

科学研究費助成事業 研究成果報告書

平成26年3月31日現在

機関番号：34310
研究種目：基盤研究 (C)
研究期間：2011～2013
課題番号：23500190
研究課題名 (和文) 階層構造モデル化に基づく多変量時系列医療データからの知識発見
研究課題名 (英文) Knowledge Discovery in Multivariate Clinical Time Series
based on Hierarchical Structure Modeling
研究代表者
大崎 美穂 (OHSAKI, Miho)
同志社大学・理工学部・准教授
研究者番号：30313927
交付決定額 (研究期間全体)：(直接経費) 2,200,000 円、(間接経費) 660,000 円

研究成果の概要 (和文)：

多変量時系列である慢性病の医療データから、病状把握や治療に役立つ知識を自動的に発見する仕組みが望まれている。そこで、本研究課題では C 型慢性肝炎の検査治療履歴を題材とし、病状のモデル化・特徴量抽出・予測とそのシステムに基づく知識発見を目指した。具体的には、自己回帰モデルと自己回帰条件付き分散不均一モデルの階層的な適用による病状記述、これらのモデルに基づく特徴量抽出、モデルの次数から必要な検査回数に関する知識を導出した。カーネルロジスティック回帰を適用して肝臓の悪化度合いを予測するシステムを開発し、さらにカーネルロジスティック回帰に識別学習を導入した新しい分類器を提案した。

研究成果の概要 (英文)：

It has been required to automatically discover useful knowledge for symptom recognition and treatment in clinical data on chronic disease, which is a set of multivariate time series. This research project aims to develop methods of modeling, feature extraction, and prediction of symptoms for clinical examination and treatment histories on chronic hepatitis C. We described symptoms by hierarchically applying autoregressive models and autoregressive conditional heteroscedastic models, extracted features using the modeling results, and derived knowledge on how much measurements are necessary for symptom recognition based on the model orders. In addition to that, we developed a system to predict the degree of liver damage using kernel logistic regression and proposed a new classifier that is based on kernel logistic regression and trains itself with discriminative learning.

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：知識発見, 多変量時系列, 検査治療履歴

1. 研究開始当初の背景

肝炎や生活習慣病のような慢性病では、医師が様々な検査値変動を見て治療を行うが、人手による多変量時系列の理解には限界がある。ゆえに、現場で蓄積された医療データから病状把握や治療に役立つ知識の発見が望まれている。

医療データは高次元・長期・数値記号混在・多欠損という分析困難な側面を持ち、小康と悪化が遷移する系の複雑さを内包する。知識

発見の実現には、このような扱い難い性質の医療データからモデルを構築し、病状を予測する仕組み、その仕組みから知識を見出す方法が必要である。

2. 研究の目的

本研究では C 型慢性肝炎の検査治療履歴を題材とし、病状のモデル化・特徴量抽出・予測とそのシステムに基づく知識発見を目的とする。従来、医療データには表出パターン記述アプローチ(表出した検査結果の波形を

記述する)が取られてきたが、我々は確率過程記述アプローチ(検査結果の背後にある確率的・動的な病状を記述する)を取る。このアプローチにより、検査治療履歴のモデル化と特徴量の抽出、特徴量に基づく知識発見、病状の予測といった一連の手法を開発する。

一連の手法を開発する中で、検査治療の項目ごとにサブモデルを構築する。そして、病状のクラスの推定精度を基準として、サブモデルからメタな階層構造モデルを構築する方法論を検討する。

3. 研究の方法

検査治療履歴のモデル化と特徴量の抽出では、自己回帰 (AR) モデルによる病状の記述と、肝臓の悪化度合いを表す肝線維化ステージの予測を目指した特徴量 (AR モデルから導出される LPC ケプストラム、平均、標準偏差の組合せ) を提案する。特徴量に関しては、肝線維化ステージの予測性能を基準とし、AR モデルの次数、言い換えると LPC ケプストラムの次数を最適化する手法も開発する。

最適化の結果は、特徴量に基づく知識発見に活用する。具体的には、最適化で得られた AR モデルの次数から、検査項目ごとに病状把握に必要な検査回数が示唆されることを示す。

次に、検査値時系列の分散変動という非定常性を記述できるように、自己回帰条件付き分散不均一 (ARCH) モデルを適用し、新たな特徴量を提案する。さらに AR モデルと ARCH モデルを階層的に適用することで、より詳細な病状の記述と特徴量の改善を図る。

