

京都大学医学部附属病院における 自然言語処理の試み

岡本 和也(京大病院 医療情報企画部)

背景

- × 診療文書の電子的な蓄積

- + テンプレートによる構造化は一部に留まっている。

- × 項目・選択肢が非常に多くなる。

- × 含まれていない項目・選択肢が出てくる可能性がある。

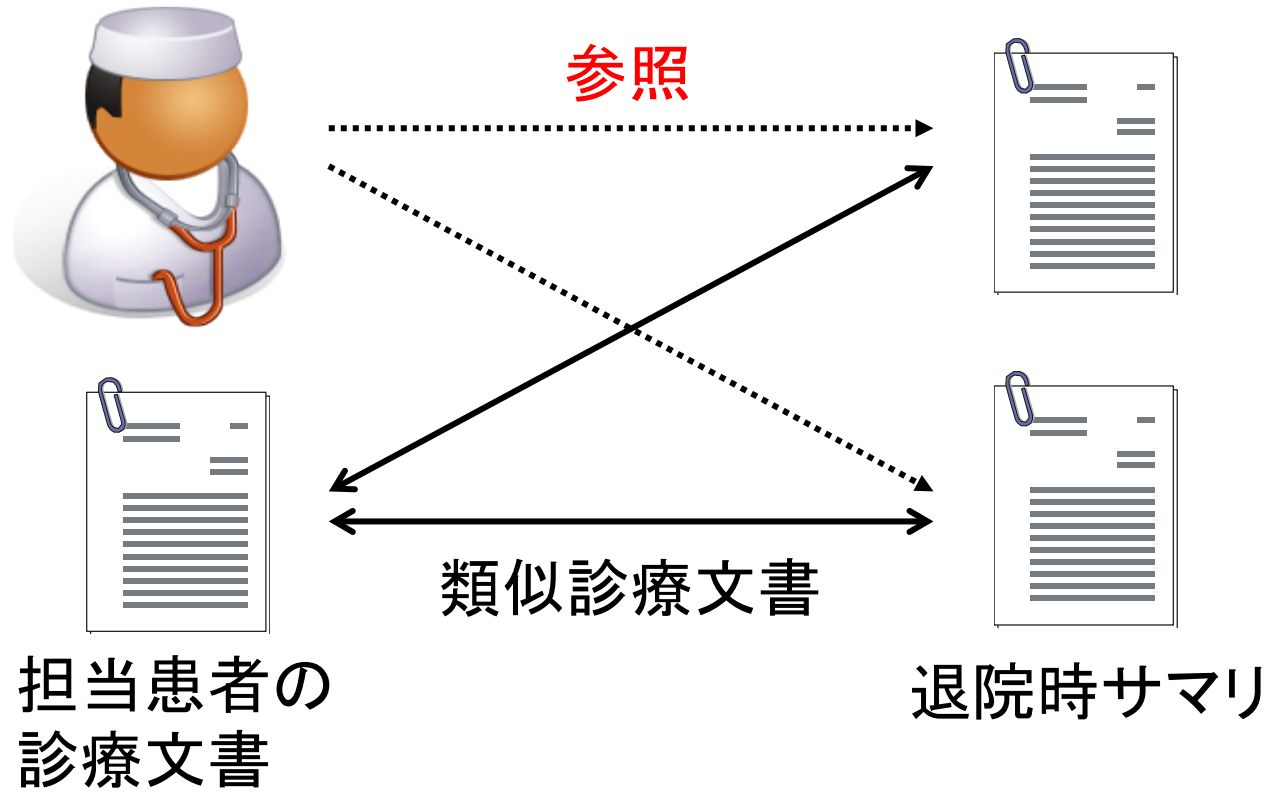
- + 自然言語記載の診療文書が多く存在する。

自然言語処理を用いた診療文書の活用

京大病院における自然言語処理の試み

- × 類似症例検索アルゴリズムの開発
- × 退院時サマリからの自動DPC分類
- × インシデントレポート作成時のコンテキスト理解による入力補助システム
- × レポート内の共起関係を可視化するテキストマイニングツールの開発

