

# クレジットカード利用者に対する 協調フィルタリングの有効性

～ SAS マーケティング研究会 ～

2003年 3月 7日

株式会社 金融エンジニアリング・グループ

## アウトライン

1. 自己紹介
2. 協調フィルタリング
3. カード利用明細データ
4. 協調フィルタリングの実験
5. 提案

## 金融エンジニアリング・グループについて

株式会社金融エンジニアリング・グループ ( F E G )

本 社：東京都港区赤坂 2 - 2 1 - 8

設 立：1 9 8 9 年 4 月

代表者：中林三平

資本金：8, 1 0 3 万円

人 員：6 4 名 ( 2003/1/31現在)

- ・金融ビジネス分野の諸問題について、数理分析、コンサルティングを通じて、問題解決にあたるコンサルティング会社
- ・信用リスク、市場リスクなどのリスク管理分野、顧客DBを活用した金融マーケティング分野に対し、データマイニング技術、OR・統計分析技術、金融工学技術、金融ビジネス・マーケティング実務知識を有機的に活用
- ・今年のKDDカップ(データマイニング世界大会)では、総合5位

## 演者について

藤本浩司 kfujim@feg.co.jp

1985年上智大学理工学部数学科卒業。

同年より、日本アップジョン社(現ファルマルシア)にて、エキスパートシステム開発、臨床・前臨床試験データ解析に従事。

1989年より、アメリカン・エクスプレス(Japan)社にて、データベースマーケティング、広告・宣伝業務、新規会員獲得業務に従事。

1999年東京農工大学大学院博士課程修了(工学博士)。

同年より(株)金融エンジニアリンググループに勤務。現在に至る。

2001年より東京農工大学工学部非常勤講師を勤める。

F E Gにおける主なプロジェクト

リスク管理：都銀、地銀、信販・リース会社 に対し信用リスクモデルを開発

マーケティング：都銀、地銀、クレジットカード、信販、専業、損保会社、ISPに対し、DBマーケティング・コンサルテーション、顧客行動予測モデルの開発

## 広告スポンサー付無料ソフト

広告スポンサー版ブラウザ(オペラ)のバナー広告



## ユーザプロフィールの設定

事前に設定された個人属性・趣味情報で、表示される広告の種類が選別される。



## 情報収集と情報フィルタリング

- 情報収集には、サーチエンジンなどの情報検索による方法とブラウジングによる方法、情報フィルタリングによる方法がある
- 情報フィルタリングとは、利用者に提供される情報を何らかの基準で優先的に収集したり、抑制したりする技術
- メール、ニュースグループ、WEB、広告などに応用（各種リコメンデーション、有害・無用コンテンツの抑制）
- コンテンツベースド・フィルタリング（キーワード）や協調フィルタリング（他利用者の評価）などがある

## 協調フィルタリングとは

- 利用者に類似する他の利用者の評価値をもとに、その利用者に対する有用度を推定
- 類似する評価・行動をとる者同士は、お互いに近い嗜好、ニーズを持っているという仮説にもとづく（テイストクローン）
- 大量の情報や商品と大量の利用者が存在する中で、各利用者に限られた数の推薦リストを提示することに適している

## アマゾンのリコメンデーション

- アマゾンのこのリコメンデーションは、非常にシンプルだが有用な協調フィルタリングの一例である。
- (アマゾンには、これ以外の協調フィルタリング推薦システムもある)



## 商用の協調フィルタリングシステム

- LikeMinds  
アンドロメディア社が開発。現在マクロメディア社
- GroupLens  
NetPerceptions社によって商用化。アマゾンがファーストクライアント
- AuraSearch (KnowledgeGauge)  
オーラライン社とFEG社の共同開発

## 協調フィルタリングのしくみ(1)

MovieLens: GroupLensの技術を使った映画推薦サイト

**Please rate more movies**

MovieLens predicts what movies you are likely to enjoy based on your opinion of previously viewed movies. In order to make recommendations, MovieLens typically needs you to rate 5 to 15 movies. Please go through the following movies and rate how you have seen. As soon as you have rated enough movies, you will automatically be taken to your personalized MovieLens home page. Thank you for using MovieLens.

