

# フルスタックJavaScriptと Python機械学習ライブラリで実践する ソーシャルビッグデータ

— 基本概念・技術から収集・分析・可視化まで —

石川 博 編著  
横山 昌平 共著  
廣田 雅春

コロナ社



ソーシャルメディア由来のソーシャルデータと実世界由来の実世界データは最近注目されるビッグデータの重要な構成要素である。さらに公共的な性格を持ったオープンデータも信頼性のあるデータとして注目される。こうしたデータを本書ではソーシャルビッグデータと総称する。

ソーシャルデータの典型は Twitter や Flickr への投稿記事や写真である。実世界データには自動車のプローブデータや JAXA の打ち上げた「はやぶさ」のデータが含まれる。また OSM 地図や Wikipedia を使いやすくした DBpedia もオープンデータの一つである。

ソーシャルビッグデータを有機的に分析できれば、新たな価値が発見できて、観光や防災、科学まで多様な分野に应用が可能となる。特にソーシャルビッグデータを時間的、空間的、意味的に分析することにより、大衆の行動を集合的に予測することができるようになる。これを本書では「イシカワ・コンセプト」と呼ぶ。イシカワ・コンセプトを実践することで、例えば訪日外国人が国内において期待以上に満足する場所はどこか、複数の楽しみが同時に体験できる観光地はどこか、観光地の目玉になる桜のような花々はその観光地でいまが旬か、災害時に現在地から避難所までの安全な経路はどこかなどがわかる。

一方、データ分析、データサイエンスで重要となる仮説そのものについては、以下のよう  
に二つの異なる見方ができる。

- ① 宣言としての仮説：伝統的なデータ分析における仮説に相当する。この種の仮説に対する主要なタスクは、仮説の検証である。
- ② 手続きとしての仮説：手順を実行し、その結果として仮説を生成する。もちろんこの場合でも生成された仮説の検証は必要である。

最近では行政やビジネス、科学のような分野では限られた予算や資源を有効に利用して意思決定を行うこと（evidence-based policy making：EBPM）ができるように、定性的で一般的な仮説でなく、定量的で具体的な仮説生成とそのための手順が求められている。

またソーシャルビッグデータの応用分野を観察すると、大量のデータ管理とそれに対するデータマイニングからなる。データ管理とデータマイニング、機械学習、人工知能は、それぞれ別々に発展してきたデジタルエコシステム（以下単にエコシステムという）である。一般にエコシステムは、製品やサービスを媒介にして複数のベンダと利用者からなる相互に依存した生態系である。つまりソーシャルビッグデータ応用の多くは、異なるエコシステムからなるハイブリッドな応用であるといえる。

さらに計算機科学あるいは情報科学も広い意味では科学であり、その手順に関して再現性のある説明が、計算機科学の応用であるソーシャルビッグデータ応用の信頼性にとっても不可欠である。

要するにソーシャルビッグデータ応用において、複数のエコシステムを利用して、再現性のある仮説生成手順の記述には、データモデルとして統一的なデータ構造とそれらに対するデータ操作を定義する必要がある。これをデータモデルアプローチという。

本書ではソーシャルビッグデータの収集、分析、可視化の基本（概念、技術）と実践（アルゴリズム、プログラム）をバランスよく学ぶことを目指す。前著『ソーシャル・ビッグデータサイエンス入門』は本書の入門編にあたり、まずは概念を理解してもらうためにあえて実践的なコードは示さなかった。そこで本書では分析に必要な基本概念、基本技術の説明に加えてソーシャルビッグデータの典型的な応用の実装にすぐに使えるコード（JavaScript や Python プログラム）を掲載し、実務家（卒論、修論などを執筆する学生、企業の技術者、研究者）にすぐに役立つ内容の提供を目指す。

まず1章ではデータ管理とデータマイニングを含む複数の事例を通して、仮説を作り出すための手順を具体的に与えている。そのためにデータモデルアプローチを導入し、手順を特定のプログラム言語とは独立な形式で記述している。また一歩踏み込んでソーシャルビッグデータ応用でよく使われる仮説生成のための方法論についても触れている。

2章ではソーシャルビッグデータ応用で必要とされるデータマイニング、機械学習の基本概念とそこでよく使われる手法、実現アルゴリズムについて説明している。分類やクラスタリング、相関ルールなどの手法に関する複数アルゴリズムの説明に加えて、人工知能として最近注目されるディープラーニング（深層学習）についても利用目的に応じてその手法を概観できるようにした。

