T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

題目(和文)	頑健かつオンライン学習可能なノンパラメトリック密度推定法
Title(English)	
著者(和文)	中村圭宏
Author(English)	Yoshihiro Nakamura
出典(和文)	学位:博士（工学）, 学位授与機関:東京工業大学, 報告番号:甲第10230号, 授与年月日:2016年3月26日, 学位の種別:課程博士, 審査員:長谷川 修,渡邊 澄夫,樺島 祥介,新田 克己,小野 功
Citation(English)	Degree:, Conferring organization: Tokyo Institute of Technology, Report number:甲第10230号, Conferred date:2016/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

Powered by T2R2 (Science Tokyo Research Repository)

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻： Department of	知能システム科学　　専攻	申請学位（専攻分野）： Academic Degree Requested	博士　（　工学　） Doctor of
学生氏名： Student's Name	中村　圭宏	指導教員（主）： Academic Advisor(main)	長谷川　修
		指導教員（副）： Academic Advisor(sub)	

要旨（和文 2000 字程度）
Thesis Summary（approx.2000 Japanese Characters ）

本論文は「頑健かつオンライン学習可能なノンパラメトリック密度推定法」と題して，全 6 章から構成される．

第 1 章では，本論文の背景と目的を述べている．近年，膨大なデータが生成されるようになってきているが，データの生成量の劇的な増加の要因の 1 つとして，実環境を継続的にセンシングするセンサーの数が増加し続けていることが挙げられる．このような継続的に生成され続けるデータはストリームデータと呼ばれている．また，今後データがより膨大に生成されるようになれば，よりデータ駆動な知的処理が重要となってくると考えられる．何故なら，全てのデータを人間が解析するのは困難であり，解析を自動化することが必要になると考えられるからである．このような背景から，データストリームから生成される大規模データ活用するための知的情報処理技術は，高速なオンライン学習，高いロバスト性，ノンパラメトリックな手法といった特徴を有する必要があると考えられる．本論文では，これら全ての性質を有する手法を提案することを目的とする．提案手法は，競合学習型ニューラルネットワークの 1 種である自己増殖型ニューラルネットワーク（SOINN）とノンパラメトリック密度推定法であるカーネル密度推定法を拡張することで，上記の 3 つの性質を実現する．

第 2 章では，代表的なノンパラメトリック確率密度推定法であるカーネル密度推定法について述べている．カーネル密度推定法は提案手法の基礎となる手法である．また，カーネル密度推定法にオンライン学習，ロバスト性を導入した最近の既存手法についても紹介している．

第 3 章では，提案手法の基礎となる手法である SOINN について，その学習原理と，確率密度推定に応用できる統計的な性質に関して述べる．SOINN は競合学習を用いたプロトタイプベースのクラスタリング手法であり，ネットワーク構造を用いて入力サンプル集合を学習する．ネットワークは同一クラスタに属するかどうかの基準である．しかし，本論文においては，SOINN のネットワークはデータの分布に関する情報を保持していると考えている．SOINN ではネットワークの形成に Competitive Hebbian Learning を用いている．これによって形成されるネットワークは，入力サンプルがある多様体に沿って分布している場合，多様体に沿った部分にのみエッジが形成される．また，各ノードは競合学習において勝ち取ったサンプルの平均に位置している．つまり，ノードは周辺のサンプルを代表するプロトタイプであり，そのノードの周辺ネットワークはノードの周辺サンプルの広がりの大きさとその方向を表し，ネットワーク全体で分布の形状を近似していると考えられる．

第 4 章では，前章で述べた SOINN のネットワークの統計的な性質を利用し，カーネル密度推定法を拡張したノンパラメトリック密度推定法を提案している．提案手法では，まず，サンプルのプロトタイプがノードであるネットワークとして，サンプルをオンラインに学習する．そして，ネットワークの各ノードの位置にガウスカーネルを配置して，その線形和としてサンプル集合の確率密度を推定する．ガウスカーネルの形状を決定する共分散行列は各ノードの周辺のネットワーク構造から適応的に決定される．この共分散行列を局所ネットワーク共分散行列と定義する．評価実験において，ロバスト性と学習時間に関して，提案手法は既存手法より高い性能を示した．

