

様々な分野で活用できる 人工知能型検索技術の ビジネスでの活用

井佐原 均

追手門学院大学心理学部 教授

豊橋技術科学大学IT活用教育センター 特命教授

自己紹介（井佐原均）

- 通商産業省工業技術院電子技術総合研究所（現、産業技術総合研究所）
- 郵政省通信総合研究所（現、情報通信研究機構 (NICT)）
- 国立大学法人豊橋技術科学大学（現在は特命教授）
- 2021年4月より追手門学院大学心理学部心理学科人工知能・認知科学専攻
- ✓ アジア太平洋機械翻訳協会会長、機械翻訳国際連盟会長、アジア自然言語処理連盟事務局長、学会理事、研究会委員長、国内会議・国際会議の各種委員長・委員を歴任
- ✓ （特非）言語資源協会理事、（特非）言語グリッドアソシエーション理事、（一社）通訳品質評議会理事、産業日本語研究会世話人会代表幹事、（一社）AIデータ活用コンソーシアム理事・副会長、知的財産高等裁判所専門委員、ISO/TC37 国内委員会委員長、その他の委員
- ✓ 日本政府観光局認定MICEアンバサダー、元人工知能アドバイザー
- 平成31年度科学技術分野の文部科学大臣表彰 科学技術賞科学技術振興部門「機械翻訳システムの社会実装技術の振興」

研究テーマ

- 自然言語処理技術全般
 - 機械翻訳
 - 質問応答システム
 - 対話
 - 観光情報
 - 医療情報
 - 英語学習支援
- データサイエンス教育（教材）

分野適応技術による自然言語 処理技術のビジネス展開



プロジェクトI
先進的AI・IoT・ビッグデータ活用技術
開発プロジェクト

- 機械翻訳、質問応答システム、対話に関する研究
- <https://youtu.be/h8PezWBNR3g>

質問応答システムに関する研究

- 令和元年度～3年度に企業群と共同研究
 - ✓地方自治体のFAQチャットボットとしての実装
 - ✓大学の研究シーズ検索システムとしての実装
- 文書（FAQ、研究紹介）の内容を普通の言葉で検索

表現が違っていても検索ができる

疑問をそのまま文にするだけで検索できる

知りたい情報と関連した情報の検索ができる

FAQ

質問部

応答部

119番は、どこにつながるのであるのか？	市内からの119番は、〇〇地区消防本部内の消防指令センターにつながります。携帯電話で119番する場合も同様です。ただし携帯電話の場合は、他の消防本部につながってしまう場合がありますが、その場合は、当消防本部に転送されます。
間違っ119番にかけたときはどうしたらいいですか？	間違っ119番へかけてしまった時は、間違いであることを必ず伝えてください。間違っ119番からといって黙って電話を切られますと、呼び返しによる確認をします。
119番の電話は、早く切らないと到着が遅れるのか？	消防指令センターでは、通報をくれた方から聞き取り中でも、状況や場所がわかり次第、必要な車両を出動させていますので遅れることはありません。係員が必要な事を聞きますので、落ち着いて質問にお答えください。
119番の日があると聞いたのですが、いつですか？	119番の日は、毎年11月9日です。昭和62年、国民の防火・防災意識の高揚を図るために「119番の日」が制定されました。

FAQチャットボット

- 二つの検索を融合

入力に対し回答が適切か？

ユーザからの
入力文

FAQの
応答部



正解候補の絞り込み

入力と類似しているか？

ユーザからの
入力文

FAQの
質問部



最も適切な候補の
ランキング

FAQの形式の統一

入力: 総合都市計画って何なのか **教えてください**

第1候補: 都市計画税について **教えてください**

第3候補: 総合計画とは何ですか

統一前	統一後
低公害車に関する補助制度や優遇措置/ はありますか	低公害車に関する補助制度や優遇措置/ について教えてください
国民年金手帳を紛失したので再発行したい場合/ はどうしたらいいですか	国民年金手帳を紛失したので再発行したい場合/ について教えてください
医療費が高額で、医療機関への支払が困難な/ のですが、何か制度はありますか	医療費が高額で、医療機関への支払が困難な/ 場合の制度について教えてください