YOUR RATING	GENRE	TITLE	REVIEWS
5	Drama	12 Angry Men (1957)	4899
5	Comedy	Backstreet (2005)	4894
5	Action, Thriller	Die Hard 2 (1990)	4893
5	Comedy	Great White Place, The (1986)	4892
5	Comedy	How to Get Ahead in Advertising (1989)	4891
5	Comedy	Strawhead (1987)	4890
5	Comedy	Yert (1974)	4889
5	Comedy	Man, Like, Don, Play, Like, Run's (Open Courts and Madras) (1988)	4888
5	Drama	Seven Nights in Beate Iselin, Ltd (1952)	4887
5	Comedy, Drama	Shogun (1980)	4886

## 協調フィルタリングのしくみ(2)

何本かの映画の評価を入力すると、推薦映画リストが表示

**Your Top 5 Recommendations**

Based on your history:

1. [Pulp Fiction \(1994\)](#) \*\*\*\*\*
2. [Lord of the Rings: The Two Towers \(2002\)](#) \*\*\*\*\*
3. [Moulin Rouge \(2001\)](#) \*\*\*\*\*
4. [Chicago \(2002\)](#) \*\*\*\*\*
5. [Nikotina: Most Wanted \(2002\)](#) \*\*\*\*\*

Based on what:

1. [My Neighbor Totoro \(1988\)](#) \*\*\*\*\*
2. [Fast Forward: The Alternative \(2002\)](#) \*\*\*\*\*
3. [Fast Forward: The Alternative \(2002\)](#) \*\*\*\*\*
4. [How to Get Ahead in Advertising \(1989\)](#) \*\*\*\*\*
5. [How to Get Ahead in Advertising \(1989\)](#) \*\*\*\*\*

Just want to rate some movies?

Get our combined recommendation for you and your friends. It doesn't matter if they're currently MovieLens users or not, MovieLens will let them know how to join your group.

- [Create a group](#)
- [Invite Old Friends you get along with](#)
- [Learn more about groups](#)

If you can't [change](#) random selections or use the "Get Recommendations" link at the top of the page to select a set of movies you would be interested in rating. Any movie you get a recommendation for, you can rate. The links below may also be of interest.

- [Get all 250 most recent additions](#) to our movie database
- [Get all 5372 ratings](#)

## 協調フィルタリングのしくみ(3)

さらに評価を続けると、推薦リストに好みのものが増えてくる

## 協調フィルタリングのしくみ(4)

相関係数によって、評価者間の類似度を定める。

[5]=Must See, [4]=Will Enjoy It, [3]=It's OK, [2]=Fairly Bad, [1]=Awful

Title	Rater 1	Rater 2	Rater 3	Mr.Fujimoto
12 Angry Men	2	4	4	5
Die Hard 2	3	3	2	3
Star Trek 94	5	5	2	3
The Net	5	3	2	3
Back to the F..	4	3	2	4
類似度	-0.73	0.06	0.88	
Star Trek 98	5	4	2	?

## 協調フィルタリングのしくみ(5)

類似度の重みつき偏差和を求め、未知の映画の評価値を推定する

Title	Rater 1	Rater 2	Rater 3	Mr.Fujimoto
12 Angry Men	2	4	4	5
Die Hard 2	3	3	2	3
Star Trek 94	5	5	2	3
The Net	5	3	2	3
Back to the F..	4	3	2	4
類似度	-0.73	0.06	0.88	
Star Trek 98	5	4	2	2.9
平均Rate	3.8	3.6	2.4	3.6
偏差	1.2	0.4	-0.4	-0.72
重みつき偏差	-0.87	0.03	-0.35	-1.20
重み	0.73	0.06	0.88	1.67

## 協調フィルタリングのしくみ(6)

- 実際の評価データには、たくさんの欠損がある。2評価者間で共通な評価値にて類似度を計算。
- 評価データ構造、類似度の尺度、推定値の方法、データベースの管理方法にバリエーションがある。
- 大量の利用者が大量の評価情報を入力しなければならないことが問題である。
- アマゾンの前述の推薦システムは、評価データが、「ある本を買った」だけであり、類似とは、「同じ本を買ったか否か」だけの単純なものだが、それでも十分に役立つ。



