

# 自然言語処理技術

～目的に応じた手法選択／精度向上手法／業務活用への提言

発刊：2020年3月 定価：40,000円 + 税 体裁：B5判ソフトカバー 180頁

執筆者一覧（敬称略）

- 鈴木潤（東北大学） ●土屋誠司（同志社大学） ●本橋和貴（Konduit（株）） ●高橋寛治（Sansan（株））
- 田村晃裕（愛媛大学） ●山田育矢（（株）Studio Ousia） ●荒木健治（北海道大学） ●森信介（京都大学）
- 渡邊信一（宇都宮大学） ●原 紳（宇都宮大学） ●水本智也（フューチャー（株）） ●清水武（（株）CINC）

## 目次

※詳細はHPにて <https://johokiko.co.jp/publishing/BC200202.php>

**第1章 導入～自然言語処理の基礎からトピックの話**

1. 自然言語処理の基礎（文章の解析、文章の生成について）
2. 自然言語処理の難しさ
3. 自然言語処理技術の最先端
3. 1 自己注意機構型NNと再帰型NNの比較
4. 現状のデファクトスタンダード
4. 1 今後の展開

**第2章 所持データ&目的に応じた手法の選定**

1. 意味理解
  1. 1 意味表現方法
  1. 2 言語資源
  1. 3 解析手法
2. 単語や話題の抽出
  2. 1 抽出対象
  2. 2 抽出方法
3. 文書分類
  3. 1 クラスタリング
  3. 2 近さの判断
4. 文書要約
  4. 1 要約手法
  4. 2 手掛かり情報
5. 情報検索
  5. 1 必要要素
  5. 2 検索モデル
  5. 3 評価手法
6. 雑談対話
  6. 1 話題の展開
  6. 2 処理方法
  6. 3 必要な機能

**第3章 ハンズオンで学ぶ文書分類モデルの作成とツールの利用方法**

1. 機械学習モデル開発のワークフローと本章で扱う内容
2. 文書分類ハンズオン
  2. 1 本章で扱う主な自然言語処理ツールの解説
    2. 1. 1 spaCy
    2. 1. 2 Transformers
  2. 2 データセットの準備
    2. 2. 1 livedoorニュースコーパス
3. 文書分類モデル
  3. 1 単語バグ（bag-of-words）
    - (1) データセットの読み込み
    - (2) 文書分類モデルの訓練
  3. 2 量み込みニューラルネットワーク
  3. 3 BERT

- (1) データの前処理
- (2) データセットの読み込み
- (3) トークンの数値化
- (4) Iteratorの設定
- (5) BERTのファインチューニングのための設定

**第4章 前処理の種類及び実装手法～処理精度向上に繋げるための登壇門**

1. 前処理の概観
2. テキストデータのクリーニング手法
  2. 1 文字コード
  2. 2 不可視文字や見た目が同じ文字に対するUnicode正規化
  2. 3 正規表現による前処理
3. 単語の処理手法
  3. 1 形態素解析
  3. 2 複合語や固有名詞を単語とする方法
  3. 3 表記ゆれ対策
4. 単語のベクトル変換手法
  4. 1 Bag-of-Words
  4. 2 分散表現
  4. 3 系列ラベリングとして取り扱うための変換
5. 前処理に活用できるライブラリやツール
  5. 1 形態素解析
  5. 2 文構造解析器
  5. 3 Pythonモジュール
  5. 4 言語資源

**第5章 アプリケーション別の応用例と活用への提案**

**第1節 ディープラーニング技術を用いたニューラル機械翻訳**

1. ニューラル機械翻訳の概要
2. 代表的なニューラル機械翻訳モデル
  2. 1 Transformer NMTモデル
  2. 2 学習モデル
3. ニューラル機械翻訳の研究開発動向
  3. 1 対訳データ不足に対する対応
  3. 2 未知語への対応
4. ニューラル機械翻訳システムの構築
5. 今後の展望

**第2節 質問応答技術の開発動向と応用事例**

1. 質問応答の主要な事例
2. 質問応答の研究動向
  2. 1 ファクトイド質問応答タスク
  2. 2 抽出型質問応答タスク
  2. 3 回答選択型質問応答タスク
  2. 4 その他の質問応答タスク
3. 質問応答の実装に使えるツールやサービス

**第3節 自然言語処理技術を用いた雑談対話技術**

1. 雑談システムの歴史と現状
2. 雑談システムの構築方法
  2. 1 ルールベースに基づく手法
  2. 2 遺伝的アルゴリズムを用いた帰納的学習による手法
  2. 3 機械学習による手法
3. 雑談システムの問題点
4. 話し上手な雑談システム構築への取り組み

**第4節 固有表現抽出技術**

1. 固有表現認識
  1. 1 機械学習に基づく固有表現認識
2. 分野特有の固有表現認識器の構築
3. 関連する自然言語処理

**第5節 感情分析を用いたクチコミ判別支援システムの開発**

1. システムの概要
2. 企業ニーズ
3. システム開発
4. 今後の展望と波及効果
  4. 1 今後の展望
  4. 2 技術開発のためのヒント
  4. 3 波及効果

**第6節 第二言語学習者のための文法誤り訂正～訂正技術・学習用データと性能評価方法**

1. 文法誤り訂正の動向
2. 文法誤り訂正技術
  2. 1 訂正手法
  2. 2 データセット
  2. 3 評価方法
3. 文法誤り訂正システムの構築
4. 文法誤り訂正性能の改善のための最新技術

**第7節 日本語センテンスを対象とした感情分類器の開発**

1. モデル開発にむけての考え方
2. 感情シグナルの定義
3. モデル構築と基本的なレイヤー
4. 日本語の符号化（入力層）
  4. 1 文字ユニグラム
5. 分類器による処理の具体的なフロー
6. 教師データの収集
7. 結果
8. アンサンブル
9. 今後の課題

FAX：03-5740-8766、または、→<https://www.johokiko.co.jp>にて

※FAX番号はくれぐれもお間違えの無い様お願い致します。

### ★書籍申込書

（書籍申し込み要領）

◎右記記入の上、FAXでお申込を承ります。

◎お申込書を確認次第、書籍、請求書および振込要領をお送りいたします。

◎未発刊の書籍をお申込の場合、申込書を確認次第、受領書をお送りいたします。発刊時に弊社より書籍、請求書および振込要領をご送付いたします（送料は弊社負担）

◎お支払いは請求日翌月末日までに、銀行振込にてお願いいたします。原則として領収証の発行はいたしません。

◎振り込み手数料はご負担ください。

★ <https://www.johokiko.co.jp/>

の申込みフォームからも承ります！

書籍名HP【BC200202】		自然言語処理技術 書籍		冊数	___冊	※記入の無い場合は1冊
会社名						
所属部課・役職等						
申込者氏名			TEL	FAX		
E-MAIL			上司役職・氏名			
住所〒						
備考						
ご案内をご希望の場合は今後の案内方法にレ印を記入下さい（複数回答可） <input type="checkbox"/> e-mail <input type="checkbox"/> FAX <input type="checkbox"/> 郵送						

ご連絡頂いた、個人情報は弊社商品の受付・運用・商品発送・アフターサービスのため利用致します。今後のご案内希望の方には、その目的でも使用致します。

今後のサービス向上のため「個人情報の取扱いに関する契約」を締結した外部委託先へ、個人情報を委託する場合があります。個人情報に関するお問合せ先policy@johokiko.co.jp