x}, y)$ のモデル化には様々な確率的分類器を用いることができるが、我々は、全疾病に対する $p(t|\mathbf{x}, y)$ を、パラメータ \mathbf{v} を用いたソフトマックスモデルを用いて下記のようにモデル化する:

$$p_v(z|\mathbf{x}, y) = \frac{\exp(\mathbf{v}_z^\top \phi(\mathbf{x}, y))}{\sum_{z'} \exp(\mathbf{v}_{z'}^\top \phi(\mathbf{x}, y))}. \quad (4)$$

ここで、パラメータベクトル \mathbf{v} は疾病固有のベクトル \mathbf{v}_z の連結ベクトルである。式 (4) では問題固有の特徴写像: $\phi(\mathbf{x}, y)$ を表すが、我々は $\phi(\mathbf{x}, y)$ をクロネッカーのデルタを用いて下記のように定義する:

$$\phi(\mathbf{x}, y) = \begin{bmatrix} \delta(y, +1)\phi(\mathbf{x}) \\ \delta(y, -1)\phi(\mathbf{x}) \end{bmatrix}. \quad (5)$$

L2 正則化を加え、我々が解くべき最適化問題は下記のように定義される:

$$\min_{\mathbf{v}} \lambda \mathbf{v}^\top \mathbf{v} - \sum_{(\mathbf{x}_i, y_i, z_i) \in D} \log p_v(z_i|\mathbf{x}_i, y_i). \quad (6)$$

ここで、 $\lambda \geq 0$ は正則化の重みを調整するハイパーパラメータである。確率モデル $p(t|\mathbf{x}, y)$ を得た後は、式 (3) を用いてリサンプリング係数を計算することができ、続いて重み付けされた訓練データに対する期待損失を評価することで式 (2) における期待損失を最小化することができる。本論文では、L2 正則化ロジスティック回帰を用いて一連のターゲット疾病のモデルを学習する。

3. 実験

3.1 実験条件

3.1.1 データセット

QIP プロジェクトで収集された一病院のデータを用いた。患者の疾病は患者の入院の契機となった疾病の ICD10 3 桁コードにより定義した。前処理として 18 歳未満の患者を除き、結果として 1,715 人の患者集合と 145 の疾病集合を得た。表 1 に、患者数に基づく分類ごとの疾病の一覧を示した。大半の疾病、具体的には 80% 以上の疾病が患者数 10 未満の最も低頻度な疾病に分類されている。患者が入院期間中に死亡した場合は患者のアウトカムを死亡とし、それ以外の場合は生存とした。患者を表現する特徴量は以下のように構成した。まず、各患者について年齢と性別を抽出し、年齢については“65 歳以下”と“65 歳より高齢”の二つの特徴量を作成した。また、患者の入院の契機となった疾病と入院時の併存症について、ICD10 4 桁コードを特徴量に加えた。最後に、入院期間中に患者に対して行われた医療費を伴う全ての介入行為を特徴量に加えた。これらは投薬、検査、手術などの患者に対して行われた介入行為であり、DPC システムによってコーディングされている。

3.1.2 予測設定

全データセットのうち 60% の患者を訓練データとしてランダムに抽出し、残りの 40% を評価データとして用いた。各患者について、退院日前日までに入手可能な全特徴量を用いた。各手法におけるハイパーパラメータは交差確認 ($k = 3$) を用いて $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0\}$ の中から決定された。サンプリング、予測、評価の一連のプロセスを 30 回繰り返し、対応のある t 検定で予測性能の平均値を比較した。評価指標としては Recall, Precision, F-measure の 3 つを採用した。これらの指標を評価するために、各疾病は一つ以上の正例 (院内死亡した患者事例) を有する必要がある。評価対象疾病を上記条件に合う疾病に絞った結果、評価対象疾病は 46 疾病となった。表 2 に、評価対象疾病について患者数に基づく分類ごとの疾病 (ICD10 の 3 桁コード) の一覧を示した。また、表 3 にデータセットの詳細についてまとめた。

表 3: データセット詳細。

患者数	1,715
疾病数	145
評価対象疾病数	46
評価対象疾病中の患者数	1,246
特徴量数	2,461

3.1.3 比較手法

まず、ICU 入室患者の死亡リスク予測のための最も標準的な手法の一つとして、全疾病に対して一つの共通モデルを構築する L2 ロジスティック回帰 (以下、common model と表記) を比較手法とした。また、疾病コンテキストを考慮した近年の手法として、疾病の分類と EHR の分類に関する医療知識を取り込んだマルチタスク学習手法 [Nori 15] (以下、multi-task

表 1: 患者数に基づく分類ごとの疾病 (ICD10 の 3 桁コード) の一覧 (本実験で学習に用いられた疾病一覧) .