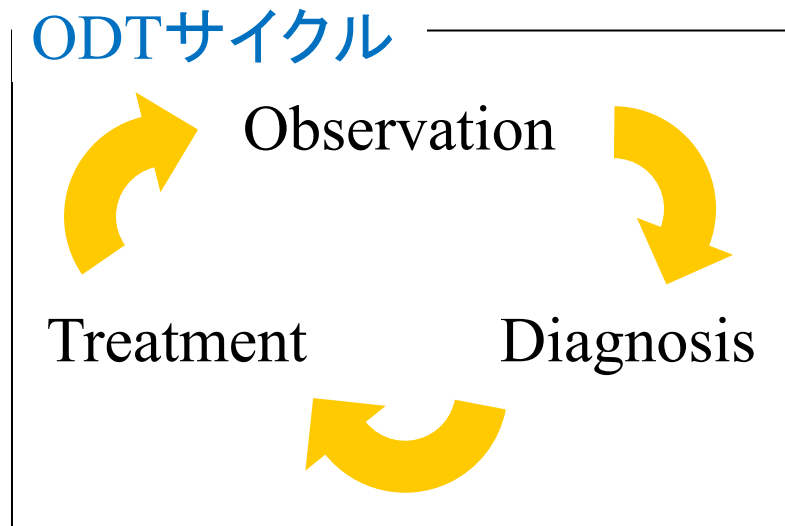
類似症例検索アルゴリズムの開発



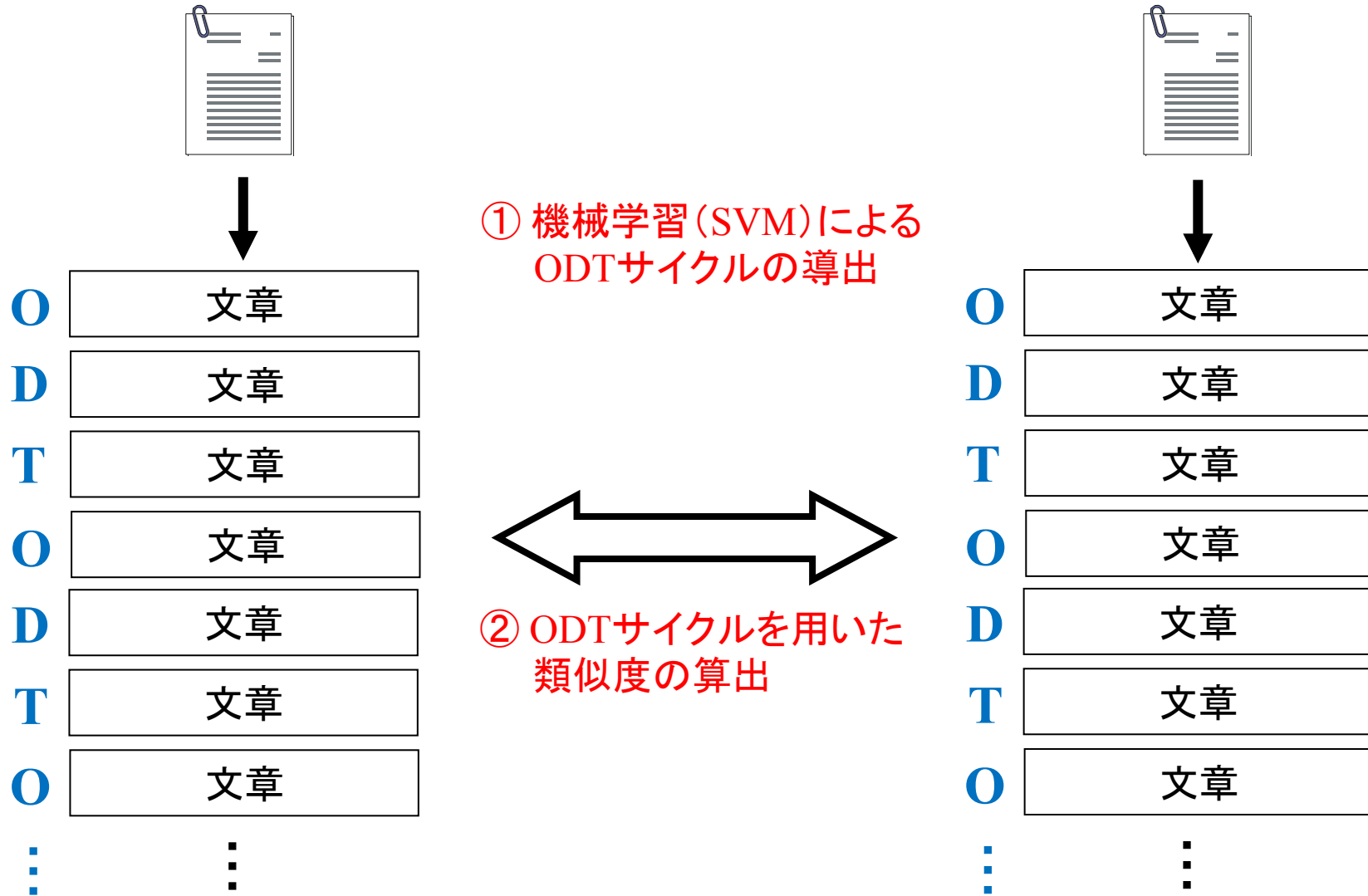
病名のみではなく、診療のプロセスに近いものを見つけたい。

類似症例検索

- × Web(一般)において類似文書検索は機能していない?
 - + 手法としてベクトル空間モデルがあるぐらい?
- × 医療の文章の特徴を活用
 - + 診療サイクルに注目した類似文書検索手法の提案

