

平成29年度

社会人向け実践教育プログラム
「先端データサイエンス実践コース」
日程およびシラバス

豊橋技術科学大学
社会連携推進センター

目次

1	はじめに	4
2	日程 (全 42H).....	5
3	シラバス	6
3.1	統計学・機械学習基礎 1,3 (6H).....	6
3.2	統計学・機械学習基礎 2,4 (6H).....	6
3.3	python 入門 1,2 (6H).....	6
3.4	Scikit-Learn 入門 (3H)	6
3.5	Chainer 入門 (3H).....	7
3.6	Chainer 実習 (12H).....	7
3.7	先端データサイエンス講演会 (6H).....	8
4	その他情報.....	9
4.1	e-Learning について.....	9
4.2	Web 会議システムについて.....	9

1 はじめに

今日、コンピュータおよびその周辺技術の急速な進展により、膨大なデータを取得、生成することが可能な時代になりました。このビッグデータの中に埋もれた有用な情報を取り出し、活用するためにはデータサイエンスの基礎理論から実践的ノウハウまで広く理解しておく必要があります。豊橋技術科学大学は、開学以来、「技術」を「科学」で裏付けし、そこから新しい技術を創造する技術科学の教育・研究を使命としています。とくに、本学の知識・情報工学系において、これまで「情報化学」などの様々なデータサイエンスの応用分野で教育・研究を行ってきた実績があります。これらのノウハウを本学学生のみならず、これからデータサイエンスの応用を目指そうとする社会人技術者にも提供したいと考えています。なお、本学で社会人向けに用意されている「計算技術科学実践教育プログラム」も「データサイエンス実践教育プログラム」と密接に関係しています。本コースでは、統計学や機械学習などのデータサイエンスの基礎理論から、「計算物質科学」等への機械学習の実践応用まで学ぶことができます。

2 日程 (全 42H)

回	タイトル	担当	日にち	時限	概要
1	統計学・機械学習基礎 1	金澤	10/5(木)	3,4	機械学習・パターン認識論
2	統計学・機械学習基礎 2	渡辺	10/5(木)	5,6	機械学習の数学的基礎
3	統計学・機械学習基礎 3	金澤	10/20(金)	3,4	機械学習・パターン認識論
4	統計学・機械学習基礎 4	渡辺	10/20(金)	5,6	機械学習の数学的基礎
5	Python 入門 1	濱田	11/1(水)	3,4	python, numpy, matplotlib
6	Python 入門 2	濱田	11/1(水)	5,6	pandas
7	Scikit-Learn 入門	濱田	11/15(水)	3,4	Scikit-Learn の基礎
8	Chainer 入門	濱田	11/15(水)	5,6	Chainer+Copy の基本概念と使い方
9	Chainer 実習 1	濱田	11/27(月)	3,4	手書き文字認識 (分類問題)
10	Chainer 実習 2	濱田	11/27(月)	5,6	単純非線形モデル系での出力予測 (回帰問題)
11	Chainer 実習 3	濱田	12/20(水)	3,4	翻訳モデル(RNN)
12	Chainer 実習 4	濱田	12/20(水)	5,6	分子活性予測
13	先端データサイエンス講演会	外部講師	1/24(水) (仮)	2,3,4,5	マテリアルズインフォマテイク等

[コメント]

時限定義

2: 10:30-12:00, 3:13:00-14:30, 4:14:40-16:10, 5:16:20-17:50, 6: 18:00-19:30 を意味する。

3 シラバス

3.1 統計学・機械学習基礎 1.3 (6H)

タイトル	担当	日にち	備考
統計学・機械学習基礎 1,3	金澤靖	10/5, 10/20	
内容			
機械学習とパターン認識に必要となる統計学や線形代数の基礎知識を学び、基本的な手法であるサポートベクトルマシン、ニューラルネットワーク、アンサンブル学習などの基本原理を学ぶ。。			

3.2 統計学・機械学習基礎 2,4 (6H)

