

スターコーディネートを応用した推薦システムの実装と評価

山形 浩平 古川 貴志[†] 増田 宏[†]

東京大学工学部

東京大学大学院 工学系研究科[†]

1. はじめに

近年、コンピュータ技術の急激な発達とインターネットの爆発的發展により、誰もが膨大な情報に接することができるようになった。またその膨大さは、人の情報収集能力をはるかに超え、“情報過負荷”を引き起こしている。このような問題を解決するために、協調フィルタリング(Collaborative Filtering)と呼ばれる手法が知られている。

しかし、既存の協調フィルタリング推薦システムには以下の問題がある。

- 推薦度に誤差が生じる
- 推薦の理由と妥当性が不透明
- リアルタイム性に欠ける
- ユーザによる操作が困難

本研究では、明示性と操作性の問題が、もっとも重大な問題であると考えた。なぜなら、いかなる推薦手法をとっても誤差を全くなくすることは困難であるが、推薦の理由が明示的で、操作が容易であれば、ユーザ自らが誤った選択を回避でき、推薦結果を簡単に変更できるからである。そこで本研究では、

情報の持つ属性ごとに協調フィルタリングを用いて推薦度を推定することで、推薦理由と妥当性の理解が容易になるようにする。

(これを属性別協調フィルタリング『SACF: Separated by Attribution Collaborative Filtering』と呼ぶことにする。)

大量のデータとそれらが持つ特徴量をインターフェース上で表現し、ユーザにインタラクティブに操作させて知識の発見を助ける視覚データマイニングインターフェースを構築する。

この ・ を本研究の目的とする。

2. SACF の構築

本研究で開発した SACF について述べる。1994 年にミネソタ大学の研究グループ GroupLens[1]が開発した協調フィルタリングを、映画の推薦に用いるとする。例えばユーザ K (他にユーザ L, M) に対する I3 という映画の推薦度は、以下のような式で表される。

$$P_{K,I3} = \bar{r}_K + \frac{\sum_{U=L,M} (r_{I3} - \bar{r}_U) w_{K,U}}{\sum_{U=L,M} |w_{K,U}|}$$

$r_{U,I3}$: ユーザ U の I3 に対する評価点

\bar{r}_K : ユーザ K の全情報に対する評価点数平均

\bar{r}_U : ユーザ U の全情報に対する評価点数平均

w_{KU} : ユーザ K とユーザ U の相関係数 (ユーザ K と U がお互いに評価している映画より計算)

ここで、映画にはジャンル (Action, Comedy など) という属性が存在するため、あるジャンルに属する映画のみで協調フィルタリングを行えば、より精度が高まると考えられる。よって、ある映画が属するジャンルごとに協調フィルタリングを行い、その平均値を推薦度として算出する。これを SACF の定義とする。例えば、映画 I3 がジャンル Action と SF に属しているとする、推薦度は

$$P_{K,I3} = \frac{\sum_{J=Action, SF} P_{K,I3,J}}{|J|} \quad \text{となる。}$$

3. 視覚データマイニング

視覚データマイニングとは、大規模データベース全体をユーザに視覚的に提示して、直感的に操作させることで、データの特徴をユーザに理解させる手法である。Kandogan が提案したスター・コーディネイト (Star Coordinates) [2] [3] は図 1 に示すように、n 次元空間の各軸を 2 次元平面のベクトルに対応させて、その和を元のデータを表す点とみなす視覚化方法である。ユーザは各軸の方向を変更させることでデータの性質を調べることができる。

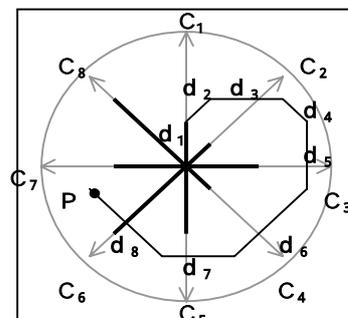


図 1 スター・コーディネイト

An Interactive Recommendation System using Star Coordinates
Kohei Yamagata Takashi HURUKAWA[†] Hiroshi MASUDA[†]
School of Engineering, The University of Tokyo
Graduate School of Engineering, The University of Tokyo[†]

本研究では、前述のSACFによって求めた映画のジャンルごとの推薦値をスター・コーディネイトの軸に対応させ、ジャンルごとに情報の位置を変更することが可能なインターフェースを実装した(図2参照)。例えばジャンルRomanceに属している映画は、Romanceの軸を移動させることにより、推薦度のRomanceに依存しているデータだけが移動するので、データの特徴がつかめる。

