

文献紹介

Social Information Filtering: Algorithm for Automating “Word of Mouth”

著者

Upendra Shardanand, Pattie Maes
MIT Media Lab.

出展

Human Factors in Computer Systems
CHI'95 Conference Proceedings

紹介者

情報システム学専攻 ロボティックス講座
9751090 福原 知宏

はじめに

- ⌘ Social Information Filtering
- ⌘ 実装システムRINGO
- ⌘ 評価手法の比較実験
- ⌘ まとめ
- ⌘ 今後の研究について

Social Information Filtering(SF)

⌘ 社会的評価に基づく情報フィルタリング

- ☑ ユーザの評価を予測
- ☑ ユーザにとって未知の情報を提示

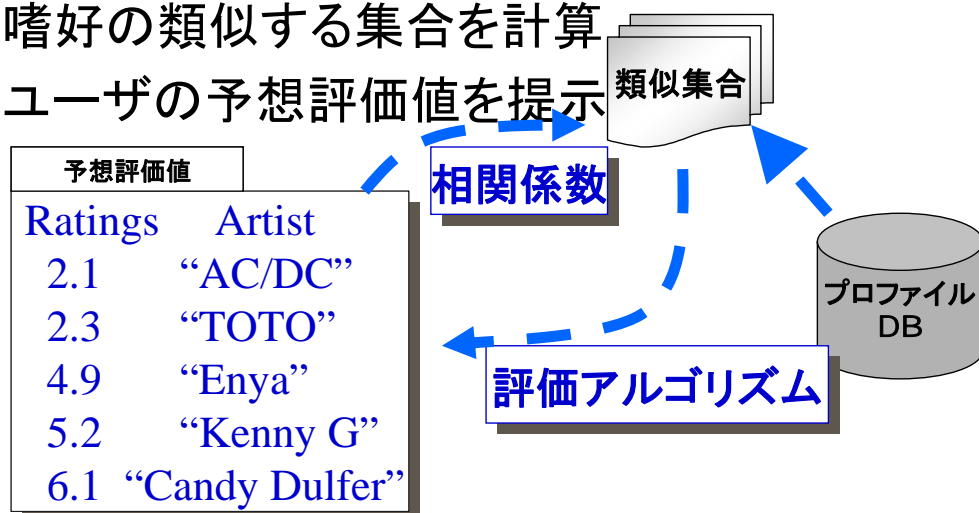
⌘ 利点

- ☑ 構文解析・意味解析を必要としない
- ☑ 精度の高い、新規性のある情報提供

RINGO: 音楽推薦システム

⌘ アルバム・アーティストの提案

- ☑ プロファイルに評価を記述(1:worst~7:best)
- ☑ 嗜好の類似する集合を計算
- ☑ ユーザの予想評価値を提示



評価実験

⌘ 対象データ

- ☑ プロファイル数1,000、データ数1,876
- ☑ 全プロファイルの80%を既知、残りを予測

⌘ 評価基準

- ☑ 平均誤差 $|\bar{E}| = \frac{\sum_{i=1}^N |e_i|}{N}$

$$\sigma = \frac{\sqrt{(\sum (E - |\bar{E}|)^2)}{N}$$

☑ 標準偏差

☑ 予測率

- ☑ 予測できた割合

$$E = \{e_1, \dots, e_N\}$$

$$= \{p_1 - r_1, \dots, p_N - r_N\}$$

評価手法の目標

⌘ 極値：6以上、2以下であるような評価

☑ 嗜好の現れる値

- ☑ 6以上：高嗜好(好き)
- ☑ 2以下：低嗜好(嫌い)

