

# 問題解決に自然言語処理と 機械学習を用いた協働学習の実践

埼玉県立川越南高等学校 春日井優

# 発表内容

---

- イン트로ダクション
- 授業について
- 授業内容
  - 知識・技能の習得の授業  
(形態素解析・特徴語抽出・機械学習)
  - 問題解決の授業
  - 生徒の質問・考えたこと
- まとめ

# 発表内容

---

- イン트로ダクション
- 授業について
- 授業内容
  - 知識・技能の習得の授業  
(形態素解析・特徴語抽出・機械学習)
  - 問題解決の授業
  - 生徒の質問・考えたこと
- まとめ

# 次期学習指導要領と解説編が 出ましたね

- 私の第一印象

びっくりに！

特に、情報Ⅱ

「仮想現実・拡張現実・複合現実のコンテンツ」

「回帰・分類・クラスタリング」「重回帰分析」

「プロジェクト・マネジメント」「プログラムを統合」

「問題発見・解決の探究」

まじか！

# 情報Ⅱって…

---

情報Ⅰでプログラミングを

教えるだけでも大変だよ

~~情報Ⅱなんて作ったけれど、~~

~~ほとんどやる学校ないよね~~

**やりましょう！**

# なぜ情報Ⅱ？ 目標①

情報に関する科学的な見方・考え方を働かせ、  
情報技術を活用して問題の発見・解決を行う  
学習活動を通して、問題の発見・解決に向けて  
情報と情報技術を  
適切かつ効果的、**創造的**に活用し、  
情報社会に主体的に参画し、  
**その発展に寄与**するための資質・能力を次のと  
おり育成することを目指す。

# なぜ情報Ⅱ？ 目標②

- (1) 多様なコミュニケーションの実現, (効果的な)  
情報システムや (コンピュータや)  
多様なデータの活用について (データの活用)  
理解を深め技能を習得するとともに,  
情報技術の発展と (情報社会と)  
社会の変化について (人の関わりに)  
理解を深めるようにする。

# なぜ情報Ⅱ？ 目標③

(2) 様々な事象を

情報とその結び付きとして捉え、

問題の発見・解決に向けて

情報と情報技術を

適切かつ効果的、**創造的**に活用する力を

養う。



# なぜ情報Ⅱ？ 目標④

- (3) 情報と情報技術を  
適切に活用するとともに、  
新たな価値の創造を目指し、  
情報社会に主体的に参画し、  
その発展に寄与する態度を養う。

# 情報Ⅰと情報Ⅱの違いは

---

情報と情報技術による**新たな価値創造**

情報社会の**発展に寄与する**

# 情報Ⅰと情報Ⅱの違いは

情報と情報技術による**新たな価値創造**

情報社会の**発展に寄与する**

2022年 15歳 → 2045年 38歳

高校生



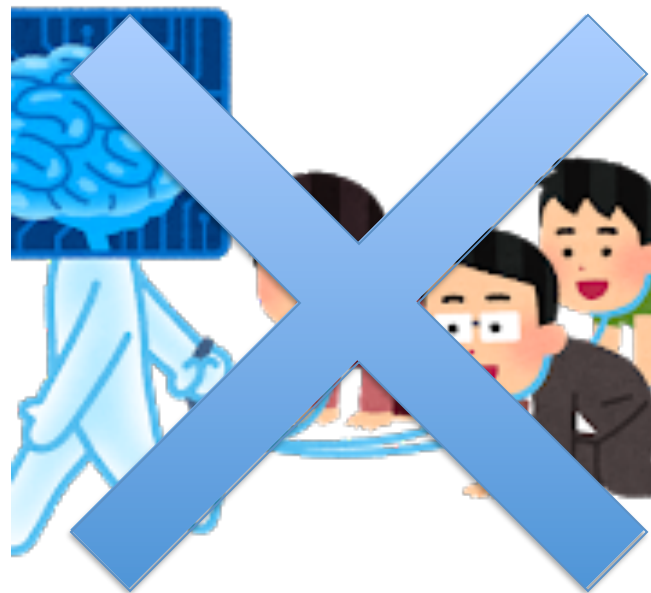
社会の中堅



# 2045年 シンギュラリティ？



or



# ということで

---

- モデル化とシミュレーション

Bag of Words モデル

→ 機械学習(単純ベイズ分類器)

を授業でやってみました。

# 発表内容

---

- イン트로ダクション
- 授業について
- 授業内容
  - 知識・技能の習得の授業  
(形態素解析・特徴語抽出・機械学習)
  - 問題解決の授業
  - 生徒の質問・考えたこと
- まとめ

