

平成 30 年 6 月 13 日現在

機関番号：17102

研究種目：基盤研究(B) (一般)

研究期間：2015～2017

課題番号：15H02778

研究課題名(和文) テキストマイニングによる医療プロセスのキーファクター抽出

研究課題名(英文) Extraction Key Factors of Medical Process by Text Mining

研究代表者

廣川 佐千男 (Hirokawa, Sachio)

九州大学・情報基盤研究開発センター・教授

研究者番号：40126785

交付決定額(研究期間全体)：(直接経費) 12,500,000円

研究成果の概要(和文)：本研究では、診療活動と患者状態を組織的に記録したクリニカルパスを対象に、問題要因発見を目指し、機械学習と可視化の研究を行った。例えば長期入院患者の記録を正例、それ以外の記録を負例として機械学習を適用し、重要因子を示す特徴語や検査値を抽出し、その識別性能の評価を行った。可視化では、解釈を容易にする対話的な因子時系列展開グラフと、診療記録のテンプレート作成を支援するシステムを実現した。

研究成果の概要(英文)：We focused on systematic clinical records (clinical paths) of clinical practice and patient condition, aiming at discovering problem factors and applied machine learning and visualization. For example, machine learning was applied by taking records of long-term inpatients as positive cases and other records as negative cases, and feature words and examination values indicating important factors were extracted, and the discrimination performance was evaluated. In the visualization, we realized an interactive factor time series evolution graph that facilitates interpretation and a system that supports creation of template of medical record.

研究分野：情報科学

キーワード：クリニカルパス 機械学習 テキストマイニング Support Vector Machine 属性選択 可視化

1. 研究開始当初の背景

日本はこれまで世界最高の平均寿命や高い保険医療水準を実現してきたが、世界に先駆けて超少子高齢化社会が到来し環境が大きく変化している。そうした変化に対応して良質で効率的な医療サービスを提供する必要性がますます高まっているため、国レベルでは医療制度の見直しや様々な事業の推進が進められている。各医療機関においても業務改善を目的としたシステム化や多くの取組みが行われており、病院情報システムの普及と医療情報の標準化により診療の電子化が進み、膨大な医療データが蓄積されてきている。医療データには、検査などの構造化数値データや画像データ、波形データ、そして医師の診察や看護師の観察により記録された非構造化テキストのデータが蓄積されている。しかし診療の本質的考察などが記録されている日々のカルテやサマリーなどの非構造化テキストデータは、十分に活用されていない。

従来の診療テキストデータに対する研究は、退院サマリーやレポート類を対象としたものがあり、重要語の抽出や語彙の関連の可視化、病名やがんステージの分類などが報告されている。また、医療情報の基礎研究として京都大学の金子周司教授らの生命科学用語オントロジーの研究や同大学の荒牧英治特定准教授らの MedNLP プロジェクトにより医療に関する言語データの基盤ができつつある。しかし、医療現場へのフィードバックは解決されてない課題であった。

2. 研究の目的

代表者は大量文書群の解析のため、特徴的語抽出と関連可視化の研究を行ってきた。表やグラフによる可視化は直感的に分り易いという利点がある反面、そ

れがどのように有効か、あるいは正しいかという定量的評価が困難である。そこで、機械学習で広く利用されている SVM に焦点を当て、属性選択として特徴語抽出を実現し、それによる推定性能の定量的評価を行ってきた。この手法の医療テキストデータへの適用について、分担者(中島)との共同研究がきっかけとなった。代表者が考案した SVM(support vector machine)における最適属性選択法は、推定性能が既存手法より大きく向上する場合がある。本研究の目的は、どんなデータでうまくいくのか、推定性能向上の理由、そして他の機械学習手法についての適用可能性を検討し発展させることであった。

3. 研究の方法

対話的分析システム開発と医療プロセス改善探索手法開発という二つの方法で研究を進めた。前者については、医療情報の二次利用を可能とする対話的分析システムを開発した。例えば回復が長引く患者の特徴分析のため、「入院期間 25 日以上」という検索を行い、該当する患者のカルテの特徴語と検査値のクロス表や、特徴語の関連図(マインドマップ)を提示するシステムの開発をめざした。後者については、改善対象を数値やキーワードとして表し検索を行い、該当するデータの特徴を抽出する。機械学習の属性選択を進展させ、医療プロセスの各段階で得られる因子(数値データやテキスト中の重要語)が、回復や退院という医療活動の最終的にどれだけ影響を与えるかを定量的に評価する手法の開発をめざした。

