

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 15 日現在

機関番号：11301

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25330002

研究課題名(和文) 計算幾何を用いたイメージ検索モデルの構築と知識抽出への応用

研究課題名(英文) Image retrieval model using computational geometry

研究代表者

全 眞嬉 (CHUN, JINHEE)

東北大学・情報科学研究科・准教授

研究者番号：80431550

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：本研究の目的は医療データにおける現在の精度限界を打破するための可視化知識抽出モデルの提案、理論研究、システムの構築である。

医療画像データ認識の為に、画像切り出し問題を、最大重み領域問題としてアルゴリズムを与えた。基本領域の分割可能な和集合で表される領域族に対する最大重み領域問題について効率的なアルゴリズムを提案した。本結果と交差を許した和集合領域の最適切り出し問題におけるNP困難性を比較すると、計算量と幾何学的性質の関連の解明として学術的に興味深い結果を示した。

研究成果の概要(英文)：In this research gives optimal approximation algorithms for medical image data and numerical data as a theoretical hierarchical segmentation. A given pixel grid and baselines, one can compute in polynomial time a maximum-weight region that can be decomposed into disjoint base-monotone regions. To finding the maximum weight region that is the union of elementary shapes for the corresponding cases. We proved to be NP-hard. This result shows improve to relation of computational complexity and geometric properties.

研究分野：計算理論

キーワード：計算理論 計算幾何 知識抽出 イメージ検索

1. 研究開始当初の背景

現行のデータマイニングシステムでは知識発見過程が医師に理解できる形で提示されておらず、医師の診断過程に利用できる客観的な意思決定をサポートしているとは言いにくい。更に、医療データマイニングで必須である数値データの取り扱いを考えると、属性間の相関関係を求めて可視化する際に、現在のグラフや相関表などの離散データ解析手法をそのまま適用することはできない。また、二値化を行うと数値データの二値化誤差から生じる情報損失のため、正確性と学習汎用性のトレードオフに関する精度限界に直面する。従って、医療データマイニングを考えると医療データに含まれる巨大数値データベースを効率的に処理する最適化アルゴリズムが必要である。

近年は社会分野のみならず医療分野でもデータマイニング技術が普及されている。医療分野では、医師の経験だけではなく実験や膨大なデータ解析を通じて客観的な根拠に基づいて診断し、治療を行う医療が重視されるようになった。このような背景により医療分野では統計に基づくデータ分析手法が広く利用されている。ところが現行のデータマイニングシステムを医療に使うことは困難である。なぜなら現行のデータマイニングシステムは抽出・発見したルールやパターンを提示するが、その知識発見過程を明示的に示し、説明することは困難である。統計手法では一般に、仮説を立ててその仮説を検証するが、仮説や事前知識をあらかじめ固定せずに、データから頻出パターンやルール、知識を抽出し、データを説明する仮説を構築するのがデータマイニングである。すなわち統計が仮説検証のための手法であるのに対し、データマイニングは仮説構築の手法である。

また、医療データには膨大な画像データがある。医療データにおける画像データはCT、MRI、レントゲン、超音波検査などの画像データがある。画像データの解析に関しても様々な研究が行われている。近年、fMRI(functional magnetic resonance imaging)が開発されており脳機能を計測・画像化する手法が開発されている。ところが、現在の手法では画像データにはノイズが多く、また膨大なデータ容量であるため、解明が困難であり、データマイニング適用に期待が寄せられている。

これらの問題を解決するために、2次記憶上の巨大数値データベースを効率的に処理する最適化アルゴリズム研究と診断過程を明示的に示す可視化システム研究が強く必要とされる。数値・画像データから知識検出の高速化と精度を上げると共にルールの可視化を行う事を目標とする。

2. 研究の目的

本研究の目的は医療データマイニングにおける現在の精度限界を打破するための可視化知識抽出モデルの提案である。

医療データには膨大な属性の数値データと膨大な画像データがある。ところが、現在の手法では画像データにはノイズが多く、また膨大なデータ容量であるため、解明が困難であり、データマイニング適用に期待が寄せられている。本研究では医療データの数値データと画像データを数値データ集合として幾何学的に扱い、計算幾何学を用いて巨大数値データベースを効率的に処理する新しい最適化アルゴリズムを与え、上記の精度限界を超える精密な手法の提案を行う。

医療画像データ処理は、画像データを数値データ集合として幾何学的に扱い、計算幾何学を用いて巨大数値データベースを効率的に処理する新しい最適化アルゴリズムを与え、上記の精度限界を超える精密な手法の提案を行う。さらに相関関係を幾何学的に可視化することにより、ユーザにとって知識発見過程が明示的であり判り易い知識抽出理論確立とシステムの設計が必要である。本研究では画像データを計算幾何学手法を利用して、構造的な圧縮データに変換し、計算時間の高速化を行う。

