

研究指導 中澤 真 准教授

評価値の偏りを考慮した情報推薦システム

下田 昂

1. はじめに

情報収集はさまざまな活動において重要な作業である。それが近年のスマートフォンの爆発的な普及やソーシャルネットワークの流行により、誰でも簡単に大量の情報へアクセスすることが可能となった。多くの情報を取り扱えるようになったことは非常に便利なことではあるが、反対に情報量の膨大さはユーザー個人の情報収集能力では手に余り、情報を効率的に探し出すことは難しい。また、ユーザーだけの力では情報の選択が困難になってしまうといった負担も生じている。そのため、膨大な情報の中からユーザーが必要とする情報のみを提供してくれる情報推薦システムの必要性が高まっている。情報推薦システムとは、ユーザーごとに購入履歴や商品評価などの嗜好データをビッグデータとして蓄積し、それらの情報を解析することによってユーザーにとって有意義な情報を抽出し提示することができるシステムである。

しかし、情報化が進み多種多様なサービスに情報推薦システムが利用されるようになった現代においても、その推薦精度は未だに不十分なところがある[1][2][3]。この原因の一つとして情報推薦に利用しているユーザーによる評価データの品質の問題が挙げられる。商品の良し悪しや好みについての評価値を入力する際に、すべての絶対的な基準でかつ適性に入力しているとは限らない。また、評価の傾向について個人差もあり、辛口な人、甘い人など偏りもある。

そこで本研究では、ベンチマークデータ¹のすべての評価値から平均評価値を差し引いた偏差や、さらにそれらを標準偏差で割り、ユーザーごとのばらつきを標準化したデータを用い、評価値の偏りを考慮することによって予測精度の向上を目指す。

2. 情報推薦システムの概要

情報推薦システムとは、ユーザーにとって有用と思われるモノやサービスを抽出し、それらをユーザーの目的に合わせた適切な形で推薦するシステムである。その手法は多岐にわたり、以下に概要を説明するとともに活用例を示す。

2.1 情報推薦システムの説明と活用例

まず、我々が普段の生活の中で大量の情報を目にしたとき日常的に行っている処理を考えると、過去の経験や直感に基づき重要な情報や有用性の高い情報を推測している。そして、それらについて情報の優先順位を設け、不要な情報の除去をしていることが多い。このよう

な処理を情報フィルタリングという[4]。この情報フィルタリングの技術を用いて作られたユーザーの情報収集を支援するシステムが情報推薦システムである。推薦の方法はさまざまであるが、現在多くの情報推薦システムには、その推薦方法に協調フィルタリングという手法が用いられている。これは興味や関心が類似しているユーザーは同じような情報を求めているという直感的かつ経験的な過程に基づいている。この手法は実用性が高いため、実装されている多くの推薦システムがこの手法に基づいている[5]。また、この手法は図1に示したように分類することができる[6]。

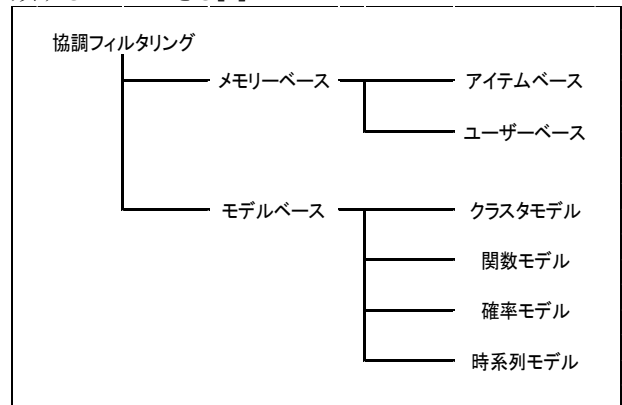


図1: 協調フィルタリングの分類

メモリーベースの協調フィルタリングについては、ユーザーベースとアイテムベースの2種類に分類することが可能である。ユーザーベース協調フィルタリングは、ユーザーの購入履歴や商品評価などの嗜好データからユーザー間の類似度を計算し、推薦するアイテムを決定する手法である。一方のアイテムベース協調フィルタリングはユーザーの購入履歴や商品評価などの嗜好データからアイテム間の類似度を計算し、類似するアイテムを推薦する点が先ほどの手法と異なっている。現在、多くの情報推薦システムに用いられている手法は後者のアイテムベース協調フィルタリングである。

