
一般口演

一般口演5 知識工学

2017年11月21日(火) 08:30 ～ 10:00 G会場 (10F 会議室1006-1007)

[2-G-1-OP5-6] クリニカルパスと入院時記録を用いた患者状態の推定

山下 貴範¹, 若田 好史¹, 副島 秀久³, 中島 直樹¹, 廣川 佐千男² (1.九州大学病院メディカル・インフォメーションセンター, 2.九州大学情報基盤研究開発センター, 3.済生会熊本病院)

【はじめに】

近年、病院情報システムに蓄積された医療文書について、テキストマイニングや自然言語処理による研究が進められている。医療文書には入院時記録や経過記録、看護記録、退院時サマリなどがあり、それぞれ診療プロセスでの診断や治療、観察に関する事項が記載されている。クリニカルパスは、フリーテキストによって記録される患者状態や診療方針がアウトカムや観察項目として、構造化型式で蓄積されている。今回、アウトカムバリエーションと入院時記録を対象にして、脳梗塞症例の重症患者の分類を行い、特徴因子による推定性能を検証した。

【方法】

急性期病院 Aにおける2014年4月から2016年1月の脳梗塞クリニカルパス適用症例（1,222例）を対象とした。12種類のアウトカムバリエーションと入院時記録中の単語群を属性として、検索エンジンを構築した。目的変数は入院時 JCS 2桁以上（重症）とした。説明変数として3つのパターン、（A）12種類のアウトカムバリエーション、（B）入院時記録、（C）A + Bを設定し、Support Vector Machineを適用し識別モデルを構築した。スコア上位 N個の変数による属性選択を行い得られたモデルについて5分割交差検定で識別性能評価を行った。

【結果・考察】

各推定値について N=6の場合、A) 正答率0.37、F値0.35、B) 正答率0.83、F値 0.56、C) 正答率0.84、F値 0.63であった。また Bと Cの性能の推移を比較すると、Cの方が少ない因子から安定した性能の推移であった。Cでは、アウトカムバリエーションが上位スコアに含まれており、それが安定推移に影響していると考えられる。脳梗塞重症患者の特徴因子として、アウトカムバリエーションからは「バイタルサイン安定、呼吸状態安定」が抽出された。

テキストの構造化は解釈や分析精度向上に寄与するため、今後は入院時記録より抽出された特徴語について検証し、クリニカルパスへの導入を検討する。

クリニカルパスと入院時記録を用いた患者状態の推定

山下貴範^{*1}、若田好史^{*1}、副島秀久^{*2}、中島直樹^{*1}、廣川佐千男^{*3}

^{*1} 九州大学病院メディカル・インフォメーションセンター、

^{*2} 済生会熊本病院、

^{*3} 九州大学情報基盤研究開発センター

Prediction of a patient condition with clinical pathway and admission record

Takanori Yamashita^{*1}, Yoshifumi Wakata^{*1}, Hidehisa Soejima^{*2}, Naoki Nakashima^{*1}, Sachio Hirokawa^{*3}

^{*1} Medical Information Center, Kyushu University Hospital,

^{*2} Saiseikai Kumamoto Hospital,

^{*3} Research Institute for Information Technology, Kyushu University

Recently, the study for medical records accumulated in hospital information system has been advanced with NLP and text mining. For text-based record that has been described in a variety of expressions, the secondary use is a difficult. However, the medical record analysis is expected to improve medical process and the clinical decision support. The present paper applied text mining to extract the feature factors of the worsening patient group were using machine learning from the outcome of clinical pathway and admission records of cerebral infarction. A Clinical pathway is structured medical process. The explanatory variables set three patterns (A: twelve kinds of outcomes, B: words of admission records, C: A+B). A performance of feature selection was better pattern “C” than pattern “B”. Some outcomes are considered to affect to high-performance of feature selection. Structuring medical text must be contributed to improving medical quality.

Keywords: clinical pathway, admission record, text mining.

