

協調フィルタリングにおけるデータの必要数の分析 Recommendation boundaries of collaborative filtering

田村 幸之介†
Konosuke Tamura

松原 仁‡
Hitoshi Matsubara

1. はじめに

近年、Web における利用者登録制のサービスの利用が増えてきている。これらのサービスは多くの提供アイテムがあるため、ユーザにとってどのアイテムが有用なものであるかをすべて判別することは難しい。そのため、推薦システムを導入して判断の負荷を軽減する試みがよく利用されている。現状では、推薦システムを導入しているサイトは膨大な規模のアイテムを保有している限られたサイトのみである。しかし、ユーザが全てのアイテムを見るのが簡単ではないと感じる量のアイテム数を保有しているサイトであれば推薦機能が利用ユーザにとって有用であると考えられる。

これらの推薦システムでよく利用されている技術の一つに協調フィルタリングと呼ばれるものがある。この技術は研究、実用ともに広く成功をおさめている。しかし、このアルゴリズムには適正な推薦を行なう条件として一定以上の嗜好データが必要であるという側面がある。

本研究では協調フィルタリングがどれほどの規模のデータ量を保持していれば適切に推薦ができるのかを実験により検証する。また、検証に当たりユーザのアイテムに対する評価率、ユーザ一人当たりの評価数、アイテム一つ当たりの評価数という3つの尺度を用いることで対象の規模を表現することを提案する。

2. 協調フィルタリングと問題点

2.1 協調フィルタリング

協調フィルタリングとはユーザの情報収集の際の行動履歴といった暗示的な情報や対象に対する明示的な評価などから興味、関心などのユーザに関する情報を大量に収集し、その中から類似したユーザやアイテムを発見することで関連性からアイテムの推薦を行なうという手法である。この手法で推薦を行なうために最低限必要な要素は、システムを利用するユーザデータ、システムの保有するアイテムデータ、ユーザがそのアイテムに対しての嗜好を評価した評価データの3つである。

協調フィルタリングは既に実用されている例が多く存在し、オンライン書店 Amazon.com などが有名である。

協調フィルタリングには大きく分けてモデルベース法とメモリベース法の2種類ある。モデルベース法とはユーザの嗜好のモデル化を行ない、そのモデルに従って推薦を行なう手法である。対してメモリベース法はユーザやアイテム同士の類似性に注目して、その類似性から推薦対象を予測する手法である。このメモリベース法は大きくユーザの類似性に注目するユーザベース手法とアイテムの類似性に注目するアイテムベース手法の2つに分けることができる。

本研究では特に後者のメモリベース法に焦点を当てて述

べることとする。メモリベース法はモデルベース法に比べシステムを構築してしまえば手軽に利用でき、ユーザの評価を得ることができれば、どんなものに対しても推薦を行なうことができる利点がある。

2.2 メモリベース法の問題点

メモリベース法では推薦を行なう際の評価データ数などが少ない場合に、システムの正確さが落ちる傾向にある[1]。その為、小規模なシステムを構築して推薦機能を付加しようとした場合や、新たに推薦システムを構築した場合にどの程度の規模のユーザデータやアイテムデータ、評価データが必要であるのかが正確な推薦を行なう上で重要になる。システムの規模が小さい場合にシステムの推薦の正確さが落ちる理由として類似ユーザ数や類似アイテム数の大きさの影響やスパースシティ(Sparsity)問題などの理由が挙げられる[2, 3]。

類似ユーザ数の不足とは、ユーザ間の類似度に基づいて判断された嗜好が似ているユーザの数が足りない状況である。類似ユーザ数が不足していると偏ったユーザによる推薦になってしまうことや、推薦自体が出来なくなる。類似アイテム数不足も同じようなことが起こると考えられる。

スパースシティ問題とはユーザが評価したアイテム数に対して、システム上の総アイテム数があまりにも大きい場合に起こるものである[3]。このようなときユーザが実際に評価するアイテムの数は総アイテム数にくらべて極めて少ないため、推薦の精度が下がってしまう。