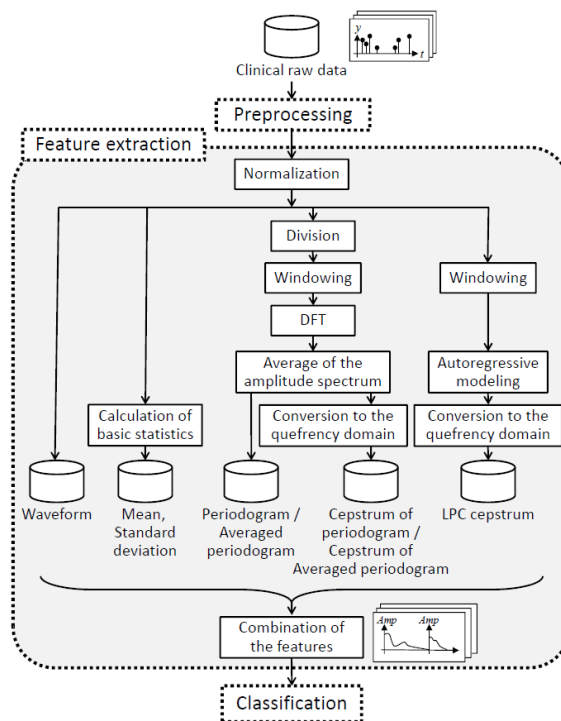
病状の予測では、肝線維化ステージ予測の分類器としてカーネルロジスティック回帰 (KLOGR) を適用することを提案し、その有効性を実験検証する。これに加えて、データの不均衡性(陰性事例が多く、陽性事例が少ない)という問題を解決するために、KLOGR に識別関数アプローチの学習を導入した新たな分類器を開発する。

4. 研究成果

検査治療履歴のモデル化と特徴量の抽出のために、前処理、モデル化・特徴量抽出、分類の処理を行う肝線維化ステージ予測のシステムを開発した(以下にシステムの構成図を掲載する)。そして、実際に各患者・各項目の検査値時系列に AR モデルを当てはめ、提案した特徴量を抽出した。

実験では、提案した動的な特徴量(モデル適用により導出した LPC ケプストラムに基づくもの)、他の動的な特徴量(モデルを適用せず

に求めたピリオドグラムに基づくもの)、従来研究の特徴量(1 時点の検査値、あるいは、表出パターン記述アプローチで導出した特徴量)を網羅的に比較した。その結果、提案した特徴量が安定した最も高い予測性能を達成した。



特に、AR モデルの次数を最適化した場合、提案した特徴量の有効性は顕著であった。また、各患者・各項目にとって最適な AR モデルの次数が得られ、患者と検査の種類に適する検査回数を知ることができた。これらの成果は、学術論文と国際会議講演論文にて公表した。

次のステップとして、検査値時系列の非定常性の記述を試みた。ARCH モデルを適用して分散変動の傾向を捕らえるとともに、ARCH モデルの係数から LPC ケプストラムに相当する量を抽出する方法、および、この量(便宜上、ARCH モデルによる LPC ケプストラムと呼ぶ)、平均、標準偏差を組み合わせた特徴量を提案した。

また、AR モデルと ARCH モデルの階層的な適用により、LPC ケプストラム、ARCH モデルによる LPC ケプストラム、平均、標準偏差の 4 つを組み合わせた特徴量も導出した。ARCH モデルに基づく特徴量を検証する実験は継続中であり、今後、有効性を明らかにしていく。

病状の予測においては、我々が適用を提案する KLOGR、従来研究で用いられた分類器である重回帰 (MLR)、ロジスティック回帰 (LOGR)、

正則化ロジスティック回帰 (RLOGR), 最近傍法 (1-NN), サポートベクターマシン (KSVM) を比較する実験を行った.

病状予測の実験結果を抜粋し, 以下の2つの表に示す. 医療データは不均衡であるため, 予測性能を複数の異なる基準 (Acc, Sens, Spec, PPV, NPV) で多面的に評価することが一般的である. そこでこれらの基準の相加平均 (AM, AM') と調和平均 (HM, HM') を求め, 1つ目の表には総合的な予測性能の順位を掲載した. 2つ目の表には, 個々の基準の値, ただし, 上位2位までの分類器の値, 最小値, 最大値に限定したものを掲載した.

Average for Two Classification Problems					
Classifiers	AM	AM'	Classifiers	HM	HM'
KLOGR	75.34 (1st)	72.91 (1st)	KLOGR	72.43 (1st)	69.85 (1st)
KSVM	70.54 (2nd)	67.86 (2nd)	KSVM	66.28 (2nd)	63.36 (2nd)
LOGR	68.08 (3rd)	66.10 (3rd)	LOGR	64.68 (3rd)	62.36 (3rd)
RLOGR	67.76 (4th)	65.41 (4th)	RLOGR	63.34 (4th)	60.63 (4th)
MLR	66.35 (5th)	63.61 (5th)	MLR	61.26 (5th)	58.24 (5th)
1-NN	61.63 (6th)	58.53 (6th)	1-NN	55.50 (6th)	52.23 (6th)
Lower Boundary	50.00	50.00	Lower Boundary	0.00	0.00

Average for Two Classification Problems					
Classifiers	Acc	Sens	Spec	PPV	NPV
KLOGR	85.03	59.71	85.21	56.49	90.24
KSVM	81.25	58.18	81.18	45.22	86.86
Maximum Value	85.03	59.71	85.21	56.49	90.24
Minimum Value	74.01	38.55	74.97	40.15	77.31

KLOGR は, 表1の AM, AM', HM, HM', 表2の Acc, Sens, Spec, PPV, NPV の全てで最も高い値を達成しており, その有効性が明らかになった. また, KLOGR は LOGR と KSVM の利点を併せ持つ点でも有用と考えられる. この成果は国際会議講演論文にて公表した.

