

数理・データサイエンス教育強化拠点コンソーシアム

NEWSLETTER

ニュースレター

vol. 4
2019/4



数理・データサイエンス教育強化を目的として
国立大学に設置されたセンターが結集して、
各大学内での数理・データサイエンス教育の
充実のための取組成果を
全国への波及させるための活動を推進し、
数理・統計・情報を基盤として
未来世界を開拓できる人材の育成を目指します。

数理・データサイエンス教育が 未来社会を開く。

| contents |

●「数理・データサイエンスと大学」インタビュー

第5回 大阪大学大学院 教授、基礎工学研究科 研究科長
大阪大学 数理・データ科学教育研究センター センター長

狩野 裕 教授

●教材分科会活動報告

・カリキュラム分科会 ・教材分科会 ・教育用データベース分科会

●インフォメーション

「数理・データサイエンスと大学」

インタビュー

第5回 大阪大学大学院 教授、基礎工学研究科 研究科長
大阪大学 数理・データ科学教育研究センター センター長

狩野 裕 教授

工夫を凝らした例題を共有化し

新入生や文系学生にも興味を持てる

魅力的な教材の開発を

文系を含むすべての学生に、大学4年間でデータ・リテラシーを身に着けさせるために、初心者向けの教材開発が急ぎ進められている。長い間、文系学生に独自の教材を使って分かりやすくデータ科学を教えてきた大阪大学の狩野裕教授に、初心者向け教材を開発した際の体験と、そこから導かれる要点を聞いた。



数式を使わずに統計を教える 文系学生を振り向かせる挑戦

——長年にわたり大学でデータ科学教育に携わってきて、現状にどのような課題を感じていますか。

痛感しているのは、「魅力的な教材」を用意する難しさです。データ科学になじみのない学部低学年では、まず身近な現象・データをきちんと咀嚼して見ることでできる力をつけるところからスタートし、そこから専門基礎としてのデータ科学へ入っていくことが重要だと、私は考えています。

しかし、「身近」で、しかも「インパクトのある」例を持ってくることが意外と大変です。難しい例を使って「難しいやろ？」と言うのは簡単ですが、「四則演算レベルで解けるけれども、データの見方を分からなかったら間違うんやで」と示すことができて、かつ、学生が「何でそうなるん？」と興味を持つような例を探すのは結構大変なのです。

でも、大学の教員はあまりそういうことに関心がありません(笑)。だからこれまで、そのようなテキストもほとんどなかったのです。とはいえ、どの先生も「1回目の授業で使うために」など、一つ二つは興味を惹く例題を用

意しておられるでしょう。そういう例題を集めてブラッシュアップし、1セットの教材に作り上げ、それを共有することが大事ではないかと思います。

——初心者向けのテキストの必要性を感じたのは、何かきっかけがあったのでしょうか。

97年に母校の大阪大学に助教授として戻ってきたとき、人間科学部という文系の学部に着任したのです。私自身、理学部の数学科出身ですし、それまでずっと理系の学生を教えてきましたが、だんだんと「同じやり方では文系の学生には響かない」ということが分かってきました。

彼らにこちらの思いを伝え、彼ら自身にプラスになるような授業にするにはどうしたらいいか——。日々考えるなかで、まず目標にしてみたのは、「数式を使わずに統計を教える」という試みです。文系の学生たちは、「統計は数学や。数学はキライ」と思っているわけです。だから、数式を使わないほうがアレルギーを起こさずにすむ。

もっとも、魅力的なテキストが必要であるという点は、文系・理系に共通した課題です。一般に、理工系の学生にものごとを教えるときは、統計を使うよりも、それぞれの分野における第一原理から演繹して説明するほうが理解しやすい。統計というのは、どうもよく分からんところを「データを取って傾向を見てみましょう」ということな

ので、理工系の学生はそういう曖昧さに違和感を持ちやすいようです。

卒業式直後のアンケート 「統計学、受けてよかった」

——学部低学年を対象とする講義では、実際にどんな話をするのですか。

私が最初の授業で学生に話すのは、こんな例です。大阪大学の人間科学部は、社会科学系の学部としては珍しく、数学と統計学が必修です。たいていの学生はそのことを知らず、入学してからびっくりする。そこで私は、卒業生の約9割が「数学・統計学の必修があってよかった」と答えている、というアンケート結果のデータを見せるのです。「今、君らは『嫌や』と言うてるけど、4年経ったら変わるよ」と。そして、「これについてどう思いますか?」と問うわけです。