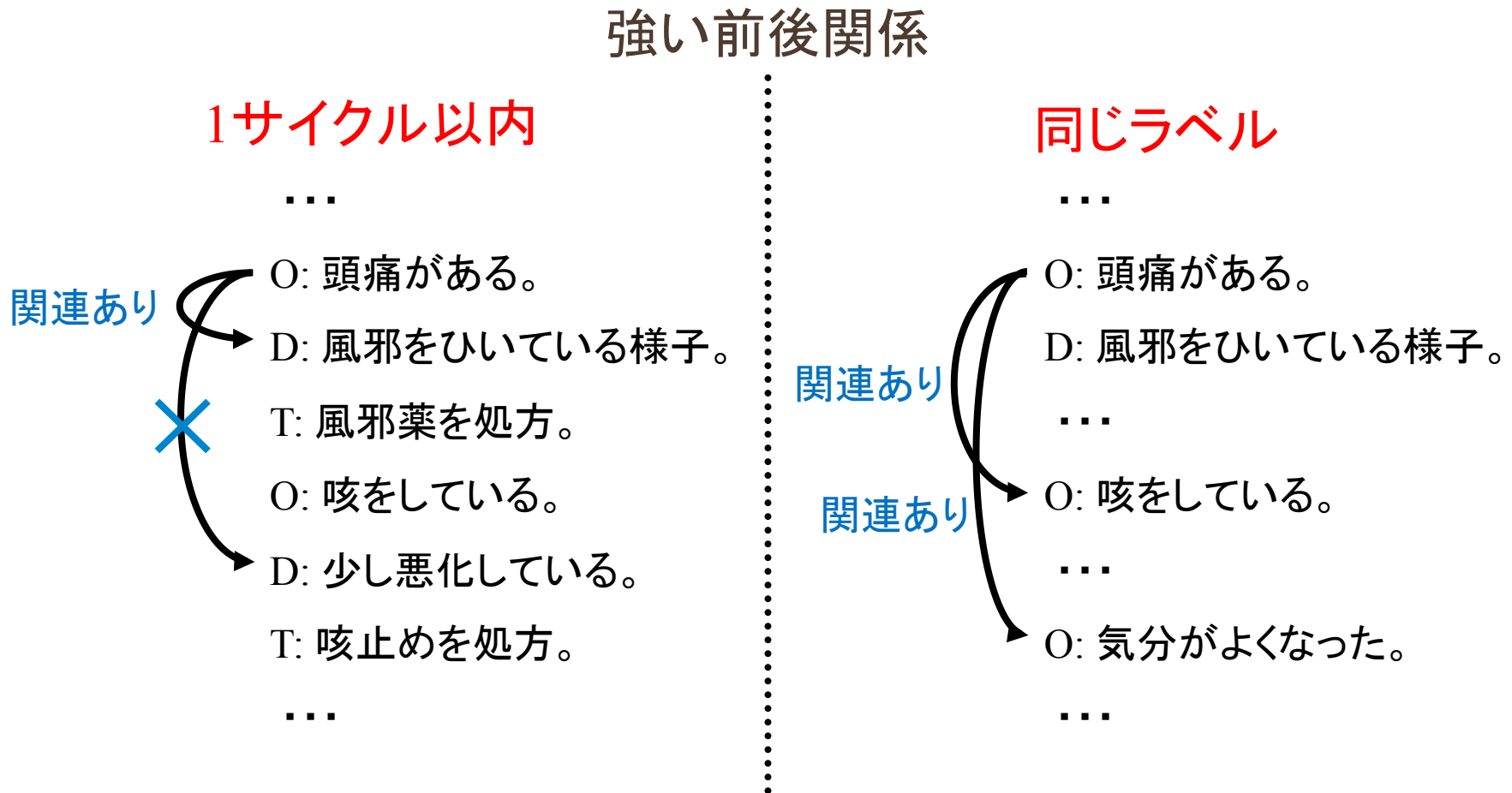


類似症例検索アルゴリズム



ODTサイクルを用いた類似度の算出

- × 「強い」前後関係の一致から類似度を算出



ODTサイクルを用いた類似度の算出

- × 「強い」前後関係の一致から類似度を算出
 - + 同じ強い前後関係を持つ場合に類似度を足し合わせる。
 - × (値の算出はTF/IDF法を用いる。)

→ T MRI,
O 腫瘍, 前頭葉, 消失,
D 深部白質, 両側, 明らか,
T
O 再発,
→ D 亜急性, 脳卒中, 疑い,
→ T PET, 推奨

→ T MRI,
O まひ, 出現,
→ D 脳卒中,
→ T PET, 推測,
O 頭蓋, 蓄積, 部位, 可能性,
D
T 精査,

類似度: 0.083

評価

× ベクトル空間モデルとの比較

+ 同じ病名の退院時サマリを抽出することができるか？