# なぜ機械学習を授業で行ったか

- 1年前に示された新科目案  
情報Ⅱ (3)情報とデータサイエンス
- 人工知能の仕組みの一端を理解  
データ → 計算結果 思考しているか？
- 人間と人工知能の関係を考える  
結果の捉え方、どのように使っていくか

# 単純ベイズ分類器を選んだ理由

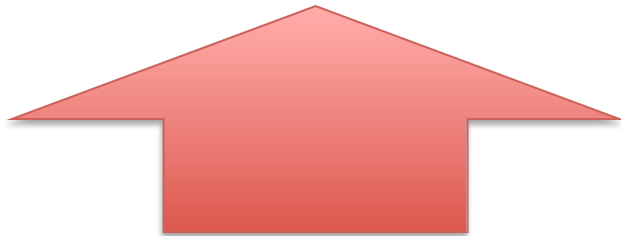
- 対象となるデータが、多様で集めやすい  
文章がデータだから
- コンピュータに文章を答えさせることができている  
人工知能っぽい  
迷惑メールフィルタでも使われている
- 仕組みが高校数学で十分理解できる  
高校数学がダメでも長方形の面積でOK！  
式に惑わされなければ、ただの面積



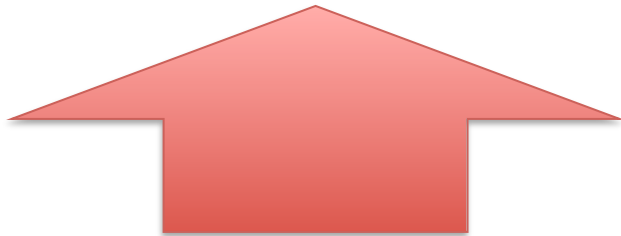
# 授業の流れ(関連内容)

---

グループでの問題解決学習



機械学習の知識と技能の習得



コンピュータが動作する仕組みと  
情報の表現の知識の習得

# 発表内容

---

- イン트로ダクション
- 授業について
- **授業内容**
  - **知識・技能の習得の授業**  
(形態素解析・特徴語抽出・機械学習)
  - 問題解決の授業
  - 生徒の質問・考えたこと
- まとめ

# 授業内容① 機械学習以前

- 文字のデジタル化 → 文字コード（割愛）
- コンピュータが動作する仕組み  
Pythonでのプログラミング

学習サイトProgate

Python I・II

変数・条件分岐

リスト・辞書・繰り返し 程度

# 授業内容② 自然言語処理

---

- 形態素解析
- tf-idf (Term Frequency-Inverse Document Frequency)
- Word Cloud (タグクラウド)

# 授業内容②-1 形態素解析

- 日本語は分かち書きされていない
  - コンピュータでは形態素解析が使える
    - 可能性 = 可能 + 性
    - ディズニーランド = ディズニー + ランド

形態素解析するWebサイト

Python + janome で経験させた

固有名詞や語句を連結した語に弱いことを経験

# 授業内容②-2 tf-idfによる特徴語1

- 難しそう ← 実は数えて分数にするだけ

tf の 発想

重要な語は何回も現れます！

繰り返します。

重要な語は何回も現れます！

繰り返します。

重要な語は何回も現れます！

## 授業内容②-3 tf-idfによる特徴語2

例) いちご : [果物 | ケーキ | ビタミン | ケーキ | 赤い]

出現語数 5語

| 出現語  | 出現回数 | tf  |
|------|------|-----|
| 果物   | 1    | 1/5 |
| ケーキ  | 2    | 2/5 |
| ビタミン |      |     |
| 赤い   |      |     |

**tf = 形態素の出現回数 / 全形態素数**

# 授業内容②-4 tf-idfによる特徴語3

idf の発想 レア

クイズ 次の語と関連するスポーツは何？

- 1) オフサイド →
- 2) スリーポイントシュート →
- 3) ホームラン →
- 4) トライ →
- 5) ボール →

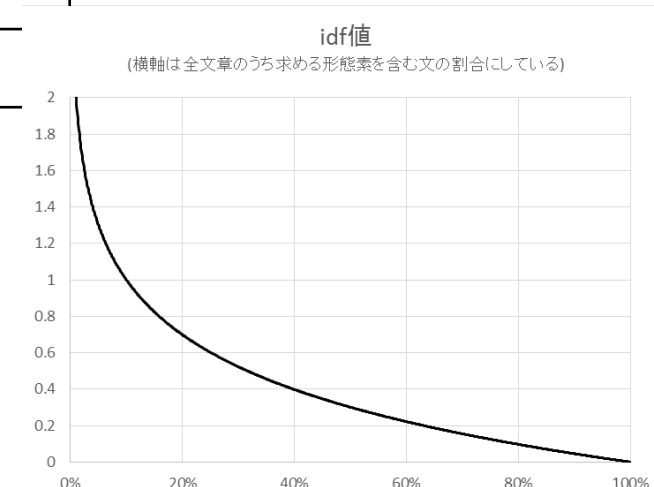


# 授業内容②-5 tf-idfによる特徴語4

例)いちご:[果物 | ケーキ | ビタミン | ケーキ | 赤い]  
りんご:[果物 | ジュース | 青森 | ビタミン | 赤い]  
キウイ:[ビタミン | 毛 | 緑 | 黄色]

