

情報フィルタリング, 協調フィルタリング, 推薦(レコメンデーション)

情報フィルタリング

Information Filtering

- フィルタリングによる情報システム
 - メッセージ(メール)やWeb情報のフィルタリング
 - プッシュ型情報配信システムのフィルタリング
 - 推薦(レコメンデーション)システム
- 情報フィルタリングの種類
 - 内容フィルタリング(Content-Based Filtering)
 - 情報の内容比較に基づくフィルタリング
 - 協調フィルタリング(Collaborative Filtering)
 - 実社会の関連(人、コミュニティなど)に基づくフィルタリング
- 推薦システム
 - 内容に基づく推薦(Content-Based Recommendation)
 - ユーザが過去に気に入ったアイテムに類似したアイテムを推薦
 - 協調型推薦(Collaborative Recommendation)
 - ユーザと好み似ている他のユーザを認識し、このような他のユーザが気に入ったアイテムを推薦

Voio が人に代るVoio
リリースことおめでとう
おめでとう
はこんなものも
買っている

2

情報フィルタリング

- Selective Dissemination of Information (SDI)
 - 図書館のSDIサービス
 - 興味のある雑誌などのタイトル等を事前に登録しておく、定期的にその雑誌の目次(TOC)情報を配信してくれるサービス
 - <http://www.cwi.nl/library/literature/dissemination/>
- メッセージのフィルタリング
 - The Information Lens: An Intelligent System for Information Sharing in Organizations (T.W.Malone, K.R.Grant, F.A.Turbak, Proc. of the ACM CHI'86, pp.1-8, April 1986.)

3

情報フィルタリング: Information Lens

- メールの内容に基づくフィルタリング
 - 種々のタイプのメッセージのフィルタリング・分類用に半構造型のテンプレート(semi-structured templates)を用意。
 - プロダクションルール(production rule)の利用
 - メッセージタイプ間のisa関連にもとづく継承

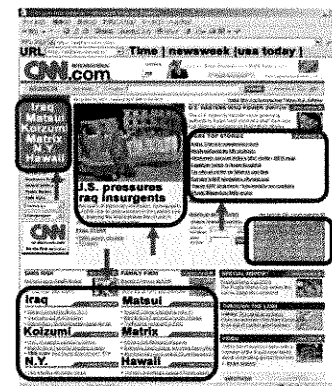
4

内容に基づくフィルタリング・推薦

- ユーザ・プロフィール(user profile)
 - ユーザの好み(選好・嗜好)の情報を記述したもの
 - 好みのキーワードやカテゴリ等の情報を、ユーザが陽に指定したり、ユーザの選んだコンテンツから自動生成。ベクトル空間モデルの特徴ベクトルとして生成することもある。
 - ユーザの選択コンテンツをユーザプロフィールに時々刻々反映させることで、ユーザプロフィールを更新。適合フィードバックに相当する。
- 処理
 - 配信されたコンテンツとユーザプロフィールの類似度を計算し、類似していると判断されたコンテンツを選択的に提示したり推薦したりする。
- 弱点
 - ユーザの好みをユーザプロフィールに十分に反映できない場合がある(映画や音楽などのマルチメディアコンテンツからのユーザプロフィール情報の抽出の困難さや、美的・感性的なものの表現)
 - 感性・印象のコンテンツからの自動抽出
 - Over-specification(ユーザは自身が選択したものに類似したものしか見ることが出来ないという)問題。
 - 遺伝アルゴリズム(交配、置換)などの適用

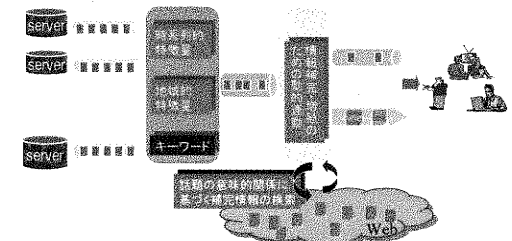
5

内容に基づくフィルタリングの例



個人用仮想ポータルサイト(MPV)

ユーザプロフィールにもとづき選択したニュース記事を使い慣れた既存ポータルを元に仮想的に定義
(河合由起子, 官上大輔, 田中克己, 個人の選好に基づく複数ニュースサイトの記事収集・閲覧システム, 情報処理学会論文誌 データベース, Vol.46, No.SIG8(TOD26), pp.14-25)



新鮮度と流行度によるフィルタリング

プッシュ型情報システムのコンテンツを、ユーザプロフィールを使用せず、新鮮度・流行度という尺度で自動的に情報フィルタリング
(馬強, 角谷和俊, 田中克己, 放送型情報配信システムのための時系列性を考慮した情報フィルタリング, 情報処理学会論文誌 データベース, Vol.41, No.SIG 6 (TOD7), pp.46-57, 2000)

