

(補足資料)  
KDDI研究所における  
レコメンデーション技術への取り組み

株式会社KDDI研究所

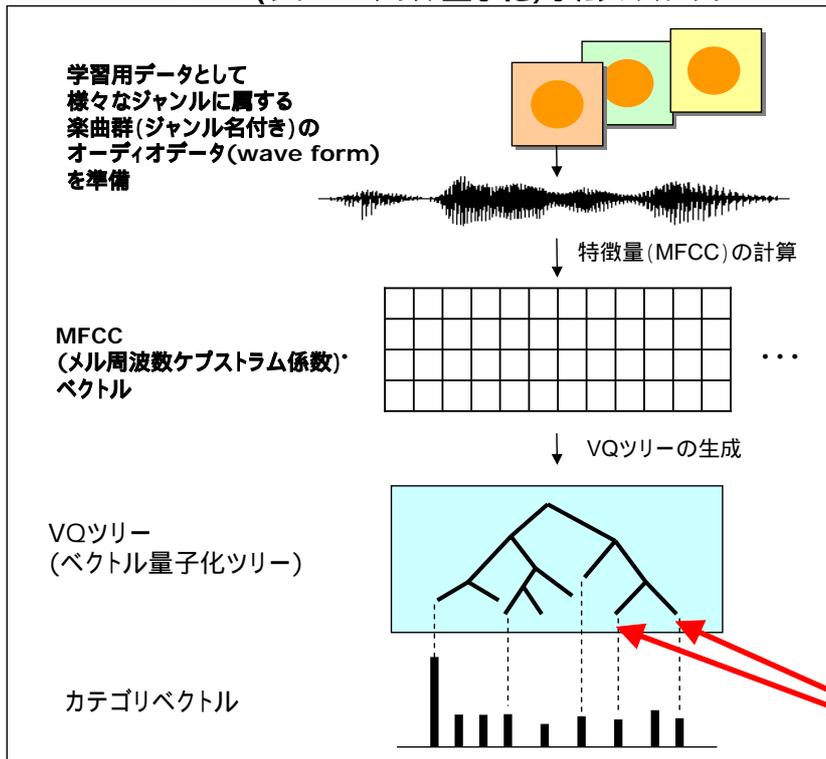
小野 智弘

2008.9.2@行動計量学会チュートリアル

## 推薦システムの基本要素

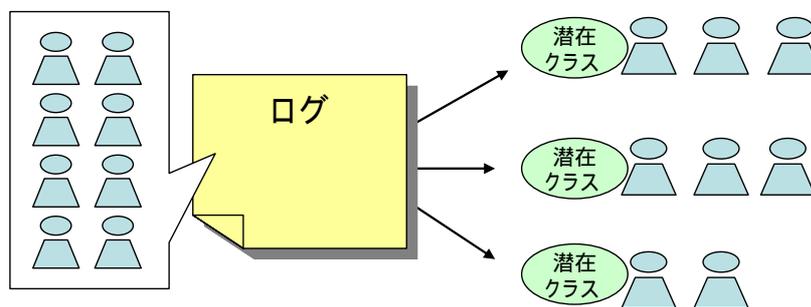
- 推薦対象のコンテンツ
  - 事例1:音楽推薦 = 楽曲
  - 事例2:ブログ推薦 = 記事
  - 事例3:クロスメディアレコメンデーション = 商品
  - 事例4:シーン別映画推薦 = 映画
- 推薦システムの動作
  1. 「推薦フロー」:ユーザからの推薦要求を受けたときの処理
  2. 「登録フロー」:新規コンテンツ登録時の特徴抽出などの処理
  3. 「内部フロー」:推薦システムの内部処理
- 推薦システムの作り方
  4. 「開発フロー」:推薦システムやそれらを構成する解析モジュール、特徴量抽出モジュール、分類モデルなどの作り方

TreeQ(ツリーベクトル量子化)手法のステップ



学習用データをそれぞれのジャンルに分類するための二分木を作成  
 各Leafが一つのカテゴリ (= ジャンル)

- 潜在クラスの抽出
  - 潜在クラスモデル[4]を適用し、分析基、レコメンド基それぞれから潜在クラスを抽出



分析基データの潜在クラス $L_a$ とレコメンド基データの潜在クラス $L_r$ は次のように表現

$$L_a = \{t_{a_1}, t_{a_2}, \dots, t_{a_c}\} \dots (1)$$

$$L_r = \{t_{r_1}, t_{r_2}, \dots, t_{r_d}\} \dots (2)$$

潜在クラス  $t$  の特徴ベクトル表現 ( $n$  次元: 共通ユーザ数分)

$$\vec{t} = (q_t^{U_1}, q_t^{U_2}, \dots, q_t^{U_n}) \dots (3)$$

$U_i$ : 分析基とレコメンド基の共通ユーザ

$q_t^{U_i}$ : ユーザ  $U_i$  が潜在クラス  $t$  に所属する確率

$L_a$  = 分析基データ内の潜在クラス

$L_r$  = レコメンド基データ内の潜在クラス

$c$  = 分析基データ内の潜在クラス数

$d$  = レコメンド基データ内の潜在クラス数

$n$  = 共通ユーザ数