4. 研究成果

本研究では、医療情報の典型的事例として主にクリニカルパスを分析対象として研究を進めた。クリニカルパスは診療活動と患者状

態の記録をスケジュール化したものである。この記録を基礎データとして、改善要因を発見することが本研究の最も重要な課題であった。

本研究では、例えば標準より長期入院となった患者の記録を正例、それ以外の患者の記録を負例として機械学習を適用し、重要因子を示す特徴語や検査値を抽出し、それらにより識別性能の評価を行った。また、得られた特徴語や特徴因子の解釈のために2つの可視化手法を提案した。一つは、非構造化テキストデータからテンプレート作成のための有向グラフで、もう一つは、構造化数値データを対象に各因子を時系列に展開する対話的グラフである。

機械学習による分析では、非構造化テキストデータである日々の診療記録と入院時記録、手術記録を対象として、それぞれの目的変数を在院日数や入院中の特異な患者状態を設定した。そして、Support Vector Machine (SVM) を適用して分類し、特徴語や特徴文を抽出した。さらに、属性選択による推定性能の評価を行った。具体的には、人工股関節置換術症例の手術記録の分析では、単語処理における辞書の有効性が実証できたが、単語のみでは臨床的解釈が困難という課題が残った。解釈の改善のために、SOAP (Subjective, Objective, Assessment, Plan) 形式で記録される同症例の日々の診療記録について、項目毎の単語出現を区別するため「タグ付単語」を導入した結果、長期在院を正例とする SVM を適用し分類した結果、Objective 項目の検査に関連する単語の重要性が分った。また、脳梗塞症例の入院時記録について、入院中の患者状態の異常を入院時点に推定することを目的とした分析を行うため、「疼痛」と「神経症状の悪化」の状態を正例として、SVM から判定された特徴語と特徴文を抽出し、性

能評価を行った。

診療データ分析の更なる解釈の改善のために2つの可視化を実現した。一つ目は、腎臓疾患の血液浄化症例のパスの診療記録を対象にテンプレート抽出を目的とした。SVM と単語 2-gram を適用し、スコアの高い単語をノードとして繋ぐ有向グラフを構築し、特徴文を抽出した。もう一つの可視化では、パスの構造化数値データを分析対象とし、人工股関節置換術症例で長期在院となった患者集合の患者状態の異常の早期発見を試みた。このシステムでは、入院日数付きの患者状態の異常を属性として、属性間の類似度には Jaccard 係数を用いた。属性をノードとする木構造グラフを描くことで、属性間の関連を可視化した。その結果、長期在院に影響している重要な患者状態として呼吸状態、疼痛、循環動態、創部問題が抽出され、それらの関連と経過が把握できた。このシステムにより、対話的にノード間の類似度や該当患者数を設定でき、試行錯誤による探索的分析ができることを確認できた。従来の記述統計や統計解析ではできなかった、変数間の作用順序や時間的重層関係の把握が日数付きの属性の導入で達成できた。

5 . 主な発表論文等

[雑誌論文] (計 1 件)

1. Takanori Yamashita, Yasuharu Nakashima, Naoki Nakashima, Yoshifumi Wakata, Yukihide Iwamoto, Sachio Hirokawa, Satoshi Hamai, Brendan Flanagan, Presumption Model for Postoperative Hospital Days from Operation Records, International Journal of Computer & Information Science 16 (1), pp.50-60,2014

[学会発表](計16件)