患者数	疾病数 (全体の患者数に対する %)	疾病一覧
1 ~ 9	118 (81 %)	A09, A41, B19, B44, C19, C23, C38, C48, C50, C54, C56, C64, C65, C66, C73, C77, C79, C85, D12, D13, D14, D15, D20, D27, D35, D37, D48, D50, D63, D64, D65, E06, E11, E14, E16, E20, E86, E87, F10, F41, G40, G51, I08, I10, I23, I26, I31, I33, I42, I46, I47, I48, I49, I61, I63, I72, I74, I77, I81, I82, I83, I84, I85, I86, J15, J20, J44, J46, J69, J80, J81, J84, J85, J86, J90, J96, J98, K22, K25, K26, K27, K29, K31, K35, K40, K41, K43, K45, K51, K52, K55, K57, K59, K72, K75, K76, K82, K85, K86, K92, L02, L98, M06, M19, M46, M84, M87, N12, N13, N18, N20, N21, N32, N39, N70, N73, N80, N82
10 ~ 99	23 (16 %)	C15, C22, C24, C25, C34, C78, I20, I21, I25, I34, I35, I50, I70, J93, K50, K56, K63, K65, K80, K81, K83, M16, M17
100 ~ 200	4 (3 %)	C16, C18, C20, I71

表 2: 患者数に基づく分類ごとの疾病 (ICD10 の 3 桁コード) の一覧 (本実験で評価に用いられた疾病一覧) .

患者数	疾病数 (全体の患者数に対する %)	疾病一覧
1 ~ 9	27 (59 %)	A41, C23, C79, D15, D65, E87, F41, I31, I46, I63, I82, J15, J20, J44, J69, J81, J84, J96, K45, K52, K55, K72, K75, K85, K92, M06, M87,
10 ~ 99	15 (32 %)	C15, C22, C24, C25, C34, C78, I20, I21, I50, I70, K50, K56, K65, K81, K83
100 ~ 200	4 (9 %)	C16, C18, C20, I71

model と表記) も比較対象とした .

3.2 結果

3.2.1 common model との比較

表 4 に , Recall, Precision, F-measure の評価指標に基づく common model と転移学習手法 (以下 transfer model と表記) の比較 (対応のある t 検定 ($p < 0.05$) を用いた評価による) を示す . 計 4 つの疾病において少なくとも 1 つの指標で common model から transfer model で有意に性能が向上した . 一方, いずれの指標においても, common model から transfer model で有意に精度が低下した疾病は 0 であった . transfer model では common model に対して各疾病で少なくとも同等かそれ以上の予測性能が確認された .

3.2.2 multi-task model との比較

表 5, 6, 7 に, それぞれ Recall, Precision, F-measure の評価指標に基づく multi-task model と transfer model の比較 (対応のある t 検定 ($p < 0.05$) を用いた評価による) を示した . Recall については計 12 の疾病において multi-task model から transfer model で有意に性能が向上し, そのうち 9 の疾病, つまり $\frac{9}{12} = 75\%$ の疾病は患者数 10 未満の低頻度な疾病であった . 一方, 計 7 の疾病において multi-task model から transfer model で有意に性能が低下し, そのうち 4 の疾病, つまり $\frac{4}{7} \approx 57\%$ の疾病は患者数 10 未満の低頻度な疾病であった . Precision については, 計 7 の疾病において multi-task model から transfer model で有意に性能が向上し, そのうち 6 の疾病, つまり $\frac{6}{7} \approx 85\%$ の疾病は患者数 10 未満の低頻度な疾病であった . 一方, 計 9 の疾病において multi-task model から transfer model で有意に性能が低下し, そのうち 3 の疾病, つまり $\frac{3}{9} \approx 33\%$ の疾病は患者数 10 未満の低頻度な疾病であった . F-measure については, 計 8 の疾病において multi-task model から transfer model で有意に性能が向上し, そのうち 7 の疾病, つまり $\frac{7}{8} \approx 87\%$ の疾病は患者数 10 未満の低頻度な疾病であった . 一方, 計 10 の疾病において multi-task model から transfer model で有意に性能が低下し, そのうち 4 の疾病, つまり $\frac{4}{10} = 40\%$ の疾病は患者数 10 未満の低頻度な疾病であった . transfer model で予測性能の向上が確認された疾病の多くが低頻度な疾病である一方, 予測性能が低下した疾病には

低頻度な疾病は多く含まれなかった .

表 4: Recall, Precision, F-measure の評価指標に基づく common model と transfer model の比較 (対応のある t 検定 ($p < 0.05$) を用いた評価による) .

(a) Common model と比較して transfer model において Recall が向上した疾病 .

疾病コード	患者数	Recall (Transfer)	Recall (Common)
D65	6	0.204	0.000
J69	9	0.220	0.065
K50	11	0.429	0.071
K65	25	0.224	0.054

(b) Common model と比較して transfer model において Precision が向上した疾病 .

疾病コード	患者数	Precision (Transfer)	Precision (Common)
D65	6	0.160	0.000
K65	25	0.170	0.040

(c) Common model と比較して transfer model において F-measure が向上した疾病 .