| 出現語  | 出現する文の数 | idf         |
|------|---------|-------------|
| 果物   | 2       | $\log(3/2)$ |
| ケーキ  | 1       | $\log(3/1)$ |
| ビタミン |         |             |
| 赤い   |         |             |

idf =  $\log(\text{全文書数} / \text{形態素を含む文の数})$   
tf-idf = tf \* idf



# 授業内容②-6 特徴語の可視化

- Excelでのグラフ化

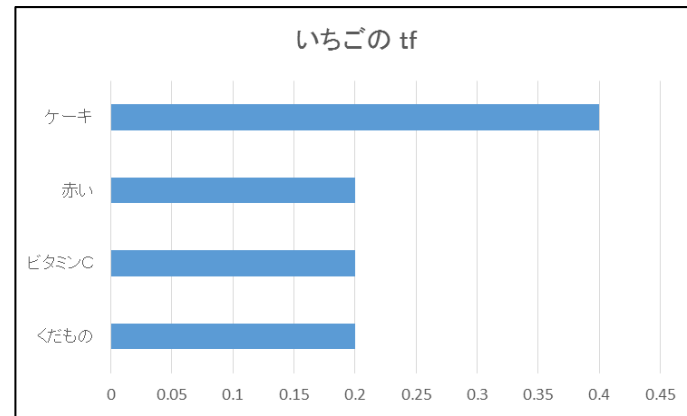
Pythonでtf-idfを計算しExcelに出力

↓

Excel上で手作業でノイズ除去

↓

Excelでグラフ化



- Word Cloud (タグクラウド)

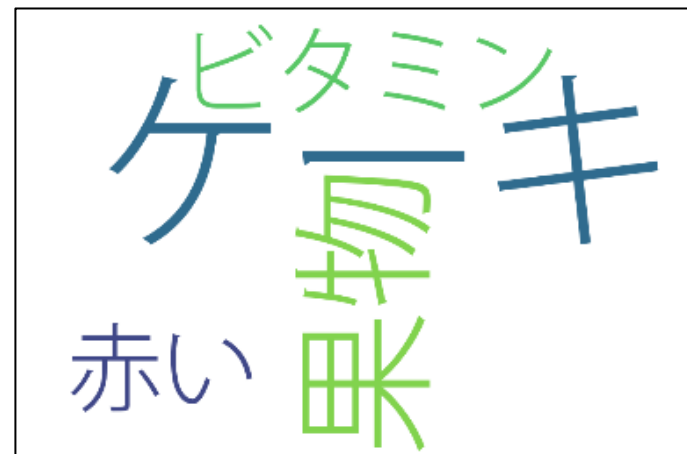
PythonでWord Cloudを出力

↓

プログラムを修正しノイズ除去

↓

再実行してWord Cloudを出力



# ここで突然「情報 I」の話です

## (4)情報通信ネットワークとデータの活用

多量のテキストから有用な情報を取り出す  
テキストマイニングの基礎やその方法を理解

単語の出現頻度について調べさせ、(中略)  
タグクラウドを作らせ、単語の重要度や  
他の単語との関係性を捉える学習活動

# 授業内容②-7 ベイズの定理

## • ベイズの定理

|                        |        |       |
|------------------------|--------|-------|
| 「振り込んでください」<br>が含まれている | 通常のメール | 迷惑メール |
| 割合 (確率)                | 0.05   | 0.8   |

