# ネットワークコンピューティング(2)
# 情報推薦

関西学院大学理工学部情報科学科
北村泰彦



# 演習問題の解答

- ベクトル空間モデルの例において，$d_3$の文書を得ようとして，“Genes and Genomes”を検索質問文として検索を行った．
  1. 類似度0.85以上の文書を検索結果とするときの，再現率と適合率を求めよ．
     再現率=　　，適合率=　　
  2. 類似度0.8以上の文書を検索結果とするときの，再現率と適合率を求めよ．
     再現率=　　，適合率=　　
  3. 類似度0.5以上の文書を検索結果とするときの，再現率と適合率を求めよ．
     再現率=　　，適合率=　　



# ベクトル空間モデル

- コサイン尺度を用いた場合の類似度計算

$$\cos(\boldsymbol{d}_1, \boldsymbol{q}) = \frac{1}{\sqrt{3}\sqrt{2}} = 0.408$$
$$\cos(\boldsymbol{d}_2, \boldsymbol{q}) = \frac{1}{\sqrt{5}\sqrt{2}} = 0.316$$
$$\cos(\boldsymbol{d}_3, \boldsymbol{q}) = \frac{2}{\sqrt{3}\sqrt{2}} = 0.816$$
$$\cos(\boldsymbol{d}_4, \boldsymbol{q}) = \frac{3}{\sqrt{6}\sqrt{2}} = 0.866$$
$$\cos(\boldsymbol{d}_5, \boldsymbol{q}) = \frac{1}{\sqrt{2}\sqrt{2}} = 0.5$$
$$\cos(\boldsymbol{d}_6, \boldsymbol{q}) = \frac{0}{\sqrt{2}\sqrt{2}} = 0$$



# 推薦システム

- 現在，インターネット上は　　(information overload)の状況に陥っており，利用者は必要な情報を見つけ出すことができない．
- 　　(recommender system)とは，利用者にとって有用と思われる対象，情報，または商品などを選び出し，それらを利用者の目的に合わせた形で提示するシステムである．

# 推薦システム

- [　　]は，利用者主導で情報を探し出すこと．検索結果は利用者の想定内．
- [　　]は，システム主導で情報を提供すること．利用者が想定しない情報を入手できることもある．例：新刊書．
- 現在，電子商取引の発展，少量多品種の消費傾向に伴い，情報推薦の重要性が高まっている．
  “If I have 3 million customers on the Web,
  I should have 3 million stores on the Web.”
  (Jeff Bezos, Amazon.com CEO)



# 推薦システムの分類
## 個人化の度合い

- [　　](no personalization)：全ての利用者に対して，同じ推薦を行う．編集者による推薦，売り上げ順位リスト．Apple Store(http://store.apple.com/jp/)
- [　　](ephemeral personalization)：システムを利用する一つのセッションで同じ振る舞いをした利用者に，同じ推薦を行う．Amazon.com
- [　　](persistent personzalization)：利用者の個人情報や過去の利用履歴に応じて異なる推薦を行う．Amazon.com



# 推薦システムの分類
## 推薦手段の分類

- [　　](broad recommendation)：全体の統計情報(「今週の売り上げランキング」)や編集者からの情報提供(「評論家が推薦する映画」)．システム初心者への推薦．
- [　　](user comments and ranking)：利用者間での相互推薦．利用者の批評文や評価レート．利用者同士の推薦の方が受け入れられやすい．



# 推薦システムの分類
## 推薦手段の分類

- [　　](notification service)：利用者がシステムを操作していないときに，電子メールなどで推薦を配送する．利用者のシステムの再利用を促す．
- [　　](item-associated recommendation)：利用者が注目しているアイテムの比較候補を示すことで，購入の判断支援や関連商品の購入を促す．
- [　　](deep personalization)：システムが利用者の情報や過去の履歴を収集し，それに基づき推薦を行う．個人向け推薦リスト．他のシステムとの差別化につながる．