実は、この調査は卒業式の直後に行ったもの。それは当然で、大学サイドは学業を全て終えた後の満足度等を知りたいのだが、大学を離れた後はデータ採取が大変だから卒業式の直後で調査となる。人間誰でも山を越えてきて、1年生の数学・統計から始まって卒論も無事に終えたら、どんなことでもいい思い出になるに決まっているんですよ、それは(笑)。だから本当は、少し間引いて考えなくてはいけない。この「間引いて考える」ことに気づくことも重要なデータの見方なのです。

もちろん、卒業のときに「数学・統計を受講してよかった」と思う学生がたくさんいることは事実。だから君たちもがんばれ、ということを新入生に伝えたいという思いが一番なのですが、同時に、「誰を対象として、どういう状況のとき、どんな質問の仕方で収集したデータなのか」をきちんと見ないと、データから正確な情報は得られない。データ採取の状況を想起しながらデータを適切に見ることができるようになってほしい」というメッセージも込めているのです。

——身近な例を使って、学生に「これまでのデータの見方は間違っていた」と気づかせるわけですね。

そうです。まずそこを認識してもらって、「では、正しく見るためにはどんな方法があるのか」という話につなげます。この例で言えば、「①在学中」と「②卒業式を終えてホッとしているとき」、「③卒業して5年後」では、同じ人でも回答が変わるかもしれないという前提でデータを観る。満足度がどのように変化していくのかを調べるためには、一つの個体(一人の学

生)を①②③と順に追っかけてデータを取っていく必要があります。これを経時データまたは縦断的データと言います。これはもう、教育の題材とか演習問題のレベルではなく研究活動の範囲ですが。

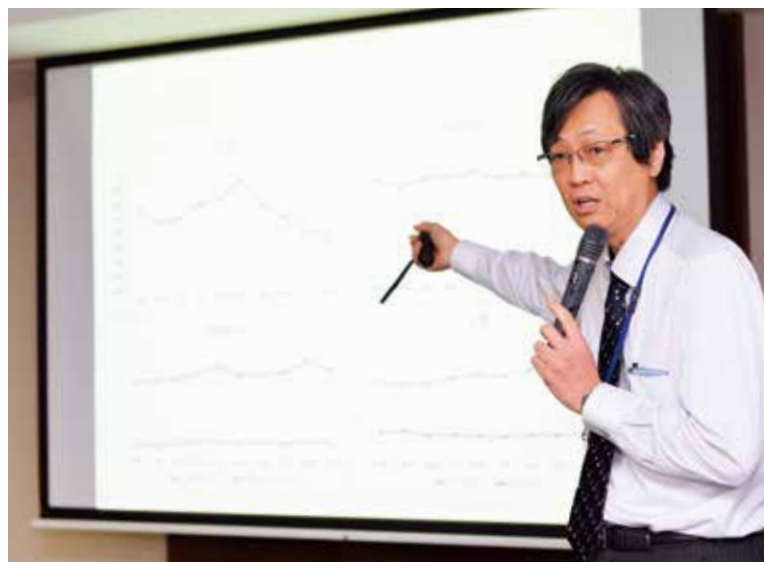
多くの学生(一般の方も)はアンケートに回答した経験があると思います。あんなに適当に回答したデータに基づいていったい何が分るんだ、と感じている方も多いのではないのでしょうか。ある日の午前に回答した内容は同日午後にはもう忘れていくような適当さです。しかし、回答の揺れ(不確実性)を誤差の確率変動とモデル化し、その背後に存する本質を「誤差を排して」捉えていく、そのための道具が統計モデルや検定・推定です。確率的な現象を正しく把握するには統計的なアプローチが必要だということが分かったら、次は勝ち負けなら「二項分布」、身長データなら「正規分布」などと、どんな現象がどんな分布をしているかということを説明していく。数学から入るのではなく、身近なところから入って、最後に数学を持ってくるようにしています。数学嫌いの学生には「なんで統計なんかやらないかんのや」という反発の気持ちがありますから、それをほぐしてやるのが大事だと思うんです。

