

2018年度 データサイエンス論 (4028)

■ 授業科目基本情報

科目区分	専門科目	教職科目	指定なし
単位数	2	選択・必修・自由	選択
授業形態	講義	主な使用言語	英語
開講時期	III	履修登録システム	使用する
履修登録期間	2018/10/02~2018/10/16	履修取消期限	2018/11/01

■ 教育プログラム別の履修区分

プログラム名	IS	CB	BS	BN	MS	CP	DS
履修区分	○	○	△	△	△	△	◎
コア科目	-	-	-	-	-	-	C
履修方法	<p>・基盤科目及び専門科目から12単位以上履修すること。</p> <p>・データサイエンスプログラムでは、コア科目であり必修科目の「データサイエンス論」を履修すること。加えて、同じくコア科目である「データ工学基礎」、「機械学習概論」及び「データマイニング」から1科目以上を履修すること。さらに、同じくコア科目である「バイオサイエンスにおけるビッグデータ」及び「マテリアルインフォマティクス特論」から1科目以上を履修すること。</p>						

■ 授業科目概要

担当責任教員	中村 哲
担当教員	中村哲、船津公人、MD.Altaf-UI-Amin、宮尾知幸、小野直亮、鈴木優、国田勝行、吉野幸一郎、田中宏季、(岩爪道昭)、(高橋哲朗)
教育目的/学修到達目標	センシング装置の小型化、高速化、ネットワークの高速化、インターネットの大規模化により、テキストからセンサーデータに至る多様で膨大なデータが出現した。これまでの仮説に基づき実験を行って検証を行うスタイルに対し、今後はデータから人工知能理論を活用し仮説を作り実験を行うデータ駆動型科学が重要となる。本講義では、多様なデータに基づくデータ駆動型サイエンスの理論、最先端の方法、実際の利用について学ぶ。
授業概要/指導方針	本授業では多様な大規模データに基づくデータ駆動型サイエンスに必要な人工知能・機械学習の理論、データサイエンスの理論を学ぶ。さらに、インターネット上のテキストデータマイニング、バイオインフォマティクス、ネットワークバイオロジー、マテリアル・ケモインフォマティクスにおけるデータ駆動型サイエンスについて学ぶ。

■ 授業計画

[1限目 9:20-10:50] [2限目 11:00-12:30] [3限目 13:30-15:00] [4限目 15:10-16:40] [5限目 16:50-18:20] [6限目 18:30-20:00]

回数	日付 [時間]	担当教員	テーマ	内容
1	11/1 [1]	-	データサイエンス解析概論・検定 (中村・田中)	ビッグデータを用いたデータ駆動推論による新たなサイエンスの創造についての考え方、方法論の基礎について理解する。また、統計的検定の考え方と方法について理解する。
2	11/1 [2]	-	大規模Web情報処理-クローリング/ランキング (岩爪)	巨大なインターネットのWEB空間におけるデータの構造、サイバーデータの収集方法、分析法について学ぶ。
3	11/5 [1]	-	データベース、並列計算1 (鈴木)	巨大なデータを扱うためのデータベースシステム、並列計算方法について学ぶ。

4	11/5 [2]	—	データベース, 並列計算2(鈴木)	実際に巨大なデータを扱う上で用いるデータフレームについて, 演習を行いつつ学ぶ。
5	11/7 [1]	—	教師あり学習と教師なし学習(田中)	データサイエンス手法の1つである機械学習について概観する。教師あり学習, 教師なし学習の違いとそれらの基本的な手法について学ぶ。
6	11/7 [2]	—	決定木, ランダムフォレスト(田中)	データサイエンス手法の教師有り学習の方法である, 決定木アルゴリズム, ランダムフォレストについて学ぶ。
7	11/9 [1]	—	パターンマイニング(吉野)	データサイエンス手法の教師なし学習の方法である, パターンマイニングについて学ぶ。
8	11/9 [2]	—	深層学習, バッチ・オンライン学習(吉野)	データサイエンス手法における深層学習の手法について学ぶ。また, 大規模データを扱う上で必要となるバッチ学習・オンライン学習の違いについて学ぶ。
9	11/13 [1]	—	社会データインフォマティクス(高橋)	ネット上のサイバーデータの一つとして社会データを取り上げ, この分析事例を通してサイバーデータの分析について学ぶ。
10	11/13 [2]	—	ネットワークバイオロジー(Amin)	要素間の関係性を把握することを目的に, 特に, 分子生物学などの例をもととしたネットワーク解析法を学ぶ。
11	11/15 [1]	—	バイオインフォマティクス(小野)	ゲノム解析や遺伝子発現解析などの網羅的な生物データの分析におけるデータサイエンス手法の応用について学ぶ。
12	11/15 [2]	—	生物データへの制御理論応用(国田)	第一原理モデリングやシステム同定などの制御理論のモデリング・シミュレーション手法の生物データへの応用を実例を通して学ぶ
13	11/19 [1]	—	PCAとPLS(有機材料の開発)(船津)	材料の特性に影響する種々のパラメータを俯瞰し, 材料とその特性との相関関係をモデル化する手法を実例を通して理解する。
14	11/19 [2]	—	GA(遺伝的アルゴリズム)とパレート最適解(触媒設計)(宮尾)	材料特性を説明するためのパラメータを合理的に選択するGA, および材料物性間のトレードオフを対象としたパレート最適解探索を触媒探索を通して理解する。
15	11/21 [1]	—	SVR(化学プラントの監視)(船津)	化学プラントの状態は常に変化する。その変化の状態をモデル化するにあたって, 非線形モデリング手法であるSVRが良く用いられる。SVRの理解とともにその実際の応用を通して活用し実際に理解する。
16	11/21 [2]	—	試験	明日の試験の問題範囲の情報です。下記の範囲で出題します。 仮説検定から一問(第1回講義) 機械学習から一問(第7, 8回講義) データ工学から一問(第3回講義) ソーシャルデータ, バイオデータ, 物質データのデータサイエンスの適用例から一つ選択して論述

■ 授業日程

[1限目 9:20-10:50] [2限目 11:00-12:30] [3限目 13:30-15:00] [4限目 15:10-16:40] [5限目 16:50-18:20] [6限目 18:30-20:00]

回数	日付	時間	講義室	備考
1	11/1	1	C109	

2	11/1	2	C109	
3	11/5	1	C109	
4	11/5	2	C109	
5	11/7	1	E207 208	
6	11/7	2	E207 208	
7	11/9	1	E207 208	
8	11/9	2	E207 208	
9	11/13	1	E207 208	
10	11/13	2	E207 208	
11	11/15	1	E207 208	
12	11/15	2	E207 208	
13	11/19	1	E207 208	
14	11/19	2	E207 208	
15	11/21	1	E207 208	
16	11/21	2	E207 208	

■ テキスト・参考書

テキスト	特になし
参考書	特になし

■ その他

履修条件	機械学習概論, データ工学基礎, パターン認識, 系列データモデリングを受講していることが望ましい
オフィスアワー	Eメールで連絡の上、日時を決める
成績評価の方法と基準	・5段階(秀・優・良・可・不可)で評価する。 ・各回の課題(50%)と最終試験(50%)の点数の合計で評価する。
関連科目	機械学習概論, データ工学基礎, パターン認識, 系列データモデリング
関連学位	工学
注意事項	特になし