通常のメールで「振り込んでください」という言葉が含まれる確率は

$$0.8 \times 0.05 = 0.04$$

迷惑メールで「振り込んでください」という言葉が含まれる確率は

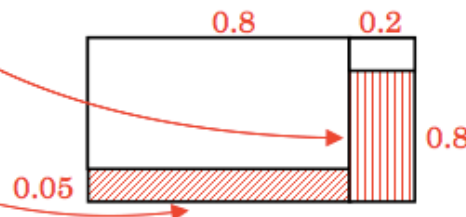
$$0.2 \times 0.8 = 0.16$$

「振り込んでください」という言葉が含まれる確率は

$$0.04 + 0.16 = 0.2$$

「振り込んでください」という言葉が含まれるもののうち、迷惑メールである確率は

$$0.16 / 0.2 = 0.8$$




「振込」から迷惑メールと推測

$$= \frac{\text{迷惑の確率} \times \text{迷惑の中で「振込」が観測される確率}}{\text{「振込」全体の確率}}$$

# 確率って 数学でやればいいんじゃないの？

情報を得て予測が変化するって

完全に「情報科」の範囲じゃないですか！！

事前の予測 (事前確率)  情報を得た後の予測 (事後確率)

その情報が条件である  
条件付き確率

何らかの情報を得る

# 授業内容②-8 単純ベイズ分類器①

## • 単純ベイズ分類器

例) **いちご**: [果物 | ケーキ | ビタミン] 、 [ケーキ | 赤い]

りんご: [果物 | ジュース | 青森 | ビタミン | 赤い]

キウイ: [ビタミン | 毛 | 緑 | 黄色]

分類データ: [果物 | ケーキ] ← **いちごに分類される**

|       | いちご | りんご | キウイ |
|-------|-----|-----|-----|
| 分類出現率 | 2/4 | 1/4 |     |

|       | いちご  | りんご  | キウイ |
|-------|------|------|-----|
| 単語数   | 5    | 5    |     |
| 果物    | 1/5  | 1/5  |     |
| ケーキ   | 2/5  | 1/5  |     |
| 単語出現率 | 2/25 | 1/25 |     |

# 授業内容②-8 単純ベイズ分類器②

## • ゼロ頻度問題 **キウイは果物じゃない?!**

例)いちご:[果物 | ケーキ | ビタミン] 、 [ケーキ | 赤い]

りんご:[果物 | ジュース | 青森 | ビタミン | 赤い]

キウイ:[ビタミン | 毛 | 緑 | 黄色] ←果物がない

分類データ:[**果物** | **ケーキ**]

|       | いちご | りんご | キウイ |
|-------|-----|-----|-----|
| 分類出現率 | 2/4 | 1/4 | 1/4 |

|       | いちご  | りんご  | キウイ |
|-------|------|------|-----|
| 単語数   | 5    | 5    |     |
| 果物    | 1/5  | 1/5  | 0   |
| ケーキ   | 2/5  | 1/5  |     |
| 単語出現率 | 2/25 | 1/25 | 0   |

# アンダーフロー対策

```
>>print(0.1**300) ←0.1の300乗  
0
```

データの扱い方や精度、計算の手順などに目を向けて改善しようとする態度を養う

$$\log(a*b) = \log(a) + \log(b)$$

を使うとよいオーダーで計算できるが  
生徒だけでここまでたどり着けるのか・・・



# 活用・探究につなげるための配慮

## 配慮したこと

- ① ブラックボックス化せず、**仕組みがわかる**
- ② 他教科の内容は**高等学校の範囲**である
- ③ **データの収集が容易**である
- ④ **問題**となる事項と**データの関連**がわかる
- ⑤ **多様な問題**への適用が可能である

# 発表内容

---

- イン트로ダクション
- 授業について
- **授業内容**
  - 知識・技能の習得の授業  
(形態素解析・特徴語抽出・機械学習)
  - **問題解決の授業**
  - 生徒の質問・考えたこと
- まとめ

ここからが本番

## 授業内容③-1 グループでの問題解決

### 課題

「tf-idf と単純ベイズ分類器を、

社会における問題発見や

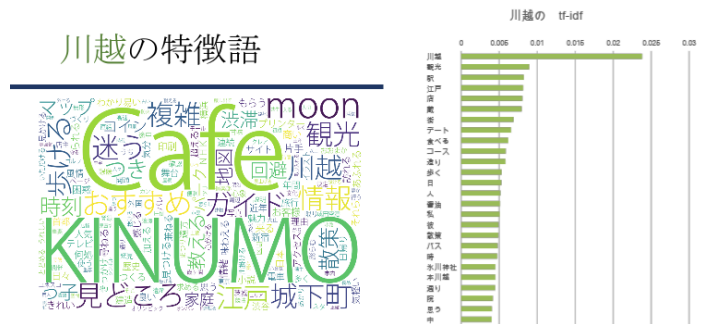
解決への使い方を考えて、

社会における問題解決を提案しなさい」

# 授業内容③-2 例示

初めての試みなのでとりあえず道を作ってみた

例)「川越」「栃木」「佐原」の3つの小江戸  
差別化を図るために特徴語を抽出  
単純ベイズ分類器で観光案内 を例示



ナイーブベイズ分類器による  
お客さんへの提案例

- ◆「東京から近くて食べ物がおいしい街はどこですか？」
  - 川越がおすすめです！
- ◆「自動車で行きたい。観光に合わせて、キャンプができればいいな。」
  - 佐原がおすすめです！
- ◆「山と川の景色を見ながら、1日ゆっくり過ごしたい。」
  - 栃木がおすすめです！

