

# トピックモデリングによる 医療読影レポート解析手法の提案

ムサシ・イメージ情報株式会社 針原英明 h\_harihara@musashi-ij.co.jp

## 背景と課題

**背景:** 日本医療安全調査機構は「画像診断報告書の確認不足」<sup>[1]</sup>を注意喚起している。情報伝達の補助手段として読影レポートから重要所見を自動的に推定することが期待される。

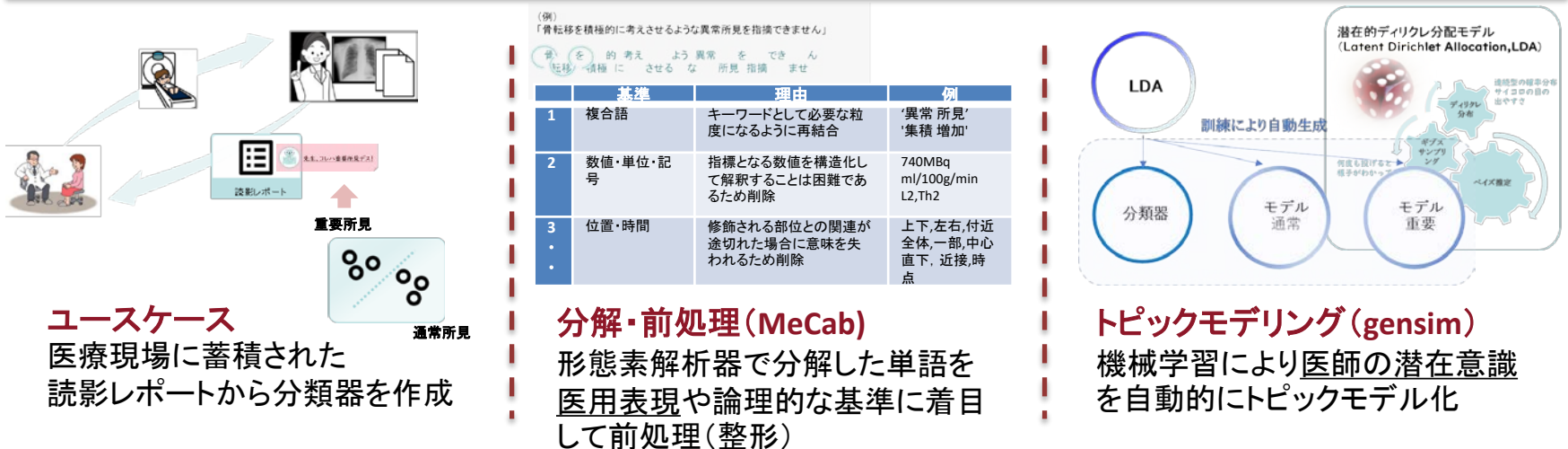
**課題:** 自由記述形式の読影レポートには医用表現(断定しない二重否定など)を含むため、内容を正しく推定することは困難である。

[1]医療安全情報 No.63 2012年2月, 2012

## 手法・ツールの適用による解決

- 適用手法:** トピックモデリングの手法として潜在的ディレクレモデル(LDA)を適用
- 適用理由:** レポートの内容を自動的にベクトル化  
医用表現に潜在する意味を分析する
- 適用手順:**
- 形態素解析器で単語に分解
  - 論理的な基準で単語を編集
  - 機械学習でトピックモデルを自動生成
  - ベクトル類似度から重要所見を推定

## 自然言語処理によるアプローチ



## 評価実験

対象読影レポート: 骨・脳血流シンチグラム

	学習データ	テストデータ	合計
通常所見	1,679	210	1,889
重要所見	665	83	748
(小計)	2,344	293	2,637

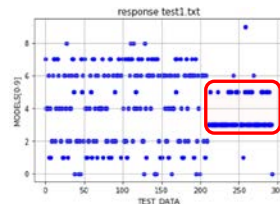
モデルの例

ID	モデル
5	0.112* "血流低下" + 0.111* "矛盾しない" + 0.003* "血流" + 0.064* "DLB" + 0.055* "閾値超え" + 0.054* "集積低下" + 0.046* "アルツハイマー病" + 0.041* "頭頂葉" + 0.035* "観察され" + 0.034* "部带状回"
4	0.175* "疑われ" + 0.140* "血流低下" + 0.086* "閾値以下" + 0.059* "血流分布" + 0.056* "集積低下" + 0.055* "部带状回" + 0.052* "頭頂葉" + 0.037* "年齢考慮" + 0.037* "脳血流量" + 0.037* "大脳半球"

医師の潜在意識

**評価実験:** 分類器が出力したベクトル値を用いてモデルとの類似度を判定した。

## 評価結果と今後の展開



読影レポートの重要・通常所見293件のうち289件が正しく分類できた。

**評価結果:** 検出率は98.6%であることから、提案手法の有効性を示すと考えられる。

**考察:** 高い検出率の要因として、学習データの不足により確率モデルが収束していない可能性がある。

**今後の展開:**

- 分類外れ値の原因・妥当性検証
- ディープラーニングなど別手法との性能比較
- 電子カルテシステムベンダーとの有効性検討