# カード利用明細と販促情報提供

販促情報は、セグメント別提供から個別提供へ

ご利用日	領収番号	ご利用明細	金額
95/07/09	324286	エフエフエフ スターバックス ショッピング	6,072
95/07/13	986195	エクスプレス・キャッシュ	
	999195	株式会社ダイエーファイナンス	10,000
	999195	エクスプレス・キャッシュ手数料	245
95/07/16	267197	アピオスタ	8,500
95/07/19	264288	サントラーゴカーブー サニーサイド 一度目録 前払証	8,137



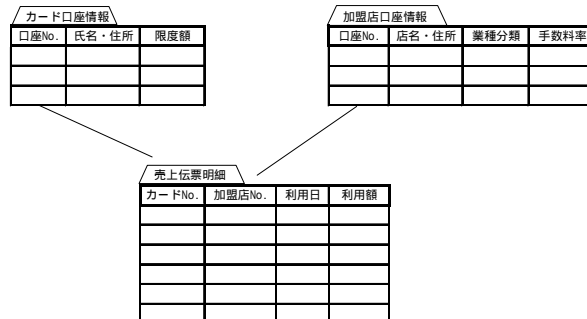
ご利用明細	ご利用金額 (円)
7月28日 紀伊国屋書店 本店 東京都 新宿区 領収番号 11227 114	5,754
7月28日 伊勢丹 東京都 新宿区 領収番号 11227 114	19,870

※7月28日～31日19日、伊勢丹新宿区にて11227Aのクレジット決済が実施され、伊勢丹の売上金額は11,117円とさせていただきます。



# カード利用明細のデータ構造

カード利用明細情報は、カード口座と加盟店口座の2つのマスターファイルと、売上傳票トランザクションファイルの3つよりなる



## カード会員がもし100人だったら..

50人は、カードをまったく利用せず  
残りの50人で、年間1500万円が使われます。

この50人からの売上傳票は、合計1000枚で、  
1枚あたりの利用単価は、1.5万円です。  
利用者ひとりあたりでは、年間30万円、20回となります。

このうちのたったひとりが、225万円を使います。  
これは、全員の利用総額のなんと15%を占めます。  
19人が1000万円(総額の70%、平均55万円)を使います。  
残る30人は、ひとり7.5万円しか使いません。

## カード会員がもし100人だったら..(2)

1000枚の売上傳票のうちの、  
1枚は、80万円ですが、  
.....きっとあの人に違いない

400枚は、3000円未満です  
.....こんな小額は昔はなかった

最初の1枚は、家電で、  
後の400枚は、プロバイダ、電話、高速道路、ガソリンスタンド  
のものです。

100枚は、海外加盟店で、会員の10人が利用しています。

## カード会員がもし100人だったら．．(3)

900枚の国内での売上傳票のうち、  
半分の450枚は、30店だけで占められています。  
売上額においても、半分は30店からのものです。

でも、100万人のカード会員だとしても、  
たった200店が半分の売上額を占めます。  
数十万の店からは、ちょうど1枚の伝票があがり、  
数百万の店からは、なにもきません。

最後に．．  
カードを使わなかった50人のうち、5人は去っていき、  
40人は使わないままです。

## 実験概要(1)

### 目的

特定加盟店でのカード利用者に対し、好みのお店の推薦に  
協調フィルタリングが有効であることを示す。

### データ

カード会員10万人の1年間の国内加盟店での売上明細

### 実験方法

1. 特定の加盟店の利用者を取り出す。(テストクローン)
2. テストクローンの利用する店を利用の多い順にならべる
3. リストされた店の全体での利用率とテストクローン内での  
利用を比較し、利用率のリフト率を確認する