**AI(BERT)の性能を十分に引き出すには
AIやデータの性質を理解し
人から歩み寄る努力が大切**

研究シーズ検索システム

研究シーズの泉

研究シーズの泉とは

豊橋技術科学大学、長岡技術科学大学、国立高等専門学校の研究シーズが結集した横断的に検索可能なサイトです。

お困りごと（技術ニーズ）を文章で入力すると、人工知能が文章を解析し、従来のキーワード検索では得られにくい、ニーズに合った研究シーズを検索できます。

検索する内容を入力してください。

— 問い合わせに至る背景

— 解決したい課題 **必須**

強制空冷用のヒートシンクを小型化・高性能化させ、冷却ファンの騒音を低減（風切音の低減）したい

— 連絡先（会社名・電話番号・メールアドレス等）

検索

- <https://aico.rac.tut.ac.jp/>

研究シーズ検索システム

研究シーズの泉

トップページ > 検索結果

豊橋技術科学大学 長岡技術科学大学 国立高等専門学校機構

解決したい課題：強制空冷用のヒートシンクを小型化・高性能化させ、冷却ファンの騒音を低減（風切音の低減）したい

検索結果 20件以上

T 高効率・低騒音なファンを目指した研究
横山 博史 [researchmap](#)

概要を表示する

U 高性能低消費電力を目指した計算機アーキテクチャ
北島 宏之 [researchmap](#)

U 高性能細径伝熱管内の新冷媒の伝熱特性
広瀬 正尚 [researchmap](#)

追加情報を入力して検索

質問応答システムに関する研究

- FAQ検索
- シーズ検索
 - テキストの情報を検索する技術
- 今後の展開の可能性
 - ✓ 自社内で社内文書や関連技術を検索
 - ✓ 顧客に向けて製品や技術の紹介

対話に関する研究

- 令和元年度～3年度に企業群と共同研究
- 介護ロボットとしての実装
- 介護施設での介護士と入居者の対話を録音しモデル化
- 今後の展開の可能性
 - ✓カウンセラーや相談者との雑談対話
 - 相手の状況を判断・把握
 - ✓人とコンピュータの間のインタフェース
 - 案内ロボット
 - ✓相手に応じた適切な発話
 - 感情、発話内容、相互の関係などを考慮



観光情報に関する研究

- 科学研究費（令和3年度～5年度、研究分担者）
- 企業との共同研究
- 旅行情報誌「るるぶ」のデータを用いた観光情報提供
- 状況に応じた情報提供
 - 位置、季節、時刻、グループ構成、予算
- 今後の計画
- ✓ ChatGPTなどの生成型人工知能の精度向上
 - 検索対象とするデータを制限
 - プロンプトによる制限



医療情報に関する研究

- 循環器系の病院との共同研究（令和3年度～5年度）
- 市民への医療情報の提供
- メディカルコンシェルジュの支援
- 医療に関する情報の検索
 - カルテ等、ウェブの情報
- 今後の計画
- ✓ 対話によって受診希望者を適切な医療機関に勧誘
 - 受診科
 - 大病院と、かかりつけ医
 - 受診希望者の満足

英語学習支援に関する研究

- 科学研究費（令和5年度～7年度、研究代表者）
- 日本人の英語学習者の能力を判定
 - ✓ 流暢性（英語らしい発声かどうか）
 - 採点済の日本人の英語発話データを用いて学習
 - ✓ 忠実性（文法や意味が正しいか）
 - 既存の音声認識システムを使用
- 今後の計画
 - ✓ 発話評価・学習システムの実装
 - 流暢性の判定ツールは完成
 - 音声認識（忠実性）との統合
 - 実証

データサイエンス教材

- 自習型オンライン教材とビデオ
 - 4 企業、5 大学で使用
 - 豊橋技術科学大学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（10月開講）
- 基礎編（15教材）
 - データの分析に必要な基本的な数学的知識から始め、機械学習で多用されるプログラミング言語 Python の基礎を学ぶ。
 - データ分析の実装で用いるライブラリまで学習を進める。
- 応用編（10教材）
 - 主としてディープラーニングに関して学ぶ。
 - 基礎となるニューラルネットワークの数学・ディープラーニングのフレームワークの実装方法を学び、さらに発展した内容として、ディープラーニングの実用について学ぶ。

