

# Simultaneous Modeling of Multiple Diseases for Mortality Prediction in Acute Hospital Care

Nozomi Nori, Hisashi Kashima, Kazuto Yamashita, Hiroshi Ikai, Yuichi Imanaka:  
Proceedings of the 21st ACM SIGKDD Conference on Knowledge Discovery and  
Data Mining (KDD 2015).

# 概要

---

- ICUにおける患者の危険リスク予測問題を複数疾病の同時モデリング（疾病をタスクの単位としたマルチタスク学習）として定式化
- 一般的な機械学習手法としての新規性はあまりないが、医療データ解析の手法として新しく有効なものを提供したと考える

# コンピュータによる医療支援： ICUは有望な領域の一つ

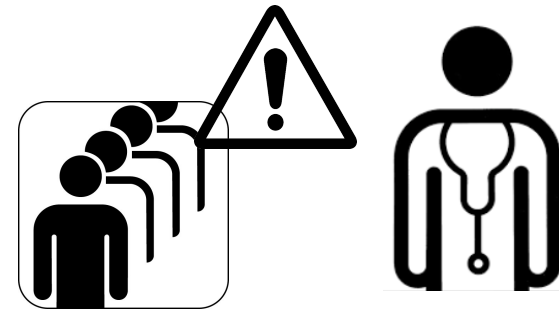


- ICUなどの急性期医療における機械による支援
  - 様々な情報源からの大量データを限られた時間で活用し意思決定する必要性
  - 機械が何かしらの形で人間の判断を支援する
    - 予測に基づき危険リスクが高い患者にアラート
    - 大量データの要約に基づく患者の病態評価支援



# ICUにおける危険リスク予測問題：予測に基づく診療支援や予測モデルに基づく医療研究の支援に繋がる

- ICUでは患者の死亡リスク（危険リスク）が患者の重症度の代替物としてしばしば利用される
- 患者の危険リスクの正確な予測
- 医療支援：医師の意思決定を助け適切な治療，患者の予後の改善に繋がりうる
- アラート/トリアージ/資源の最適化
- 医療研究支援：予測モデルに基づく分析
- ICU間パフォーマンス比較等



# しかし従来は疾病毎に危険リスクを説明するルールが異なるという疾病の個別性が考慮されてこなかった

- 背景：ICUでは様々な疾病の患者が対象になるが、従来は疾病毎に区別されずに分析が行われてきた
- 疾病の個別性：疾病毎に危険リスクを説明するルールが異なる
  - 例：同じ薬が異なる疾病に用いられる



既に出血している胃潰瘍の治療



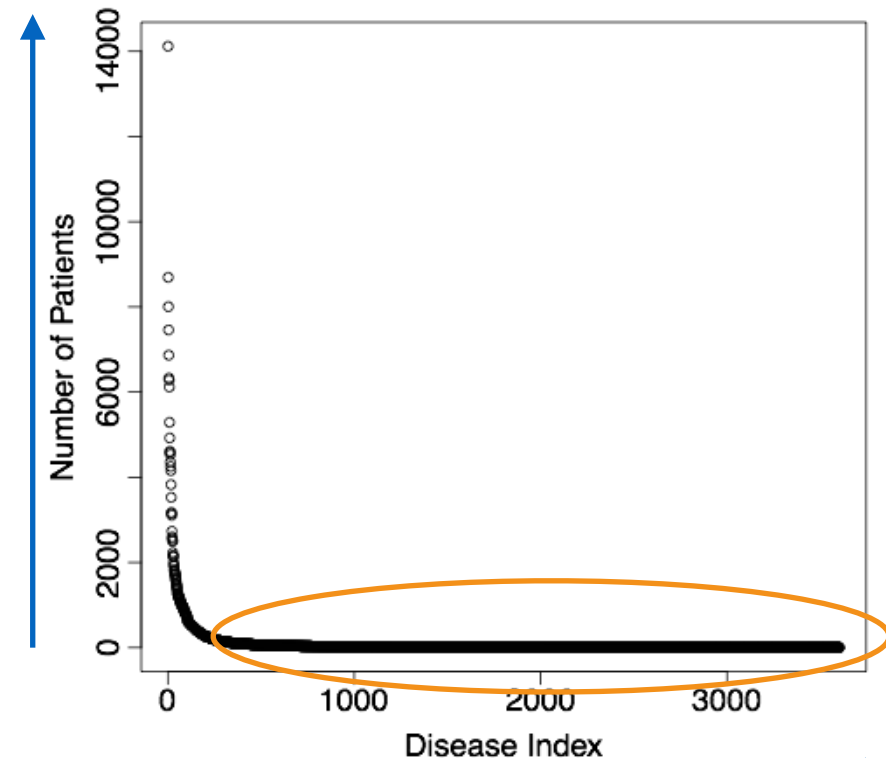
人工呼吸を受けた患者に対する胃潰瘍予防  
(人工呼吸を受けると胃潰瘍になりやすいため)