大学で文系を選んだとしても、高校入試を真面目に取り組んだ学生なら、数学はけっこうできる。中学・高校程度の数学の知識が身につけていれば、かなりいろいろなデータ分析が使えます。その経験から高等数学の必要性を感じたらその段階で勉強すればよい。自分からシャットアウトしてその知識を使わないのでは、その学生は先の人生、ずいぶん損をすることになる。

データと分析のコンテキストを セットにした教材が理想

——6大学コンソーシアムでは、データサイエンス入門教科書シリーズの編集が進んでいます。

そうですね。ただ、私から見ると、まだ従来の延長上を





脱していない気がします。もっと皆さんが授業で実践しているひと工夫、ふた工夫を一挙に集めて教材に整え、例題のデータベースを構築するのがいいと思うのですけどね。みなさん、「こんなものを出すのは恥ずかしい」と謙遜されているのかもしれない。

というのは、ひとつの統計的な手法を説明するために挙げた例は、厳密に言えば正しくないとか、他の手法で解くと別の答えになるかもしれない、ということがあるからです。初心者向けにシンプルに説明しようとすればするほど、別の角度から見ると違う、ということが起りがちで、突っ込みどころはいくらでもある。それを突っ込まれないように書くから面白味がなくなるわけですが、まあこれは大学教員の習性かもしれません。

—— 初心者向けの教材として最も重要なことは何でしょうか。

大事なのは、コンテキストです。「このデータは、最初是这样見えたが、分析していったら最後はこんなドンデン返しが起こった！」となったら、面白いじゃないですか(笑)。学生の興味を引きつけながら、こうやってデータを見ていくんですよ、と教えることができる。

しかし、これまで統計学のトレーニング用の教材は、「こんなデータがあります。どうぞ分析してください」と、ただデータだけが置いてある状態で、分析の仕方が少しでも説明してあればいいほうでした。これでは、食材だけを示して料理の手順が書いていないレシピのようなものです。

本来ならば、そこからいろいろなストーリーが考えられるはず。1セットのデータがあれば、例えばPower Pointで20シート分ぐらいは、分析する観点があるのです。そういう、素材と料理の仕方を一緒にしたレシピを100種類ぐらい用意する。教員は、たとえば「この部分はこう変えて、後半は省こう」など、自由にアレンジできるようにすればいいと思います。

—— 狩野先生自身は、これまでどうやって「コンテキストのある教材」を作っていましたか。

実際には、そう簡単にコンテキストの描けるデータと分析が見つかるわけではありません。私の場合ラッキーだったのは、人間科学部で7年間教えていたときに、学生の卒業論文の審査などで、さまざまな実社会の研究テーマに接する機会を得たことです。データ科学の授業で使えそうだった論文やデータをもらってきて、自分でアレンジを加えました。

でも、実はここからが大変なのです。コンテキストの面白さを残したまま、教えた統計学のポイントを強調しつつ、しかも90分の授業で説明できるシンプルなシチュエーションに凝縮して、教材として仕上げなければいけません。これはものすごく手間のかかる作業です。だからわれわれはつい、シンプルすぎてコンテキストのない無味乾燥の例題を使ってしまいがちなのです。

しかし、データ科学とはそもそも、現実社会の課題に対し、統計学的手法を使って分析とフィードバックのPDCAサイクルを回す、その全体の制度設計そのものです。現実とかけ離れた例題では、そのダイナミズムが学生に伝わるはずがありません。教材づくりは大変ですが、私は「教材とは、最初からあるものではなく、努力して作り上げるものだ」と今では思っています。