データサイエンス教材



tkbasic011.ipynb - Colaboratory

colab.research.google.com/drive/1g0xPiB6FAaO9I-ydYCLnOOwRyMU1A6XM#scrollTo=dehoANTINPN

tkbasic011.ipynb

ファイル 編集 表示 挿入 ランタイム ツール ヘルプ 全ての変更を保存しました

+ コード + テキスト

RAM ディスク

Matplotlibによるデータ可視化の基礎

(*) データフレームを見ただけでは直感的にデータの特徴を理解することは簡単ではありません。データを直感的に理解するにはデータの可視化が有効です。Pythonでのデータの可視化にはよく **Matplotlib** というパッケージを用いることが多いです。Matplotlibを用いてのデータの可視化を行きましょう。Matplotlibは `matplotlib.pyplot` を `plt` という別名をつけて読み込むのが一般的です。

```
import matplotlib.pyplot as plt
```

散布図

散布図 (**scatter**) は変数間の相関を視覚的に確認したり、データのばらつきや値の範囲を視覚的に確認するのに便利なものです。Matplotlibでは与えられた配列から散布図を作成する `plt.scatter()` が用意されています。

まずは、`median_income` 列のデータと `median_house_value` 列のデータをそれぞれ横軸、縦軸に取った散布図を描画してみましょう。

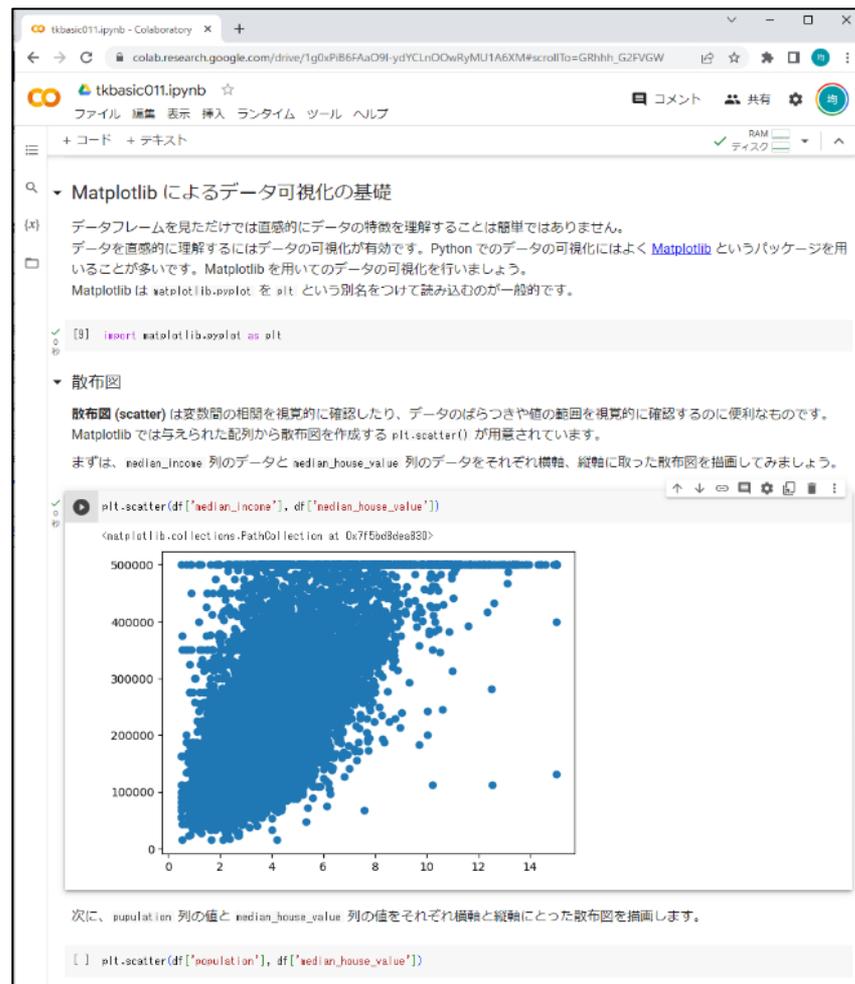
```
plt.scatter(df["median_income"], df["median_house_value"])
```

次に、`population` 列の値と `median_house_value` 列の値をそれぞれ横軸と縦軸にとった散布図を描画します。

```
plt.scatter(df["population"], df["median_house_value"])
```