3. 協調フィルタリングの推薦における許容データ数の割り出し手法

3.1 推薦のためのデータ規模

利用ユーザ数とアイテム数の比率は予測精度に影響するといわれている[4]。そのため、推薦システムに必要なユーザデータ、アイテムデータ、評価データの3データの数に注目し、その3つの比率として、アイテムに対する評価率、ユーザ一人当たりの評価数、アイテム一つ当たりの評価数の3つをデータ規模の尺度とすることを提案する。この尺度は対象サイトなどの保持するデータ数の大小に関わらず評価数、ユーザ数、アイテム数のバランスと大きさを評価できると考えられる。

● ユーザのアイテムに対する評価率

これはユーザがアイテムに対しての全体な評価している割合であり、実際に評価されている評価数を全ユーザと全アイテムの積で除算することで算出する。

● ユーザ一人当たりの平均評価数

これは実際に評価されている全評価数を全ユーザ数で除算して算出する。

† 公立はこだて未来大学 システム情報科学研究科

‡ 公立はこだて未来大学

		Per Item (8/3 = 2.66 ratings)			Rate (11/16)
		A	B	C	
users	items				
1	1	Non	3	4	Per User (8/4 = 2 ratings)
2	1	2	5	Non	
3	1	5	Non	5	
4	1	Non	3	4	

Rate: ユーザのアイテムに対する評価率
Per User: ユーザー当たりの評価数
Per Item: アイテム一個当たりの評価数

Five point scale(1-5)

図1 データ規模となる3つの尺度

- アイテム一つ当たりの平均評価数

これは実際に評価されている全評価数を全アイテム数で除算して算出する。

単純にユーザ数が多ければ類似ユーザ数も増える可能性が高まるためユーザ数がシステムの推薦の正確さに関係しているといえる。さらに、一定数の嗜好データが必要であることや類似ユーザ数や類似アイテム数の変化によって推薦の正確さが変化することから全体の評価値数とアイテム数が関連している。

また、オスレを行なうには単純にこれらの数が多ければいいというわけではない。スパースティ問題からユーザが評価したアイテムの数に対して、アイテムの総数があるよりも大きい場合にも、推薦精度が落ちてしまう。

3.2 データ数変動によるデータ規模推定手法

推薦システムの必要データ規模の推定手法として提案するのが、データ数変動による判別である。

必要なデータは次の3データである。

- 評価データ
- ユーザデータ
- アイテムデータ

データ数変動とは、ユーザ数、アイテム数、評価数のうち2値の数値を固定し残りの1つの値を動かすという操作を3つの値それぞれに適用させる手法である。これは類似ユーザ数やアイテム数がある一定規模を境に推薦の正確さが急激に変動することを利用したものである[2, 3]。

各データ数を十分に用意し徐々に減少させていけば、それに伴い類似ユーザ数が減少し近傍形成に影響を与える。更には推薦の正確さも急激に損なわれると考えられることから、その境界点を求めることで適切な推薦に必要なデータ規模を推定することができる。

本研究での適切な推薦が出来るとは、システムが安定してユーザへの評価の予測を行なえる状況とし適切な推薦ができないとは、システムの評価の予測がバラツキ不安定になることとする。その為、必要データ規模とは MAE 値がフラットであるときのデータ数を指すこととする。また、ユーザの嗜好とは違った推薦が多く推薦予測の正確さが低い状況を指すこととする。

システムの推薦予測の正確さの指標には平均絶対誤差(MAE)を用いる。MAEは以下の式(1)で計算される。(p: 実際の評価値, q: 予測評価値, N: 評価アイテム数)

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \quad (1)$$

これはシステムの予測の正確さを統計的に測る有効な手段である。この MAE の値が小さければ小さいほど、システムの予測値は正確なものである。

4. データ規模推定実験・実験結果

4.1 必要データ規模推定実験

システムが適切な推薦を行なうために必要なデータ規模の測定が可能であるか、実際に実験を行なった。

実験は協調フィルタリングの代表的なアルゴリズムである GroupLens[5]のアルゴリズムを用いた推薦システムを実装し行なった。これはメモリベース手法の中でもユーザベース手法と呼ばれる推薦手法を用いている。これまでの研究でより優れたアルゴリズムは多くあるが、多くは GroupLens から派生した手法であるため比較されることが多く必要なデータ規模を出す上では良い基準となると考えられる。