Profile



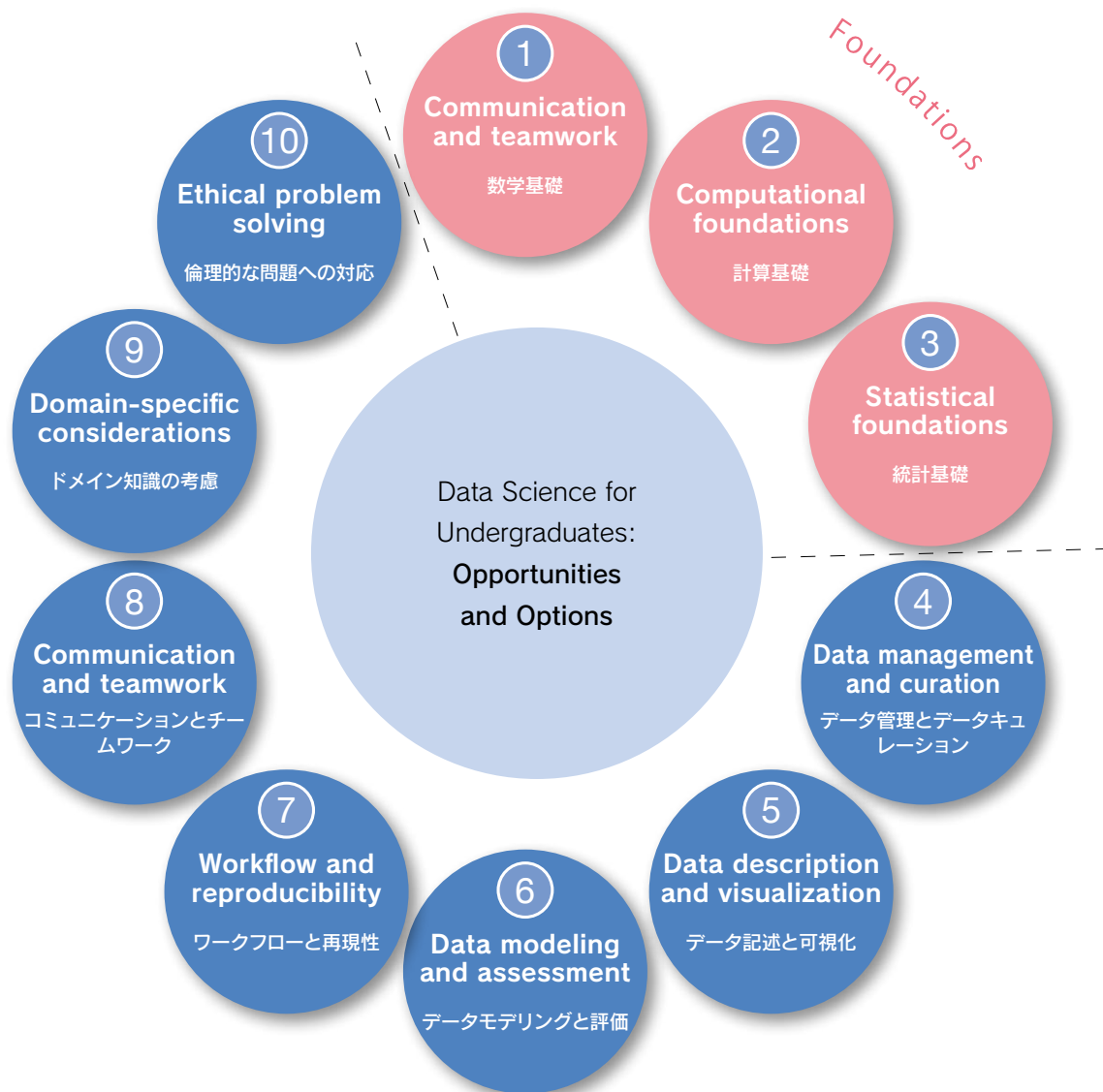
狩野 裕 教授

1981年、大阪大学理学部数学科卒業。84年、大阪大学大学院基礎工学研究科博士後期課程退学。同年に運輸省海技大学校教養科助手となる。86年、工学博士。87年に大阪大学基礎工学部助手。大阪府立大学工学部講師、筑波大学数学系助教授を経て97年、大阪大学人間科学部助教授。04年から同大学院基礎工学研究科教授。18年度に第23回日本統計学会賞受賞

カリキュラム分科会

カリキュラム分科会は全国的なモデルとなる標準カリキュラムを協働して作成・普及に取り組むことになっています。

カリキュラム分科会は全国的なモデルとなる標準カリキュラムを協働して作成・普及に取り組むことになっています。その作成に当たり、データサイエンス教育で先行するアメリカの取り組みを調査したところ、2018年5月に The National Academies of Sciences, Engineering, and Medicine において、プロジェクト Data Science for Undergraduates: Opportunities and Options の最終報告書が公開されたところでした。報告書においては、以下の10分野が重要であると提言しています。