ビデオを参考にしつつ
教材に入力し課題を実行



tkbasic011.ipynb - Colaboratory

colab.research.google.com/drive/1g0xPiB6FAaO9I-ydYCLnOOwRyMU1A6XM#scrollTo=GRhhG2FVGW

tkbasic011.ipynb

ファイル 編集 表示 挿入 ランタイム ツール ヘルプ

+ コード + テキスト

RAM ディスク

Matplotlibによるデータ可視化の基礎

(*) データフレームを見ただけでは直感的にデータの特徴を理解することは簡単ではありません。データを直感的に理解するにはデータの可視化が有効です。Pythonでのデータの可視化にはよく **Matplotlib** というパッケージを用いることが多いです。Matplotlibを用いてのデータの可視化を行きましょう。Matplotlibは `matplotlib.pyplot` を `plt` という別名をつけて読み込むのが一般的です。

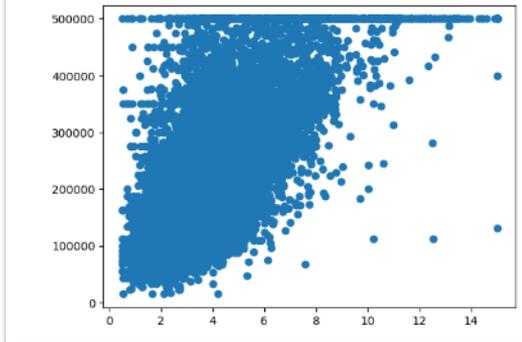
```
import matplotlib.pyplot as plt
```

散布図

散布図 (**scatter**) は変数間の相関を視覚的に確認したり、データのばらつきや値の範囲を視覚的に確認するのに便利なものです。Matplotlibでは与えられた配列から散布図を作成する `plt.scatter()` が用意されています。

まずは、`median_income` 列のデータと `median_house_value` 列のデータをそれぞれ横軸、縦軸に取った散布図を描画してみましょう。

```
plt.scatter(df["median_income"], df["median_house_value"])
```



次に、`population` 列の値と `median_house_value` 列の値をそれぞれ横軸と縦軸にとった散布図を描画します。

```
plt.scatter(df["population"], df["median_house_value"])
```

昨年度の例

【日程】 2022年10月5日(水)～2023年2月8日(水) 全時間 54時間

(オンデマンド講義自習 42時間+オンラインライブ講義 9回 12時間)

※全てご自身のPCから接続頂くオンライン講座となります。

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、本学の情報・知能工学系において、これまで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。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、機械学習・ディープラーニングの実践応用まで学ぶことができます。今年度は裏面プログラムの内容でオンラインで開催いたします。多くの方々にご参加いただきたくご案内申し上げます。

【内容】 1) 統計学・機械学習基礎 2) Pythonの基礎 3) 数値計算とデータ処理・可視化
4) 機械学習実践入門 5) ニューラルネットワークの数学/実装
6) 画像処理とディープラーニング 7) 時系列処理とディープラーニング

以上のオンデマンド講義コンテンツの自習、及び対応する科目ごとにオンライン授業
(ガイダンスとQ&A)を行います。

詳細については「**先端データサイエンス実践コース 開講スケジュール**」(チラシ裏面)
をご確認ください。各講義ではそれ以前に行う講義の内容を理解されているとして進行します。

【自習教材】 本学IT活用教育センター 井佐原 均 特任教授と株式会社キカガクの
共同開発の教材(テキスト及びビデオ講習)によりオンライン自習を進めて
頂きます。本学の学生に分かりやすいと定評のある教材です。

【対象者】 一般の技術者(大学理工系学部卒業程度以上が望ましい)
(大学レベルの数学(解析学、線形代数)の学習経験、
1ヵ月程度以上のプログラミング経験)



井佐原 特命教授

自然言語処理技術

- 今回紹介した研究開発は自然言語処理技術の応用
- 基盤となる技術があれば、
 - ✓個々の企業のニーズに基づいてタスクを決め
 - ✓適切なデータを収集し学習データとすることで
実用化を進めることが可能。
- 例：研究シーズ検索は、FAQ検索の技術を流用
- 何なりと、ご相談ください。