

ユーザーの予算を考慮した, マーケットバスケット分析による 商品推薦システム

知能情報学実験及び演習 T1/関先生

目次

- 想定状況，やりたいこと
- 提案手法
 - 概要
 - 定義
 - 愚直解法との比較
- デモ
- 振り返り
 - 結果と考察
 - 苦しんだ点
- 今後の課題
- まとめ

想定状況，やりたいこと

■ 想定状況

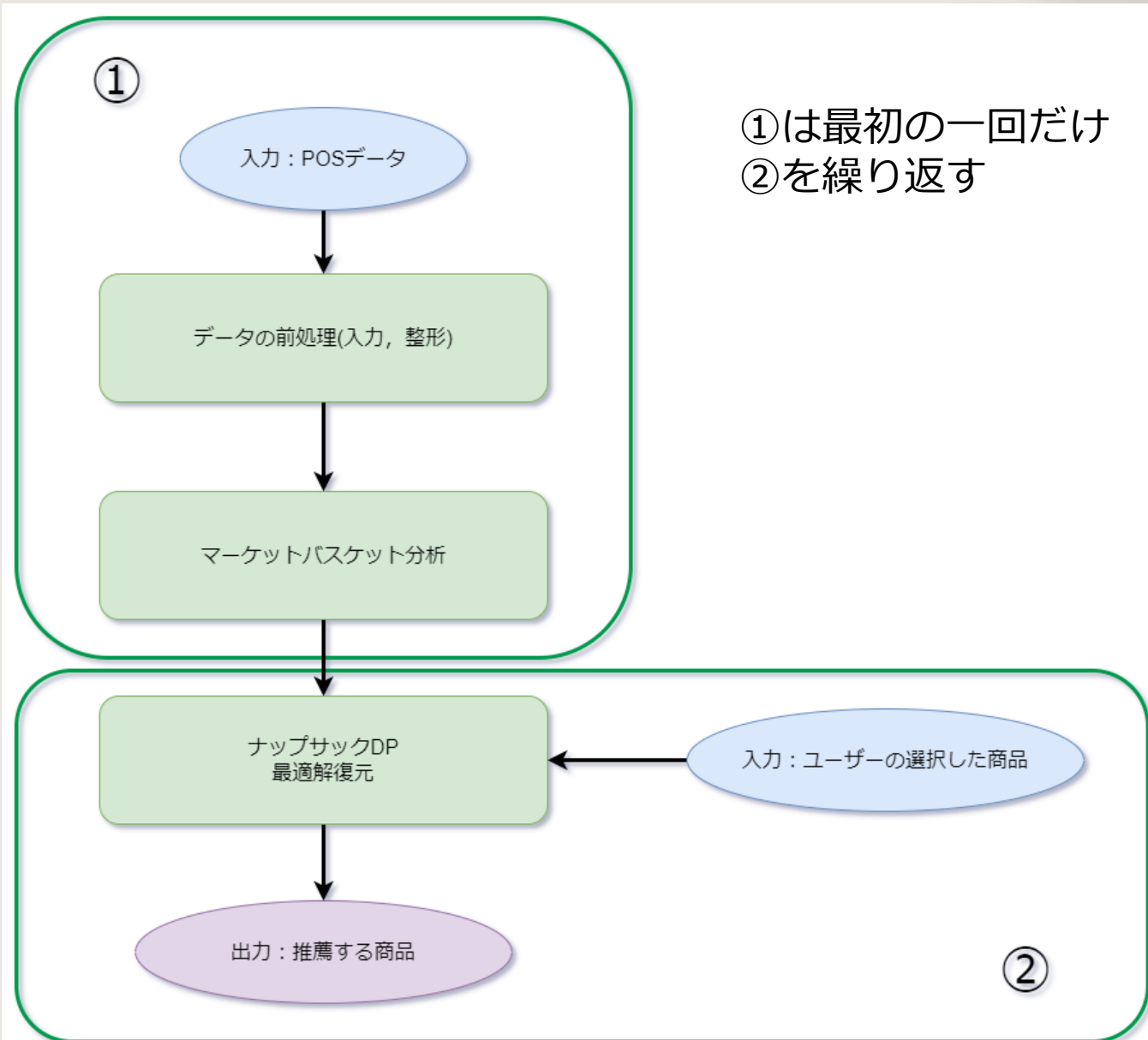
- コンビニに来たユーザーが一つ商品を選んだ
- ユーザーは手持ちのお金であと少し何か買いたい