さらに3章以下の第II部は、ソーシャルビッグデータの収集や分析、可視化まで含めた包括的なアプリケーションをフルスタック JavaScript にて実装する実践的な内容となっている。フルスタック JavaScript とはデータベース問い合わせ・サーバ・クライアントの3層すべてを JavaScript で開発する方法論で、本書ではそれに加えて Python の著名な機械学習ライブラリ scikit-learn との連携手法も示している。なお、本書で実装するプログラムはすべて GitHub リポジトリ (<https://github.com/abarth500/sobig> (2018年7月現在)) として公開している。

最後に本書の執筆にあたりご協力いただいた職業能力開発総合大学校の遠藤雅樹氏ならびにコロナ社で企画から編集、出版までをお手伝いいただいた関係者の皆様に深く感謝する。

2018年12月

著 者



## 第I部 基礎編

### 1. ビッグデータ

1.1	ビッグデータの特徴	1
1.2	実世界データ	2
1.3	オープンデータ	3
1.4	ソーシャルデータ	6
1.5	ソーシャルビッグデータ	11
1.5.1	実世界データとソーシャルデータのインタラクション	11
1.5.2	ユニバーサルキー	12
1.6	イシカワ・コンセプト, そしてジオソーシャルビッグ	12
1.6.1	ソーシャルビッグデータ利用の一般的な流れ	14
1.6.2	ソーシャルビッグデータの4番目のv	15
1.7	ソーシャルビッグデータ応用 (ケーススタディ)	20
1.7.1	ケース1—観光応用: 掘り出し物スポット発見 (ソーシャルデータ × 印象評価)	20
1.7.2	ケース2—観光応用: トピックの地域的影響 (異種ソーシャルデータ × クラスタリング&グラフモデル)	22
1.7.3	ケース3—観光応用: 観光資源名の意味 (ソーシャルデータ&オープンデータ × Word2Vec)	25
1.7.4	ケース4—観光応用: 見頃推定 (ソーシャルデータ&オープンデータ × 時系列分析)	27
1.7.5	ケース5—観光応用: FreeWi-Fi スポット設置候補 (異種ソーシャルデータ × 統合分析)	31
1.7.6	ケース6—防災応用: 危険避難路発見 (ソーシャルデータ&オープンデータ × グラフモデル)	34

1.7.7 ケース 7—科学応用：中央丘クレーターの発見 (オープンデータ (実世界データ) × ディープラーニング) .....	36
1.8 本書で学んでほしいこと .....	39
1.8.1 データサイエンティストとデータエンジニア .....	39
1.8.2 データサイエンティスト .....	41

## 2. データマイニング・機械学習の基本技術

2.1 概 論 .....	46
2.2 データマイニングとは .....	46
2.2.1 データマイニングの細分化 .....	47
2.3 データマイニングと周辺分野の関連 .....	49
2.3.1 データベース .....	49
2.3.2 統計解析 .....	50
2.3.3 機械学習 .....	50
2.4 データマイニングのプロセス .....	51
2.4.1 データの選択 .....	51
2.4.2 データの前処理 .....	53
2.4.3 データの変換 .....	62
2.4.4 パターン、知識の発見 .....	65
2.4.5 解釈と評価 .....	68
2.5 クラスタリング .....	69
2.5.1 概 要 .....	69
2.5.2 階層的クラスタリング .....	71
2.5.3 k-means .....	77
2.5.4 DBSCAN .....	79
2.6 分 類 .....	80
2.6.1 概 要 .....	80
2.6.2 k 近 傍 法 .....	80
2.6.3 決 定 木 .....	83
2.6.4 SVM .....	86
2.6.5 ディープラーニング .....	89
2.7 その他の手法 .....	90
2.7.1 アンサンブル学習 .....	90