# 推薦システム設計の要素
# 推薦の評価尺度

- [　　]：推薦したアイテムに利用者がどの程度関心を持つか．適合率と再現率．
- [　　](serendipity)：利用者が知っているアイテムを推薦しても意味がない．セレンディピティとは目新しさ，思いがけなさ，意外性を表す．
- [　　](coverage)：全アイテムのうち，推薦評価値の予測が可能なアイテムの割合．



# 推薦システム設計の要素
# 推薦の評価尺度

- [　　](learning rate)：嗜好データの増加に伴って予測精度は向上する．その向上の度合いを学習率と呼ぶ．実用的な予測精度に達するまでに必要な嗜好データの数．



# 推薦システムの実行過程

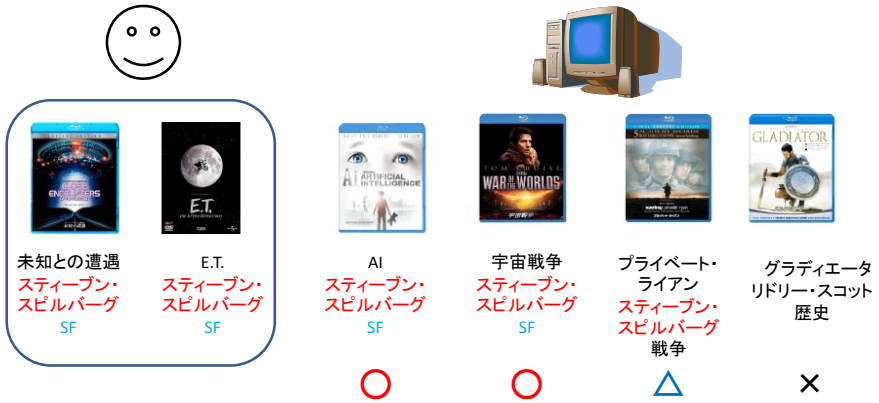
1. [　　]：推薦システムを利用して推薦を受けようとする人を推薦利用者と呼ぶ．推薦利用者は自身の嗜好データ(preference data)を推薦システムに入力する．嗜好データとはいろいろなアイテムについての関心や好みの度合いを数値化したデータである．
2. [　　]推薦利用者の嗜好データに加えて，収集しておいた他の利用者の情報やアイテムの情報を利用して，推薦利用者がまだ知らないアイテムへの嗜好を予測する．
3. [　　]予測した嗜好に基づいて，目的に応じた適切な形式で，推薦結果を推薦利用者に提示する．



# 嗜好の予測

- [　　](content-based filtering)：推薦利用者の嗜好データと推薦アイテムを直接比較して，嗜好データと類似性の高いアイテムを推薦する．
- 映画を推薦する場合，推薦利用者に好きな監督・俳優やジャンルを尋ねてから，その条件にあった映画を推薦する．

