

レセプト情報を用いた機械学習による死亡確定手法の検討

奈良県立医科大学
MBT学 博士1年 大井川

背景

- レセプト(診療報酬明細書)
 - 診療内容が記載
 - 近年電子化(CSV形式)
- 巨大なコホートとして期待
- NDB・KDBが構築

格納件数:100億件超え



しんじょう医院
http://www.shinjo-clinic.com/kobetoppupeji.html

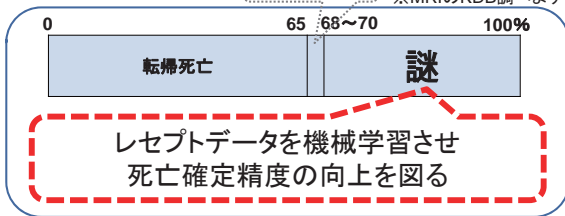
レセプト蓄積
= 身体の経年変化 (死亡経緯)の可視化

目的

DBの構築とその先にある分析のために

- 個人毎のレセプト上にイベントのフラグを明記
 - 発病
 - 入院
 - 死亡 (本課題)

ターミナルケア加算
看取り加算
死亡診断加算 ※MRIのKDB調べより



死亡確定の指針

- 材料
 - 国保データベース(KDB)システム
 - 奈良県における死者情報とその組となっているレセプト
 - 全国のレセプト
 - 手法
 - 機械学習 (= 深層学習)
- 教師データ: 奈良県レセプトデータ → 機械学習 → テストデータ: 全国のレセプトデータ
- 年間死者数(2016) 約130万人(全国) 約1万4千人(奈良)
- 検証
 - 全国の死亡転帰・ターミナルケア加算などの死亡確定レセプトを対象に識別, 正答率算出

調査 1/2

- 対象: 奈良県のKDB縮小版(1/20)データ
- 期間: 2013.3~2015.2(2年分)

	全患者	転帰死亡者
人数	19,999	513
平均年齢	62.3 (±22.6)	84.4 (±10.2)
医薬品レコード数(対象484人)	1,239,126	97,527
医科診療行為レコード数(対象513人)	5,408,395	268,436

調査 2/2

転帰死亡者 = 死亡確定者

- 転帰死亡までの診療期間別の人数算出
- 転帰死亡記載時の診療年月-調査期間中初めの診療年月
 - 死亡に関係する時系列期間を半年に設定

半年	3ヶ月	1ヶ月	0
395(313)人	463(393)人	513(484)人	死亡

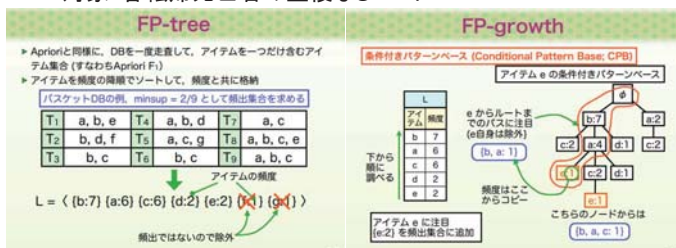
診療行為(医薬品)

- 各転帰死亡者の医薬品・医科診療行為コードの抽出

最大コード数(平均)	医薬品総量	診療行為総量	医薬品種類の数	診療行為種類の数
1ヶ月分	305(28.4)	337(49.0)	65(10.8)	105(32.6)
3ヶ月分	1126(106.1)	1637(185.0)	93(22.8)	156(61.3)
6ヶ月分	1989(168.9)	3183(335.5)	148(31.6)	196(78.8)

実験

- 機械学習: FP-Growth
- 対象: 各転帰死亡者の重複なしコード



引用: Toshihiro Kamishima 頻出パターンマイニング
http://www.kamishima.net/archive/freqpat.pdf

どの医科診療行為あるいは医薬品コードの組合せが頻出か導き出す

結果 1/2

- 医薬品
- 頻度 ≥ 10(回) のコードあるいはその組合せ

	該当コード数(最大頻度)
1ヶ月分 [484]	133(95)
3ヶ月分 [393]	654(125)
6ヶ月分 [313]	1099(122)

- (21・2) 生理食塩液20mL注射液 / ラシックス注20mg
 - 18 (20・2) 生理食塩液20mL注射液 / キシロカインゼリー2%
 - (20・2) 生理食塩液20mL注射液 / ビーフリード輸液 500mL
 - (17・3) 生理食塩液20mL注射液 / ラシックス注20mg / ラシックス錠20mg
 - 36 (16・3) キシロカインゼリー2% / ラシックス注20mg / キシロカイン注ポリアンブ10mg 10mL
 - (13・4) キシロカインゼリー2% / キシロカインポンプスプレー8% / キシロカインビスカス2% / ガスコンドロップ内用液2%
 - 88 (13・4) 生理食塩液20mL注射液 / キシロカインゼリー2% / ビーフリード輸液 500mL / 生理食塩液 100mL
- (頻度・組合せ数)項目例

