
情報IIを想定した階層型 クラスタリングの授業実践

— 岸本有生 本多佑希 兼宗進 —

追手門学院大学

四天王寺大学

大阪電気通信大学

背景

- 高等学校「情報II」のデータサイエンスの単元
- 教科書3冊中、全ての教科書で階層型クラスタリングが取り扱われている
 - 語句説明
 - アルゴリズムの説明
 - Pythonを使った実習
- **データ活用の視点で利用するにはPythonの理解が必須**
- **難しすぎると、生徒にとってつまらないものになる**

```
import pandas as pd
from sklearn.preprocessing import StandardScaler
from scipy.cluster.hierarchy import linkage, dendrogram
import matplotlib.pyplot as plt
import japanize_matplotlib

file_path = 'sample.csv'
data = pd.read_csv(file_path, index_col=0)
scaler = StandardScaler()
data_scaled = scaler.fit_transform(data)
linkage_result = linkage(data_scaled, method='centroid',
metric='euclidean')

plt.figure(figsize=(10, 7))
dendrogram(linkage_result, labels=data.index, leaf_rotation=90, leaf_font_size=12)

plt.title('階層型クラスタリングのデンドログラム', fontsize=16)
plt.xlabel('データサンプル', fontsize=12)
plt.ylabel('クラスタ間の距離', fontsize=12)
plt.grid(axis='y')
plt.tight_layout()
plt.show()
```

クラスタリングとは

クラスタリングとは似たもの同士をグループにすることを指します

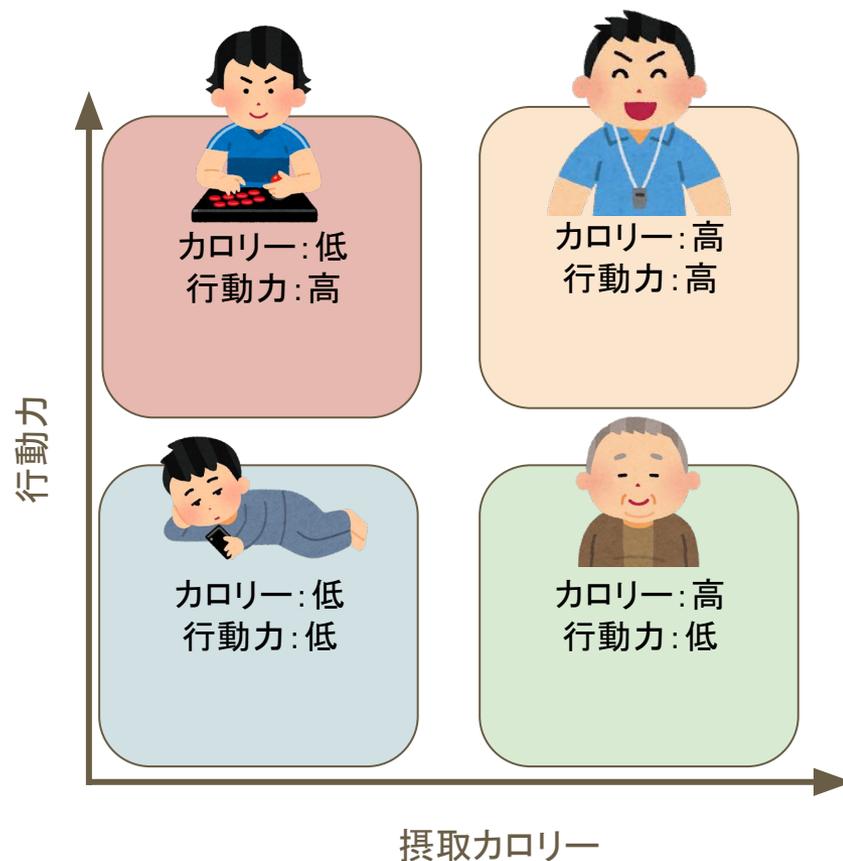
(例)