実験は評価数に関して、ユーザ数に関して、アイテム数に関しての3つの実験を2段階行なった。1段階目はデータ数の減少幅を大きくした上で必要データ規模となる境界点を求め、2段階目ではデータの境界点付近を更に細かく検証を行い、必要データ規模を推定した。

4.1.1 実験環境

実験に使用したデータセットは、映画を1~5の5段階で評価を行なった MovieLens のデータセットを利用した。実験の計算速度の関係で評価数変動実験のみユーザ数を3020人、評価数は50万評価で行なった。この実験では映画がアイテムに当たる。

- 評価数変動実験, アイテム数変動実験

ユーザ数: 約3000人
アイテム数: 約4000アイテム
評価数: 約50万評価

- ユーザ数変動実験

ユーザ数: 約6000人
アイテム数: 約4000アイテム
評価数: 約100万評価

なお、評価数変動実験のみ約100万評価のデータセットを二分割し、二度の実験を行なった。

4.1.2 データ数減少方法

ランダム抽出で減少させたデータセットを作成し、計測は複数回行なってその減少数での結果の平均値を最終結果とした。また、計測回数はデータの減少数に伴い計算回数を増やし、より平均的な結果を得られるようにした。2段階目ではより詳細な結果を求めるために、評価数は1万評価単位の測定、ユーザ数は50ユーザ単位の測定、アイテム数は100アイテム単位の測定を行い、1段階目よりも3倍以上の試行回数を重ねた。

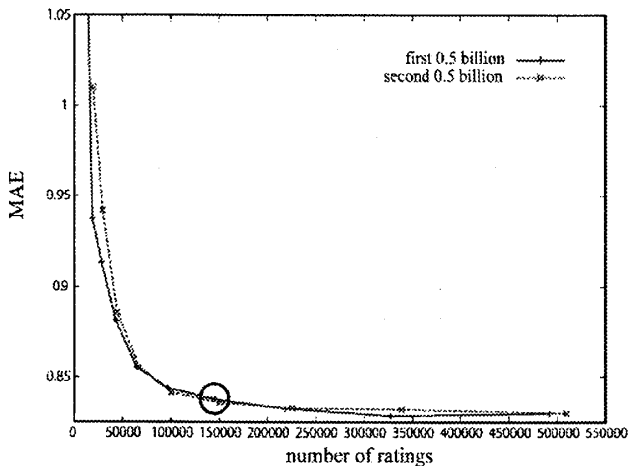


図 2.1 評価数と MAE の推移

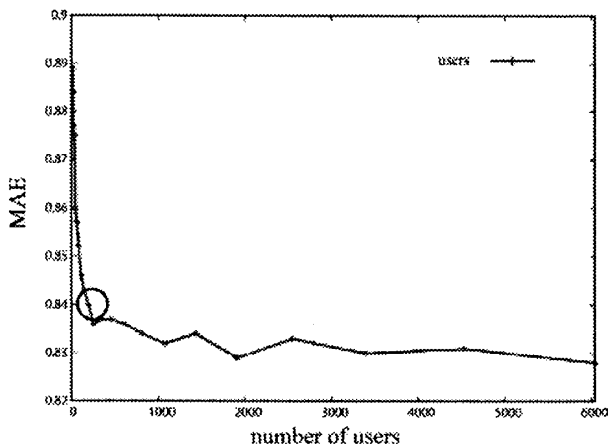


図 2.2 ユーザ数と MAE の推移

4.1.3 必要データ規模の判断基準

本実験では適切なデータ規模となる MAE 値の判別は全体における記録値から急激に変化する点を基準となる境界点とする。境界点を越えない段階でのデータ数を境界におけるデータ数とする。

4.2 実験結果と考察

4.2.1.1 段階目

評価数変動の結果を図 2.1, ユーザ数変動実験の結果を図 2.2, アイテム数変動実験の結果を図 2.3 に示す。3つの図はどれもデータ数が大きいときに MAE がフラットな状態となり、小さいときに急激に上昇している。図のなかの丸で囲っている点は必要データ数を判別する境界点である。