# 疾病毎にモデルを個別化しようとする際の課題： データ過疎への対処

## 1. 大抵の疾病では患者数は少ない

- 疾病毎にその疾病の患者データのみを用いた場合，予測モデルを構築するためのデータが十分に得られないという問題に直面する

患者数 Disease-Patients Distribution Plot



疾病のインデックス  
(ICD10の4階層目までのコーディング)

※約170病院での約20万人の  
ICU入室患者の疾病データ

# 疾病毎にモデルを個別化しようとする際の課題： データの疎性への対処

## 2. EHR (Electronic Health Record) に特有の疎性

- 診療行為や薬は複数の観点から細かく分類されており、そのうちの一つの性質でも異なれば異なるコードが割り振られる
  - 薬：薬効/成分/剤形等， 医科診療行為：手術の部位/方式等
- 患者数が少ない場合などデータが限られる場合には，ほとんど観測されないような薬/診療行為が一定数存在するような状況が生じる



薬価基準収載医薬品コード

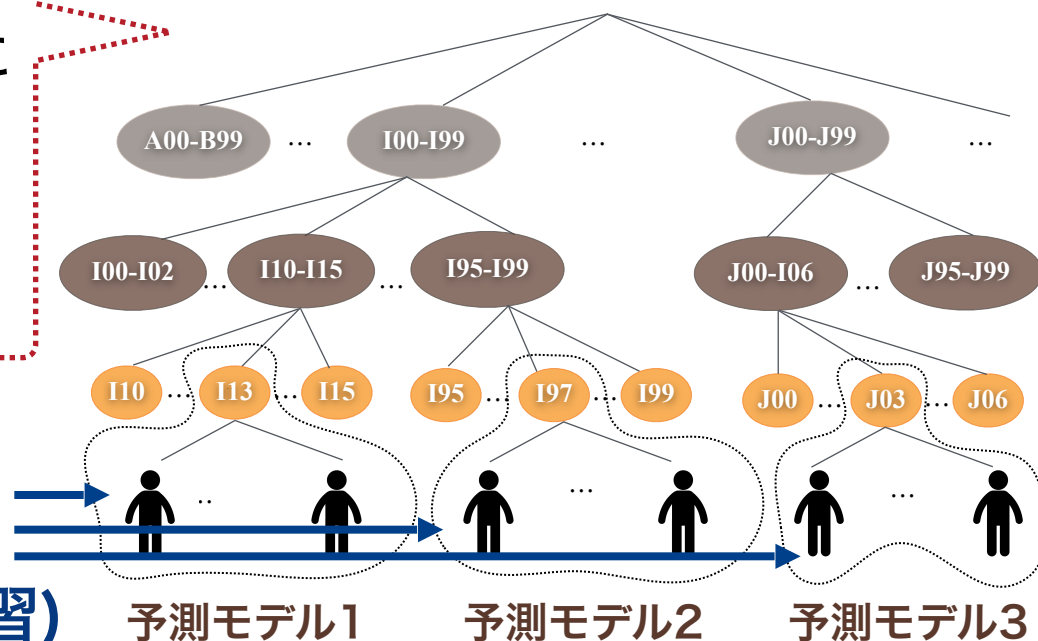
1143001X1015

薬効 成分 剤形 ...

# 解決策：疾病の分類やEHRの分類といった医療知識を取り込み，疾病ごとの予測モデルを同時に学習

1. 疾病の階層分類上類似した疾病間でモデルパラメータが類似するような正則化

ICD10に基づく疾病の階層分類



疾病間で情報を共有し疾病毎の予測モデルを同時に学習

(医療知識を取り込むマルチタスク学習)