1. はじめに

近年、病院情報システムに蓄積された医療文書について、テキストマイニングや自然言語処理による研究が進められている。医療文書には入院時記録や経過記録、看護記録、退院時サマリなどがあり、それぞれ診療プロセスでの診断や治療、観察に関する事項や患者状態がフリーテキストで記載されている。フリーテキストの場合、記述内容が統一されておらず、個人により多様な表現で記述されているため、二次利用は困難な状況である。しかし、診療の本質的な内容が含まれる重要なデータであり、医療の質改善や意思決定支援に期待されている。¹⁾²⁾

著者らはこれまで、手術記録を対象に機械学習と属性選択を用いて、長期在院を特徴付ける単語や文章を抽出し、その推定性能を評価している。³⁾⁴⁾ 5)では、同様の手法で入院時記録から、クリニカルパスのアウトカムバリエーションを特徴付ける単語と文を抽出している。今回は、5)と同じ脳梗塞症例のデータ(入院時記録とクリニカルパスのアウトカムバリエーション)を用いて、JCS (Japan Coma Scale)を基準に患者の分類を行い、状態悪化した患者の特徴抽出とその推定性能を検証した。

2. クリニカルパスアウトカム・入院時記録

急性期病院 A における 2014 年 4 月から 2016 年 1 月の脳梗塞クリニカルパス適用症例(1,222 例)を対象とした。クリニカルパスは、フリーテキストによって記録される患者状態や診療方針がアウトカムや観察項目として、構造化データとして蓄積されている。12 種類のアウトカムバリエーションと入院時記録中の単語群を属性として、検索エンジンを構築した。検索エンジンの構築は NII で公開されている GETA を利用した。⁶⁾

表 1 は、12 種類のアウトカムバリエーションについて、入院期間中に一度でもバリエーションと判定された患者数を示す。「食事摂取ができる」、「バイタルサインが安定している」、「疼痛のコントロールができてい

る。

目的変数は、入院中に一度でも JCS 2 ケタ以上を記録した場合を悪化として定義した。悪化は 250 名、状態不変または改善は 972 名であった。説明変数の属性は 3 つのパターンを設定した(パターン A:12 種類のアウトカムバリエーション。パターン B:入院時記録から抽出した単語。パターン C:A+B)。

表 1 バリエーション判定されたアウトカムとその患者数

アウトカム	バリエーション数
食事摂取ができる	522
バイタルサインが安定している	513
疼痛のコントロールができてい	456
神経症状・所見の悪化がない	356
循環動態が安定している	157
排尿障害の症状・所見がない	133
呼吸状態が安定している	122
肺炎の症状・所見がない	14
副作用の症状・所見がない	12
安静が守られている	6
瞳孔異常の症状・所見がない	4
不均衡症候群の症状・所見がない	1

3. 分析方法

入院時記録全文書に出現する単語について、形態素解析を行い、Bag of Words に変換した(単語のベクトル化)。記号など解釈が難しい文字に対して stop words を用いて除外し、7,433 個の単語を対象にした。そして機械学習手法の一つである Support Vector Machine をパターンごとに適用し、悪化群の特徴付けを試みた。具体的には、線形カーネルでモデ

ルを構築し、その属性だけを含む仮想的な文書を考え、その属性のスコアを算出した。このモデルを適用した際、悪化群に該当する場合は **Positive**、そうでなければ **Negative** とした。

そして、**Positive**と**Negative**スコア(w.o)の上位N個の変数による属性選択を行い、得られたモデルについて5分割交差検定を適用し、その平均値の **Accuracy**、**F-measure** による推定性能の評価を行った。評価指標について、スコアと文書頻度の積(d.o)、さらにその対数(l.o)も追加で設定した。⁷⁾

3. 結果と考察

表2、表3は、パターンC(アウトカムバリエーション+入院時記録)で悪化群として分類され、特徴付けられた **Positive** と **Negative** の上位10の属性を示す。**Positive** 属性として、クリニカルパスのアウトカムからは「バイタルサイン安定、排尿障害の症状・所見がない、呼吸状態安定、循環動態安定」、入院時記録からは「意識障害、MCA、発語」などの単語が抽出された。**Negative** 属性として、クリニカルパスのアウトカムからは「疼痛コントロール」、入院時記録からは「動揺、考え、構音障害」などの単語が抽出された。「動揺」は歩行時動揺に関する文章であった。