⌘ 極値における平均誤差・標準偏差の改善

評価アルゴリズム

- ⌘ 平均
- ⌘ 最小2乗誤差法
 - ⊠ 先行研究で採用
- ⌘ ピアソン積率相関係数
- ⌘ アーティスト間相関係数

精度向上のため
閾値Lを採用

平均

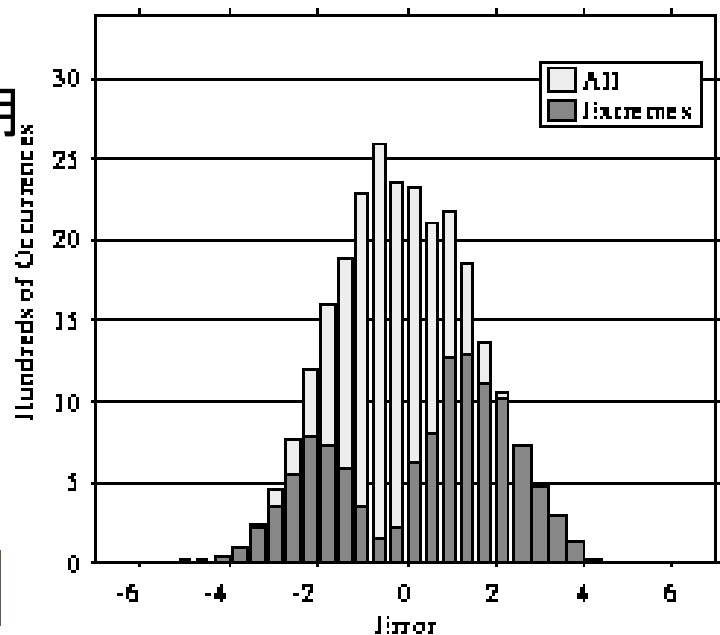
- ⌘ 各アーティストの平均評価値を採用

⌘ 結果

- ⊠ 平均誤差1.8
- ⊠ 標準偏差2.0
- ⊠ 予測率90%
- ⊠ 極値に適用不可

最小2乗誤差法

→ 平均
最小2乗誤差法
ピアソン積率相関係数
アーティスト間相関係数



最小2乗誤差法

平均
最小2乗誤差法
ピアソン積率相関係数
アーティスト間相関係数

⌘ ユーザ間の
非類似度を求める

⌘ 結果(L=2.0)