最後に, 医療データに顕著である不均衡性の問題解決に向け, 識別関数アプローチにより KLOGR を再学習する仕組みを持つ分類器を提案した. 現時点では予備実験の段階ではあるが, この新しい分類器が KLOGR よりも高い予測性能を持つことを確認している. 今後, 多くのデータを用いて詳細な評価を行う予定である.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計2件)

(1) 大崎美穂, 宮崎淳文, 谷口恵梨, 片桐滋, 横井英人, 高林克日己, C型慢性肝炎の肝線維化ステージ推定を目指した検査値時系列の動的特徴量の提案, 医療情報学, 査読有, vol. 33, no. 2, 2013, pp. 79-98

(2) 大崎美穂, 音量の好みを用いた評価値予測によるインタラクティブ EC フィッティングの改善手法, 知能と情報, 査読有, vol. 24, no. 2, 2012, pp. 680-690

[学会発表] (計11件)

(1) Akifumi Miyazaki, Miho Ohsaki, Eri Taniguchi, Shigeru Katagiri, Hideto Yokoi, Katsuhiko Takabayashi, A Feature Consisting of Mean, Standard Deviation, and LPC Cepstrum for the Prediction of Liver Fibrosis Stages in Chronic Hepatitis C, Joint International Conference on Soft Computing and Intelligent Systems and International Symposium on Advanced Intelligent Systems SCIS&ISIS-2012, 査読有, 2012, pp. 2295-2298

(2) Kenji Matsuda, Miho Ohsaki, Shigeru Katagiri, Hideto Yokoi, Katsuhiko Takabayashi, Application of Kernel Logistic Regression to the Prediction of Liver Fibrosis Stages in Chronic Hepatitis C, Joint International Conference on Soft Computing and Intelligent Systems and International Symposium on Advanced Intelligent Systems SCIS&ISIS-2012, 2012, 査読有, pp. 780-784

(3) Akifumi Miyazaki, Miho Ohsaki, Eri Taniguchi, Shigeru Katagiri, Hideto Yokoi, Katsuhiko Takabayashi, Feature Extraction for the Prediction of Liver Fibrosis Stages in Chronic Hepatitis C, IEEE Region 10 Conference TENCON-2012, 2012, 査読有, paper-no. 1569625253

(4) 安本亮介, 宮崎淳文, 大崎美穂, 片桐滋, 横井英人, 高林克日己, C型慢性肝炎の肝線維化ステージ推定を目指した検査値時系列区分化の影響の検証, 情報処理学会研究報告, vol. 2013-BI0-36, no. 14, 2013, pp. 1-6

(5) 宮崎淳文, 大崎美穂, 安本亮介, 片桐滋, 横井英人, 高林克日己, C型慢性肝炎の肝線維化ステージ推定を目指したスペクトラム推定量に基づく特徴量の比較, 情報処理学会研究報告, vol. 2013-BI0-36, no. 13, 2013, pp. 1-6

(6) Akifumi Miyazaki, Miho Ohsaki, Shigeru Katagiri, Modeling and Feature Extraction of Clinical Time Series Aiming at the Diagnosis Prediction of Chronic Hepatitis C, Joint Symposium between Chonnam National University and Doshisha

University, 2013, no.PD14

(7) 安本亮介, 大崎美穂, ファジィアプローチとベイズアプローチの関係の調査, 第 29 回ファジィシステムシンポジウム, 2013, poster-no. 20

(8) 大崎美穂, 宮崎淳文, 松田健司, 片桐滋, 横井英人, 高林克日己, C 型慢性肝炎の肝線維化ステージ予測への取り組み, パーティクルフィルタ研究会, 2013, paper-no. 3

(9) 谷口恵梨, 大崎美穂, 片桐滋, 横井英人, 高林克日己, C 型慢性肝炎の肝線維化ステージ予測に適する特徴量の検討, 第 39 回知能システムシンポジウム, 2012, pp. 305-308

(10) 谷口恵梨, 大崎美穂, 片桐滋, C 型慢性肝炎の肝線維化ステージ予測に役立つ特徴量の提案, 第 32 回東海ファジィ研究会, 2012, pp. 24. 1-24. 4

(11) 松田健司, 大崎美穂, 片桐滋, C 型慢性肝炎の肝線維化ステージ分類手法の比較, 第 32 回東海ファジィ研究会, 2012, pp. 23. 1-23. 4

6. 研究組織

(1) 研究代表者

大崎 美穂 (OHSAKI, Miho)

同志社大学・理工学部・准教授

研究者番号: 30313927