

授業科目名 <英訳>	ビッグデータの計算科学 Computational Science for Big Data	担当者所属・ 職名・氏名	情報学研究科 教授 中村 佳正 情報学研究科 特定准教授 木村 欣司 情報学研究科 助教 関戸 啓人 学術情報メディアセンター 教授 小山田 耕二
---------------	---	-----------------	--

配当 学年	1回生以上	単位数	2	開講年度・ 開講期	2017・ 後期	曜時限	水5	授業 形態	講義	使用 言語	日本語
----------	-------	-----	---	--------------	-------------	-----	----	----------	----	----------	-----

授業種別 プロジェクト科目

[授業の概要・目的]

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。特に、C言語を利用して、大次元の疎行列に対するデータ分析の演習を行う。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、最も一般的かつ普遍的な手法は、特異値分解を行うことである。それ以外にも、特異値分解は、解析したいデータがはじめから表や行列として表現されている問題への幅広い応用も可能で、最小2乗法、主成分分析といった多変量解析にもよく用いられる。そこで、本科目は、受講者が特異値分解をおこなうプログラムをソースコードのレベルから作成することにより、大規模データを分析するための基本的な技術を習得することを目的とする。ソースコードのレベルからプログラムを作成することは、プログラミング技術を習得することにもつながる。本科目では、C言語の基本文法などの基礎的な話題から演習を開始する。よって、過去にC言語を学んだことのない学生の受講も歓迎する。

Because of the recent progress in computers and information infrastructure technology, large-scale data are generated from the social activity performed through the Internet such as cloud computing and obtained through the computer simulation which is an important technique of computational science, and the increase of the quantity of the data becomes bigger and bigger every day. It is the purpose of this course to study the technique for analyzing and visualizing those big data. In particular, a C language program for the data analysis to the large sparse matrix is written as an exercise.

A large sparse matrix has the capability to express a weighted directed graph through the adjacency matrix of a graph. Thus, it is possible to express various objects for analysis. When extracting the feature quantity of the matrix, that is the object for analysis, the most general and universal technique is performing a singular value decomposition. Besides, a singular value decomposition is also applicable to the problem in which data are expressed in the term of a table or a matrix originally. Thus, it is often used for multivariate statistics such as a least squares method, principal component analysis. The aims of this course are mastering the fundamental technology for analyzing large-scale data by writing the program code for a singular value decomposition. Writing a program code leads also to mastering a programming technique. In this course, an exercise is started from learning fundamental subjects, such as a basic statement of the C language. Thus, the students who had not studied the C language in the past are also welcomed.

ビッグデータの計算科学(2)へ続く

ビッグデータの計算科学(2)

【到達目標】

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2乗法、主成分分析の内容を理解する。加えて、大次元疎行列に対する主成分分析を行うプログラムを作成できる程度のC言語のプログラミング能力を習得する。

【授業計画と内容】

- ガイダンス(木村欣司/1回 講義)
計算科学は、数学的モデルとその定量的評価法を構築し、計算機を駆使して科学技術上の問題を解決する学問分野である。計算科学概論、計算科学の応用について講述する
クラウドコンピューティング入門(關戸啓人/1回 講義)
クラウドコンピューティングの基本的な話題について解説を行う
ビッグデータの可視化(小山田耕二/3回 講義)
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について(中村佳正、關戸啓人/4回 講義)
- (1)線形代数入門
 - (2)大次元疎行列(大次元隣接行列)と重み付き有向グラフの関係についての解説
 - (3)行列計算を利用した重み付き有向グラフの解析
 - (4)データを分析するための統計的手法についての解説
- I. 最小2乗法
II. 主成分分析
大次元疎行列の特異値分解法(木村欣司/6回 講義と演習)
- (1)C言語の基本的な文法などを解説
 - (2)C言語を用いた、複数の特異ベクトルを求めるための直交化付きべき乗法の実装

All schedules are as follows.

- Guidance (Kinji Kimura/1 time Lecture)
Introduction to cloud computing (Hiroto Sekido/1 time Lecture)
Visualization of big data (Koji Koyamada/3 times Lecture)
Singular value decomposition for data matrices (Yoshimasa Nakamura, Hiroto Sekido /4 times Lecture)
- (1)Introduction to linear algebra
 - (2)A relationship between sparse matrices of large scale and weighted directed graphs
 - (3)Analysis of weighted directed graphs using matrix computations
 - (4)Statistical approaches for analyzing date: I. Least squares method II. Principal component analysis
Singular value decomposition for large sparse matrices (Kinji Kimura/6 times Lecture and Exercise)
- (1) On C programming language
 - (2) Implementation of the power method with orthogonalization by using C programming language

【履修要件】

特になし

【成績評価の方法・観点及び達成度】

レポート試験の成績(80%) 平常点評価(20%)
「ビッグデータの可視化」(配点25点)、「密行列の特異値分解法」(配点25点)、「大次元疎行列係数の特異値分解法」(配点30点)で、それぞれ1つずつのレポート課題を出題します。

ビッグデータの計算科学(3)へ続く

ビッグデータの計算科学(3)

「大次元疎行列係数の特異値分解法」は、プログラムを作成することを課題とするレポートであり、独自の工夫がみられるものについては、高い点を与えます。

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。

The understanding level of the each content of the lecture about "Visualization of big data", "Singular value decomposition for dense matrices", and "Singular value decomposition for large sparse matrices" is evaluated by a report, respectively.

"Visualization of big data": it is worth 25 points.

"Singular value decomposition for dense matrices": it is worth 25 points.

"Singular value decomposition for large sparse matrices": it is worth 30 points.

Attendance and active participation through questions is worth 20 points.

[教科書]

講義資料を配布
特に定めない

Handouts to be distributed

Not specified

[参考書等]

(参考書)

小山田耕二, 坂本尚久 『粒子ポ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: ISBN:978-4-339-02449-4 (See http://www.coronasha.co.jp/np/detail.do?goods_id=2726)

[授業外学習(予習・復習)等]

統計に重要な数値線形代数の知識は、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。プログラミング言語Cについては、授業時間内のみでは完全な習得が困難であるため、予習と復習を授業と並行して行うことを期待する。

Please do preparation and review about basic knowledge of linear algebra and statistics including principal component analysis. In addition, it is expected that students learn programming language C in preparation and review.

(その他(オフィスアワー等))

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。

木村欣司:kkimur@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。

See KULASIS data for office-hour information.

Kinji Kimura : kkimur@amp.i.kyoto-u.ac.jp

Hiroto Sekido : sekido@amp.i.kyoto-u.ac.jp

Please send mail to the above-mentioned address to contact outside the class time.

オフィスアワーの詳細については、KULASISで確認してください。