- 摂取カロリーの数値データ
- 行動力の数値データ

「摂取カロリーが高くて、行動力が高い」

「摂取カロリーが低くて、行動力が低い」

などのようにグループ分けできます



授業提案のポイント

Connect DB

- データ分析学習システムConnect DBに階層型クラスタリングの機能を追加

Point 1: マウスをクリックするだけで操作ができる

Point 2: 授業でそのまま使える、身近なサンプルデータを用意されている

Point 3: 生徒の意思で「納得いくグループ分け」を行える

身近なサンプルデータ

- 飲み物のクラスタリング
 - 飲み物の名前(コーラ、コーヒー、カフェオレ、お茶、スポーツドリンク、エナジードリンク、乳酸菌飲料、ヨーグルトなど)
 - 炭水化物、たんぱく質、脂質
- ガジェットのクラスタリング
 - ガジェットの名前(スマートフォン、スマートウォッチ、イヤホン、タブレット、ノートPC、ゲーム機、電子書籍リーダーなど)
 - 重量、平均消費電力、駆動時間

ConnectDBの使用

Connect DB

- データ分析学習システムConnect DB
- URL : <https://cdb.eplang.jp/>



- サンプルデータ: 飲み物(クラスタリング)
- 実際に見ていただくと助かります

実演をします

階層型クラスタリングの実行

- 以下のものにチェック

- 飲み物
- 炭水化物
- たんぱく質
- 脂質

- 「階層型クラスタリング」を選択

方法:
階層型クラスタリング

クラスターの距離:
重心法

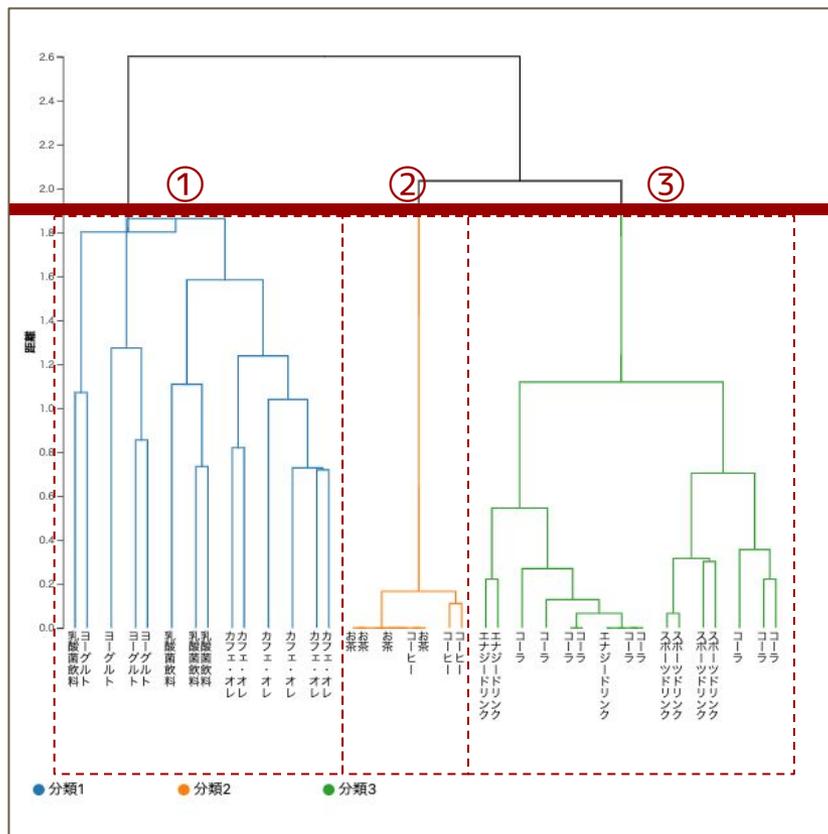
表示

データ削除 入れ替え 型

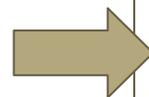
スクロールしてテーブルを見てね

<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> 飲み物	<input checked="" type="checkbox"/> 炭水化物(g/100mL)	<input checked="" type="checkbox"/> たんぱく質(g/100mL)	<input checked="" type="checkbox"/> 脂質(g/100mL)
<input type="checkbox"/>	ヨーグルト	15	3.20	1.70
<input type="checkbox"/>	乳酸菌飲料	14.10	1.50	0.50
<input type="checkbox"/>	エナジードリンク	14	0	0