以上のような情報推薦システムは、Netflix²のおすすめ機能やiTunes³の楽曲推薦など多種多様なサービスで用いられている。情報推薦システムの利点は、ユーザーの要求を満たすと思われる情報をユーザーに代わって探してくれることである。これにより、自身が膨大にある情報から手探りで必要な情報を探す必要がなくなるのである。そのため、情報と結びつきが強い現代において情報推薦システムは欠かすことのできないものとなっている。

¹ 実際のデータのことである

² netflix.com [https://www.netflix.com/jp/]

³ apple.com [https://www.apple.com/jp/itunes/]

2.2 情報推薦システムの評価指標

情報推薦システムの評価指標として、いかにユーザーが未知であったアイテムを推薦するかを評価する新規性や単純な方法では予測が困難な意外なアイテムの推薦を評価する意外性、意外性があるおかつ、有用なアイテムの推薦を評価するセレンディピティなどがある[7]。上記以外にも数多くの評価指標があるが、その中でも最も重要なものとして精度があげられる。

精度の評価指標として、システムが推薦した情報に対してユーザーの要求を満たす情報を含む割合である適合率や、ユーザーの要求を満たしている情報のうち実際に推薦された情報の割合である再現率などさまざまあるが、一般的に絶対平均誤差(MAE)⁴で定義されることが多い。そのため、本研究においても情報推薦システムの精度の評価基準としてMAEを用い、精度の向上を目指す。MAEは予測対象となる全てのアイテム数を N 、 i 番目のアイテムの実測の評価値を y_i 、 i 番目のアイテムの評価の予測値を \hat{y}_i と置くと以下の式で表される。

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (1)$$

3. 先行研究の概要

先行研究[1][2][3]では、アイテムベースの協調フィルタリングをWeightedSum法で行っている。この手法は、ユーザー u の未評価アイテム i の評価値を、式(2)の計算によって予測するものである。予測値 $p(u)_i$ は、ユーザー u の評価済みアイテムの評価値に基づいて計算され、予測対象の未評価アイテム i との類似度が高い評価済みアイテムの評価値ほど、予測値に強い影響を与えるように重み付けをするような計算をしている。

$$p(u)_i = \frac{\sum_{j \in I} (s_{i,j} * R_{u,j})}{\sum_{j \in I} |s_{i,j}|} \quad (2)$$

式(2)で、 $R_{u,j}$ はユーザー u のアイテム j に対する評価値、 $s_{i,j}$ はアイテム i と j の調整コサイン類似度、 I はアイテムの集合とする。さらに、アイテム i と j の調整コサイン類似度 $s_{i,j}$ は次の式で求められる。

$$s_{i,j} = \frac{\sum_{u \in U_S} (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U_S} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U_S} (R_{u,j} - \bar{R}_u)^2}} \quad (3)$$

また、 U_S はアイテム i と j のどちらにも評価したユーザー集合を、 \bar{R}_u はユーザー u の平均評価値を表す。

これに対し、工藤の手法[5]では高い精度が得られない原因を、式(2)において信頼性が低い評価データまですべてを使って計算していることが影響していると考えた。そのため式(2)の計算過程において、予測精度を悪化させていると考えられる評価値を加重和の計算から除外することにより精度の向上を図っている。しかし、それでもMAEは4.3ほどあり、精度が高いとはいえない。そのため、本研究では予測計算内においてユーザーごとの評価値の偏りを考慮したうえで、予測精度を悪化させていると思われる要素を排除することによって精度向上を目指す。

4. 提案手法

本研究ではアイテム間類似度に基づく加重和によって、未評価アイテムの評価値を予測する方法をベースとしている。先行研究[5]では、閾値を設けることにより計算に用いる評価値を限定して予測値を算出していたが、本研究では偏差の値や標準化した値を用いることでユーザーごとの評価値の偏りを考慮した予測値の計算をする。また、工藤の手法は類似度に基づく評価値を求める式(2)の計算過程で不必要な評価値を限定することで予測値を算出していたが、この考えに基づくのであれば類似度を求める式(3)の計算過程においても、使用する評価値に制限をかけるべきだと考える。そのため本研究では、工藤の手法にこの修正を加えた予測値の算出方法についても提案する。以下にその詳細を述べる。

一つ目の手法はベンチマークデータの評価値を偏差に置き換えた上で、類似度を算出し加重和を計算する手法である。評価の傾向について個人差もあり、辛口な人、甘い人など偏りが生じている。これは予測精度をかえって悪化させると仮説を立て、全ての評価値について平均評価値を差し引いた偏差を求め、この値を用いて類似度を算出し予測計算を行った。この手法を「手法A」とする。

