

実務者のためのデータベース研究講座 その3

MDVを用いた研究の実際(2)

使えたらいいな、をカタチへ 機械学習とBIG DATAを活用した 医薬品のリスク評価に向けた取り組み

2020.1.20

慶應義塾大学

薬学部 医薬品開発規制科学講座 漆原 尚巳, 種村 菜奈枝

理工学部 システムデザイン工学科 西川景太, 矢向 高弘

背景

- 医療情報のIT化、SNSの普及に伴い、爆発的にデータ量が増加
- General Purpose GPUの出現による計算機の処理能力の向上

様々な分野において、ビッグデータ分析によるデータの利活用が活発化

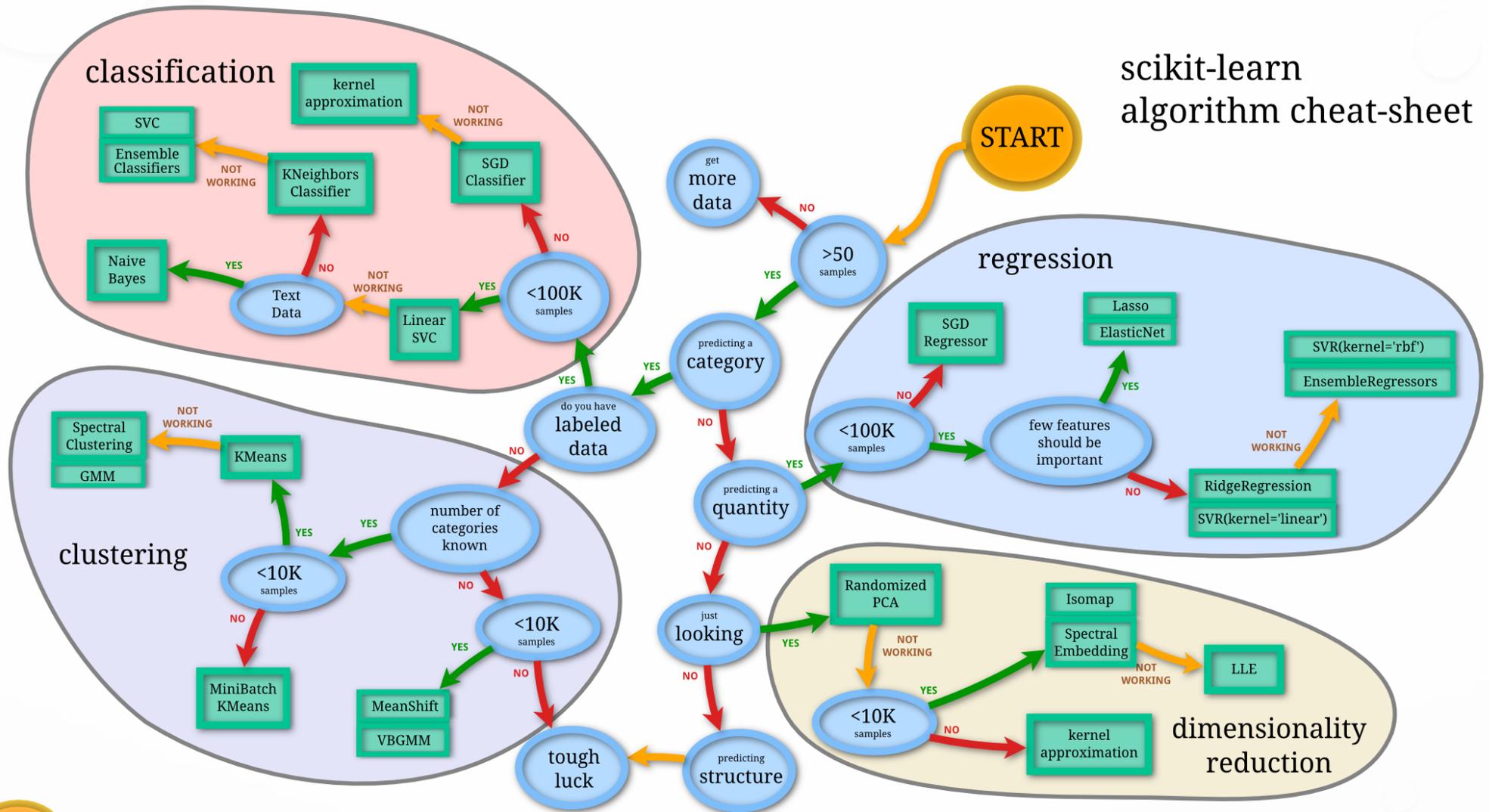
- ・マーケティング分野: 購買傾向の調査、在庫予測
- ・経済分野: 過去の時系列データからの株価予測
- ・医療分野: 画像診断、診断補助