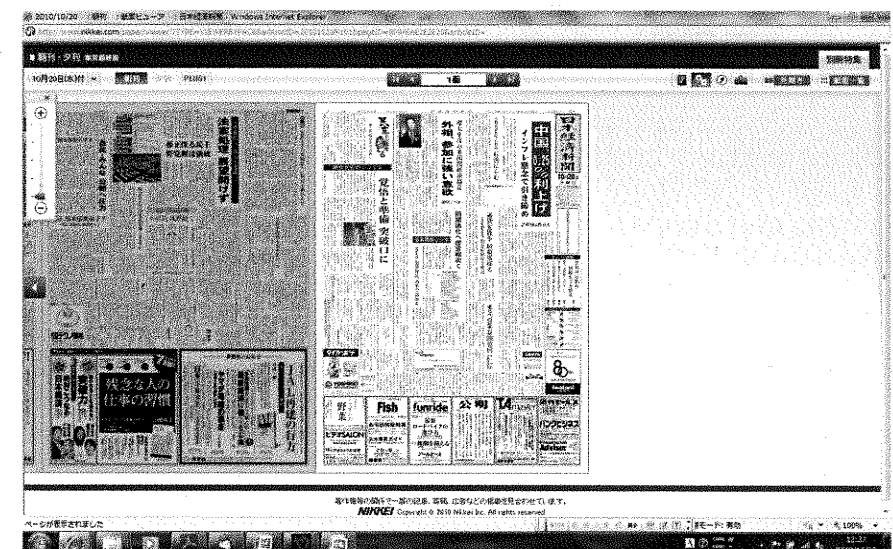
6

電子新聞の例(日本経済新聞)



7

電子新聞の例(日本経済新聞)



8

フィルタリングの例



9

フィルタリングの例(My Yahoo!)



10

協調フィルタリングと推薦

In content-based recommendation one tries to recommend items similar to those a given user has liked in the past, whereas in collaborative recommendation one identifies users whose tastes are similar to those of the given user and recommends items they have liked.

M.Balabanovic and Y.Shoham, "Content-Based, Collaborative Recommendation", CACM Vol.40, No.3, pp.66-72, March 1997より

協調フィルタリング

- 効率的な情報収集
 - 各ユーザが持つ問題解決のための情報を自動的に収集
 - 同じ問題を持つユーザに提供
 - 欠点は、新規アイテムがデータベースに登録された場合や、ユーザの好みが極端に他と異なる場合に、ユーザに推薦できないことと、
- アプリケーション例
 - WWWページの推薦、ニュースのフィルタリング、音楽・映画の推薦システム
- アーキテクチャ
 - User Profile: 取捨選択の基準を記述
 - Recommendation: ユーザが情報収集で得た情報
 - User Modeler: ユーザの特徴を抽出する処理単位
 - Mach Maker: ユーザ間の類似度を計算する処理単位
 - Score Predictor: 個々のユーザに応じて情報のランキングを計算する単位

11

12

協調フィルタリング: GroupLens

GroupLens (米国ミネソタ大学)

- Resnick, P., Iacovou, N., Sushak, M., Bergstrom, P., and Riedl, J.
GroupLens: An open architecture for collaborative filtering of netnews.
Proceedings of the 1994 Computer Supported Collaborative Work
Conference. (1994).

<http://www.grouplens.org/index.html>

入力

推奨情報, ユーザプロフィール

出力

個々のユーザに応じて評価された推奨情報

手順

- ユーザが情報収集で得た情報を蓄積
- ユーザプロフィールの獲得
- ユーザ間の類似度を計算
- ユーザに応じて情報のランキングを他のユーザのプロファイル(アイテムへの評価値)を計算

$$K_{pred} = \bar{K} + \frac{r_{KM}(M_6 - \bar{M}) + r_{KL}(L_6 - \bar{L})}{|r_{KM}| + |r_{KL}| + |r_{KN}|}$$

評価予測

ユーザ間の類似度からあるユーザが未評価のアイテムの評価値を予測

相関が存在するユーザの評価値を利用

KenのMessage 6の評価予測:

$$K_{6pred} = \bar{K} + \frac{\sum (J_6 - \bar{J}) r_{KJ}}{\sum |r_{KJ}|} = 3 + \frac{2r_{KM} - r_{KL}}{|r_{KM}| + |r_{KL}|} = 3 + \frac{2 - (-0.8)}{|1| + |-0.8|} = 4.56$$

J: 相関が存在するユーザの集合(ユーザL,M,N)

raters: 相関が存在するユーザ(ユーザL,M,N)

J: 各ratersの評価の平均(L,M,N=3)

類似度計算

(例) メールメッセージ

message #	Ken	Lee	Meg	Nan
1	1	4	2	2
2	5	2	4	4
3	3		3	
4	2	5		5
5	4	1		1
6	?	2	5	?

