

CONTENTS

第1章

人工知能と機械学習

この章では、人工知能とは、ビッグデータとは、データサイエンスとはなんなのか、その概要を説明し、以降の章の内容とその読み方を紹介します。

●1-1	人工知能とは？	
1-1-1	現在の妖怪「人工知能」	002
	◎コラム：ゲームをさらに進化させる人工知能	006
●1-2	BIはAIの夢を見るのか？	
1-2-1	ビジネスを支えるAIと機械学習エンジニア	007
●1-3	いまの人工知能＝知能？	
1-3-1	人工知能がデータ分析を行う日	011
●1-4	まとめ	014

第2章

データ分析処理のための基盤技術

Hadoopはインターネットサービス企業を中心に採用されてきましたが、近年は非常に多種多様な企業でも採用されつつあります。本章では、Hadoopを中心とした大規模分散データ処理環境の概要とその構築方法について解説していきます。

●2-1	Hadoop ～大規模データを処理するインフラ～	
2-1-1	Hadoopとは？	016
2-1-2	MapReduce処理	017
●2-2	Hive ～Hadoop上でより容易にデータ処理を実現～	
2-2-1	Hiveとは？	021
●2-3	Presto ～高速なデータ処理を実現～	
2-3-1	Prestoとは？	026

●2-4	Spark ～より高速な分散計算処理環境～	
2-4-1	Sparkとは？	029
●2-5	その他のHadoopに関連するミドルウェア	
2-5-1	その他のHadoopミドルウェア	032
●2-6	さまざまなHadoopディストリビューション	
2-6-1	Hadoopディストリビューションとは？	036
●2-7	BIツール ～処理されたデータを可視化～	
2-7-1	BIのあれこれ	039
2-7-2	Pentaho	040
2-7-3	Zeppelin	043
●2-8	まとめ	045
	◎コラム：機械学習や人工知能でできること	046

第3章

リアルタイムにデータを分析する ～データから「いま」を知る技術～

Hadoopは大量なデータをバッチ処理することに長けており、これまで想像もできなかった量のデータを処理することができるようになりました。しかし、大量なデータを処理することができるようになると、次はいかに早くデータを処理するかが要求されます。この章では、データを適宜処理するリアルタイムデータ処理について説明します。

●3-1	Fluentd	
3-1-1	データをリアルタイムに収集する	048
●3-2	Norikra	
3-2-1	データをリアルタイムに集計する	053
●3-3	事例 ～Twitterの人名のリアルタイム計測～	
3-3-1	リアルタイムで人名を計測する	058
3-3-2	Ruby環境を整える	059
	◎コラム：単語分割の鍵を握る技術「形態素解析」	061
3-3-3	Fluentdのセットアップ	062
3-3-4	Twitter Streaming APIとの連携	064
	◎コラム：Twitter API	066
	◎コラム：Twitterの開発者用のサイト	066
	◎コラム：アプリケーションの登録	067

◎コラム：アプリケーションのアクセストークン	067
3-3-5 FluentdのMeCabプラグインの作成	068
3-3-6 Norikraのセットアップ	073
3-3-7 NorikraとFluentdの接続	074
3-3-8 Norikraにキーワード集計用のクエリを登録	076
3-3-9 ElasticSearchとKibanaによる可視化	078
3-3-10 Norikura、Fluentd、ElasticSearchとKibanaの連携	081
●3-4 まとめ	086

第4章

機械学習アルゴリズム ～データから「未来」を知る技術～

この章では、過去のデータの傾向から未知の事柄を推定したり、未来の動向を予測したりする技術について説明します。このようなデータから直接読み取ることができない情報を獲得するためには、「機械学習」と呼ばれる技術を用います。

●4-1 機械学習とは	
4-1-1 「レストラン予約」で機械学習を考える	088
●4-2 機械学習アルゴリズム	
4-2-1 機械学習アルゴリズムの概要	091
4-2-2 機械学習アルゴリズムの分類	094
◎コラム：matplotlib	097
●4-3 ナイーブベイズ	
4-3-1 ナイーブベイズとは	098
4-3-2 ナイーブベイズの理論	099
4-3-3 ナイーブベイズの具体例	101
4-3-4 ナイーブベイズの実装	110
4-3-5 ナイーブベイズのまとめ	113
◎コラム：macOS用のソフトウェア管理ツール「Homebrew」	114
◎コラム：Homebrewのインストール	114
◎コラム：HomebrewでMeCabをインストール	114
●4-4 単純パーセプトロン	
4-4-1 単純パーセプトロンとは？	115
4-4-2 パーセプトロンの学習	117
4-4-3 パーセプトロンの実装	119

4-4-4 パーセプトロンの問題点	121
4-4-5 パーセプトロンのまとめ	121
●4-5 オンライン学習	
4-5-1 オンライン学習とは	122
4-5-2 オンライン学習版パーセプトロン	123
4-5-3 Passive Aggressiveアルゴリズム	125
4-5-4 オンライン学習のまとめ	129
●4-6 サポートベクトルマシン	
4-6-1 サポートベクトルマシンとは？	130
4-6-2 サポートベクトルマシンの理論	131
●4-7 線形回帰	
4-7-1 「回帰」のアルゴリズムについて	143
●4-8 自己回帰	
4-8-1 過去のデータから目的関数を推定する手法	146
◎コラム：さまざまな機械学習用ライブラリ	148
●4-9 クラスタリング	
4-9-1 クラスタリングのアルゴリズムについて	149
●4-10 MapReduceと機械学習	
4-10-1 機械学習アルゴリズムの大規模データへの適用	157
●4-11 サンプル事例	
4-11-1 Python環境のセットアップ	160
4-11-2 IRISデータによる機械学習の例	163
4-11-3 digitsデータによる機械学習の例	185
4-11-4 Bostonデータによる機械学習の例	192
●4-12 まとめ	201
◎コラム：チャットボット	202

第5章

ディープラーニング

この章では、近年話題になっているディープラーニングについて、これまでの機械学習アルゴリズムを踏まえて、その違いや性能について紹介していきます。

●5-1	単純パーセプトロンとその仲間たち	
5-1-1	単純パーセプトロン（おさらい）	204
5-1-2	ロジスティック回帰とは	206
5-1-3	多項ロジスティック回帰	208
●5-2	ニューラルネットワーク	
5-2-1	単純パーセプトロンとニューラルネットワークの違い	210
5-2-2	ニューラルネットワーク	212
5-2-3	ユニット	213
5-2-4	誤差逆伝播法	217
5-2-5	プログラムの例	222
5-2-6	単純パーセプトロン、ロジスティック回帰、 多項ロジスティック回帰、ニューラルネットワーク	224
●5-3	畳み込みニューラルネットワーク	
5-3-1	畳み込みニューラルネットワークとは？	225
5-3-2	畳み込み層（convolution layer）	226
5-3-3	プーリング層（pooling layer）	229
5-3-4	多層ニューラルネットワーク	231
5-3-5	畳み込みニューラルネットワークの学習	231
●5-4	再帰型ニューラルネットワーク	
5-4-1	再帰型ニューラルネットワークとは	233
5-4-2	再帰型ニューラルネットワークの学習	237
●5-5	TensorFlow	
5-5-1	インストール	238
5-5-2	多項ロジスティック回帰	241
5-5-3	多層ニューラルネットワーク	249
5-5-4	畳み込みニューラルネットワーク	252
5-5-5	続・畳み込みニューラルネットワーク	257
●5-6	まとめ	265
●用語集		266
	主要参考図書／URL／論文	273
	INDEX	274