二つ目は工藤の手法2を類似度の算出をする際にも適応するというものである。工藤の手法2は、 $|R_{u,j}| \geq \beta$ の条件を満たす項のみを用いて、加重和を計算する手法である。これは中間的な評価値は曖昧な情報を多く含んでおり、予測計算の精度に悪影響を及ぼす可能性が高いという考えに基づいている。しかし、類似度の算出の時にも評価値を用いる際に不必要な評価値を除外しないのは矛盾が生じている。したがって、類似度計算においても使用する評価値に制限を加えればさらなる精度向上につながると仮説を立てた。この手法を「手法B」とする。

三つ目は工藤の手法3を類似度の算出をする際にも適応するというものである。工藤の手法3は、 $|R_{u,j}| \leq \beta$ の条件を満たす項のみを用いて、加重和を計算する手法である。これは極端な評価値は予測計算への影響が大きく、その評価が適切でない場合は強い悪影響を及ぼすという考えに基づいている。これもまた手法2と同様、類似度の算出の際に不必要な評価値を切り捨てないのは矛盾が生じていると考え、類似度計算に用いる評価値に制限を加えればさらなる精度向上が見込めると仮説を立てた。この手法を「手法C」とする。

四つ目はベンチマークデータの全ての評価値を標準化し、それをもとに類似度を算出し加重和を計算する方法である。これは手法1の考えや仮説に基づいている。ユーザーごとに評価の傾向に偏りが生じていて、それが予測計算に悪影響を及ぼすのであれば、標準偏差で割り、個人差によるばらつきを是正することでより正確な類似度、およびより正確な予測値の算出が可能となり、予測精度の向上につながると仮説を立てた。この手法

⁴ この値が小さければ小さいほど精度が高いといえる

を「手法D」とする。

5. シミュレーションによる評価検証

先にあげた4つの手法について、工藤の手法と比較検証するために、ベンチマークデータを用いてシミュレーションをする。

5.1 使用するデータ

本研究では、ベンチマークデータであるJester Jokeを用いる[8]。このデータは、24980人のユーザーが100個のジョークに対して評価をしており、評価値は-10点から10点までの範囲をとる。表1はその一部である。

表1: データのサンプル(一部)

	アイテム1	アイテム2	アイテム3	アイテム4	アイテム5	アイテム6	アイテム7	アイテム8
ユーザー1	-7.82	8.79	-9.66	-8.16	-7.52	-8.5	-9.85	4.17
ユーザー2	4.08	-0.29	6.36	4.37	-2.38	-9.66	-0.73	-5.34
ユーザー3					9.03	9.27	9.03	9.27
ユーザー4		8.35				1.8	8.16	-2.82
ユーザー5	8.5	4.61	-4.17	-5.39	1.36	1.6	7.04	4.61
ユーザー6	-6.17	-3.54	0.44	-8.5	-7.09	-4.32	-8.69	-0.87
ユーザー7					8.59	-9.85	7.72	8.79
ユーザー8	6.84	3.16	9.17	-6.21	-8.16	-1.7	9.27	1.41
ユーザー9	-3.79	-3.54	-9.42	-6.89	-8.74	-0.29	-5.29	-8.93
ユーザー10	3.01	5.15	5.15	3.01	6.41	5.15	8.93	2.52
ユーザー11	-2.91	4.08			-5.73		2.48	-5.29
ユーザー12	1.31	1.8	2.57	-2.38	0.73	0.73	-0.97	5
ユーザー13					5.87		5.88	0.53
ユーザー14	9.22	9.27	9.22	8.3	7.43	0.44	3.5	8.16

5.2 シミュレーション方法

このベンチマークデータを用いて、工藤の手法と先に示した本研究の4つの手法について、それぞれシミュレーションによる評価検証をする。

シミュレーション方法はランダムにユーザー u^* を一人選び出し、そのユーザーの評価済みアイテムの中から一つ評価値 R_{u^*i} を未評価の状態へと置き換えて、予測値の計算をする。このとき、置き換え前の評価値 R_{u^*i} から予測値を差し引いた値が、ユーザー u^* に対してアイテム i についての評価値予測をした誤差となる。これをすべての評価済みアイテムに対して計算し、平均を求めたものがユーザー u^* に対するMAEとなる。シミュレーションではランダムに選び出した100人のユーザーに対して計算した平均値をその手法のMAEとしている。