2.7.2 相 関 ル ー ル	93
-----------------	----

## 第 II 部 実践編

### 3. ソーシャルビッグデータ分析を支える Web 技術

3.1 フルスタック JavaScript	97
3.1.1 サーバ上の JavaScript	97
3.1.2 データベース上の JavaScript	99
3.1.3 本書で実装するシステムの構成	100
3.2 環 境 構 築	102
3.2.1 Node.js のインストール	103
3.2.2 MongoDB のインストール	105
3.2.3 実装ディレクトリの準備	107
3.3 Web 3 層アーキテクチャの実装	107
3.3.1 Web サーバ	108
3.3.2 WebSocket	109
3.3.3 WebSocket によるチャットプログラム	111
3.3.4 JavaScript object notation (JSON)	114
3.3.5 GeoJSON	116
3.3.6 MongoDB の利用	116
3.3.7 Node.js から MongoDB へのアクセス	118
3.3.8 非同期処理環境における処理フローの記述	121
3.3.9 MongoDB と WebSocket によるチャットプログラム	124

### 4. データを集める

4.1 ソーシャルビッグデータを知る	128
4.2 ソーシャルビッグデータの収集	129
4.2.1 検索による Twitter データの収集	130
4.2.2 新着監視による Twitter データの収集	131
4.2.3 検索による Flickr データの収集	132
4.2.4 新着監視による Flickr データの収集	136

4.2.5 DBpedia を用いた Wikipedia データの収集 .....	137
4.3 ジオソーシャルビッグデータの収集 .....	138
4.3.1 ジオタグ付きツイートの収集 .....	139
4.3.2 ジオタグ付き写真の収集 .....	144
4.3.3 OpenStreetMap への問い合わせ .....	145
4.4 クローラの実装と運用 .....	147
4.4.1 クローラ機能のパッケージ化 .....	148
4.4.2 さらに高度な実装のために .....	151

## 5. データを可視化する

5.1 可視化ライブラリのインストール .....	152
5.1.1 チャート描画 .....	153
5.1.2 カラーパレット .....	154
5.2 ソーシャルデータ分析可視化環境の準備 .....	155
5.2.1 プログラムのインストール .....	155
5.2.2 分析対象ソーシャルデータのクロール .....	156
5.3 散 布 図 .....	156
5.4 ワードクラウド .....	158
5.5 地 図 描 画 .....	161
5.6 ヒートマップ .....	162

## 6. データを分析する

6.1 準 備 .....	164
6.2 線 形 回 帰 .....	165
6.3 k-means .....	167
6.4 DBSCAN .....	171
6.5 機 械 学 習 .....	171
6.6 TF-IDF .....	173
6.7 お わ り に .....	174

引用・参考文献 .....	175
---------------	-----

索 引 .....	181
-----------	-----

## 第I部 基礎編

# 1 ビッグデータ

本章ではソーシャルビッグデータとその統合分析の基本概念を説明する。まずソーシャルビッグデータを構成する実世界データ、およびオープンデータ、ソーシャルデータの概要と実例について説明する。続いて実世界データとソーシャルデータのインタラクションを説明した後、それに基づく統合分析の基本概念である「イシカワ・コンセプト」を導入する。さらに基本概念に沿った統合分析の流れを説明した後、統合分析のためのデータモデルアプローチを導入する。それに基づき、複数のケースを通して統合仮説と統合分析について具体的に説明する。最後に本書で学ぶべき項目を列挙し、それらについて説明する。

### 1.1 ビッグデータの特徴

まずはじめにビッグデータの特徴についてまとめて説明する<sup>1)†</sup>。一般にビッグデータは以下のように、vではじまる四つのワードで説明できる。

- 生成されるデータの総量 (volume) が大きい。
- 生成されるデータの速度 (velocity) が速い。
- 生成されるデータの種類 (variety) が多い。
- 生成されるデータにあいまいさ (vagueness) がある。

最初の三つのvについては、改めて説明の必要はないだろう。そこで四つ目のvについてコメントしておく。例えば複数のデータがあれば、データの欠損、関連するデータ間の矛盾など、データのあいまいさが引き起こされる。また生成されたデータの使用に関しても、その説明に現状ではあいまいさがある。後者のあいまいさについては1.6.2項で詳しく説明する。

ビッグデータは、おおよそ実世界データおよびオープンデータ、ソーシャルデータに分類される。以下それぞれについて、具体例とともに詳しく見てみよう。

<sup>†</sup> 肩付数字は、巻末の引用・参考文献を表す。



## 1.2 実世界データ

まず実世界データといわれるものには、人間が実世界（フィジカルワールド, physical world）で行動、活動することによって発生するさまざまなデータが含まれる。最近では加速度、ジャイロ、心拍数などの測定装置の小型化が進んできた。そのためスマートウォッチなどのウェアラブル端末で、活動量、心拍数など人間の諸活動に関するデータが簡単に記録できるようになった。また、GPS (global positioning system) 装置の小型化により、人間の生成するさまざまなデータに位置情報が付与されるようになった。