属性選択について、各パターンの推定値を N=6 で比較すると、パターンAは、Accuracy:0.37、F-measure:0.35。パターンBは、Accuracy:0.83、F-measure:0.56。パターンCは、Accuracy:0.84、F-measure:0.63であった。そして、パターンBとパターンCの推定性能の推移を比較すると、パターンCの方が少ない属性から安定した性能の推移であった(図1、図2)。両図の d.o を見ると、その差が顕著である。またパターンCの方が、Nが1~2の差ではあるが少ない属性でピークに達している(図2)。パターンAは属性数が12と少ないため推定性能の比較はできない。

パターンCは、アウトカムが上位スコアに含まれており、それが安定推移に影響していると考えられる。

4. まとめ

状態悪化患者を基準にして、クリニカルパスのアウトカムバリエーションと入院時記録に対して機械学習を適用し、特徴抽出を行った。臨床的評価はこれからであるが、そのために解釈できる属性が抽出されたと考える。属性選択では、構造化と非構造化を組合せたことで、推定性能が高く安定した性能を達成できた。目的変数の設定は異なるが、5)の属性選択よりも少ない属性数で高い性能が達成された。アウトカムは属性選択に有効であることが分かった。

今後は医療の質改善に向けて、表現の標準化や分析精度の向上を目的としたフリーテキストの構造化を進める予定である。

参考文献

- 1) Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research Yearbook of medical informatics 2008. 128-144.
- 2) Zhua F, Patumcharoenpolc P, Zhanga C, et al. Biomedical text mining and its applications in cancer research Journal of Biomedical Informatics 2013. 46(2): 200-211.
- 3) 山下貴範, Brendan Flanagan, 若田好史ら. テキストマイニングで抽出された重要単語を利用した重要文の評価. 第35回医療情報学連合大会論文集 2015. 35: 520-523.
- 4) Yamashita T, Wakata Y, Nakashima N, et al. Presumption Model for Postoperative Hospital Days from Operation Records.

International Journal of Computer & Information Science 2015: 16: 50-59.

- 5) Yamashita T, Wakata Y, Soejima H, Nakashima N, Hirokawa S. Prediction of Key Patient Outcome from Sentence and Word of Medical Text Records. Proceedings of Clinical Natural Language Processing Workshop at COLING 2016: 86-90.
- 6) <http://geta.ex.nii.ac.jp/geta.html>
- 7) Sakai T, Hirokawa S. Feature Words that Classify Problem Sentence in Scientific Article Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services 2012. 360-367.

表2 SVM分類されたPositive属性 (Top10)

※下線はアウトカム

SVM スコア	頻度	属性
0.2818	200	意識障害
0.2788	513	<u>バイタルサインが安定している</u>
0.2391	314	MCA (中大脳動脈)
0.2196	113	<u>排尿障害の症状・所見がない</u>
0.2009	181	発語
0.1875	122	<u>呼吸状態が安定している</u>
0.1717	89	評価不能
0.1595	157	<u>循環動態が安定している</u>
0.1553	52	レバミピド
0.1550	30	変動

表3 SVM分類されたNegative属性 (Top10)

※下線はアウトカム

SVM スコア	頻度	属性
-0.1810	456	<u>疼痛のコントロールができています</u>
-0.1588	456	動揺
-0.1404	316	考え
-0.1363	745	異常
-0.1360	756	構音障害
-0.1295	210	病型
-0.1244	30	顆粒
-0.1236	539	正常
-0.1197	176	口角下垂
-0.1194	194	未同定

