

集合知プログラミング

第2章 推薦を行う

江口晃

はじめに

- 推薦の利用

- オンラインショップでの製品の推薦
- 面白いWebサイトへのリンクの提案
- 音楽や映画を探す際の手助け
- 好みそうなコンサートのアラート

協調フィルタリング

- 協調フィルタリング
 - 大規模な集団を検索
 - 好み似た小集団を発見
 - 小集団の好きなものから推薦リストを作成

嗜好の収集

- 嗜好の情報を表現
 - Pythonのディクショナリをネストして使う
 - 利用者の行動に数値を割り当てる
- **Pythonでの例**

```
animal={'Lisa': {'dog': 2, 'cat': 5},  
        'Jack': {'cat': 3, 'pig': 5} }
```

似ているユーザを探し出す

- 類似性スコア
 - どの程度似ているかを表すスコア
- スコアの計算方法
 - ユークリッド距離
 - ピアソン相関

ユークリッド距離によるスコア(1)

- ユークリッド距離

- 評価したアイテムを軸にとってグラフを表示し、グラフ上に人を配置する。この距離が近いほど嗜好は近くなる。この図を嗜好空間と言う。

- ユークリッド距離の計算式

- 比較するアイテム数をnとすると、以下の計算式で求められる

$$d(a, b) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$$

ユークリッド距離によるスコア(2)

- **Pythonでのユークリッド距離**
 - 二人とも評価しているアイテムのリストを得る
 - すべての差の二乗を足し合わせる
 - ユークリッド距離は値が小さいほど似ている。似ているほど大きい値を返す関数にするため、1を加え(0乗算を防ぐため)逆数を取る。
 - この関数は常に0から1の間の値を返す。値が1の場合、二人の好みは全く同じであることを表す。

ピアソン相関によるスコア(1)

- ピアソン相関係数
 - 二つのデータセットがある直線にどの程度沿っているかを示す
- ピアソン相関係数とユークリッド距離の比較
 - ユークリッド距離よりも数式が複雑
 - 正規化されていないデータでは、ユークリッド距離よりもよい結果を得られることが多い

ピアソン相関によるスコア(2)

- ピアソン相関係数の計算式

- 比較するアイテム数を n とすると、以下の計算式で求められる

$$num = \sum_{i=1}^n a_i b_i - \sum_{i=1}^n a_i \sum_{i=1}^n b_i / n$$

$$den = \sqrt{\left(\sum_{i=1}^n a_i^2 - \left(\sum_{i=1}^n a_i\right)^2 / n\right) \left(\sum_{i=1}^n b_i^2 - \left(\sum_{i=1}^n b_i\right)^2 / n\right)}$$

$$r = num / den$$

ピアソン相関によるスコア(3)

- **Pythonでのピアソン相関**

- 二人とも評価しているアイテムのリストを得る
- 要素の数を調べる
- ピアソンによるスコアを計算する
- この関数は-1から1の間の値を返す。値が1なら両者はすべてのアイテムに対して全く同じ評価である。

どちらの類似性尺度を利用すべきなのか？

- 類似性の使い分け

- 類似性を計る方法は数多く存在する。作るアプリケーションに合った方法に切り替える。

- 例：Jaccard係数、マンハッタン距離など

評価者をランキングする

- 評者のランキング
 - ある人に対するすべて人々のスコアを算出し、もっとも近い組を探す。
- **Python**でのランキング
 - 特定の人に対するすべて人々のスコアを算出する。
 - スコアの結果を降順にソートする。
 - ソートされた結果の中から最初のn個の結果を返す。

アイテムを推薦する(1)

- **評者のランキングの欠点**
 - 似た嗜好の人を探し、そこから欲しいアイテムを選ぶので手間が掛かる。
 - 嗜好に合うはずのアイテムを得ていない評者を選んでしまう可能性がある。
 - ランキングで選ばれた評者だけが、他の評者と違う評価をしている可能性がある。
- **アイテムの推薦**
 - アイテムの評価と評者との類似性をかけあわせて、アイテムにスコアをつける。嗜好の似ている人の評点の方が、全体に大きく影響を与えるようになる。

アイテムを推薦する(2)

- **Pythonでのアイテム推薦**
 - 特定の人に対するすべて人々のスコアを算出する。
 - 特定の人はまだ得ていないアイテムを選ぶ
 - 類似度 × アイテムを足し合わせる
 - 類似度を足し合わせる
 - 類似度 × アイテムの合計を、類似度の合計で割って、リストを作る。
 - リストを降順にソートしてリストを返す

アイテムを推薦する(3)

- アイテム推薦のその他の特徴
 - それぞれのアイテムに対する、特定の人の評点の予想を得られる。
 - 結果はどの類似性尺度を選んでもあまり大差がない