相関係数 r_{KL}

ユーザ間の類似度

- 例: KenとLeeの類似度 r_{KL}
- 各ユーザの評価値より計算
- 幾何学的には、ベクトルK-K, L-Lのなす角をθとすると、 r_{KL} は $\cos\theta$ となる。

<http://web.sophia.ac.jp/~t-dama/correlation.doc>

GoupLensの場合

σ_K, σ_L : 標準偏差
K, L: 評価の平均
(アイテム1,2,4,5)
Cov(K,L)は共分散

$$r_{KL} = \frac{\text{Cov}(K, L)}{\sigma_K \sigma_L} = \frac{\sum (K_i - \bar{K})(L_i - \bar{L})}{\sqrt{\sum (K_i - \bar{K})^2} \sqrt{\sum (L_i - \bar{L})^2}} = \frac{-2 - 2 - 2 - 2}{\sqrt{10} \sqrt{10}} = -0.8$$

ユーザはすべてのアイテム(メッセージ)についての評価を行うわけではない

$$r_{KL} = -0.8$$

$$r_{KM} = +1$$

$$r_{KN} = 0$$

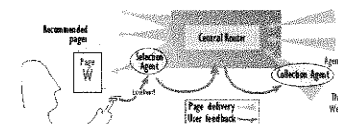
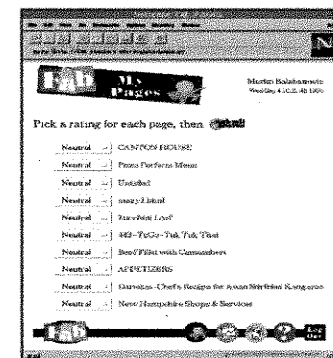
推薦システム Fab

(Content-based, collaborative recommendation)

Fab

- 内容に基づくフィルタリングと協調フィルタリングの双方を組み合わせた推薦システム。双方の欠点を克服。
- ユーザプロフィールに類似し、かつ、類似のプロファイルを持つ他のユーザにも高くランクされているものを推薦する。
- スタンフォード大学のデジタルライブラリプロジェクトの一貫で開発。
- Content-Based, Collaborative Recommendation, Marko Balabanovic and Yoav Shoham (CACM, Vol.40, No.3, pp.66-72, March 1997).

<http://maya.cs.depaul.edu/~classes/ds575/papers/Fab.pdf>



推薦システム PHOAKS

(People Helping One Another Know Stuff)

■ PHOAKS

- 自動的にNetnewsのニュースグループ記事群から、推薦Webページ(URL)を抽出する。URLを推薦している記事およびFAQを自動検出。
- 実験では、適合率88%、再現率87%を達成。
- あるURLを推薦している記事数を、推薦のランキングに使用。
- PHOAKS: a system for sharing recommendations. Loren Terveen, Will Hill, Brian Amento, David McDonald and Josh Creter (CACM, Vol.40, No.3, pp.59-62, March 1997.)
- http://www.ischool.utexas.edu/~i385d/readings/Terveen_PHOAKS_97.pdf

PHOAKS: Resources for rec.music.dylan

People Helping One Another Know Stuff
"Together, we know it all."

More resources: [Page 2](#) [Page 3](#) [Page 4](#)

[Top Posters](#) [Volunteered Resources](#) [Newsgroup Info](#) [Feedback](#) [Help](#)

Navigate Up: [PHOAKS Entry Page: rec.music.dylan](#)

Resources for rec.music.dylan

Resource Title	Distinct Posters	Click on Bars for Message Contact(s)
(1) Bob Dylan	18	████████████████████
(2) Terveen Update	11	██████████████████
(3) The Book of Bob	7	██████████
(4) Expecting Rain	7	██████████
(5) A Cat's Bob Dylan Links	6	████████
(6) Mailing List WWW Gateway	6	████████
(7) CDNew: welcome	5	██████
(8) Bob Dylan Church	5	██████
(9) Federal Communications Law Jour...	4	████
(10) Bob Dylan	4	████

*Note: each square represents the posting of one resource (e.g., URL) by one person. The taller the square, the more recent the post. Click on a square to view messages when this resource was mentioned. Posting a web resource does not necessarily imply endorsing that resource. Sometimes it may actually mean the opposite. Consult the relevant newsgroup messages to obtain context.