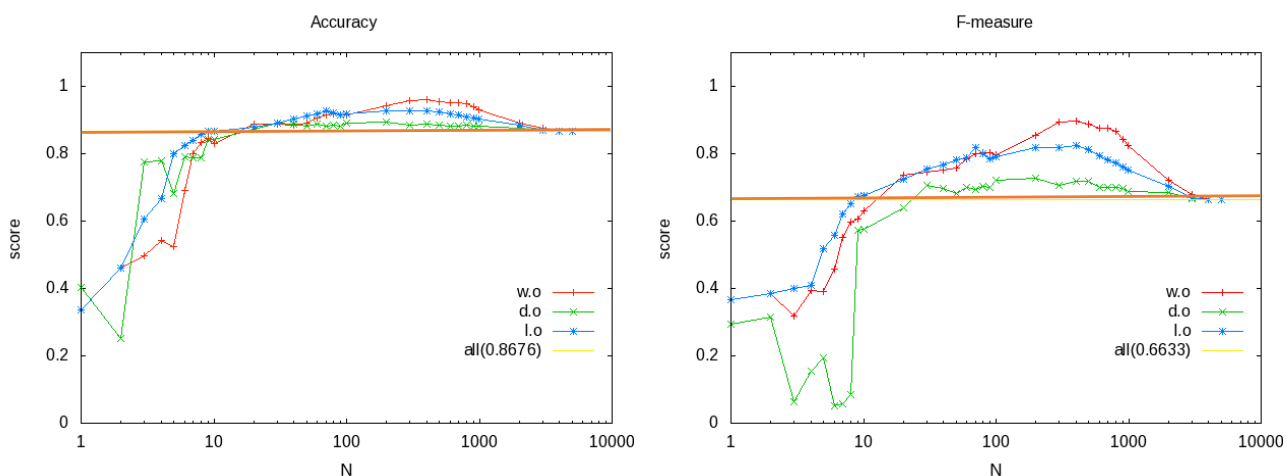


図 1 入院時記録のみの推定性能(パターン B)

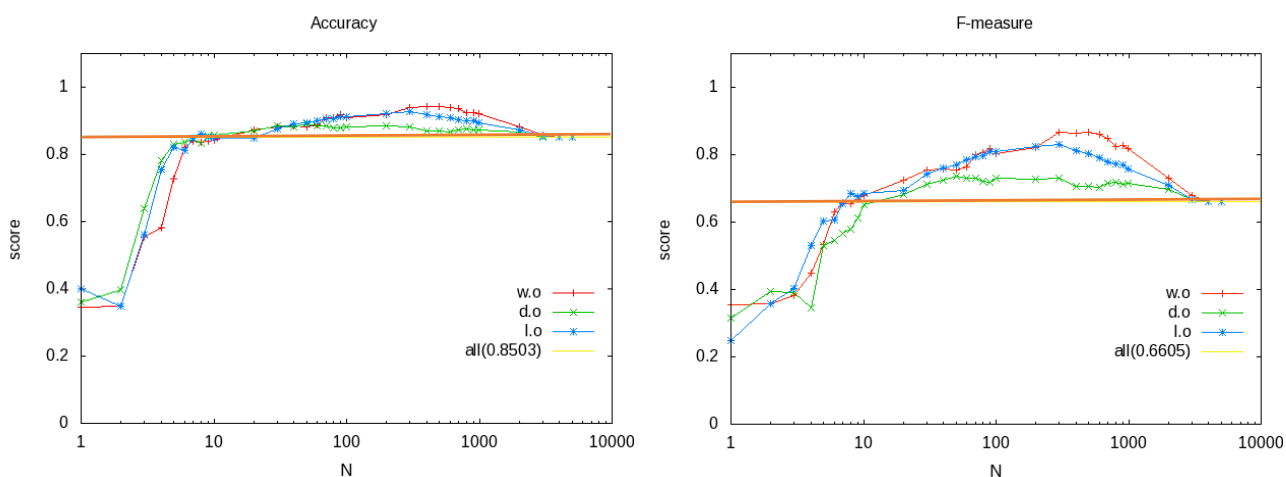


図 2 入院時記録+アウトカムバリエーションの推定性能(パターン C)