本研究では、応募者が持つ離散アルゴリズムを用いた計算幾何学的な関数の最適近似の独自技法を利用し、データ分布を、ユーザに説明しやすい関数を用いて近似することで数値データベースからルールを抽出し、更にそれを組み立てて知識発見過程が医師にとって明確に理解できるシステムの構築を行う。これは、過去のデータマイニング技法に計算幾何学技法と幾何学的可視化を取り入れた画期的なシステム構築となる。

3. 研究の方法

本研究の最終的目的地は、提案する数値データに対する階層化ルール理論のアルゴリズムの高速化と改良、医療画像データに対する知識抽出の際に、画像切り出し問題を、最大重み領域問題として扱い、医療画像の切り出しを行い、新しい最適化アルゴリズムを与え、原稿の医療診断の精度限界を超える精密な手法の提案である。

新しい領域族を階層最適化し、より高次元のルールの効率的なアルゴリズムの設計を行い、その結果を結合ルール生成だけでなく、データの視覚化及びデータマイニングへの幾何学的なアプローチにおいても有効に応用を行った。

自動的に抽出し表示された知識形態は、ユーザにより意志決定等の補助として用いられる。重要な条件は、抽出した知識形態がシンプルであり(単純性)、正確にデータの特性を記述すること、知識としてサンプルに依存しない汎用性を持つ事さらにユーザにとって説得力があり、検証が容易であること

(透明性)である。単純性と透明性の観点から、結合ルール及びそれを用いた決定木は有力な手法である。

本研究で提案する確率的な非決定性決定木構造を用いた階層構造は、現行の判定システムにおいて主流になっている決定論的な決定ルールに比較して、強いルールの影響を縮小する方法を適用する。本研究で提案する医療画像データを数値データとして扱い階層化ルールを用いることで拘束力の弱いルールで判定を行う、即ち非決定性を持たせた柔軟な決定システムの構築を行った。

4. 研究成果

本研究の目的は医療データにおける現行の知識発見システムにおける現在の精度限界を打破するための可視化知識抽出モデルの提案、理論研究、システムの構築である。本研究で達成するシステムは、視覚的に判りやすく、かつ精度及び効率性の高いものであり、医療分野における意思決定支援という現代社会の必須技術として大きな実用インパクトを持ち現代社会に寄与できる。

本研究で対象とする医療データは、医療検査の数値データと画像データである。本研究では数値データと画像データを、離散アルゴリズムを用いた計算幾何学的な関数の最適近似の独自技法を利用し、データ分布を、ユーザに説明しやすい関数を用いて近似することで数値データベースからルールを抽出し、更にそれを組み立てて知識発見過程が医師にとって明確に理解できるシステムの構築を行うための理論確率を行った。これは、過去のデータマイニング技法に計算幾何学技法と幾何学的可視化を取り入れた画期的なシステムである。提案手法は計算幾何学手法を用いた医療画像データから自動的画像切り出しであり、切り出された画像を知識として提示することにより、医療画像データ(図2)解析の際に、知識として提案すると云った、医療画像データ知識抽出理論構築の成果を得た。

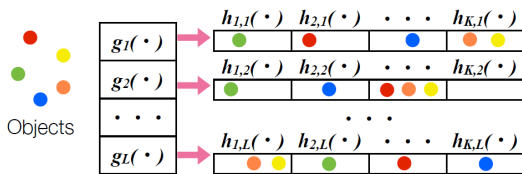


図1. Locality Sensitive Hash (LSH)

本研究で提案した医療知識発見システムの総合的実装と性能評価実験を行った。提案した計算アルゴリズム理論・学習アルゴリズム理論の確かさを明らかにする比較実験を行った。本研究の目的を達成するために下記の研究を実施した。データから知識獲得の概念を数理モデル化し、複数の数値属性に関し

ての高次元高速アルゴリズム設計と解析を行い、最適なデータ近似手法の計算理論を確率するために(1)医療画像データ診断システムの構築実験:1)提案した理論の実装を行い、計算機実験より画像データから画像診断を行う実験を実施し、提案システムの精度を調べた。2)ハッシュを使うことにより提案アルゴリズム高速化の改良を行った(図1)。(2)提案アルゴリズムの実装:1)実験データを使いシステムの性能評価を行い、提案手法の妥当性、有効性、精度を調べた。2)実装実験より既存手法と比較し、過学習の回避と予測制度の向上を実験より示した。



図2. 入力医療画像

| Grid y , Depth d | Room-edge Region | Output |
|-------------------------|------------------|--------|
| $y = 7,$ $d = 3$ | | |
| $y = 7,$ $d = 4$ | | |
| $y = 15,$ $d = 3$ | | |
| $y = 15,$ $d = 4$ | | |