タイトル	担当	日にち	備考
統計学・機械学習基礎 2,4	渡辺一帆	10/5,10/20	
内容			
機械学習で重要となる潜在変数を持つ学習モデルやベイズ推測の枠組みを学ぶ。高専・大学初等の数学的知識を用いた、学習法の導出を通じて、確率モデルを用いた機械学習への理解を深める。			

3.3 python 入門 1,2 (6H)

タイトル	担当	日にち	備考
python 入門 1,2	濱田信次	11/1	
内容			
データサイエンスで中心的に使用される言語である python 言語(ここでは python3 を使用)の基本的使い方を習得する。また python におけるもっとも多用されるパッケージである numpy(配列処理)と matplotlib(可視化)の基本的使い方を習得する。さらにはデータサイエンスにおいてとくに重要なパッケージである pandas も習熟する。			

3.4 Scikit-Learn 入門 (3H)

タイトル	担当	日にち	備考
Scikit-Learn 入門	濱田信次	11/15	
内容			

Scikit-Learn は使い勝手の良い、機械学習用 Python パッケージである。Scikit-Learn を利用したいいくつかのサンプルプログラムを動かすことによって、「統計学・機械学習基礎」で学んだ内容をより具体的に理解する。

3.5 Chainer 入門 (3H)

タイトル	担当	日にち	備考
Chainer 入門	濱田信次	11/15	
内容			
<p>Chainer の簡単なサンプルプログラムを通じて、Chainer の基本概念、利用方法について理解する。 それと同時にニューラルネットワークの動作原理についても理解する。 さらに、chainer から利用される cupy (Python から cuda を簡単に利用するためのライブラリ) の基本概念、利用方法についても理解する。</p>			

3.6 Chainer 実習 (12H)

タイトル	担当	日にち	備考
Chainer 実習	濱田信次	11/29, 12/20	
内容			
<p>いくつかのサンプルプログラムを動作させることにより、Chainer の実践的な知識を理解する。前半の2つは基本概念、基本手法の理解を目的とする。手書き文字認識はニューラルネットワークにおける Hello World プログラムと呼ばれるものであるが、まずこのプログラムを完全に理解し、さらに各種パラメータを変えると認識性能がどう変わるかも実験する。次に、回帰問題の例として、多入力多出力の単純な非線形モデル系における出力予測を取り扱う。ここでも各種パラメータと予測性能との関係を実験する。また、確定系モデルと確率系モデルとの本質的違い（過学習など）について理解する。後半の2つは応用上より興味深い問題を取り扱う。翻訳モデルは、再帰型ニューラルネットワークの例である。分子活性予測は、ディスクリプタと呼ばれる実験的あるいは理論的に得られた多くの数値から、分子のもつ様々な活性値を予測するものであり創薬における非常に重要な技術である。</p>			

3.7 先端データサイエンス講演会 (6H)

タイトル	担当	日にち	備考
先端データサイエンス講演会	濱田信次	1/24(水)(仮)	
内容			
先端データサイエンス分野で活躍されている外部講師を招き講演をしていただく。 内容は現時点で未定ですが決定次第 Web にてお知らせします。(マテリアルズインフォマティクス等)			

4 その他情報

4.1 e-Learning について

e-Learning とは教育用 Web システム(Moodle)を通じて学習することです。Moodle 上には講義資料、講義ビデオ、サンプルプログラム、小テストなどが置かれており、登録ユーザーは自宅等から任意の時間にアクセスして学習することができます。本コースはすべて来学していただき受講されることを原則としますが、参加できなかった受講者は後日 e-Learnig で受講していただけます。

4.2 Web 会議システムについて

本プログラムではでは来学が困難な受講者のために Web 会議システム(OmniJoin)を用いることで授業のライブ配信を行うことを予定しています。自宅 PC(Win/MAC)から接続して、その場で質問等が可能となります。Web 会議システムでの参加を希望される方は前日 17 時までに担当 (濱田 0532-44-6881, hamada@adsim.tut.ac.jp) までご連絡ください。接続可能数に限りがありますので御希望に添えない場合もございます。なお Web カメラ、ヘッドセットも貸し出し可能ですが、これも台数に限りがございます。