疾病コード	患者数	F-measure (Transfer)	F-measure (Common)
D65	6	0.173	0.000
K65	25	0.176	0.040

4. おわりに

本論文では, 低頻度な疾病に対して個別化された予測モデルを構築するにあたって課題となるデータの希少性に対処するために転移学習手法を用いることを提案し, 国内病院の実データを用いた初期実験の結果を報告した . 既存手法の中で最も標準的な手法の一つである, 全疾病に対して一つの共通モデルを構築する L2 ロジスティック回帰と比較して, 転移学習手法では各疾病で同等かそれ以上の予測性能が確認された . また, 疾病の分類や EHR の分類に関する医療知識を取り込んだマルチタ

表 5: Recall の評価指標に基づく multi-task model と transfer model の比較 (対応のある t 検定 ($p < 0.05$) を用いた評価による) .

(a) Multi-task model と比較して transfer model において Recall が向上した疾病 .

疾病コード	患者数	Recall (Transfer)	Recall (Multi-task)
E87	1	0.333	0.000
F41	1	0.167	0.000
K45	3	0.200	0.000
K72	3	0.222	0.000
M06	3	0.294	0.000
J96	5	0.444	0.000
M87	7	0.211	0.000
I63	9	0.365	0.154
J69	9	0.220	0.036
K50	11	0.429	0.000
I70	89	0.175	0.000
I71	133	0.385	0.000

(b) Multi-task model と比較して transfer model において Recall が低下した疾病 .

疾病コード	患者数	Recall (Transfer)	Recall (Multi-task)
K52	1	0.000	0.800
K55	3	0.150	0.617
D65	6	0.204	0.574
K92	9	0.200	0.520
C24	12	0.194	0.528
K81	13	0.059	1.000
K56	59	0.143	0.488

表 6: Precision の評価指標に基づく multi-task model と transfer model の比較 (対応のある t 検定 ($p < 0.05$) を用いた評価による) .

(a) Multi-task model と比較して transfer model において Precision が向上した疾病 .

疾病コード	患者数	Precision (Transfer)	Precision (Multi-task)
E87	1	0.333	0.000
F41	1	0.167	0.000
K72	3	0.194	0.000
M06	3	0.235	0.000
J96	5	0.389	0.000
M87	7	0.096	0.000
K50	11	0.264	0.000

(b) Multi-task model と比較して transfer model において Precision が低下した疾病 .

疾病コード	患者数	Precision (Transfer)	Precision (Multi-task)
K52	1	0.000	0.800
K55	3	0.233	0.933
K92	9	0.150	0.640
C24	12	0.085	0.583
K81	13	0.020	1.00
K65	25	0.170	0.481
I21	50	0.046	0.417
I50	57	0.354	0.600
K56	59	0.051	0.589

表 7: F-measure の評価指標に基づく multi-task model と transfer model の比較 (対応のある t 検定 ($p < 0.05$) を用いた評価による) .

(a) Multi-task model と比較して transfer model において F-measure が向上した疾病 .

疾病コード	患者数	F-measure (Transfer)	F-measure (Multi-task)
E87	1	0.333	0.000
F41	1	0.167	0.000
K45	3	0.167	0.000
K72	3	0.204	0.000
M06	3	0.255	0.000
J96	5	0.407	0.000
M87	7	0.132	0.000
K50	11	0.310	0.000

(b) Multi-task model と比較して transfer model において F-measure が低下した疾病 .

疾病コード	患者数	F-measure (Transfer)	F-measure (Multi-task)
K52	1	0.000	0.800
K55	3	0.178	0.722
D65	6	0.173	0.377
K92	9	0.156	0.560
C24	12	0.111	0.546
K81	13	0.029	1.000
K65	25	0.176	0.381
I21	50	0.060	0.305
I50	57	0.280	0.404
K56	59	0.070	0.508

スク学習手法と比較した場合, 特に低頻度疾病に対して予測性能が向上する傾向が確認された. 以上の結果は, 低頻度疾病に対する個別化予測モデルの構築に転移学習が有効であることを示唆すると考えられる .

参考文献

- [Bickel 08] Bickel, S., Bogojeska, J., Lengauer, T., and Scheffer, T.: Multi-task Learning for HIV Therapy Screening, in *Proceedings of the 25th International Conference on Machine Learning*, pp. 56–63 (2008)
- [Ghassemi 15] Ghassemi, M., Celi, L. A., and Stone, D. J.: State of the art review: the data revolution in critical care, *Critical Care*, Vol. 19, p. 118 (2015)
- [Griggs 09] Griggs, R. C., Batshaw, M., Dunkle, M., Gopal-Srivastava, R., Kaye, E., Krischer, J., Nguyen, T., Paulus, K., and Merkel, P. A.: Clinical research for rare disease: opportunities, challenges, and solutions, *Molecular Genetics and Metabolism*, Vol. 96, No. 1, pp. 20–26 (2009)
- [Nori 15] Nori, N., Kashima, H., Yamashita, K., Ikai, H., and Imanaka, Y.: Simultaneous Modeling of Multiple Diseases for Mortality Prediction in Acute Hospital Care, in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 855–864 (2015)