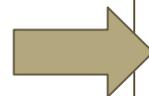
各グループの特徴の分析



平均より
高い



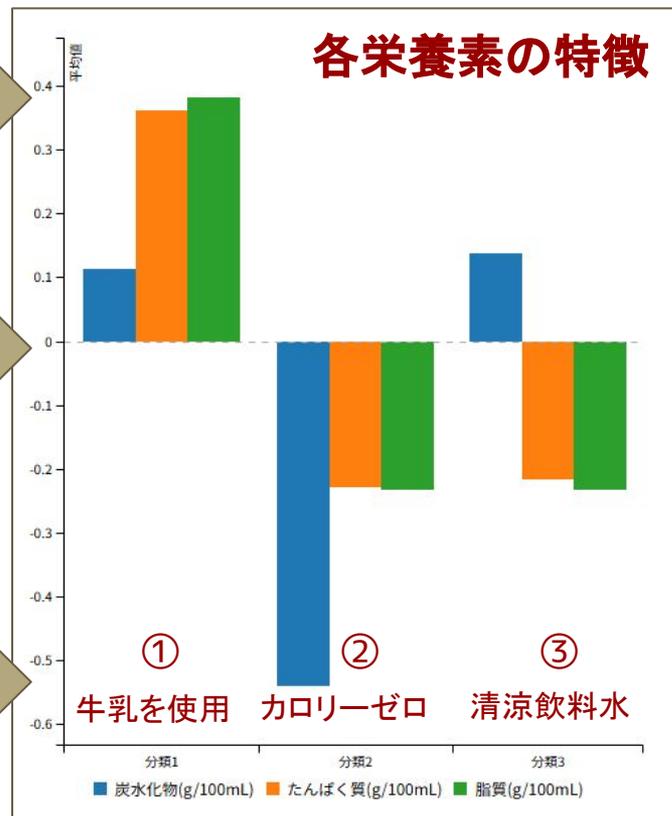
平均



平均より
低い



各栄養素の特徴



飲み物の実践例

三つの栄養素

- 食事に含まれる炭水化物(g)、タンパク質(g)、脂質(g)からエネルギー(kcal)を算出する

家庭基礎などでも取り上げている

$$\begin{array}{|c|} \hline \text{エネルギー} \\ \text{(kcal)} \\ \hline \end{array} = \begin{array}{|c|} \hline \text{炭水化物} \times 4 \\ \hline \end{array} + \begin{array}{|c|} \hline \text{タンパク質} \times 4 \\ \hline \end{array} + \begin{array}{|c|} \hline \text{脂質} \times 9 \\ \hline \end{array}$$