カリキュラム分科会においては、この提言を参考にして10分野についてそれぞれ学修目標やスキルセットを整備していくこととしました。2018年度は、主に最初の3つの基礎 (foundations) について検討しました。具体的には「数学基礎 (大分類) -- 線形代数 (中分類) -- ベクトル (小分類)」のように階層構造を設けます。そして中分類に対して、データサイエンスの観点から概要を説明し、小分類にはいくつかの学修内容やスキルセットとして整理します。さらに、文系を含む全大学に対するデータサイエンス教育の普及のために高大接続を意識し、例えば高校「数学B」の内容である二次元ベクトルがこのセットに含まれています。いくつかの学術機関、関係省庁、データサイエンス関連団体との調整を経て、2019年度の前半に一般公開予定です。もちろん、4以降の分野についても項目出しを行い、早期に一般公開出来るよう作業を進めて参ります。

教材分科会

教材分科会では、全国的なモデルとなる教材を協働して
作成・普及に取り組むことを目的として活動しています。

審議事項は、●教科書シリーズの企画・編纂 ●各大学のeラーニング教材、講義動画等の統合的配信方法の検討 ●普及方法やその他の可能性の検討です。目標とするアウトプットは、●データサイエンスの叢書 ●eラーニング教材および講義動画等のポータルサイトです。

1. データサイエンスの叢書

6 拠点校の教員が編集委員となり、教科書シリーズ「データサイエンス入門シリーズ（講談社）」を3期に分けて刊行予定です。

第1期（2019年9月刊行予定）

『データサイエンスのための数学』椎名洋・姫野哲人・保科架風（著）清水昌平（編）
『データサイエンスの基礎』濱田悦生（著）狩野裕（編）
『最適化手法入門』寒野善博（著）駒木文保（編）

第2期（2019年冬刊行予定）

『統計モデルと推測』松井秀俊・小泉和之（著）竹村彰通（編）
『Pythonで学ぶアルゴリズムとデータ構造』辻真吾（著）下平英寿（編）
『Rで学ぶ統計的データ解析の基本』林賢一（著）下平英寿（編）
『データサイエンスのためのデータベース』村井哲也・吉岡真治（著）水田正弘（編）

第3期（2020年春刊行予定）

『スパース回帰分析とパターン認識』西井龍映・梅津佑太・上田勇祐（著）
『モンテカルロ統計計算』鎌谷研吾（著）駒木文保（編）
『テキスト・画像・音声データ分析』西川仁・佐藤智和・市川治（著）清水昌平（編）



2. eラーニング教材・講義動画等のポータルサイト

数理・データサイエンス教育強化拠点コンソーシアムに、各拠点大学のeラーニング教材・講義動画等のポータルサイトへのリンクを集約するページを作成しました。各教材の利用条件についての詳細は各拠点大学にお問い合わせください。一般公開されている教材や動画もあります。

<http://www.mi.u-tokyo.ac.jp/consortium/e-learning.html>



- 数理・データサイエンス教育用データの収集・公開
- 既存の公開済データに関する情報
- 教育に利用可能なオープンソースツール等の情報

※	サービス名	提供形態	提供開始時期	提供内容・概要
※	サービス名	提供形態	提供開始時期	提供内容・概要
11	Facebook コーポレート	無料	2018年12月	Facebookの広告掲載人数を削減する実験プログラム（試験、自動停止、バリエーション、停止）により、Facebookの広告掲載人数を削減する。
12	Reata Corp. 提供	無料	2019年1月	米国で約600人の人脈を世界向けで提供したCSVファイル
13	日本自動車データセット	無料	2019年2月	全国の車両、7ヶ所の、販売店、自動車メーカーのデータ
14	LINE OSU	無料	2019年3月	LINEのOSUゲームのスコアを収集し、分析する
15	LINE JAP	無料	2019年3月	LINEのJAPゲームのスコアを収集し、分析する
16	STL 10 dataset	無料	2019年3月	STL 10 datasetのデータを収集し、分析する
17	CB AH 10	無料	2019年3月	CB AH 10のデータを収集し、分析する
18	MINST	無料	2019年3月	MINSTのデータを収集し、分析する
19	1000個画像データセット	無料	2019年3月	1000個画像データセットのデータを収集し、分析する
20	Labeled on AWS	無料	2019年3月	Labeled on AWSのデータを収集し、分析する
21	NEURAC on AWS	無料	2019年3月	NEURAC on AWSのデータを収集し、分析する
22	自動運転システムに関するデータ	無料	2019年3月	自動運転システムに関するデータのデータを収集し、分析する
23	investing	無料	2019年3月	investingのデータを収集し、分析する
24	CMU Graphics Lab Mt. Fuji Dataset	無料	2019年3月	CMU Graphics Lab Mt. Fuji Datasetのデータを収集し、分析する
25	the KIT Motor Body Motion Database	無料	2019年3月	the KIT Motor Body Motion Databaseのデータを収集し、分析する
26	自動運転データセット	無料	2019年3月	自動運転データセットのデータを収集し、分析する
27	自動運転データ	無料	2019年3月	自動運転データのデータを収集し、分析する