同様に人間が機械やシステム、メディアと関わることで蓄積されるデータも実世界データの一つと考えられる。また、人間の状態に限らず、生産現場における設備機械の状態や宇宙における探査機の状態が各種センサによって測定されて生み出されるデータも実世界データの一つである。一般にインターネットなどのネットワークにつながれたセンサを持つ機器をIoT (internet of things)<sup>2)</sup> 機器という。

さらに降雨や桜の開花など自然現象などを観測、観察することによって発生する気象データや科学的データ (JAXA の月惑星科学データ<sup>3)</sup> など)、自動車の車両走行データ (プローブデータ)<sup>4)</sup> も実世界データの一つである。

図 1.1 に JAXA の DARTS (Data Archives and Transmission System) サイト<sup>5)</sup> にある

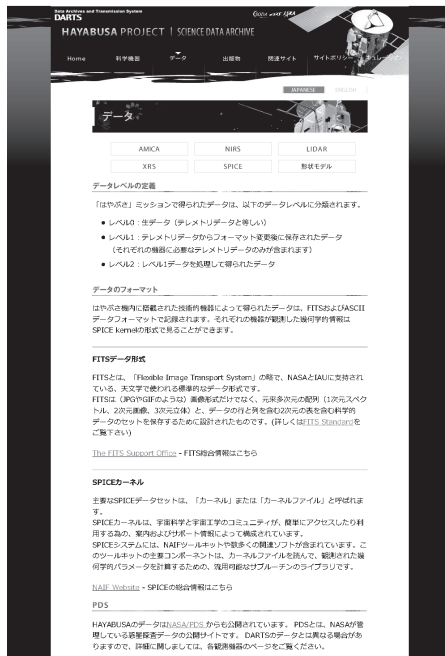


図 1.1 DARTS 「はやぶさ」

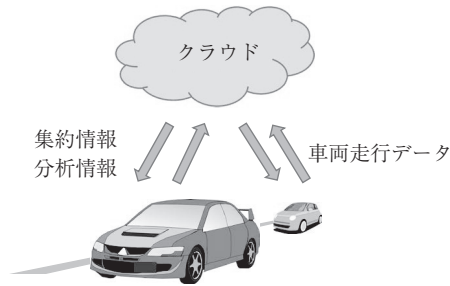


図 1.2 車両走行データ

「はやぶさ」のデータに関するページを示す。また、図 1.2 に車両走行データの利用形態の一例を示す。

他にも、交通系 IC カードのデータ、各種メディアの視聴ログデータなどが挙げられる。

### 1.3 オープンデータ

〔1〕 オープンデータの定義 つぎにオープンデータについて説明しておこう。オープンデータとは政府や自治体、研究機関など公共性の高い団体が収集し、整理したものを、すべての人がアクセスできるように公表したデータである<sup>6)</sup>。

まず人間が見るといふことだけを前提としたこれまでのデータは HTML (hypertext markup language) 形式や PDF (portable document format) 形式での公表が主であった。それに対して、オープンデータはその利用できる形式に基づいて以下のように分類される。

- ① 機械処理が可能なデータ形式：XLS, DOC のようなプロプライエタリ (proprietary, 独占的) ライセンスのもとで提供されたプログラムで利用できるデータ形式
- ② 二次利用がしやすいデータ形式：CSV, XML などオープンな形式
- ③ 外部連携や検索が可能なデータ形式：LOD (linked open data)<sup>7)</sup>, RDF (resource description framework)

ここでオープンデータとしては、①が最も初歩的レベルで、②、③となるにつれて、より高度なレベルになる。また②は①の特徴を、③は②の特徴をそれぞれ引き継ぐ。

〔2〕 オープンデータの例 国や地方公共団体などの組織から提供されるオープンデータには、それらの組織が収集した避難施設などの防災情報や人口動態などの統計情報が含まれる。学術資源である論文も、特にオープンアクセスジャーナルに掲載されるものは、だれもがアクセスできるという意味でオープンデータの一つである。そうしたオープンアクセスジャーナルを用いた学術トレンドの発見を目指した研究も行われている<sup>8)</sup>。