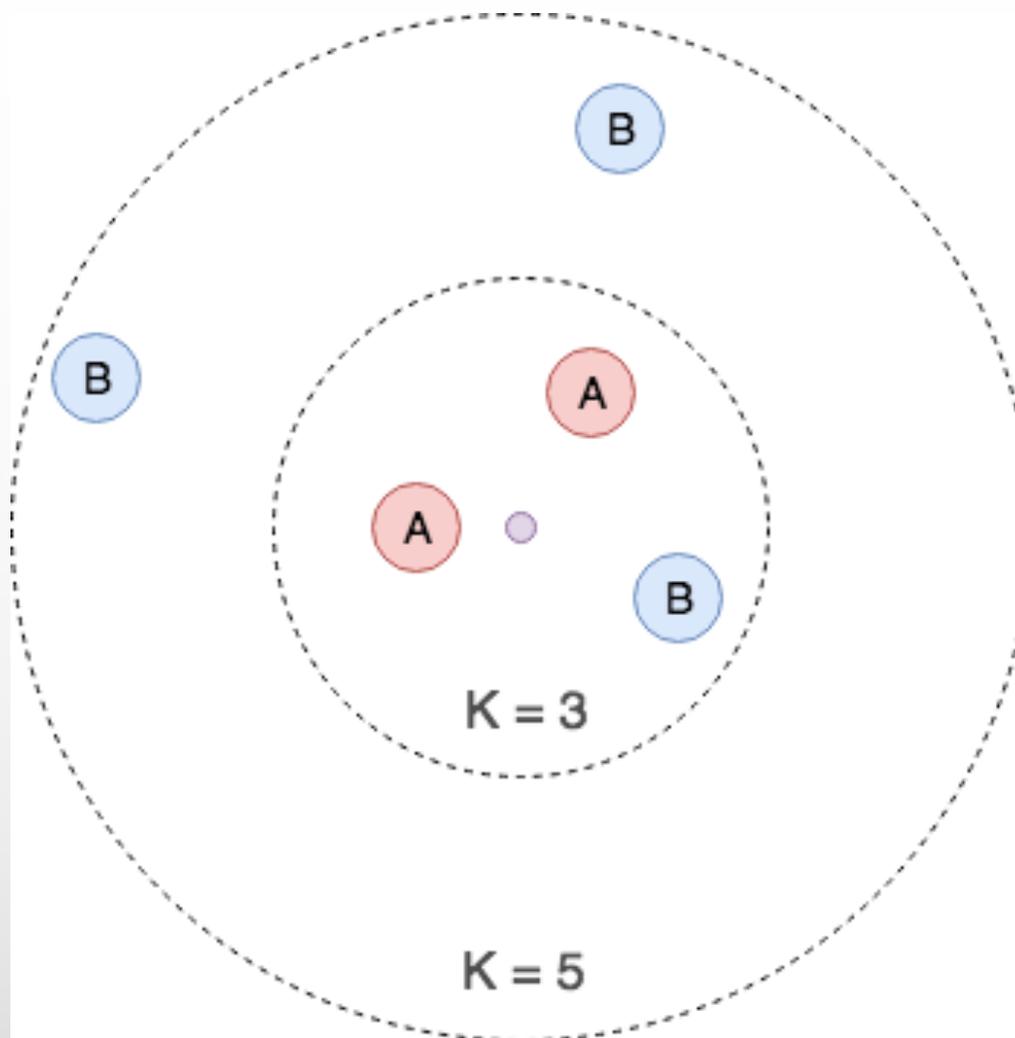
医療データの蓄積



医療分野においてもビッグデータの活用が進んでいる

scikit-learn algorithm cheat-sheet





k近傍法とk平均法の違いと詳細 @NoriakiOshita

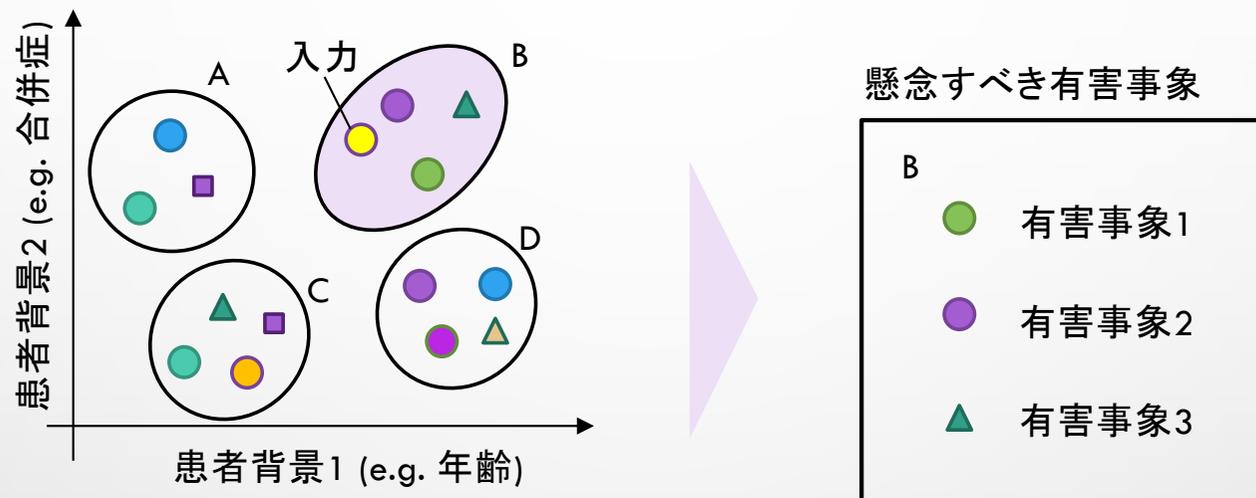
<https://qiita.com/NoriakiOshita/items/698056cb74819624461f>

目的

医療ビッグデータを用いて開発時における医薬品リスクを予測し、安全性評価の指標とする

目的

患者背景(合併症, 年齢, 性別)を入力とし,
どの有害事象が起こりうるかを予測, 懸念すべき有害事象を提示



対象薬剤の有害事象を抽出し、
カテゴライズ

今回は予測モデルの評価のため、
ICD-10大分類を用いてカテゴライズ

複数の有害事象をカテゴライズし、投薬後に発症するリスクのある
有害事象を提示することで安全性評価の指標とする

研究方法

データソース

メディカル・データ・ビジョン社から得られた
大規模医療病院事務・DPCデータ

使用項目

病名データ:

患者ID・受診月・年齢・性別・ICD10分類コード

薬剤データ:

患者ID・受診年月日・ATCコード

調査対象期間

2016年1月から2017年12月

対象患者

0歳～15歳児の小児糖尿病患者

研究方法

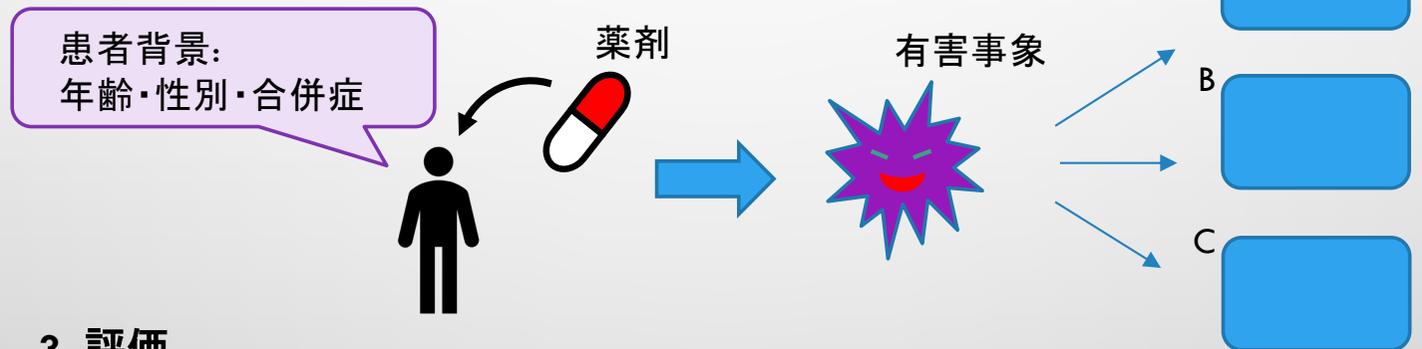
1. データクレンジング

- 患者背景・有害事象を抽出、データセット作成
- 学習用/テスト用データに分割

2. モデル作成

学習用データより予測モデルを作成

入力: 患者情報 出力: 有害事象の器官別大分類



3. 評価

テスト用データを用いて、予測精度を算出

教師データ

e.g.糖尿病 (ATCコード: A10xx)
(ICD-10コード: E10xx)

薬剤・疾患を時系列順にソート

ターゲットの薬剤(糖尿病用薬)を決定

ターゲット薬剤の対象疾患(糖尿病)を抽出

投薬前の対象疾患を除いた傷病を合併症、
投薬後の傷病を有害事象とする

KJcc 解剖治療化学分類 (ATC分類) 薬剤の抽出

[Brite menu | Download htext | Download json]

医療用医薬品のATC分類

- ▼ ▼ ▼ ▼ ▼ One-click mode
- ▼ A 消化管と代謝作用
 - ▶ A01 口腔科用製剤
 - ▶ A02 胃酸関連疾患用薬
 - ▶ A03 機能的胃腸疾患用薬
 - ▶ A04 鎮吐薬と制嘔吐薬
 - ▶ A05 胆汁と肝治療
 - ▶ A06 便秘薬
 - ▶ A07 止痢薬、腸内抗炎症薬・抗
 - ▶ A08 抗肥満用製剤、食事制限用
 - ▶ A09 消化薬、酵素を含む
 - ▼ A10 糖尿病用薬
 - ▶ A10A インスリンと類似体
 - ▶ A10B 血糖値降下薬、インス
 - ▶ A10X その他の糖尿病用薬

KJcc ICD10対応標準病名

[Brite menu]