2. EHRの分類上類似した説明変数間でモデルパラメータが類似するような正則化

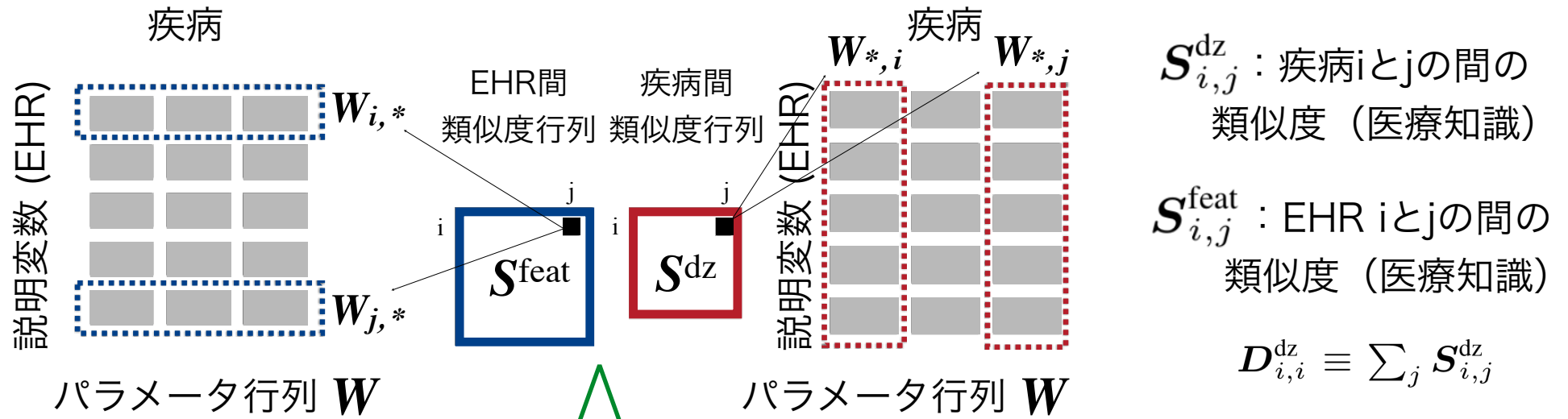


1143001X1015

薬効 成分 剤形 ...



# 提案手法：リスク予測問題を最適化問題として定式化する際に正則化項でgraph Laplacianを用い医療知識を取り込む



医療知識上類似した疾病間/EHR間でモデルパラメータを類似させる

疾病間  
類似度

$$\Omega^{\text{dz}}(\mathbf{W}) \equiv \frac{1}{4} \sum_{i=1}^T \sum_{j=1}^T \mathbf{S}_{i,j}^{\text{dz}} \left\| \frac{\mathbf{W}_{*,i}}{\sqrt{\mathbf{D}_{i,i}^{\text{dz}}}} - \frac{\mathbf{W}_{*,j}}{\sqrt{\mathbf{D}_{j,j}^{\text{dz}}}} \right\|^2 = \frac{1}{2} \text{Tr}(\mathbf{W} \mathbf{L}^{\text{dz}} \mathbf{W}^{\text{T}})$$

EHR間  
類似度

$$\Omega^{\text{feat}}(\mathbf{W}) \equiv \frac{1}{4} \sum_{i=1}^M \sum_{j=1}^M \mathbf{S}_{i,j}^{\text{feat}} \left\| \frac{\mathbf{W}_{i,*}}{\sqrt{\mathbf{D}_{i,i}^{\text{feat}}}} - \frac{\mathbf{W}_{j,*}}{\sqrt{\mathbf{D}_{j,j}^{\text{feat}}}} \right\|^2 = \frac{1}{2} \text{Tr}(\mathbf{W}^{\text{T}} \mathbf{L}^{\text{feat}} \mathbf{W})$$

# 実験結果1：予測精度の向上 (病院の実データを用いた入院中死亡予測問題で評価)

Table: Comparison of averaged AUCs among various methods with Wilcoxon signed rank test

	Pre-ICU discharge prediction (2 days before)	Pre-ICU discharge prediction (1 day before)	Retrospective prediction
<b>Proposed</b>	<b>0.763</b>	<b>0.795</b>	<b>0.914</b>
Proposed-feat	0.692	0.717	0.837
Proposed-dz	0.758	0.793	0.913
non-MTL-1 (separate)	0.692	0.714	0.836
non-MTL-2 (common)	0.760	0.783	0.886
MTL-1 ( $l_{21}$ )	0.717	0.723	0.819
MTL-2 (Trace)	0.724	0.739	0.837

※ 提案手法はグレー表示以外の他の手法を上回る予測精度

全予測設定を通じて提案手法と同等以上の手法はない

疾病の分類とEHRの分類は共に予測性能の向上に有効であることを示唆

## 実験結果2： 疾病毎の予測ルールの学習

---

- 概ね、医師が見ても納得できるものだった；疾病全体で一つの共通のモデルを構築するよりも疾病の特性を反映した予測モデルになっていた
- 一方、解釈は可能だが、医療分野でどのような意味を持つかを考えるには、今後医療分野での調査検証を要するような示唆もいくつか確認された；見方を変えると「統計的仮説の構築に繋がる示唆」とも言える

# まとめ

---

- ICU入室患者の危険リスク予測において”疾病によって患者の危険リスクを説明する予測ルールが異なる”という疾病の個別性を考慮するために，疾病の分類とEHRの分類に関する医療知識を取り込むマルチタスク学習手法を提案
- 成果
  - 既存手法に対して予測精度の向上
  - 疾病毎の予測ルールの学習が可能に
- 患者の疾病の個別性を考慮した医療/医療研究支援のための統計的機械学習手法の実験的応用の一つのさきがけ