

RES230 データサイエンスⅡ

2年 3,4クォーター

担当教員 井上 博行

授業形態 講義

単位数 2

曜日・時限 木曜日・3時限

授業概要

近年様々なデータがインターネット上で使えるようになってきている。また、ビッグデータと呼ばれる大規模なデータが活用されるようになってきている。このようなデータを正しく扱え、必要な情報を取り出し、活用することは、地域社会や国際社会の問題を扱う上で重要であり、有効的に利用できる能力が求められている。本科目では、データベース技術とデータの分析手法を中心に、データの扱い方や、処理の方法について学び、データを加工でき、情報を読み取る能力を身に付ける。まず、データの表現方法などデータベースの基礎を学び、SQLを用いてデータベースより情報抽出する方法を身に付ける。応用として、具体的に政府統計などのインターネット上から入手したデータをデータベースにし、必要な情報を取り出すことをおこなう。次に、多変量解析を中心に統計的にデータを解析し、データを読み取ることを行う。具体的には、重回帰分析、判別分析、クラスター分析、主成分分析、因子分析を実際のデータに適用し、それぞれの特徴を理解したうえで、情報としてどのように解釈し、利用するのかを学ぶ。また、データマイニングの手法についても触れる。以上のことは、実際にPC上でツールソフトを操作しながら学び、身に付ける。

到達目標

学生は、

- (1) データの取り扱いについて理解する。
- (2) データベースから情報を取り出し、読み取る方法を身に付ける。
- (3) インターネット上にある統計データについて調べる。
- (4) 集めたデータをデータベース化する方法を学ぶ。
- (5) 具体的な統計データ得られる情報について考察する。
- (6) データ分析を行うツールの利用方法を理解する。
- (7) 学生自身が興味のあるデータを集め、データを分析することを体験する。

先修科目

情報処理基礎、情報技術基礎、データサイエンスⅠ

教科書・参考資料等

◎教科書

速水治夫, 宮崎収兄, 山崎晴明「データベース」オーム社 (2006)

◎参考資料

プリントとして配布

授業の方法

この授業は、講義とPCを用いた実習により行う。担当教員は講義を行い、学生はその内容について実際にPCで演習を行いながら理解する。各項目での演習課題を課し、授業内や授業外に行うようにする。

成績評価

演習課題を提出する。

成績

40% 演習課題

60% 最終課題とそのレポート

授業スケジュール

第1回: データベース, データ分析の基礎

授業全体で学ぶ内容のガイダンスを行うとともに、データベースの基礎と社会的な必要性を学ぶ。また、データ分析の実例などを紹介する。

第2回: リレーショナルデータベースの基礎

リレーショナルデータベースの基礎となる、リレーショナルモデルと主キーや外部キーなどの概念を学ぶ。

第3回: データベースの設計

第1正規形, 第2正規形, 第3正規形などデータベースの設計に必要な理論を学ぶ。

第4回: SQL を用いた情報抽出 (1)

Microsoft Access を用いてデータベース操作言語である SQL を学ぶ。特に SELECT 文の基礎を中心に学習する。実際にデータを扱いながら学ぶ。

第5回: SQL を用いた情報抽出 (2)

前回に引き続き SQL の基礎を学ぶ。SELECT 文の GROUP BY 句や副問い合わせなどを学習する。実際にデータベースより情報を抽出しながら学ぶ。

第6回: 政府統計などの取得とデータベース化

総務省の政府統計のポータルサイトより、政府統計の取得方法を学ぶ。必要なデータを集めて、データベース化を試みる。

第7回: ソフトウェアを用いたデータベース検索

Microsoft Access のクエリ作成機能を学び、簡単にデータベースから情報を取り出す方法を学ぶ。

第8回: 地域データや国際データを用いたデータベースの演習課題

ここまで学んだデータベースに関する演習課題に取り組む。こちらで用意した演習問題だけでなく、学生自身が収集したデータで作成したデータベースからの情報抽出も行う。

第9回: データと多変量解析

様々な大規模データを加工する方法を解説し、データ解析の理解を深める。特に多変量解析を注目し、その概要を理解するとともに、ソフトウェアの使い方を理解する。

第10回: 重回帰分析と判別分析によるデータ分析

重回帰分析と判別分析について講義をし、実際のデータを用いて、重回帰分析と判別分析の演習を行う。

第11回: クラスタ分析によるデータ分析

クラスタ分析について講義をし、実際のデータを用いて、クラスタ分析の演習を行う。

第12回: 主成分分析と因子分析

主成分分析と因子分析について講義をし、実際のデータを用いて、主成分分析と因子分析の演習を行う。

第13回: データマイニングの手法 (テキストマイニングを含む)

データマイニングで使われている手法について解説する。具体的には、決定木分析や、自己組織化マップ、ニューラルネットワークなどの他に、テキストマイニングについても触れる。

第14回: 地域データや国際データを用いたデータ分析の演習課題

ここまで学んだ多変量解析を中心とするデータ分析についての演習課題に取り組む。こちらで用意した演習問題だけでなく、学生自身が収集したデータに関しても解析を行わせる。

第15回: 授業のまとめ, 振り返りと最終課題

この授業の行ったことの振り返りを行い。最終課題について取り組む。

事前・事後学習

予習: 各回の授業内容に示しているキーワードを一通り調べる。教科書・事前に配布されている資料を一読すること。

復習: 授業内容を復習し、授業で扱ったデータや関連データを用いて、繰り返し手法の演習を行うこと。