図3. Quadtree decompositionを用いた heterogeneous components セグメンテーション結果

データから知識獲得の概念を数理モデル化し、複数の数値属性に関しての高次元高速アルゴリズム設計と解析を行い、最適なデータ近似手法の計算理論を確立するために(1)数値データに対する階層化ルール理論に関する研究として、階層化ルールを医療診断に用いるためのアルゴリズムの最適化を行い、階層化ルールの整備を行った。(2)計算幾何学手法を用いて医療画像データ(図2)から知識抽出理論研究として画像データを数値データの手法として幾何学的に扱い、計算幾何学を用いて巨大数値データベースを効率的に処理する新しいアルゴリズム設計を行った。(3)画像データを構造的な圧縮データに変換し、計算時間の高速化とマイニング精度を上げるアルゴリズム理論の設計を行い、システム構築・改良を行った。図3、は入力医療画像データ(図1)を入力とし Quadtree decomposition を用いた heterogeneous components セグメンテーション結果実験結果である。図4は k-baselines decomposition、 optimal k-baseline location、 and quadtree decomposition を用いたセグメンテーション結果である。

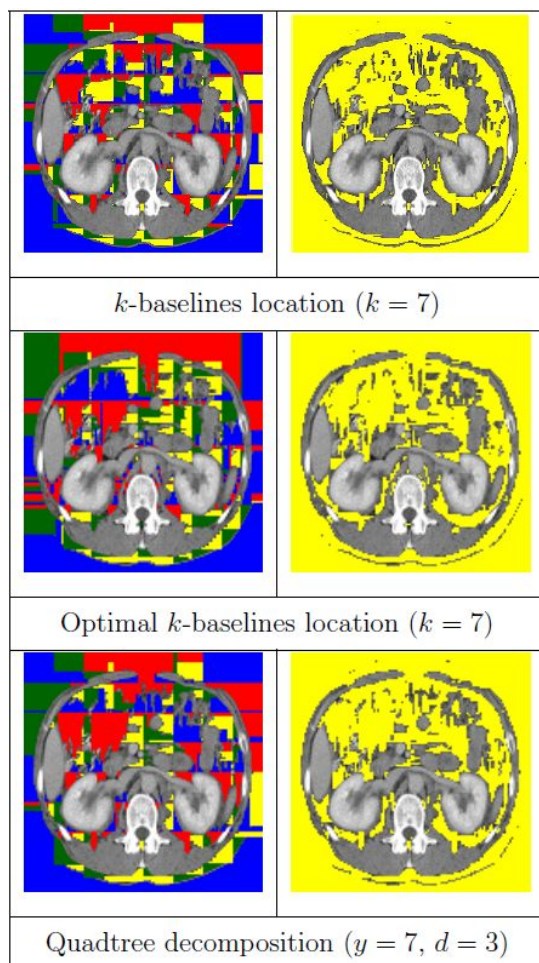


図4 . k-baselines decomposition、 optimal k-baseline location、 and quadtree decomposition を用いたセグメンテーション結果

本研究の主な結果として、医療画像データ認識の為に、画像切り出し問題を、最大重み領域問題としてアルゴリズムを与えた。基本領域の分割可能な和集合で表される領域族に対する最大重み領域問題について効率的なアルゴリズムを提案した。本論文における結果と交差を許した和集合領域の最適切り出し問題における NP 困難性を比較すると、計算量と幾何学的性質の関連の解明として興味深い成果と考えられ、学術的に興味深い結果[雑誌論文 1,7,9]を示した。

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9件)

1. Natsuda Kaothanthong, Jinhee Chun, Takeshi Tokuyama, Distance Interior Ratio: a new shape signature for 2D shape retrieval, Pattern Recognition Letters, Volume 78, pp.14-21(2016), DOI:10.1016/j.patrec.2016.03.029, 査読有
2. Jinhee Chun, Xiang Yu, Takeshi Tokuyama, A Reliable Top-k Dominating Query on Uncertain Big Database, Proceedings of the 18th Korea-Japan Joint Workshop on Algorithms and Computation, pp.59-64(2015), 査読有
3. Jinhee Chun, Xiang Yu, Takeshi Tokuyama, Reliable Top-k Dominating Query on Uncertain Big Data with x-Relation, Proceedings of the Young Researcher Workshop on Automata, Languages and Programming (YR-ICALP2015), pp.33-41(2015), 査読有
4. Jinhee Chun, Akiyoshi Shioura, Truong Minh Tien, Takeshi Tokuyama, A Unified View to Greedy Geometric Routing Algorithms in Ad Hoc Networks, IEICE Transactions, Vol.97-A, pp.1220-1230(2014), 査読有
5. Jinhee Chun, Takashi Horiyama, Takehiro Ito, Natsuda Kaothanthong, Hirotaka Ono, Yota Otachi, Takeshi Tokuyama, Ryuhei Uehara, Takeaki Uno, Base-object location problems for base-monotone regions, Theoretical Computer Science, Vol.555, pp.71-84(2014), 査読有, DOI: 10.1016/j.tcs.2013.11.030
6. Jinhee Chun, Ricardo Gonzalo and Takeshi Tokuyama, Space-efficient and data-sensitive polygon reconstruction algorithms from visibility angle information, Proceedings of the 24th International Symposium on Algorithms and Computation, Vol. 8283, pp.