# 授業内容③-3 テーマ決め

- 個人での問題発見



自分の考えを持つ

問題に適用できる確信が持てない

- グループで意見交換・問題発見



他者と話すことで視点が整理される

- テーマを発表（発表者を交替しながら）

# 授業内容③-4 実際に適用する

クローリング

- Webサイトをめぐり、ページを見つける(手作業)

スクレイピング

- データとして必要な部分を抽出(手作業)

データ分析

- 配布したプログラムを用いてデータを処理
- 得られたデータを見ながら特徴を見つける

資料作成

- 問題点や提案などを発表資料にまとめる

# 授業内容③-5 生徒が発見した問題①

- 旅行でどこに行けばよいか？  
（観光地の特徴・旅先案内システム）
- どのような商品を買えばよいか？  
（商品の特徴・購入商品提案システム）
- どの店に行けばよいか？  
（店の特徴・店紹介システム）

## 授業内容③-6 生徒が発見した問題②

- 花粉症の原因はどの花粉か？  
（症状の特徴・症状診断システム）
- 自分の好きな異性をタレントに例えると？  
（タレントの特徴・異性紹介システム）
- 料理にあうさつまいもの品種はどれ？  
（品種の特徴・食材推奨システム）

など





# 発表内容

---

- イン트로ダクション
- 授業について
- **授業内容**
  - 知識・技能の習得の授業  
(形態素解析・特徴語抽出・機械学習)
  - 問題解決の授業
  - 生徒の質問・考えたこと
- まとめ

# 単純ベイズ 生徒の質問例①

- 旅行先で「沖縄」は出るけど、「札幌」が出ないのはどうしてですか？

データの行数を事前確率とするプログラム

人気のものはデータが多いはず

沖縄のデータ 3桁 vs 札幌のデータ 2桁

→ 理由不十分の原則

データの行数を揃えて事前確率を等確率に

# 単純ベイズ 生徒の質問例②

- 「使いやすいペン」と「使いやすすくないペン」の結果はなぜ同じなんですか？

使う / やすい / ペン

使う / やすい / ない / ペン

現れる形態素がほとんど同じだから

→ どうやって工夫したらいいか考えてみよう

# 単純ベイズ 生徒の質問例③

- 朝のおすすめ番組を答えるシステム  
「あさチャン！」と入れたら「ZIP」と答えた

あさチャン！のデータに「あさチャン！」という言葉が入っていないからですね

と生徒自身で解決

# 学習を通して生徒が考えたこと

- 集めるデータ量をなるべく同じくらいにする
- 年代・男女など幅広く情報を集め、  
情報のかたよりをなくす
- なにげなく使っている文房具などの身近なものにも  
寡占状態などを引き起こす可能性があることに  
気づいた。

機械学習のうち1つを経験したことで  
定性的な理解が十分に深まったと考えられる

→ 仕組みの理解 + メディア・リテラシー

# 発表内容

---

- イン트로ダクション
- 授業について
- 授業内容
  - 知識・技能の習得の授業  
(形態素解析・特徴語抽出・機械学習)
  - 問題解決の授業
  - 生徒の質問・考えたこと
- まとめ

# 今後 考えたいこと

- アルゴリズムが決まっているものを  
プログラミングでどう授業で扱うか  
正しくないと動かない・正しい結果が出ない  
Wekaなどのツール活用でもよいのでは？
- 人工知能・機械学習を巡る法整備のあり方  
事故の際の責任所在、著作権等
- 人工知能・機械学習と人との関係  
ビッグデータとプライバシー  
適用範囲(人事評価などへの適用)



# 最後に改めて今日の主張

- **情報Ⅱを開講しましょう！**

(少なくとも生徒が選択できる余地を作りましょう！)

- 情報を深く理解するには、  
他の教科も重要です。

**(特に数学は避けられません！)**

- これまでの「**社会と情報**」の内容も  
**仕組みを知る**と、ものすごく**深まります**！

# 発表内容

---

- イン트로ダクション
- 授業について
- 授業内容
  - 知識・技能の習得の授業  
(形態素解析・特徴語抽出・機械学習)
  - 問題解決の授業
  - 生徒の質問・考えたこと
- まとめ