必要データ規模の基準境界点として、フラットとなる範囲は、記録した MAE 値の全体の 7%とし、MAE の最低値 0.828 から 0.841 までとした。したがって、MAE が 0.841 を越える値を記録したデータ数の前の段階でのデータ数を必要データ規模と定めた。

各実験でのデータ規模は評価数では 145564, ユーザ数は 191, アイテム数は 937 である。なお、ここで示している評価数は前半 50 万評価のデータセットを用いた数値である。

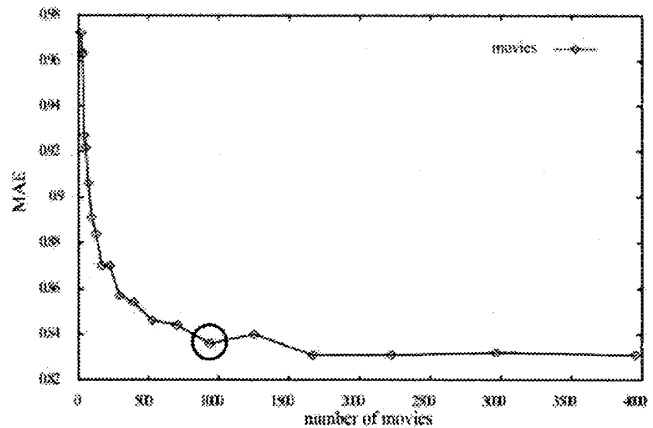


図 2.3 アイテム数と MAE の推移

4.2.2 段階目

1 段階目の結果より、境界付近でのより細かなデータ数の減少幅をとり、計測回数を増やすことでより正確な結果を測定し、その結果を元に評価率、ユーザー人当たりの平均評価数、アイテム一つ当たりの平均評価数を計算した。評価数を減少させて記録した結果を表 2、ユーザー数を減少させて記録した結果を表 3、アイテム数を減少させて記録した結果を表 4 に示す。

必要データ規模の基準境界点は 1 段階目の結果を用いた結果は必要数となる評価数は 80000 評価、ユーザー数 20 アイテム数 900 である。必要データ規模の結果として表 1 に示す通り、評価率は 0.67%, ユーザー人当たりの平均評価数は 29.43 評価、アイテム一つ当たりの平均評価数は 9.9 評価である。

4.3 実験の考察

メモリベース手法の中でもユーザーベースに分類される手法において MAE が急激に上昇することが観測でき、必要データ規模の判断基準を定めることで、適切な推薦に必要な必要データ規模を推定可能なことが確認できた。

今後の課題として、今回は必要データ規模の基準値を記録した MAE 値の全体の割合を用いたため、データ数が少ないときは図 3 に示すように MAE の標準偏差が大きくなるなど挙動が安定しなかった。そのため、MAE の最大値の測定は難しく信頼性が低いという問題点があり、別の手法を確立する必要がある。

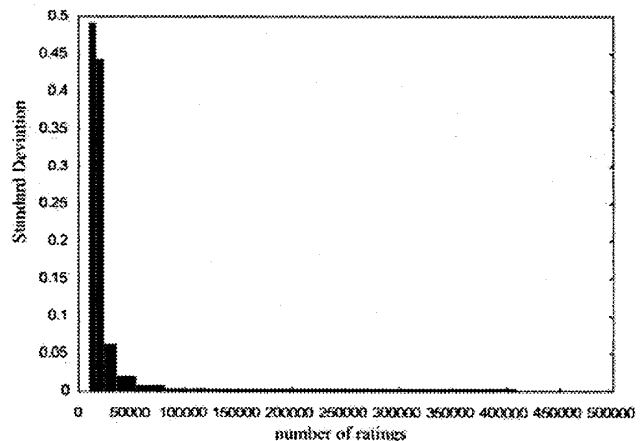


図3 評価数と MAE の標準偏差

表1 必要データ規模結果

評価率	0.0067
ユーザー一人当たりの平均評価数	29.43 評価
アイテム一つ当たりの平均評価数	9.9 評価

5. まとめ

本論文では、推薦システムの要素技術としての協調フィルタリングにおける必要データ規模の分析を行なった。

メモリアベースの協調フィルタリングの推薦システムでは類似ユーザ数や類似アイテム数が一定数以上では推薦の正確さが一定値になることや、利用ユーザ数とアイテム数の比率が推薦の正確さに影響を与えることに注目し、評価数やユーザ数、アイテム数を減少させることで必要データ規模の推定を行なった。本研究では実際にデータ規模の推定法の有効性を検証するために推薦システムを実装し、実験を行なった。