## 実験概要 (2)

海外加盟店、電話・保険・通信などの売上を除外した。

巨大200加盟店は、除外した。サンプル10件を以下に示す。

ヨドバシカメラ	高島屋新宿	西友	ソニープラザ
JR東日本	ミキモト	DHC	エネオス
田崎真珠	ユニー		

カード利用者数: 91,232  
 売上傳票数: 1,005,473  
 加盟店数: 67,060

## 実験結果 (1)

ある家具店のテストクローンでは、この家具店の次によく利用される店もやはり家具店であった。

上位10店のリフト率は、4 - 10倍といずれも高かった。

OBS	加盟店業種	利用者数 (テストクローン)	利用者数 (全体)	利用率 (テストクローン)	利用率 (全体)	リフト率
1	家具	917	917	100.000	1.00513	99.4896
2	家具	63	738	6.870	0.80893	8.4930
3	ホテル	36	792	3.926	0.86812	4.5223
4	中華料理	34	562	3.708	0.61601	6.0189
5	家具	32	363	3.490	0.39789	8.7704
6	家電	31	338	3.381	0.37048	9.1248
7	ショッピングセンター	28	398	3.053	0.43625	6.9993
8	家電	28	513	3.053	0.56230	5.4302
9	ショッピングセンター	27	338	2.944	0.37048	7.9474
10	ホテル	27	677	2.944	0.74206	3.9678
11	家具	27	266	2.944	0.29156	10.0986
		(a)	(b)	(c)=a/917	(d)=b/91232	(e)=c/d

## 実験結果 (2)

比較のため、ランダムに選んだ8,897会員を、テストクローンとみなして、リフト率を計測。  
利用率のリフトは見られなかった。

OBS	加盟店業種	利用者数 (テストクローン)	利用者数 (全体)	利用率 (テストクローン)	利用率 (全体)	リフト率
1	家具	92	917	1.03406	1.00513	1.02878
2	スーパーマーケット	85	752	0.95338	0.82427	1.15906
3	家電	82	612	0.92166	0.67082	1.37393
4	紳士服	81	797	0.91042	0.87360	1.04215
5	ショッピングセンター	78	736	0.87670	0.80673	1.08673
6	紳士服	78	784	0.87670	0.85935	1.02019
7	家具	77	730	0.86546	0.80016	1.08161
8	デパート	76	685	0.85422	0.75083	1.13770
9	カメラ	75	703	0.84298	0.77056	1.09398
10	家具	75	738	0.84298	0.80893	1.04210
11	ショッピングセンター	73	703	0.82050	0.77056	1.06481
		(a)	(b)	(c)=a/8897	(d)=b/91232	(e)=c/d

## 実験結果 (3)

10店を選び、それぞれのテストクローンからのトップ推奨店に対する利用率をみると、6 - 20倍もの高いリフトがみられた。

OBS	シード店	トップ推奨店	推奨店利用率 クローン	全体	リフト率
1	家具	家具	6.9	0.8	8.5
2	紳士服	家電	4.9	0.7	7.3
3	ホテル	ホテル	15.4	0.7	20.8
4	紳士服	時計宝飾	4.0	0.6	6.5
5	スーパーマーケット	デパート	7.6	0.6	13.2
6	家具	家具	8.5	1.0	8.5
7	ショッピングセンター	デパート	8.4	0.4	20.9
8	家具	家電	5.3	0.7	7.2
9	ショッピングセンター	靴	6.9	0.5	15.2
10	カメラ	家電	10.7	0.5	19.9

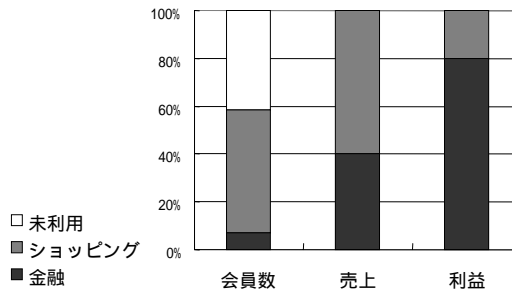
## 実験ーネクストステップ

この実験は、非常に簡易なものであった。  
さらに、次の点についての有効性を検証すべきと考える。

- 会員属性などによるセグメント化
- 加盟店情報の活用
- 評価値の高度化

## これは、クレジットカードなのか？

加盟店手数料率が正味3%を下回り、利益の大部分は一握りの金融商品利用者からの、高金利な金利収入から成り立っている。  
これは、クレジットカードなのか、それともローンカードなのか？





## 提案

加盟店利用を健全化するために、 . . .