

Course number		G-LAS12 80014 LJ11					
Course title (and course title in English)	ビッグデータの計算科学 Computational Science for Big Data			Instructor's name, job title, and department of affiliation	Graduate School of Informatics Professor,YAMASHITA NOBUO Graduate School of Informatics Program-Specific Associate Professor,SATO HIROYUKI Institute for Liberal Arts and Sciences Program-Specific Senior Lecturer,SEKIDO HIROTO Academic Center for Computing and Media Studies Professor,KOYAMADA KOUJI Academic Center for Computing and Media Studies Program-Specific Senior Lecturer,NATSUKAWA HIROAKI		
Group	Interdisciplinary Graduate Courses			Field(Classification)	Statistics, Informatics and Data Science		
Language of instruction	Japanese			Old group		Number of credits	2
Number of weekly time blocks	1	Class style	Lecture (Face-to-face course)		Year/semesters	2024・Second semester	
Days and periods	Tue.3		Target year	Graduate students		Eligible students	For science students
(Students of Graduate School of Informatics cannot take this course as liberal arts and general education course. Please register the course with your department.)							
[Overview and purpose of the course]							
<p>近年のコンピュータの進歩や情報基盤技術の整備に伴って、クラウドコンピューティングなどのインターネットを介して行われる社会活動から生成されるデータの量、あるいは、計算科学の重要な技法であるコンピュータシミュレーションを通じて得られるデータの量は、日々増加の一途をたどっている。それらのビッグデータを分析、可視化するための手法を学ぶことが、この科目の目的である。</p> <p>大次元疎行列は、隣接行列と解釈することで大規模な有向グラフを表現することができ、多様な分析対象を表現することが可能である。その行列の特徴量、すなわち、分析対象の特徴量を抽出する際に、最も一般的でかつ普遍的な手法は、固有値分解、もしくは、特異値分解を行うことである。そこで、データ解析手法について、多変量解析の基礎である最小二乗法と主成分分析からはじめ、グラフのスペクトラルクラスタリングや行列の欠損値推定のためのEMアルゴリズムなどの固有値分解や特異値分解を用いて行う様々なデータ解析手法について教授する。</p> <p>また、データ解析手法を実際に適用する際には最適化問題が頻出であり、たとえば、最小二乗法・主成分分析・スペクトラルクラスタリング・行列の欠損値推定はいずれも最適化問題として定式化される。こうした最適化問題は線形代数に基づく計算を用いて解ける場合もあるが、一般的には最適化問題を解くためのアルゴリズムが必要となる。たとえば、行列の欠損値の推定は、小規模密行列の場合は特異値分解によって達成できるが、大規模疎行列の場合は特異値分解では時間がかかりすぎるため実用的ではない。よって、この講義では大規模疎行列の欠損値を推定するための最適化アルゴリズムを題材として、ビッグデータに対する最適化アルゴリズムを解説する。</p> <p>【大学院横断型教育の概要・目的】 「ビッグデータの可視化」、「固有値・特異値分解を用いたデータ解析手法」、「ビッグデータのための最適化手法」を通じて、大規模データ(ビッグデータ)から重要な情報を取り出す分析能力、可視化する能力を身につけることを目指す科目である。大規模データ(ビッグデータ)を分析する能力は、研究分野を問わず重要である。</p>							

Continue to ビッグデータの計算科学(2)							

ビッグデータの計算科学(2)

[Course objectives]

ビッグデータが、重み付き有向グラフや大次元疎行列の形式で与えられたとき、それらの解析手法を理解する。特に、特異値分解を利用したグラフのカットを行う技法を理解する。さらに、基本的な統計解析手法である最小二乗法、主成分分析の内容を理解する。加えて、ビッグデータに対する最適化手法について理解する。

[Course schedule and contents]

・ガイダンス(1回)

・ビッグデータの可視化 (3回)

ビッグデータを視覚的に理解するための技法について解説する

・データ行列の特異値分解と固有値分解、特異値分解を用いるデータ解析手法 (6回)

特異値分解の定義をはじめとする線形代数の基礎を講義する。また、基本的なデータ解析手法である最小二乗法、および、特異値分解を用いてデータ分析する上で基本的な考え方となる主成分分析について解説する。その後、グラフのスペクトラルクラスタリングなどの固有値分解や特異値分解を用いて行う様々なデータ解析手法について概説する。

・ビッグデータのための最適化手法 (5回)

ビッグデータを解析する際に現れる大規模な最適化問題へのアプローチを習得するために、最適化アルゴリズムの基礎的な事柄からはじめ、スパース推定のためのLasso回帰において現れる最適化問題や大規模疎行列の欠損値の推定問題を解説する。

[Course requirements]

None

[Evaluation methods and policy]

レポート試験の成績(75%)・平常点評価(25%)

「ビッグデータの可視化」、「固有値・特異値分解を用いたデータ解析手法」、「ビッグデータのための最適化手法」について、それぞれ1つずつのレポート課題を出題します(それぞれ配点25点)

。平常点評価には、出席状況と質問など通した授業への積極的な参加を評価します。

[Textbooks]

講義資料を配布する。
教科書は特に定めない。

[References, etc.]

(References, etc.)

小山田耕二、坂本尚久 『粒子ボリュームレンダリング-理論とプログラミング』(コロナ社) ISBN: 978-4-339-02449-4 (See http://www.coronasha.co.jp/np/detail.do?goods_id=2726)

[Study outside of class (preparation and review)]

統計に重要な線形代数の知識は、授業内でも解説を行うが、予習あるいは復習することを期待する。さらに、統計の基礎知識、特に、主成分分析などの知識を予習あるいは復習し、受講されることを期待する。

Continue to ビッグデータの計算科学(3)

ビッグデータの計算科学(3)

[Other information (office hours, etc.)]

オフィスアワーについては担当教員の KULASIS 登録情報を参照すること。

佐藤寛之: hsato@amp.i.kyoto-u.ac.jp

關戸啓人: sekido@amp.i.kyoto-u.ac.jp

授業時間外で質問がある場合には、あらかじめ、上記のアドレスにメールをすること。