

平成29年度研究科横断型教育プログラム（Aタイプ）授業科目

開講方式	Aタイプ (研究科 開講型)	研究科名	情報学研究科	カテゴリー	情報活用・計算科学 科目群	横断 区分	理系横断型				
授業科目名 (英訳)		ビッグデータの計算科学 (Computational Science for Big Data)		講義担当者 所属・職名・ 氏名	情報学研究科・ 特定准教授・木村欣司 情報学研究科・ 特定助教・關戸啓人 学術情報メディアセンター・ 教授・小山田耕二 情報学研究科・教授・ 中村佳正	開講 場所	吉田南 学術情報メ ディアセン ター南館 203				
配当 学年	修士 博士後期 専門職	単位 数	2単位	開講期	後期	曜時限	水5限 (16:30-18:00)	授業 形態	講義 (演習 を含 む)	使用 言語	日本語
【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、最も一般的で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。											
【研究科横断型教育の概要・目的】											
「ビッグデータの可視化」、「密行列の特異値分解」、「疎行列の特異値分解」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。											
【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)											
(3) 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											

<p>〔履修要件〕</p> <p>特に定めない</p>
<p>〔成績評価の方法・観点及び達成度〕</p> <p>レポート試験の成績(80%) 平常点評価(20%)</p> <p>「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。</p> <p>「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。</p> <p>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。</p>
<p>〔教科書〕</p> <p>講義資料を配布する。</p> <p>教科書は特に定めない。</p>
<p>〔参考書等〕</p> <p>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)</p>
<p>〔授業外学修(予習・復習)等〕</p> <p>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。</p>
<p>〔その他(授業外学習の指示・オフィスアワー等)〕</p> <p>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。</p> <p>木村欣司:kkimur@amp.i.kyoto-u.ac.jp</p> <p>關戸啓人:sekido@amp.i.kyoto-u.ac.jp</p> <p>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。</p>