こうしたオープンデータと実世界データとは必ずしも排他的ではない。実世界データの中

にも、オープンデータは存在する。例えば 1.7.4 項で後述するような気象庁の提供する生物季節観測<sup>9)</sup>に関するデータや JAXA が DARTS を通して提供する月惑星科学データのように、観察、観測によって得られ公表された時間・位置情報付きのデータなどもオープンデータの一つであると考えられる。

図 1.3 に W3C<sup>10)</sup>のサイトにある RDF データの例を示す。例えば、book1 (主語) は、"SPARQL Tutorial" (目的語) を title (述語) として持つ。図 1.4 には、この RDF データに対して SPARQL という言語による問い合わせとその実行結果の例を示す。

```
@prefix dc:<http://purl.org/dc/elements/1.1/>.
@prefix:<http://example.org/book/>.
@prefix ns:<http://example.org/ns#>.

:book1 dc:title"SPARQL Tutorial".
:book1 ns:price 42.
:book2 dc:title"The Semantic Web".
:book2 ns:price 23.
```

図 1.3 RDF データ

```
PREFIX dc:<http://purl.org/dc/elements/1.1/>
SELECT ?title
WHERE{ ?x dc:title ?title
      FILTER regex(?title,"^SPARQL")
}
```

title
"SPARQL Tutorial"

(a) 文字列の一致

```
PREFIX dc:<http://purl.org/dc/elements/1.1/>
PREFIX ns:<http://example.org/ns#>
SELECT ?title ?price
WHERE{ ?x ns:price ?price.
      FILTER (?price<30.5)?x dc:title ?title.
}
```

title	price
"The Semantic Web"	23

(b) 数値の比較

図 1.4 PARQL による問い合わせと結果

さらにみんなで作り上げたような、地図データ (OpenStreetMap : OSM) や百科事典 Wikipedia (ウィキペディア) も、だれもがアクセスできるという意味でオープンデータに含まれる。

(1) **OSM** OSM<sup>11),12)</sup>は、編集可能な世界地図を作成するための共同プロジェクトである。Wikipedia<sup>13)</sup>が成功したことに加えて、英国をはじめとして各国において商用地図データが支配的になったことや安価な GPS の出現に触発されて、2004 年に英国のスティーブ・コースト (Steve Coast) によって作成された。それ以来 OSM は、人手による調査、GPS 機器利用、航空写真撮影、その他の無料情報源の利用を通してデータを収集できる、200 万人以上の登録ユーザを持つプロジェクトにまで成長した。OSM は、ボランティアベースで作成された地理情報の顕著な例の一つである。このようにしてクラウドソース (crowdsourcing) されたデータは ODbL (open database license)<sup>14)</sup> というライセンスで利用可能になる。

OSM には、例えば以下のようなデータが含まれる。

- POI (point of interest) : 自然物 (山, 川など), 施設 (水族館, 美術館など), 店舗 (レストラン, ハンバーガー店など), 神社仏閣など
- 道路情報 : 道路, 橋, トンネルなど

図 1.5 は OSM プロジェクトで作成された新宿御苑周辺の地図である。



図 1.5 OSM 地図

(2) **DBpedia** DBpedia<sup>15),16)</sup> (「DB」は「database」を意味する) は、Wikipedia プロジェクトで作成された情報から、構造化コンテンツを抽出することを目的としたプロジェクトである。こうして構造化された情報は、world wide web (WWW) 上で利用可能になる。利用者には、Wikipedia 内のデータに関する属性や関係に対する意味的な問い合わせ、言い換えればデータの構造を意識した問い合わせを許す。DBpedia では SPARQL を使って Wikipedia の内容の検索ができる。WWW の発明者である Tim Berners-Lee<sup>17)</sup> は、DBpedia を LOD に関する取り組みの中で最も著名なものの一つとしている。