Ex: 「弁当」を選んだユーザーが，予算500円以内で何か買いたい

■ やりたいこと

- ユーザーの予算内で，おすすめの商品を推薦する
- 「おすすめの商品」を適当ではなく，
データマイニング/データサイエンスを使って求めたい

提案手法 / 概要



提案手法 / 概要

■ 使用したデータ

- コンビニのPOSデータ(レシート of データ)1年分
 - ▶ 約580万件
- 商品名ではなく, 分類名を使用した

Ex: 「冬限定おーいお茶500ml」 → 「ペットボトル入り緑茶飲料」

- ▶ 表記揺れを防ぐ

提案手法 / 定義

■ 各商品のおすすめ度(得られる利益)を

「ユーザーの選択した商品とのconfidence値」と定義する

- confidenceの求め方は以下(mlxtend/frequent_patterns/apriori/)

```
- support(A->C) = support(A+C) [aka 'support'], range: [0, 1]
- confidence(A->C) = support(A+C) / support(A), range: [0, 1]
- lift(A->C) = confidence(A->C) / support(C), range: [0, inf]
```

$$confidence(x \rightarrow y) = \frac{|x| \cap |y|}{|x|}$$

x : x (ユーザーの選択した商品)を買った人の数
y : y (x以外の商品)を買った人の数

■ DPテーブルを以下のように定義する

- DP[i][j] := i番目までの商品考えた時, 予算jで得られる最大の利益
DPで最適値(最大の利益)を求めた後, 最適解(選ぶ商品)を復元する

提案手法 / 愚直解法と比較

- 愚直解法：全探索(bit全探索などで実装可能)
- 計算量を比べてみる

商品の候補数:N / 予算:M とする

	ナップサック DP	bit 全探索
時間計算量	$O(NM)$	$O(2^N)$
	多項式時間アルゴリズム	指数時間アルゴリズム
Ex:N=50,M=500	$O(50 * 500) = O(10^4)$	$O(2^{50}) = O(10^{15})$



デモ

- Google Colabとローカルで実演します
- 時間の都合上
 - 前処理部分は1ヵ月分のデータ(Colab)
 - 推薦システムは事前に用意した1年分のデータ(LOCAL)で動かします

振り返り / 結果と考察

■ 結果

- 思っていたよりも大分手間取ったが、最低限思っていたシステムは構築することが出来た
- すこし結果が微妙な場合もあった

■ 考察

- Support/Confidence/Lift値への理解を深めることが大切(そうしないとパラメータチューニングできない)
- 結果が微妙な原因として、データ数の少なさも考えられる。せめて5年分くらいは試したかったが、1年分でも全体で50分くらいかかったので、これ以上はマシンをパワーアップするしかない？

振り返り / 苦しんだ点

- mlxtendのドキュメントやstackoverflowなどの英語を読むのが大変だった
- でっかいデータ(GB単位)を扱うのに慣れておらず、一つの計算に30分くらいかかるのが初めての経験で色々大変だった

対策案：プログレスバー, pickle, logging, etc...

今後の課題

- 一対一だけでなく、一対多の関係も考慮したい
 - マーケットバスケット分析：length辺りを考えたらいけそうだけど…？
 - DP：そもそも漸化式が立てられない
- 商品名→分類名にしてましになったとはいえ、同じ商品が複数推薦されることがある

Ex:紙パック入りコーヒー飲料とペットボトル入りコーヒー飲料

 - マーケットバスケット分析する前の前処理の段階で、同じ意味のものを一つにまとめる処理(単語同士の意味を比較する自然言語処理)をすれば良さそう？

まとめ

- 予算を考慮した商品推薦システムを作った
- Pythonは使えるつもりだったけど、色々とな
新たな発見があり無知を痛感した
- データ分析楽しかった

参考文献

- [商品分析の手法（ABC分析、アソシエーション分析） | データ分析基礎知識 \(albert2005.co.jp\)](#)
- [Apriori - mlxtend \(rasbt.github.io\)](#)
- [Association rules - mlxtend \(rasbt.github.io\)](#)
- [<https://docs.python.org/ja/3/howto/logging.html>](#)
- [\[https://pandas.pydata.org/docs/user_guide/index.html\]\(https://pandas.pydata.org/docs/user_guide/index.html\)](#)
- [<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html>](#)
- [<https://docs.python.org/ja/3/library/pickle.html>](#)
- [<https://stackoverflow.com/>](#)
- [<https://texclip.marutank.net/>](#)

以上です．ありがとうございました．