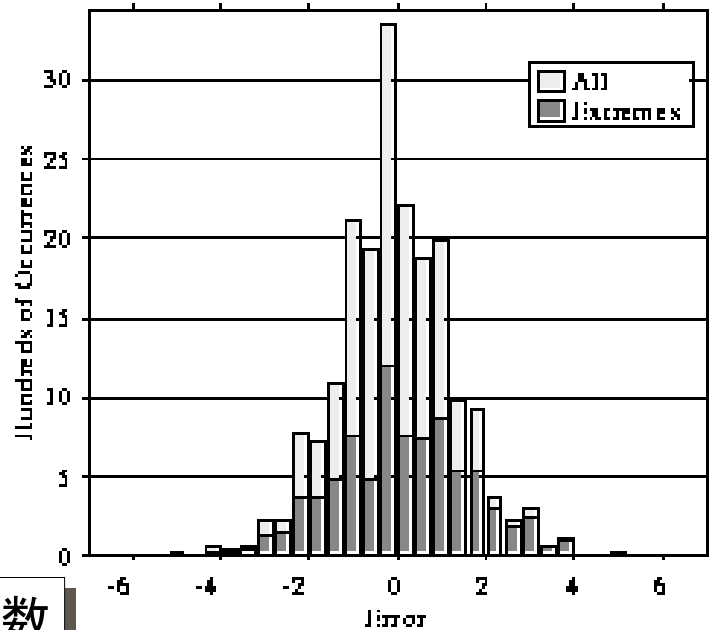
⊠ 平均誤差1.2

⊠ 標準偏差1.6

⊠ 予測率70%

⊠ 正・負の相関の
欠如

ピアソン積率相関係数



ピアソン積率相関係数

平均
最小2乗誤差法
ピアソン積率相関係数
アーティスト間相関係数

⌘ 正・負の相関の考慮

⊠ 相関係数は
-1.0~+1.0

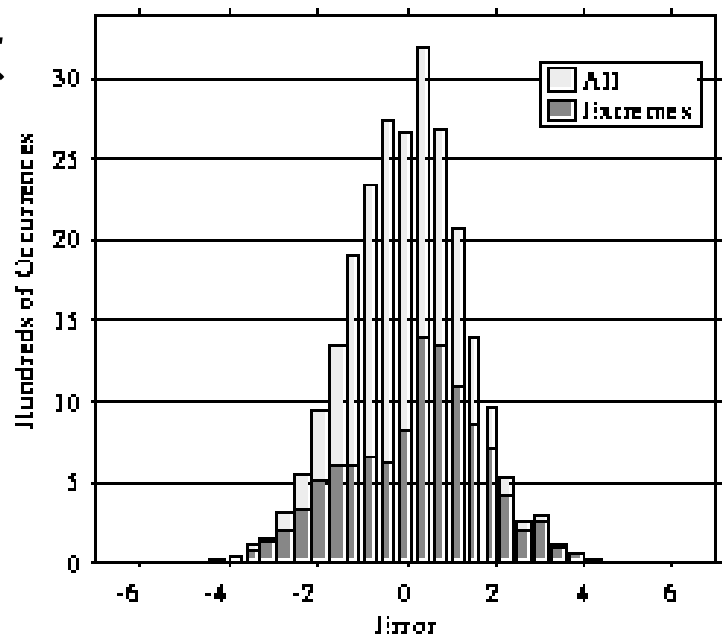
⌘ 結果(L=0.5)

⊠ 平均誤差1.3

⊠ 標準偏差1.6

⊠ 予測率95%

⊠ 対称評価の欠如



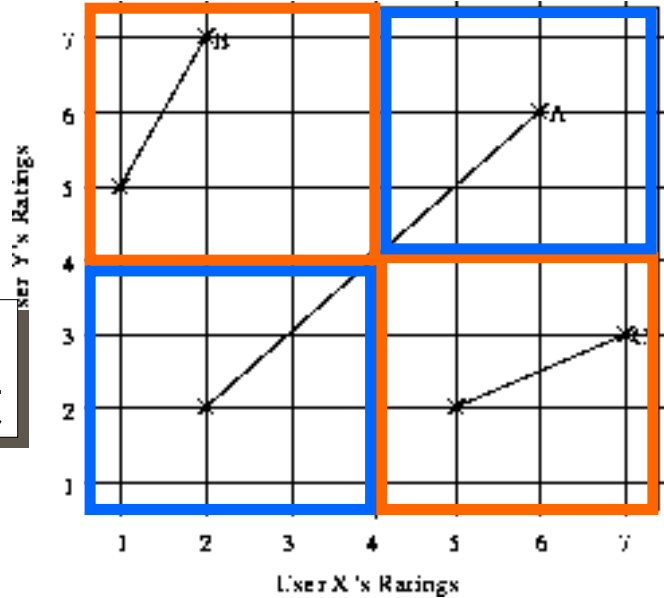
ピアソン積率 相関係数(cnt'd)