<b>【あ】</b>		観光応用	20, 22, 25, 27, 31	次元の呪い	63
あいまいさ	1, 15	<b>【き】</b>		支持超平面	87
アプリオリ	95	機械学習	50	支持度	94
アンサンブル学習	90	強学習器	90	実世界データ	2, 36
<b>【い】</b>		教師あり学習	65	弱学習器	90
イシカワ・コンセプト	13	教師なし学習	66, 69	情報銀行	15
異常値	58	共有サービス	6	シリアル分析	14
——の発見	58	距離	66	シンクロソロ分析	13
位置情報	59	<b>【く】</b>		シンクロ分析	12
印象評価	20	クラウドソーシング	7	<b>【す】</b>	
インスタントメッセージング		クラス	65	数値データ	57
サービス	6	クラスターリング	22, 69	スケーラビリティ	43
<b>【う】</b>		グラフモデル	22, 34	<b>【せ】</b>	
後向き法	64	クリギング	29	正規化	62
<b>【え】</b>		群平均法	75	生物季節観測	4, 27
枝刈り	85	<b>【け】</b>		線形判別分析	86
<b>【お】</b>		欠損値	60	線形分離可能	86
オッカムの剃刀	84	決定木	83	線形分離超平面	86
オープンデータ		決定木学習	171	<b>【そ】</b>	
	3, 25, 27, 34, 36, 55	<b>【こ】</b>		関連ルール	93
<b>【か】</b>		コア点	79	想像力	41
回帰直線	165	交差検定	172	族	18
階層的凝集法	71	行動経済学	23	ソーシャルゲーミング	7
階層的クラスターリング	71	コラボレーション	7	ソーシャルサーチ	7
階層的データ構造	18	<b>【さ】</b>		ソーシャルデータ	6, 20, 22, 25, 27, 31, 34
外的分離	69	最近傍法	81	ソーシャルニュース	7
科学応用	36	再現性	17	ソーシャルビッグデータ	11
過学習	85	最短距離法	73	ソーシャルビッグデータ	
学習器	65	最長距離法	74	研究センター	21
確信度	94	差分	33	ソフトクラスターリング	71
かぐや	36	サポートベクトル	87	ソフトマージン SVM	88
過剰適合	85	散布図	153	<b>【た】</b>	
仮説の多様性	17	<b>【し】</b>		タプル	19
カテゴリデータ	57	ジオソーシャルビッグ	13	ダミー変数	58
カーネルトリック	88	ジオタグ	59, 139, 161, 167	単位の変換	56
カラーパレット	154	時系列分析	27, 28		

<b>【ち】</b>		<b>【の】</b>		<b>【ま】</b>	
チェイニング	74	ノイズ	58	マイクロプロギング	6
チェビシェフ距離	68			前処理	53
中央丘クレーター	36	<b>【は】</b>		前向き法	64
超平面	86	排他的クラスタリング	71	マハラノビス距離	68
		バギング	91	マルチクラス分類	66
<b>【て】</b>		外れ値	58, 79	マルチラベル分類	66
ディズニールランド	149, 156, 170	ハードクラスタリング	71	マンハッタン距離	67
ディープラーニング	36, 89	パラレル分析	14		
デジタルエコシステム	16	汎化能力	85	<b>【み】</b>	
データ				密度到達可能点	79
——の選択	51	<b>【ひ】</b>		ミンコフスキー距離	67
データウェアハウス	45	ビッグデータ	1		
データエンジニア	39	ビデオコミュニケーション	7	<b>【め】</b>	
データ管理	16	ヒートマップ	162	メタデータ	59
データ形式の変換	53	非排他的クラスタリング	71	メッシュ	18
データサイエンティスト	39, 41	標準化	63		
データフレーム	19			<b>【ゆ】</b>	
データベース	49	<b>【ふ】</b>		ユークリッド距離	67
データマイニング	17	フィルタ法	64	ユニバーサルキー	12
——のプロセス	51	ブースティング	92		
データモデル	16	ブートストラップサンプリング	92	<b>【ら】</b>	
デンドログラム	71	フルスタック JavaScript	100, 152, 164	ラッパ法	64
		プロギング	6	ラムダアーキテクチャ	43
<b>【と】</b>		プログラミング言語	44	ランダムフォレスト	92
統計解析	50	分割最適化クラスタリング	77		
特徴選択	63	分類	80	<b>【る】</b>	
特徴抽出	64	分類器	65	類似度	66
特徴量	65				
トランザクション	93	<b>【ほ】</b>		<b>【れ】</b>	
トレーニングデータ	50, 65	防災応用	34	レビューサービス	7
		補間	60		
<b>【な】</b>				<b>【わ】</b>	
内的結合	69			ワードクラウド	158

