

【第134回生涯教育講座】

機械学習による退院時要約からのDPC分類の推測

木村知広¹⁾ 津本周作²⁾
ひらのしょうじ²⁾

キーワード：テキストマイニング，退院時要約，DPC，深層学習，Random Forest

要 旨

電子化した退院時要約からテキストマイニングによって疾患別の重要語を抽出し、これを元に退院時要約の文章から疾患名を特定できるかについて検討した。

キーワードとターゲットとの対応付けが弱く、分類精度がそれほど上がらないという、従来のテキストマイニングによる分類学習の問題に対して、本研究では、対応分析によるDPC毎のキーワード選定を行い、機械学習で用いられる手法を適用、キーワードによるDPCコーディングを行う分類器を構築した。

反復交差検証法により分類の正答率を比較し、その検証結果から、SVM、決定木が最大78%程度の正答率であるのに対し、深層学習が86%以上、Random Forest、XGBoostでは90%以上の正答率を示した。

構築した分類器は高い正答率を示しており、診療文書の記載内容からテキストマイニングにより疾患の予測が可能となることが示唆された。

1. 背景および目的

電子カルテシステムに蓄積された診療情報の電子的データを用いて、新たな規則や知識を発見し、診療・研究・病院管理に役立てることが可能になってきている。しかしその研究の多くは、入力・保存や統計処理が容易な数値データや画像

データであり、医師のカルテ記録などの診療文書の解析は、従来の自然言語処理手法が診療文書の特殊性に対応できなかったことなどもあり、その数は少ない^{1,2)}。医療文書の分析法を確立することは、医療プロセスの意思決定や予防医療の支援に期待されており、医療の質向上に大きく寄与するものと考えられる。筆者らは、その分析の一つとして電子化した退院時要約からテキストマイニングによって疾患別の重要語を抽出し、これを元に退院時要約の文章から疾患名を特定できるかについて検討した。

Shusaku TSUMOTO et al.

1) 島根大学医学部医療サービス課

2) 同 医学科医療情報学

連絡先：〒693-8501 島根県出雲市塩冶町89-1

島根大学医学部医学科医療情報学