1. Takanori Yamashita, Yoshifumi Wakata, Satoshi Hamai, Yasuharu Nakashima, Yukihide Iwamoto, Brendan Franagan, Naoki Nakashima, Sachio Hirokawa, Visualization of Key factor Relation in Clinical Pathway, Proc. KES2015, 342-351, 2015
2. Takanori Yamashita, Yoshifumi Wakata, Satoshi Hamai, Yasuharu Nakashima, Yukihide Iwamoto, Brendan Franagan, Naoki Nakashima, Sachio Hirokawa, Temporal Relation Extraction in Outcome Variance of Clinical Pathway, poster, MEDINFO2015, pp. 1077, 2015
3. 久保永, 山下貴範, ブレンダン フラナガン, 若田好史, 中島直樹, 副島秀久, 廣川佐千男, 半構造化された診療記録に対する長期在院患者の推定電気・情報関係学会九州支部連合大会, 10-1P-12, 2015-07-27
4. 足立裕介, 山下貴範, ブレンダン フラナガン, 若田好史, 中島直樹, 副島秀久, 廣川佐千男, 長期在院患者の推定と2つの機械学習手法の比較電気・情報関係学会九州支部連合大会, 10-1P-13, 2015-07-27
5. Naoya Onimura, Takanori Yamashita, Naoki Nakajima, Hidehisa Soejima, Sachio Hirokawa, Generation of Sentence Template Graph from SOAPFormat Medical Documents, Proc. CSCI2016, pp.159-162, 2016
6. Takanori Yamashita, Yoshifumi Wakata, Hidehisa Soejima, Naoki Nakashima, Sachio Hirokawa, Structuralization of Variance Text Records in Clinical Pathway, Proc. APAMI2016, p.85
7. Takanori Yamashita, Yoshifumi Wakata, Yasunobu Nohara, Satoshi Hamai, Yasuharu Nakashima, Yukihide Iwamoto, Naoki Nakashima, Sachio Hirokawa, Performance Evaluation of Predicting the Length of Hospitalization by Operation Record, Proc. IEEE International Conference on Biomedical and Health Informatics, 2016
8. Haruka Kubo, Takanori Yamashita, Yoshifumi Wakata, Hidehisa Soejima, Naoki Nakashima, Sachio Hirokawa, Effect of Synonym on Prediction Performance for Postoperative Hospital Stay by Text Mining, Proc. AROB2016, pp.409-413, 2016
9. Yuusuke Adachi, Takanori Yamashita, Yoshifumi Wakata, Hidehisa Soejima, Naoki Nakashima, Sachio Hirokawa, Comparison of SVM and Decision Tree for Prediction of Postoperative Hospital Stay Proc. AROB2016, pp.414-419, 2016
10. Naoya Onimura, Brendan Flanagan, Takanori Yamashita, Sachio Hirokawa, Performance Effect of Feature Selection on Support Vector Machine Proc. AROB2016, pp.420-424, 2016
11. Adachi, Y., Onimura, N., Yamashita, T., Hirokawa, S., Standard measure and SVM measure for feature selection and their performance effect for text classification, Proc. iiWAS 2016, pp. 262-266, 2016
12. Takanori Yamashita, Naoya Onimura, Hidehisa Soejima, Naoki Nakashima, Sachio Hirokawa, Graph Clustering System for Text based Records in a Clinical Pathway, Studies in Health Technology and Informatics, vol.245,

pp.649-652, 2017
DOI10.3233/978-1-61499-830-3-649
Proc. MEDINFO2017

13. Yusuke Adachi, Naoya Onimura, Takanori Yamashita, Sachio Hirokawa, Classification of Imbalanced Documents by Feature Selection, Proc. ICCDA 2017, pp.228-232, 2017
14. N.Onimura, T.Yamashita, N.Nakashima, H.Soejima, S.Hirokawa, Interactive Visualization of Template Graph for Daily Clinical Notes, Proc. of the Eighth International Conference On Advances in Computing, Electronics and Electrical Technology - CEET 2018, pp.19-24, 2018
15. Naoya Onimura, Takanori Yamashita, Naoki Nakashima, Hidehisa Soejima, Sachio Hirokawa, Machine Learning Support for Template Design of Clinical Notes, Proc. BHI 2018,2018
16. Takanori Yamashita, Naoki Nakashima, Sachio Hirokawa, Classification and Feature Extraction for Text-based Drug Incident Report, Proceedings of the 2018 6th International Conference on Bioinformatics and Computational Biology pp.145-149, 2018

6 . 研究組織

(1)研究代表者

廣川 佐千男 (HIROKAWA, Sachio)
九州大学・情報基盤研究開発センター・教授
研究者番号 : 40126785

(2)研究分担者

中藤 哲也 (NAKATOH, Tetsuya)
九州大学・情報基盤研究開発センター・

助教

研究者番号 : 20253502

中島 直樹 (NAKAJIMA, Naoki)

九州大学・大学病院・教授

研究者番号 : (60325529)

(3)連携研究者

(4)研究協力者

副島 秀久 (SOEJIMA, Hidehisa)

済世会熊本病院・院長

山下 貴範 (YAMASHITA, Takanori)

九州大学・大学病院・助教