第 5 章では，学習アルゴリズムを改良することで，より密度推定に適したネットワーク構造を形成するようにし，性能向上を試みている．SOINN において最も重要な部分である閾値領域に関して考察し，より不要なエッジが形成されないように改良している．SOINN の閾値領域は各ノードを中心とする超球であり，各ノード周辺のサンプルは等方的ガウス分布に従うと仮定している．これは Competitive Hebbian Learning によって形成されたネットワークと矛盾しており，密度を表さないエッジが形成されてしまう．そこで，閾値領域を局所ネットワーク共分散行列が規定するガウス分布の形状になるようにすることで，分布を表さないネットワークの形成を抑えるよう改良している．評価実験によって，ロバスト性，高次元データに対する密度推定，実データに対する推定精度に関して大きく性能が向上していることを確認できる．

第 6 章では，本論文をまとめ，課題および今後の展望について述べている．

以上のように，本論文では，実環境のデータストリームから生成される大規模データ活用するための知的情報処理技術として，頑健かつオンライン学習可能なノンパラメトリック密度推定法を提案した．
ノイズを含む実環境からのストリームデータの分布を推定できるようになれば，確率的な予測や推論がより実際的なアプリケーションとして構築でき，その応用範囲は広いと考えている．

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。
Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800

Words (English).

注意：論文要旨は、東工大リサーチリポジトリ（T2R2）にてインターネット公表されますので、公表可能な範囲の内容で作成してください。
Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻： Department of	知能システム科学　　専攻	申請学位（専攻分野）： Academic Degree Requested	博士　（　工学　） Doctor of
学生氏名： Student's Name	中村　圭宏	指導教員（主）： Academic Advisor(main)	長谷川　修
		指導教員（副）： Academic Advisor(sub)	

要旨（英文 300 語程度）
Thesis Summary（approx.300 English Words ）

This dissertation proposes the robust fast online non-parametric density estimation as a basis of intelligent information processing to handle data streams from the real environment. Here is the summary of the document.

Chapter 1 presents the background and objective of this research. Recently, massive amount of data is generated. One of the causes of it is the increasing number of sensors that are sensing real environment continuously. Data generated continuously is called stream data. As more massively data will be generated, it becomes more difficult to analyze all data by experts and the automation of data analysis will be necessary. So, the intelligent information processing to handle data stream from real environment have to have these characteristics: fast online learning, high robustness, and non-parametric approach.

Chapter 2 introduces a typical approach to non-parametric density estimation called Kernel Density Estimation (KDE), which is a basis of our proposed method.

Chapter 3 describes Self-Organizing Incremental Neural Network (SOINN), which is another basis of our proposed method. SOINN is a prototype-based clustering method that learns samples as networks. We determined that the network structure has information regarding the distribution of underlying observed samples.

Chapter 4 presents our proposed method, which is a non-parametric density estimator based on the SOINN's statistical characteristics presented in the previous chapter. It learns samples as networks whose nodes are prototypes of the samples in an online manner. Then it determines the shape and size of each kernel at each node on the basis of the local structure of the network around the node and estimates the probability density function of the samples as a summation of the kernels. In experiments, our proposed method outperforms existing methods from the aspect of robustness and training time.

Chapter 5 presents improvements of our proposed method to improve its performance. Considering threshold regions, which is the most important part of SOINN, improves its algorithm not to create inappropriate edges. In experiments, the improvement brings better performances in terms of robustness, high dimensionality and estimation accuracy.

Chapter 6 concludes and summarizes this dissertation, and shows the future directions of this research.

備考 ： 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。
Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ（T2R2）にてインターネット公表されますので、公表可能な範囲の内容で作成してください。
Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).