平均
最小2乗誤差法
ピアソン積率相関係数
アーティスト間相関係数

⌘ 対称評価

- ⊡ 対称的な評価でも
同じ相関係数をとる
- ⊡ すべて+1.0を取る

制約付き
ピアソン積率相関係数



制約付きピアソン 積率相関係数

平均
最小2乗誤差法
ピアソン積率相関係数
アーティスト間相関係数

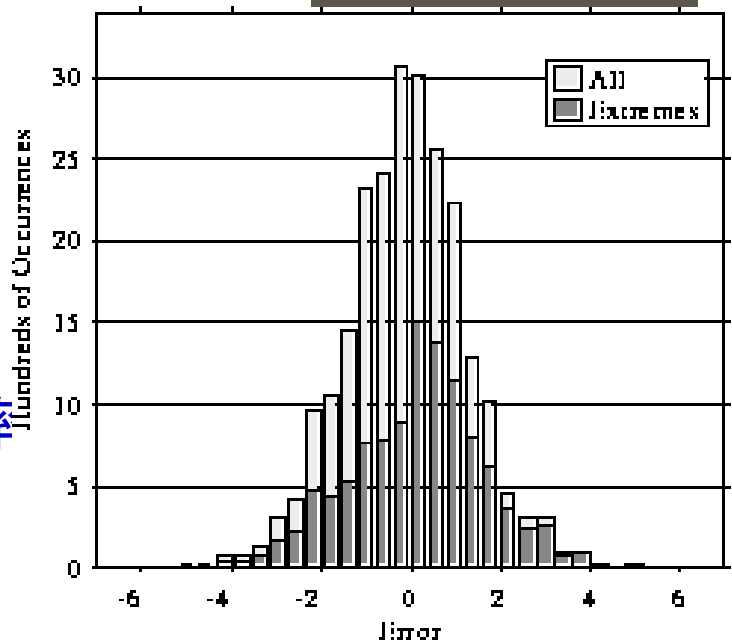
⌘ 対称評価の考慮

⌘ 結果(L=0.6)

- ⊡ 平均誤差1.2
- ⊡ 標準偏差1.6
- ⊡ 予測率91%

⌘ 全体的に高い予測率

- ⊡ 97% for L=0.5
- ⊡ 91% for L=0.6
- ⊡ 70% for L=0.7



アーティスト間相関係数

平均
最小2乗誤差法
ピアソン積率相関係数
アーティスト間相関係数

⌘ ユーザ間の類似度

- ⊠ 平均
- ⊠ 最小2乗誤差法
- ⊠ ピアソン積率相関係数

⌘ 対象間の類似度

- ⊠ アーティスト・アルバム間の相関

アーティスト間相関係数

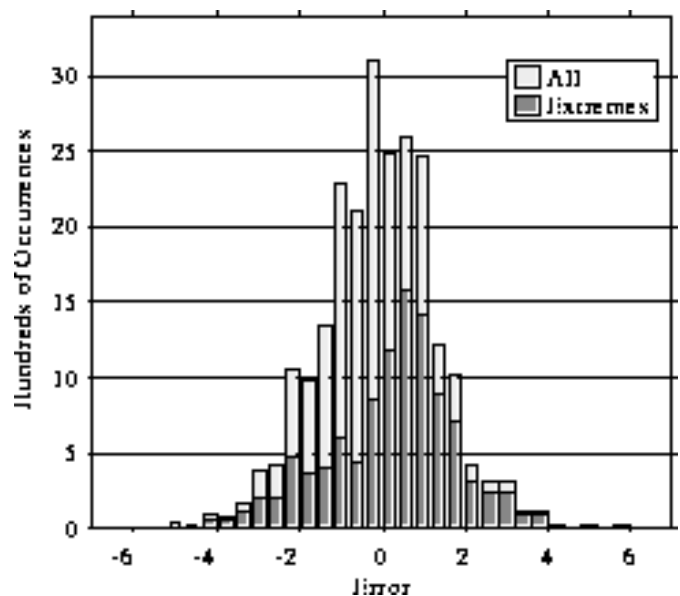
アーティスト間相関係数 (cnt'd)

平均
最小2乗誤差法
ピアソン積率相関係数
アーティスト間相関係数

⌘ アーティスト間の類似度を採用

⌘ 結果(L=0.6)

- ⊠ 平均誤差1.3
- ⊠ 標準偏差1.6
- ⊠ 予測率89%



考察

⌘ 最適手法

- ⊡ 制約付きピアソン積率相関係数

- ⊡ 先行研究(最小2乗誤差法)を予測率の面で改善

⌘ 予測率・精度のトレードオフ

- ⊡ 対象領域ごとに設定

⌘ 課題

- ⊡ 閾値設定の最適化

- ⊡ アルゴリズム計算時間の短縮

- ⊡ アーティスト・アルバムのクラスタリング

おわりに

- ⌘ Social Information Filteringについて述べた

- ⌘ RINGOシステムを作成

- ⌘ 評価アルゴリズムの比較

- ⌘ 制約付きピアソン相関の有効性