- × 5病名(2型糖尿病、肺癌、狭心症、子宮筋腫、脳梗塞)、20症例

提案手法: 72.0%, ベクトル空間モデル: 77.6%

+ 同一疾患のサマリの中で類似症例を抽出できるか？

- × 肺癌の病名の退院時サマリ22例に対して、医師の判断と比較
 - * ランキング方式: 1位3点、2位2点、3位1点

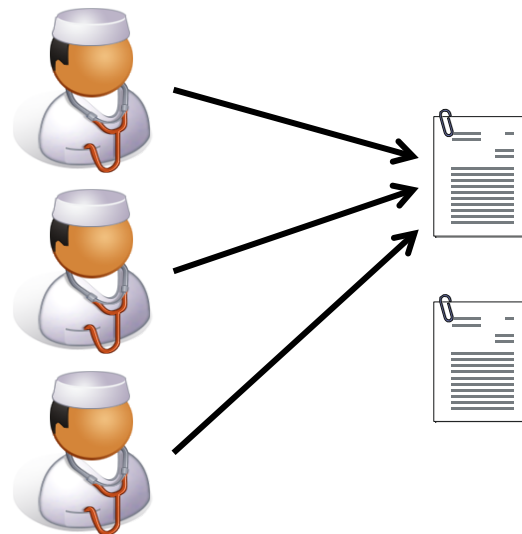
提案手法: 36点, ベクトル空間モデル: 22点

診療の流れを把握した類似症例検索

(ちょっと逸れて)

閲覧履歴を用いた診療文書の重要度評価

- ✕ そもそもコンピュータが自然言語を理解するのは難しい。
 - + PageRankもHITSも文書の周辺情報(リンク)を活用している。
 - ✕ 逆に内部情報を利用して成功している例はほとんどない。
- ✕ HISにカルテへの閲覧履歴が蓄積されている。
 - + 閲覧履歴という周辺情報を活用して重要度を評価する。