ICD10対応標準病名

▼ ▼ ▼ ▼ ▼ One-click mode

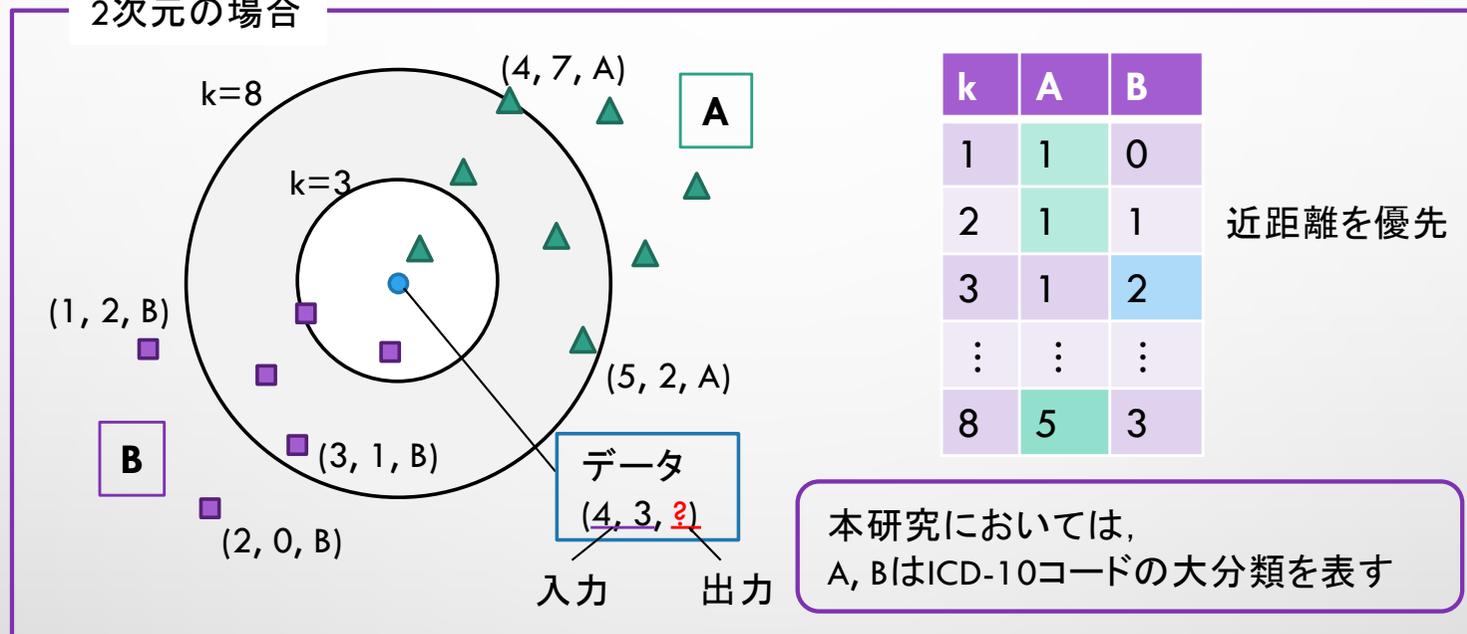
- ▶ 1. 感染症及び寄生虫症 (A00-B99)
- ▶ 2. 新生物 (C00-D48)
- ▶ 3. 血液及び造血器の疾患並びに免疫機構の障害 (D50-D89)
- ▼ 4. 内分泌、栄養及び代謝疾患 (E00-E90)
 - ▶ E00-E07 甲状腺障害
 - ▼ E10-E14 糖尿病
 - ▼ E10 1型糖尿病
 - 20083022 1型糖尿病
 - 20090028 緩徐進行1型糖尿病
 - 20075098 不安定型糖尿病
 - ▶ E100 1型糖尿病: 昏睡を伴うもの
 - ▼ E101 1型糖尿病: ケトアシドーシスを伴うもの
 - 20085409 1型糖尿病・ケトアシドーシス合併あり
 - 20093027 1型糖尿病性アシドーシス
 - 20093028 1型糖尿病性アセトン血症

対象疾患の抽出

K近傍法

入力に近いk個のデータから、多数決によって分類を行う手法

2次元の場合



本研究では 入力: 名義尺度で数値化した年齢・性別・合併症
出力: 名義尺度で数値化した有害事象のカテゴリ とした

結論と今後の課題

□ 結論

目的

- 複数の有害事象を分類し、投薬後に発症するリスクのある有害事象群を予測・提示することで安全性評価の指標とする

方法

- 簡易的にICD-10コードの大分類を用いて分類を行った.
- k-NNを用いて患者背景から投薬後の有害事象の分類予測を行い、予測精度の評価を行った.

□ 今後の課題

- クラスタリング手法等を用いてICD-10コードに依存しない新たな分類の作成
- 精度向上に向け、入力パラメータの多次元化や他のアルゴリズムの使用