# 内容ベースフィルタリング



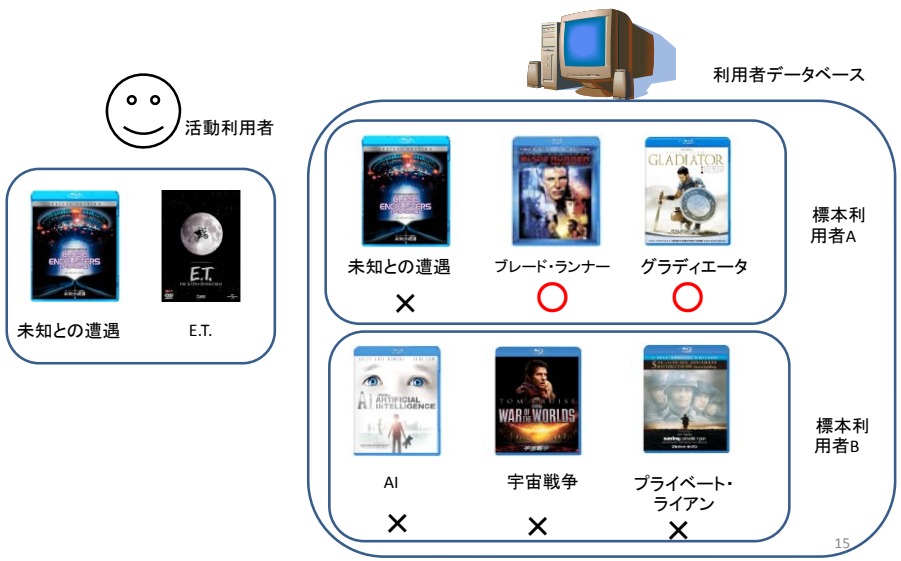



# 嗜好の予測

- [　　](collaborative filtering)：推薦利用者の嗜好データと類似している別の利用者を見つけ出し，推薦利用者が好むアイテムを推薦する．
- 映画を推薦する場合，映画の趣味があう知り合いに映画を推薦してもらう．



# 協調フィルタリング






# 協調フィルタリングと内容ベースフィルタリングの比較

| | 協調フィルタリング | 内容ベースフィルタリング |
|---|---|---|
| セレンディピティ | ○ | × |
| ドメイン知識が不要 | ○（アイテムに関する知識が不要） | × |
| Cold-start問題（新しい利用者やアイテム）への対応 | × | ○ |
| 少ない利用者数 | × | ○ |
| 被覆率 | ×（評価されていないアイテムを推薦できない） | ○ |
| 類似アイテム | ×（異なる色の商品は異なる商品とされる） | ○ |
| 少数派の利用者 | ×（少数派の嗜好パターンは無視される） | ○ |

# 協調フィルタリング

1. [　　]: 利用者データベースの各利用者と推薦利用者の嗜好の類似度を求める. 類似度とは，嗜好パターンがどれほど似ているかを定量化したものである.
2. [　　]: 推薦利用者が知らないアイテムについて，それらのアイテムに対する利用者の好みと，その利用者と推薦利用者の間の類似度に基づいて，推薦利用者がそのアイテムをどのくらい好むかを予測する.



# 協調フィルタリング

- $n$人の全利用者の集合を$X = \{1, \cdots, n\}$, $m$種類のアイテムの集合を$Y = \{1, \cdots, m\}$とする. 評価値行列$S$は利用者$i \in X$のアイテム$j \in Y$への評価値$s_{ij}$を要素とする行列である. $s_{ij}$は評価済みなら評価値の定義域Rのいずれかの値を取り，未評価なら欠損値*をとる. 推薦利用者を $a$ で表す. すなわち，$s_{aj}$は推薦利用者のアイテム$j$への評価値である. 利用者$i$と推薦利用者が評価済みのアイテムの集合を，それぞれ$Y_i = \{j | s_{ij} \neq *\}$と$Y_a$で表す.



# 協調フィルタリング

- 推薦利用者と利用者$i$の類似度は，共通に評価しているアイテムについてのPearson相関で測る.

$$\rho_{ai} = \frac{\sum_{k \in Y_{ai}} (s_{ak} - \bar{s}'_a)(s_{ik} - \bar{s}'_i)}{\sqrt{\sum_{k \in Y_{ai}} (s_{ak} - \bar{s}'_a)^2} \sqrt{\sum_{k \in Y_{ai}} (s_{ik} - \bar{s}'_i)^2}}$$

- ここで，$Y_{ai}$は二人が共通に評価したアイテムの集合，すなわち$Y_{ai} = Y_a \cap Y_i$.また$\bar{s}'_i = \sum_{j \in Y_{ai}} s_{ij} / |Y_{ai}|$である. なお，$|Y_{ai}| \leq 1$ならば，$\rho_{ai} = 0$とする.



# 協調フィルタリング

- アイテム$j \notin Y_a$の評価式は$\rho_{ai}$で重み付けした，各利用者のアイテム$j$への評価値の加重平均で予測する.

$$\hat{s}_{aj} = \bar{s}_a + \frac{\sum_{i \in X_j} \rho_{ai} (s_{ij} - \bar{s}'_i)}{\sum_{i \in X_j} |\rho_{ai}|}$$

- ただし$X_j$はアイテム$j$を評価済みの利用者の集合で，$\bar{s}_a = \sum_{j \in Y_a} s_{aj} / |Y_a|$である.