炭水化物

米、小麦、とうもろこし、イモ、**砂糖**

タンパク質

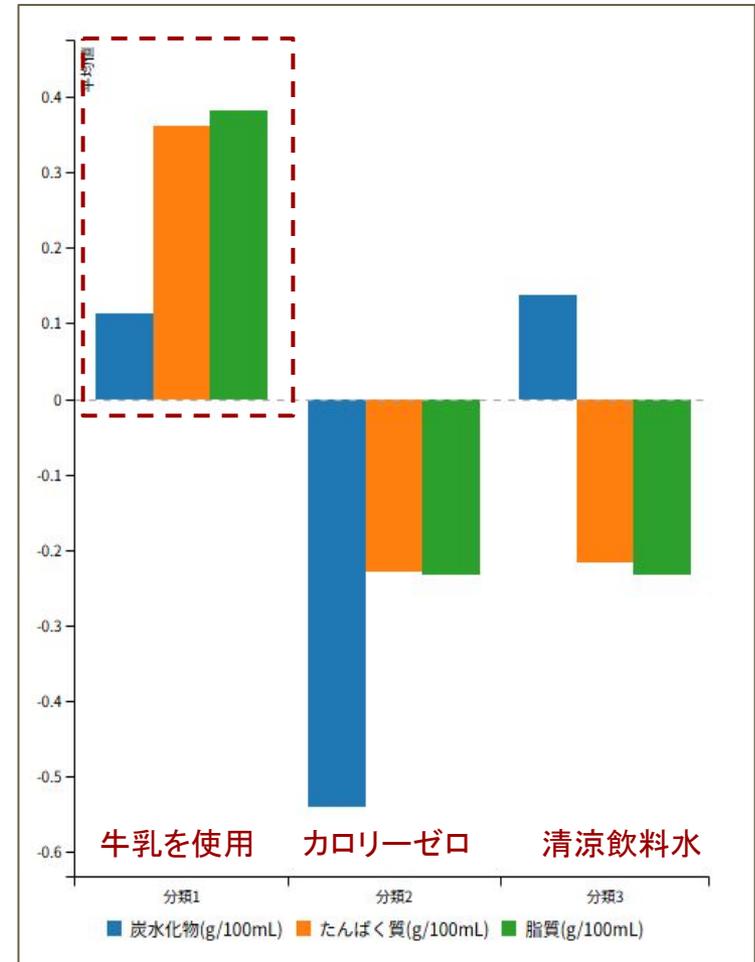
油、**牛乳**、カカオ

脂質

牛乳、卵、大豆、小麦、ゼラチン

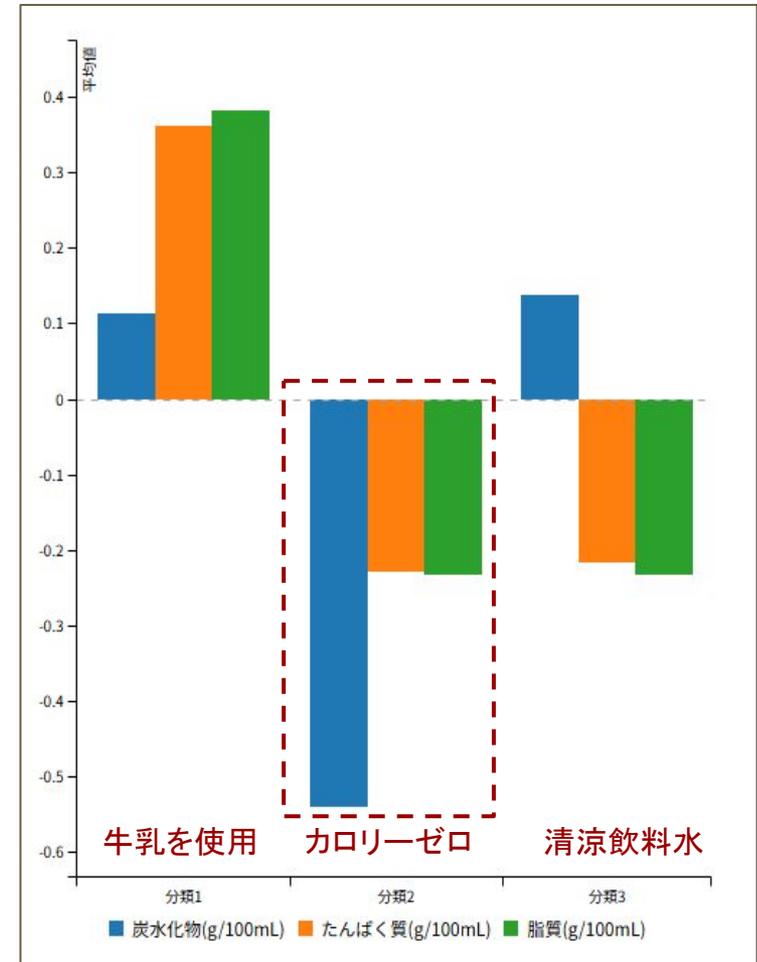
飲み物のクラスタリングの説明例

- 牛乳を使用するグループは、
炭水化物、タンパク質、脂質とも
に平均よりも大きい



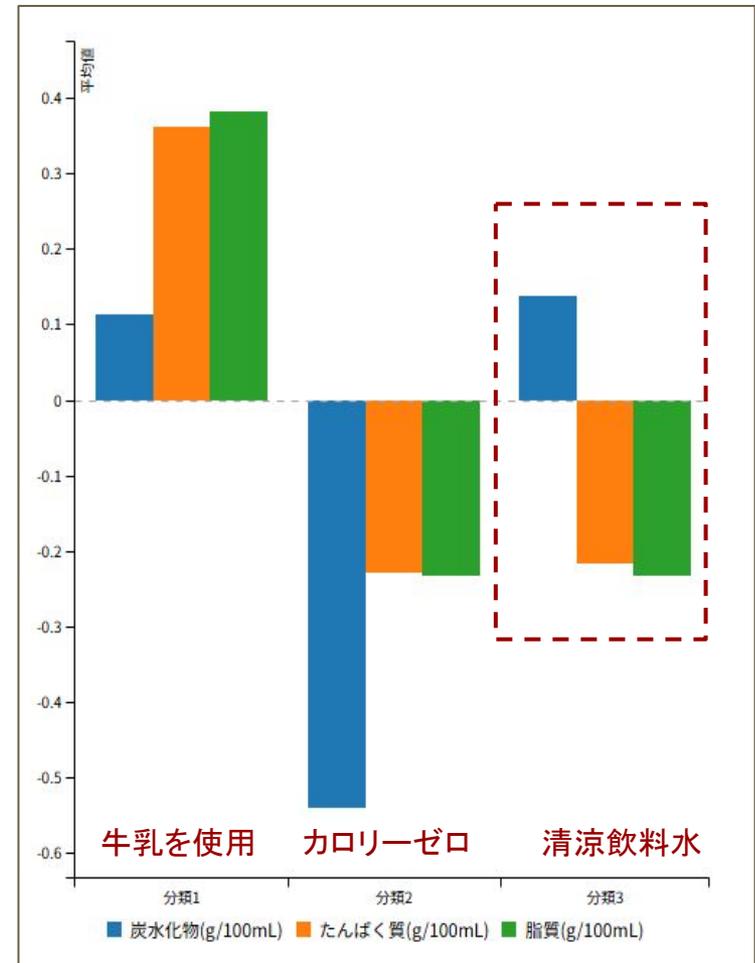
飲み物のクラスタリングの説明例

- 牛乳を使用するグループは、炭水化物、タンパク質、脂質ともに平均よりも大きい
- カロリーゼログループは、**炭水化物、タンパク質、脂質が共に少ない**



飲み物のクラスタリングの説明例

- 牛乳を使用するグループは、炭水化物、タンパク質、脂質ともに平均よりも大きい
- カロリーゼログループは、炭水化物、タンパク質、脂質が共に少ない
- 清涼飲料水は砂糖が使われているため、**炭水化物だけが平均よりも大きい**



大学での実践例

大学での授業実践例

- 今日の発表の内容で説明し、教員が説明をしながら、飲み物のサンプルデータを使って実習を行った
- その後、違うサンプルデータ(ガジェットのクラスタリング)を使って学生が課題に取り組んだ

ガジェットではどんなことがわかる？

- ノートPCは他の商品に比べて重い
- ゲーム機は消費電力が高い → ゲームはGPUを使うから？
- 電子書籍リーダーは駆動時間が長い → Kindleとか？

学生からの感想

- 意外とイヤホンとスマートウォッチは似ていた
- 欲しいものと全体の性能を比較するのに使えそう
- 実際の数値から意外な気づきを発見した

まとめ

- Connect DBを利用すると階層型クラスタリングの授業が可能になる
- 身近なサンプルデータで階層型クラスタリングを利用したデータ活用の授業実践を行った
- その結果、他のグループとの違いを比較することができた
- さらに、これまで考えていなかった気付きも発見した