- 22-32(2013) , 査読有,
DOI: 10.1007/978-3-642-45030-3_3
7. Jinhee Chun, Takashi Horiyama, Takehiro Ito, Natsuda Kaothanthong, Hiroataka Ono, Yota Otachi, Takeshi Tokuyama, Ryuhei Uehara, and Takeaki Uno, Base-Object Location Problems for Base-Monotone, Theoretical Computer Science, Vol. 555, pp. 71-84(2014) , 査読有,
DOI:10.1016/j.tcs.2013.11.030
 8. Tran Phuong Nhung, Cam-Tu Nguyen, Jinhee Chun, Ha Vu Le and Takeshi Tokuyama, A Multiple Instance Learning Approach to Image Annotation with Saliency Map, Proceedings of the 5th International Conference on Knowledge Discovery and Information Retrieval, pp.152-159(2013) , 査読有
 9. Jinhee Chun, Natsuda Kaothanthong, Takeshi Tokuyama, Classified-Distance Based Shape Descriptor for Application to Image Retrieval, Proceedings of the 15th International Conference on Computer Analysis of Image and Patterns, Vol.2, pp.1-8(2013) , 査読有

〔学会発表〕(計8件)

1. Jinhee Chun, Xiang Yu, Takeshi Tokuyama, Efficient Top-k Dominating Query on Uncertain Database, 第14回 FIT 情報科学技術フォーラム 2015, 2015年9月15日, 愛媛
2. Jinhee Chun, Xiang Yu, Takeshi Tokuyama, A Reliable Top-k Dominating Query on Uncertain Big Database, The 18th Korea-Japan Joint Workshop on Algorithms and Computation, 2015年8月21日, Incheon, Korea
3. Jinhee Chun, Xiang Yu, Takeshi Tokuyama, Reliable Top-k Dominating Query on Uncertain Big Data with x-Relation, Young Researcher Workshop on Automata, Languages and Programming, 2015年7月5日, Kyoto, Japan
4. Jinhee Chun, Ricardo Gonzalo and Takeshi Tokuyama, Space-efficient and data-sensitive polygon reconstruction algorithms from visibility angle information, The 24th International Symposium on Algorithms and Computation(ISAAC2013), 2013年12月16日, Hongkong, China
5. Tran Phuong Nhung, Cam-Tu Nguyen, Jinhee Chun, Ha Vu Le and Takeshi Tokuyama: A Multiple Instance Learning Approach to Image Annotation with Saliency Map, The 5th International Conference on Knowledge Discovery and

- Information Retrieval (KDIR 2013), 2013年9月21日, Vilamoura, Portugal
6. Jinhee Chun, Natsuda Kaothanthong, Takeshi Tokuyama, Classified-Distance Based Shape Descriptor for Application to Image Retrieval, The 15th International Conference on Computer Analysis of Image and Patterns(CAIP2013), 2013年8月23日, York, United Kingdom
 7. Jinhee Chun, Natsuda Kaothanthong and Takeshi Tokuyama, Correspondence Finder using Classified Distance Distribution for Efficient Shape Retrieval, The 6th Korea-Japan Joint Workshop on Algorithms and Computation (WAAC 2013), 2013年7月12日, Tokyo, Japan
 8. Jinhee Chun, Natsuda Kaothanthong and Takeshi Tokuyama, Shape Description using Classified Distances, The 6th Annual Meeting of Asian Association for Algorithms and Computation (AAAC2013), 2013年4月20日, Matsushima, Japan

〔図書〕(計0件)

〔産業財産権〕
出願状況(計0件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
国内外の別:

取得状況(計0件)

名称:
発明者:
権利者:
種類:
番号:
取得年月日:
国内外の別:

〔その他〕
ホームページ等

6. 研究組織

(1)研究代表者

全 眞嬉 (CHUN JINHEE)

東北大学・大学院情報科学研究科・准教授
研究者番号: 8 0 4 3 1 5 5 0

(2)研究分担者

()

研究者番号：

(3)連携研究者 ()

研究者番号：