# 協調フィルタリング

| | 1:親子丼 | 2:牛丼 | 3:海鮮丼 | 4:カツ丼 |
|---|---|---|---|---|
| 1:山田 | 1 | 3 | * | 3 |
| 2:田中 | * | 1 | 3 | * |
| 3:佐藤 | 2 | 1 | 3 | 1 |
| 4:鈴木 | 1 | 3 | 2 | * |

上の表は，$R=\{1,2,3\}$とする評価値行列$S$である．推薦利用者を2:田中($a=2$)としたとき，2:田中の親子丼への推定評価値$\hat{s}_{2,1}$を求めよ．



# 協調フィルタリング

- 親子丼を評価済みの利用者（$X_1$に含まれる利用者）と推薦利用者の間の相関係数を求める．
- 1:山田，3:佐藤，4:鈴木の3人とも親子丼を評価済みなので，$X_1=$ [ ] である．
- 2:田中と1:山田の相関$\rho_{2,1}$は，共通に評価しているアイテムが2:牛丼だけなので，$\rho_{2,1}=$ [ ] である．



# 協調フィルタリング

- 次に，2:田中と3:佐藤の相関を計算する．この二人がともに評価しているアイテムは2:牛丼と3:海鮮丼なので，$Y_{2,3}=$ [ ] となる．これらのアイテムについての$Y_{2,3}$上の平均評価値はそれぞれ以下の通りである．

$$\bar{s}'_2=\frac{\sum_{k=2,3} s_{2,k}}{2}=$$ [ ]

$$\bar{s}'_3=\frac{\sum_{k=2,3} s_{3,k}}{2}=$$ [ ]



# 協調フィルタリング

- したがって相関は

$$\rho_{2,3}=\frac{\sum_{k=2,3}(s_{2,k}-\bar{s}'_2)(s_{3,k}-\bar{s}'_3)}{\sqrt{\sum_{k=2,3}(s_{2,k}-\bar{s}'_2)^2}\sqrt{\sum_{k=2,3}(s_{3,k}-\bar{s}'_3)^2}}$$

$=$ [ ]

$=$ [ ]

- 同様に計算すると2:田中と4:鈴木の相関は$\rho_{2,4}=$ [ ] となる．

# 協調フィルタリング（追加）

- 同様に計算すると2:田中と4:鈴木の相関は

$$\rho_{2,4} = \frac{\sum_{k=2,3}(s_{2,k} - \bar{s}'_2)(s_{4,k} - \bar{s}'_4)}{\sqrt{\sum_{k=2,3}(s_{2,k} - \bar{s}'_2)^2}\sqrt{\sum_{k=2,3}(s_{4,k} - \bar{s}'_4)^2}}$$

$$= \frac{(1-2)\left(3-\frac{5}{2}\right) + (3-2)(2-\frac{5}{2})}{\sqrt{(1-2)^2+(3-2)^2}\sqrt{(3-\frac{5}{2})^2+(2-\frac{5}{2})^2}}$$

$$= -1$$

- ここで$\bar{s}'_4 = \frac{5}{2}$



# 協調フィルタリング

- 次に推定評価値を計算する．まず，2:田中の全評価済みアイテム上の平均評価値を求める．

$$\bar{s}_2 = \frac{\sum_{k=2,3} s_{2,k}}{2} =$$

- したがって，

$$\hat{s}_{2,1} = \bar{s}_2 + \frac{\sum_{i=1,3,4}\rho_{2,i}(s_{i,1} - \bar{s}'_i)}{\sum_{i=1,3,4}|\rho_{2,i}|}$$

$$=$$

$$=$$

- よって2:田中は1:親子丼が好きであると予測される．



# 参考文献


- 神嶌敏弘：推薦システムのアルゴリズム(1)，人工知能学会誌，22(6):826-837, 2007.
- 神嶌敏弘：推薦システムのアルゴリズム(2)，人工知能学会誌，23(1):89-103, 2008.
- 神嶌敏弘：推薦システムのアルゴリズム(3)，人工知能学会誌，23(2):248-263, 2008.