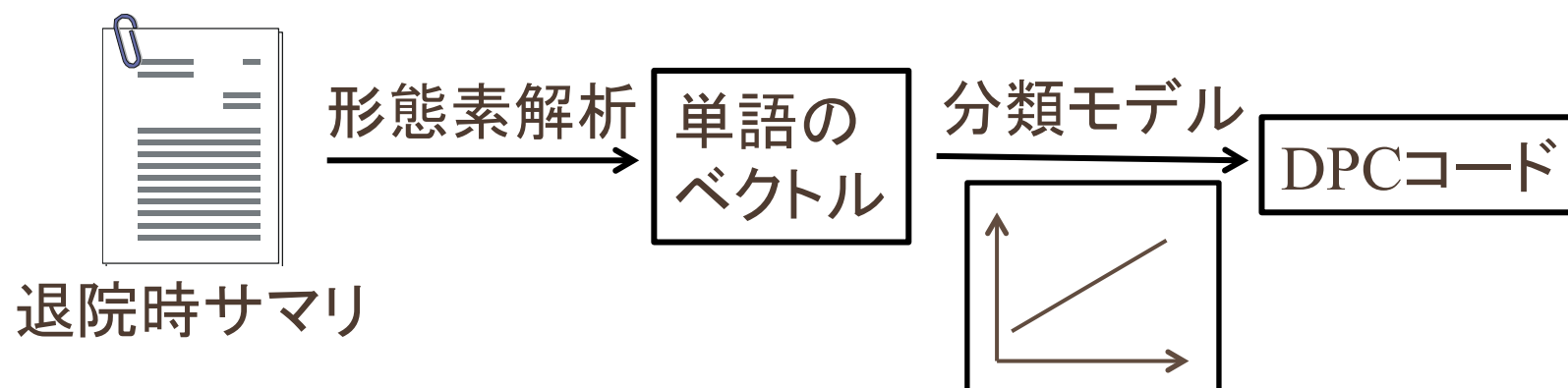


重要？

自然言語処理ではないが、
自然言語の診療文書を
扱うための1手法の提案


退院時サマリからの自動DPC分類

- ✕ 文書そのものからメタな情報を取り出す。
- ✕ 機械学習により人間の判断を学習し、クラス分類を自動化。
 - + 156種類のDPCに対して、76.1%の精度で自動分類可能。
 - ✕ 各入院症例に1つのDPCコードが割り当てられている。



医療ならではの感がなく面白みに欠ける...

インシデントレポート作成時の コンテキスト理解による入力補助システム

- ✖ 良いインシデントレポートには
 - + 患者プロフィール
 - + 事故発生プロセス
 - + 観察・診断
 - + 処置
 - + 潜在理由
 - + 事故対策が含まれている。
- 
- ✖ インシデントレポートの入力補助システム
 - + どの情報が含まれているか判定し、
 - + 含まれていない情報を記載するように促す。

インシデントレポート作成時の コンテキスト理解による入力補助システム

IAReport Input Guide

何時? 何が起きたのか? 何が起こりそうだったのか?

何年 何月 何日
何時 何分

どこで?

診療科
病棟
部門・部署

あなたの職種は?
職種

あなたの部門・部署の
リスクマネージャの名前は?
RM

あなたの部門・部署の
リスクマネージャに連絡していますか?
YES/NO

患者プロフィール 事故発生プロセス 観察診断情報 処置情報 潜在的事由 対策情報

今日入院した77歳の患者。ベッドから立ち上がろうとしたところふらつき転倒。頭を強く打ち、出血。担当医師に報告。消毒後にガーゼを巻き、様子観察となった。

現在「処置情報」を記載中。
次は「潜在的事由」を記載して下さい。
残りの項目：潜在的事由、対策情報
不足項目：患者プロフィール

レポート内の共起関係を可視化する テキストマイニングツールの開発

- ✖ 大量のデータに全て目を通し、理解することは大変。



データを凝縮し、可視化することで知識を抽出
(例えば、強い共起関係)

ある特定の疾患に関する病理カンファレンスレポートを用いて、
各レポート内の単語の共起関係を可視化すると...

テキストマイニングの難しさ

- × 何か出るかなと期待してもなかなか出ない...。
 - + 医師は「面白いですね。」とか「何も知識を足さなくてもここまでできるのですね。」というものの...

Future Work

- × ユーザと対話ができる仕組みを入れる。
 - + そもそも可視化するということはユーザに期待している。
 - + ユーザフレンドリー？なシステムが必要。
- × 医療の要素を取り入れる。

京大病院における自然言語処理の試み

- × 類似症例検索アルゴリズムの開発
- × 退院時サマリからの自動DPC分類
- × インシデントレポート作成時のコンテキスト理解による入力補助システム
- × レポート内の共起関係を可視化するテキストマイニングツールの開発