結果 2/2

・医科診療行為

- 頻度 ≥ 閾値 のコードあるいはその組合せ
- ≤ 閾値 はメモリ不足
- 又はファイルが開けなかった

	閾値	該当コード数 (最大頻度)
1ヶ月分 [513]	200	14223(261)
3ヶ月分 [463]	300	22345(355)
6ヶ月分 [395]	300	10943(329)

組合せ最多・頻度最大項目

(208・12) 生化学的検査(1)判断料
血液学的検査判断料
免疫学的検査判断料
4 個
AST BUN クレアチニン ALT
カリウム ナトリウム及びクロール CRP
LD BIL/総

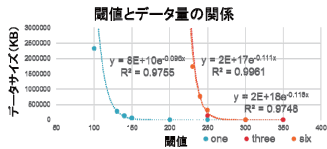
(308・12) 生化学的検査(1)判断料
血液学的検査判断料
免疫学的検査判断料
7 個
AST BUN クレアチニン ALT
カリウム グルコース
ナトリウム及びクロール CRP LD

(305・11) 生化学的検査(1)判断料
血液学的検査判断料
免疫学的検査判断料
8 個
AST BUN クレアチニン ALT
カリウム グルコース
ナトリウム及びクロール CRP

考察 2/3

・医科診療行為

- 生化学的・血液学・免疫学的検査が上位を占め、グルコース・LD・BIL/総の組合せが期間で異なる
- 最大頻度が対象者数の半数を超え、組合せも多数あるが、適切な閾値(範囲)を設定する必要がある
 - 頻出集合の数に比例して計算量・出力容量が増大
 - 上位には、本来死亡確定の材料にしよとしたコード(看取り加算など)が絡んでいない



呼吸心拍監視(14日超)	53
医療機器安全管理料(生感)	52
在宅患者訪問診療料(同一)	52
処方料(その他)	50
BIL/総	50
シ-値	50
電子検出加算2(200床)	50
薬剤料(内服薬・投薬薬)	49
一般患者入院期間加算(1)	47
診療録管理体制作加算	47
薬剤料(注射・点・膏)	48
人工呼吸	48
看護記録加算(200床)	48
ICU治療管理料	49
診療情報提供料(1)	45
診療情報提供料(2)	45
入院時生活療養(1)費	44
外来診療料	43
救命のための気管内挿管	43
特殊検査管理加算(4)	43
療養医療管理加算1	43
救急(継続法)	42
呼吸心拍監視(7日超14日)	42
患者対応加算2	42
中心静脈注射用カテーテル	40
看護	40
時間外緊急院内検査加算	40
看護補助加算1	40
調剤検査料(薬)	39
HbA1c	39
HbA1c測定性・測定費	39
ICUリソース総合評価	38
基幹型臨床研修病院入院	38
感染防止対策加算1	38
感染防止対策増量加算	38
最新情報提供料	37
特定疾患治療管理料(診療)	37
2項目以降検査CT、MRI	37
CT撮影(64列以上マルチ)	37
2591(急性重症脳神経症)	37
調剤料(外用薬)	36
調剤料(内用薬)	35
血液交叉加算	34
血液検査加算(登録料)	33
死亡記録加算(登録料)	33

代案

- 全対象者の医薬品・医科診療行為レコードから抽出された特徴を学習し死亡確定する方法

→深層学習の導入

- GPUと処理のブラックボックス化による高速化 (これまで:メモリ活用)
- 学習
 - 頻度
 - 各対象者の特徴を表現(word2vec)
 - 時系列
 - 複数項目の量や回数の連続的变化を表現(LSTM)



引用: Google Official Blog
<https://googleblog.blogspot.jp/2012/06/using-large-scale-brain-simulations-for.html>

考察 1/3

・医薬品

- 上位は固定(生理食塩液20mL注射液・キシロカインゼリー2%)だが、期間によって変化
 - 短期になるにつれ
 - ソルデム3A輸液減少
 - ビーフリード輸液増加
 - アドレナリン注0.1シリンジ出現 など
- 対象期間が短くなるほど使用される医薬品は限定されるが、頻出パターンによる単純な死亡確定は困難
 - 最大頻度は対象者の約1/5~1/3
 - 共通する医薬品の組合せは最大4(6ヶ月分)

考察 3/3

- 死亡確定者と一般患者の比較が必要
 - 死亡確定者だけに見られる頻度なのか否か
 - 最新の1/20データでは死亡確定者を識別
- 医科診療行為と医薬品の組合せによる頻度に期待
 - 双方の情報の有効活用
- 時系列変化に着目
 - 投薬量変化や回数(パタリと特定医薬品の投与がとまった)

KDB・NDBに発展させた場合、本手法は恐らく使用困難(計算量と計算機の能力に不安)

結語

- レセプト情報を用いた死亡確定手法の検討
 - KDB/1/20データによる転帰死亡患者を対象にした事前調査
 - FP-Growth(頻度計算)
 - 医薬品コード
 - 医科診療コード
 - 頻度集合の大きさがデータが膨大になる
 - 全対象者の医薬品・医科診療行為レコードを学習し死亡確定する方法への切り替え
- 今後
 - 深層学習での解析
 - データ規模の拡張(KDB)