実験からデータ規模の尺度として、ユーザのアイテムに対する評価率、ユーザー一人当たりの平均評価数、アイテム一つ当たりの平均評価数によってデータ規模を推定することが出来た。今回の実験では適切な推薦が行なえる範囲を MAE 値からフラットな範囲とし、それを MAE の最低値から 7% までとした。

実験の結果、推定できた推薦に必要なデータ規模は評価率が 0.67% 以上、ユーザー一人当たりの平均評価数を 29.43 評価以上、アイテム一つ当たりの平均評価数を 9.9 評価以上保有することが適切な推薦を行なうために必要であるという結果となった。

本研究ではユーザベースの協調フィルタリングでの検証しか行なっていないが、メモリアベースでの手法であれば今回用いた手法が利用できると考えられるため、今後はアイテムベース手法での適用を目指すと同時に、必要データ規模の基準値確定手法の確立を目指す。

参考文献

- [1] David Goldberg, David Nichols, Brian M. Oki and Douglas Terry "Social Information Filtering: Algorithms for Automating "Word of Mouth"" Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'95).
- [2] Jonathan L. Herlocker, Joseph A. Konstan, Al Borchers, and John Riedl, "An Algorithmic Framework for Performing Collaborative Filtering" Proceedings of the 1999 SIGIR Conference on Research and Development in Information Retrieval, (1999).
- [3] Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl "ItemBased Collaborative Filtering Recommendation Algorithms" Proceedings of the 10th international conference on World Wide Web, Hong Kong, (2001).
- [4] J. Herlocker, J. Konstan, L. Terveen and J. Riedl, "Evaluating Collaborative Filtering Recommender Systems", ACM Transactions on Information Systems 22 (2004), ACM Press, 5-53.
- [5] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom and John Riedl "GroupLens: An Open Architecture for Collaborative Filtering of Netnews" CACM. 40(3), March 1997.

表2 評価数減少結果

訓練数	MAE	評価率	標準偏差
150000	0.835	1.26	0.0035
140000	0.835	1.17	0.0034
130000	0.836	1.09	0.0037
120000	0.836	1.01	0.0036
110000	0.837	0.92	0.0036
100000	0.838	0.84	0.0036
90000	0.839	0.75	0.0044
80000	0.841	0.67	0.0042
70000	0.844	0.59	0.0056
60000	0.849	0.50	0.0067

表3 ユーザ数減少結果

ユーザ数	訓練数	MAE	perItem	標準偏差
600	79567	0.834	20.13	0.0117
550	72815	0.834	18.42	0.0110
500	66283	0.834	16.77	0.0112
450	59921	0.837	15.16	0.0123
400	53467	0.837	13.53	0.0119
350	47021	0.837	11.90	0.0147
300	39131	0.841	9.90	0.0160
250	32980	0.838	8.35	0.0157
200	26607	0.839	6.73	0.0209
150	20066	0.844	5.08	0.0246

表4 アイテム数減少結果

アイテム数	訓練数	MAE	perUser	標準偏差
1700	168772	0.832	55.88	0.0052
1600	158153	0.835	52.37	0.0060
1500	149348	0.835	49.45	0.0059
1400	139520	0.835	46.20	0.0052
1300	128984	0.839	42.71	0.0060
1200	119236	0.835	39.48	0.0071
1100	109736	0.838	36.34	0.0088
1000	99613	0.837	32.98	0.0066
900	88873	0.839	29.43	0.0092
800	79383	0.842	26.29	0.0084
700	70026	0.845	23.19	0.0084
600	59466	0.845	19.69	0.0104