データ整理項目

- データ名
- データ概要
- 有料/無料
- 提供団体等
- 提供国
- 分野 (学間分野, 産業分野)
- 出典 (URL等)
- データ種類
- 対象地域
- データ規模

The image displays three overlapping screenshots of the Hokkaido University Data Commons website. The top screenshot shows the '人口総数：総務省『国勢調査』' (Total Population: Statistics of Japan) page. The middle screenshot shows the 'データセット' (Data Set) page with a search bar and a list of datasets. The bottom screenshot shows the 'データセット検索' (Data Set Search) page with a search bar and a list of datasets.

6

20 大学が協力校として選定されました

平成31年1月に文部科学省から下記の20大学が「大学における数理・データサイエンス教育の全国展開」の協力校として選定されたことが公表されました。

北見工業大学、東北大学、山形大学、筑波大学、宇都宮大学、群馬大学、千葉大学、お茶の水女子大学、新潟大学、長岡技術科学大学、静岡大学、名古屋大学、豊橋技術科学大学、神戸大学、島根大学、岡山大学、広島大学、愛媛大学、宮崎大学、琉球大学

今後は、6拠点校と20協力校が6つのブロックを分担し、ブロック会議の開催、ブロック別ワークショップの開催、拠点・協力校連絡会合の開催などを通して、数理・データサイエンス教育の全国展開を推進することになります。

3 大学合同シンポジウムを後援しました

平成31年1月23日に早稲田大学大隈講堂で東京大学、横浜市立大学、早稲田大学合同シンポジウム「大学におけるデータサイエンス教育の現状と未来」が開催されました。シンポジウムでは早稲田大学田中総長の開会あいさつ、国公私立大学という異なった特徴を持つ3つの大学の取り組み状況の報告の後、群馬大学、武蔵大学、立教大学の先生も加わって、データサイエンス教育の課題に関するパネルディスカッションが行われ、活発で有益なシンポジウムとなりました。このシンポジウムの詳細については下記のサイトをご覧ください。

<https://www.waseda.jp/inst/cds/news/553>

教育用データベース分科会主催のセッションが開催されました

平成31年2月19日に北海道大学で開催されたデータサイエンス公開シンポジウムで、コンソーシアムの教育用データベース分科会主催セッション「データサイエンス教育とデータ活用について」が開催されました。このセッションでは、コンソーシアムの目的、カリキュラム分科会、教材分科会、教育用DB分科会の取り組み内容が紹介されました。また、パネルディスカッションには3分科会から委員が出席し、「データサイエンス教育とデータ活用について」をセッションテーマとして、データサイエンス教育用データや教育の普及・展開に関する活発な議論が展開されました。



第5回コンソーシアム会合が開催されました

平成31年3月15日に東京大学で数理・データサイエンス教育強化拠点コンソーシアムの第5回会合が開催されました。

主な議題

- ・ 文部科学省から「コンソーシアムへの依頼事項」
- ・ 阪大狩野センター長の講演「データ科学教材例」
- ・ 平成30年度活動報告(6大学, 3分科会, コンソーシアム全体)
- ・ 全国展開へ向けた取り組みについて



数理・データサイエンス教育強化拠点コンソーシアム事務局
東京大学大学院情報理工学系研究科 情報理工学教育研究センター