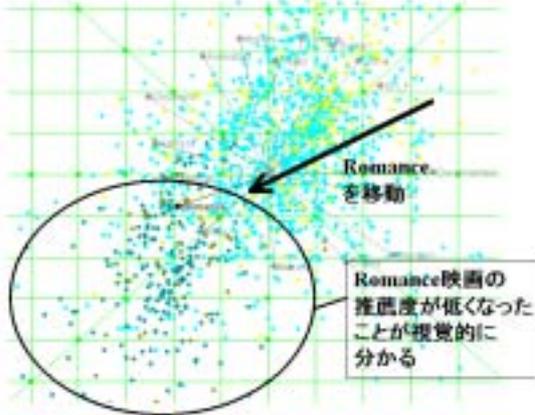


図2 インターフェースの実装例

ただし、本研究では、スター・コーディネイトのように単純にベクトルの和を求めるのではなく、2次元平面の各点のXとYに推薦度に依存する量を初期値として与える。この場合、軸の操作によって、特定の属性を持つデータ群のみが平行移動することになる。

X軸には全属性のSACFの平均を計算し、あるユーザへの推薦度を表現する。また、推薦理由として他のユーザ達の評価点数を提示することは、かなり有効であることが分かっている^[4]ので、Y軸には他ユーザ達の推薦度の平均を表現する(図3参照)。これによって、図2に示すように、特定の属性の影響度をインタラクティブに変更することができる、視覚的に分かりやすいインターフェースが実現できる。

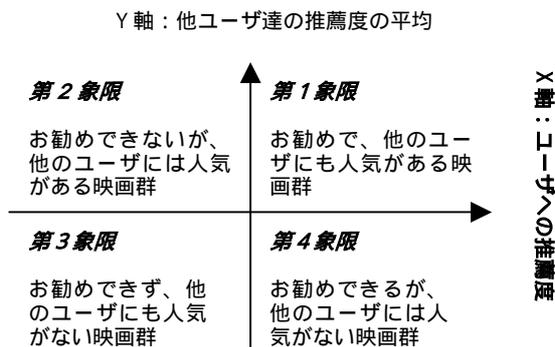


図3 2次元平面の意味付け

4. 評価実験

本研究では前述のインターフェース(「IF1」とする)と、従来の協調フィルタリングの結果のみを表示したインターフェース(「IF2」とする)の比較実験を実施した。両インターフェースとも、Java 2のSwingを用いて実装されている。被験者は東京大学工学部の理系学生男女計1

3人である。また、順序効果を無くすために、半数ずつIF1とIF2の実験順を変えた。

インターフェースを実際に操作してもらい、操作に十分習熟した段階で主観的なアンケートに答えてもらった。その結果をまとめた表が表1である。この結果より「推薦理由の明示性」・「推薦理由の妥当性」・「推薦結果の信頼性」が従来のシステムよりも優れていることが分かる。

表1 アンケート結果

(全ての項目に【悪い】1・2・3・4・5【良い】で回答、その平均点を小数点第2位で算出)

項目	IF1	IF2
操作性	3.15	3.69
推薦理由の明示性	2.23	1.46
推薦理由の妥当性	2.15	1.69
推薦結果の信頼性	2.85	2.31
学習性	3.85	3.92
推薦システムの満足度	3.54	2.85

またIF1のみ質問した「属性による分類表示は意味があるか」、「他ユーザの推薦度表示(y軸)が参考になったか」の問いに対してそれぞれ4.54点、4.08点とほとんどのユーザがその効果を認めている。よってSACFと視覚データマイニングインターフェースによって推薦システムがより優良になったと言えるだろう。

その代わり、「操作性」と「学習性」については従来のシステムよりも点が落ちている。これは結果が表示されるだけのシステムよりも操作が煩雑になったため、やむを得ない。今後は2次元インターフェースにおける情報の扱いやすさを考慮していく必要がある。しかし、推薦結果の訂正をリアルタイムに行える点は被験者によって高く評価された。

5. まとめ

- SACFによって、既存の協調フィルタリングより推薦精度が向上させることができた。
- 情報の操作がインタラクティブに行えることで、推薦結果の訂正と嗜好の変化に伴う操作をリアルタイムに行えるようになった。
- 情報データマイニングを応用して、インターフェース上で情報を視覚化して、ユーザに推薦理由を直感的に理解させることができた。

参考文献

[1] Resnick, P., et al., GroupLens: An Open Architecture for Collaborative Filtering of Netnews, CSCW'94, 1994.

[2] Kandogan, E., Star Coordinates: A Multi-dimensional Visualization Technique with Uniform Treatment of Dimensions, I EEE Visualization, 2000.

[3] Kandogan, E., Visualizing Multi-dimensional Clusters, Trends, and Outliers using Star Coordinates, KDD, 2001.

[4] Herlocker, et al., Explaining Collaborative Filtering Recommendations, CCSW, pp. 241-250, 2000.