<b>【A】</b>		<b>【D】</b>		<b>【G】</b>	
AdaBoost	93	DBpedia	5, 128, 137	GeoJSON	116, 141, 165, 167
		DBSCAN	23, 79, 171	GIGO	53
<b>【B】</b>				GitHub	44
biasedLexRank	24	<b>【E】</b>		GPS	59
		EBPM	37		
<b>【C】</b>		ETL	44	<b>【H】</b>	
C10K 問題	98	EXIF	134	hot-deck imputation	61
CART	85				
CASAP サイクル	129	<b>【F】</b>		<b>【I】</b>	
CNN	39, 89	Flickr	10	ID3	85
				IoT	2

	<b>[J]</b>		<b>[T]</b>
JAXA	4, 36	NoSQL	49, 99, 105, 116
	<b>[K]</b>	<b>[O]</b>	TF-IDF
k 近傍法 (KNN)	80, 171	OpenStreetMap	173
k-means	77, 167	OSM	8
kd-tree	82	Overpass API	128, 146
KDD	47		<b>[W]</b>
	<b>[L]</b>	<b>[P]</b>	Ward 法
LAMP	153	POI	75
Lance-Williams の更新式	76		WebSocket
linked open data (LOD)	128	<b>[R]</b>	110, 111, 124, 155
LSH	82	RDF	97, 108, 116, 153, 164
	<b>[N]</b>	<b>[S]</b>	Wikipedia
Non-blocking	98, 121	scikit-learn	128, 137
		SNS	101
		SPARQL	25, 26
		SVM	25, 26
			~~~~~
			<b>[数字]</b>
			2 値分類
			66

石川 博 (いしかわ ひろし)  
1979年 東京大学理学部情報科学科卒業  
1979年 株式会社富士通研究所勤務  
1992年 博士(理学)(東京大学)  
2000年 東京都立大学教授  
2006年 静岡大学教授  
2013年 首都大学東京教授  
現在に至る

横山 昌平 (よこやま しょうへい)  
1996~ 株式会社オリエンタルランド勤務  
2008年  
2001年 東京都立大学工学部電子情報工学科卒業  
2003年 東京都立大学大学院工学研究科修士課程修了(電気工学専攻)  
2006年 東京都立大学大学院工学研究科博士課程修了(電気工学専攻)  
博士(工学)  
2006年 産業技術総合研究所特別研究員  
2008年 静岡大学助教  
2012年 静岡大学講師  
2016年 静岡大学准教授  
2016年 フランス Université de Pau et des Pays de l'Adour 客員研究員  
2018年 首都大学東京准教授  
現在に至る

廣田 雅春 (ひろた まさはる)  
2010年 静岡大学情報学部情報科学科卒業  
2011年 静岡大学大学院情報学研究科修了(情報学専攻)  
2014年 静岡大学創造科学技術大学院自然科学系教育部修了(情報科学専攻)博士(情報学)  
2015年 大分工業高等専門学校助教  
2017年 岡山理科大学講師  
現在に至る

## フルスタック JavaScript と Python 機械学習ライブラリで実践するソーシャルビッグデータ — 基本概念・技術から収集・分析・可視化まで —

Social Big Data in Practice with Full Stack JavaScript  
and Python Library for Machine Learning

© Hiroshi Ishikawa, Shohei Yokoyama, Masaharu Hirota 2019

2019年2月18日 初版第1刷発行



検印省略

編著者 石川 博  
著者 横山 昌平  
廣田 雅春  
発行者 株式会社 コロナ社  
代表者 牛来真也  
印刷所 三美印刷株式会社  
製本所 有限会社 愛千製本所

112-0011 東京都文京区千石 4-46-10  
発行所 株式会社 コロナ社  
CORONA PUBLISHING CO., LTD.  
Tokyo Japan  
振替 00140-8-14844・電話(03)3941-3131(代)  
ホームページ <http://www.coronasha.co.jp>

ISBN 978-4-339-02889-8 C3055 Printed in Japan

(三上)



JCOPY <出版者著作権管理機構 委託出版物>

本書の無断複製は著作権法上での例外を除き禁じられています。複製される場合は、そのつど事前に、出版者著作権管理機構(電話 03-5244-5088, FAX 03-5244-5089, e-mail: info@jcopy.or.jp)の許諾を得てください。

本書のコピー、スキャン、デジタル化等の無断複製・転載は著作権法上での例外を除き禁じられています。購入者以外の第三者による本書の電子データ化及び電子書籍化は、いかなる場合も認めていません。落丁・乱丁はお取替えいたします。