続いて、提案手法の閾値について説明する。まず、手法Bである $|R_{u,j}| \geq \beta$ を満たす項のみで類似度の算出と加重和の計算をする手法では、 $\beta = 1$ の場合でシミュレーションしている。閾値を $\beta = 1$ に設定した根拠としては、工藤の手法で最も良い結果であったからである。次に手法Cの、 $|R_{u,j}| \leq \beta$ を満たす項のみで類似度の算出と加重和の計算をする手法は、 $\beta = 8$ の場合でシミュレーションする。その根拠としては手法Bと同様に、工藤の手法において最も良い結果であったためである。

5.3 シミュレーション結果

まず、手法Aの検証結果を図2に示す。図中の手法1、手法2、手法3は工藤の手法を指している。また、手法Aと手法1、手法Aと手法2、手法Aと手法3は、それぞれ手法Aによってデータを偏差に置き換えたのち、工藤の各手法を用いていることを指している。なお、グラフに用いている数値は閾値を変化させたときに最小となるMAEの値を用いている。

このグラフからわかるように、手法2と手法3においてMAEの改善が見られた。また、手法3は工藤の手法の最良の方法であるため、それをほんのわずかでも改善

できたのは評価すべき点である。しかし、手法1に関しては、かえってMAEが大きくなってしまった。

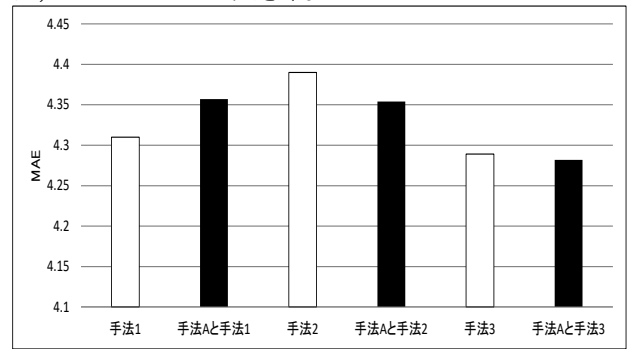


図2: 工藤の手法と手法Aの比較

次に、手法Bの検証結果を図3に示す。グラフからわかるように、MAEがわずかであるが増加して精度が悪化している。したがって、この手法は精度を高めるための有効な手段とはいえない。

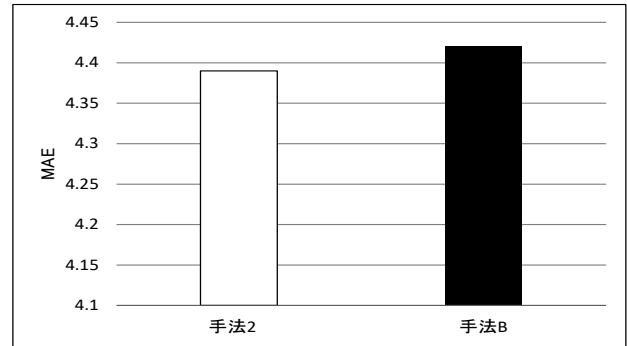


図3: 工藤の手法と手法Bの比較

続いて、手法Cの検証結果を図4に示す。グラフから読み取れるように、MAEが大きく増加してしまっている。これは精度の大幅な悪化を意味している。よって、この手法も精度を高めるための有効な手段とはいえない。

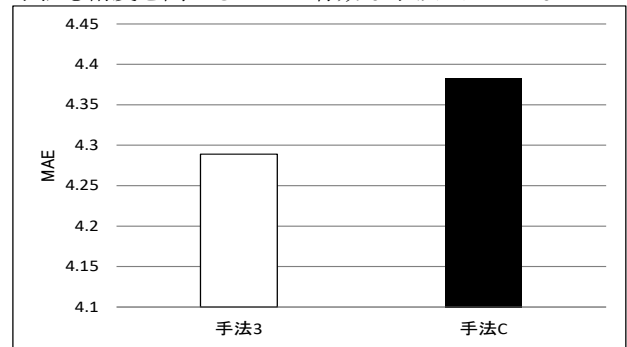


図4: 工藤の手法と手法Cの比較

最後に、手法Dの検証結果を図5に示す。図中の手法1、手法2、手法3は工藤の手法を指している。また、手法Dと手法1、手法Dと手法2、手法Dと手法3は、それぞれ手法Dによってデータを標準化したのち、工藤の各手法を用いていることを指している。なお、グラフに用いている数値は閾値を変化させたとき最も結果がよかったものである。

このグラフからわかるように、手法Dと組み合わせた手法2と手法3では、MAEの改善が見られたものの、その変化は非常に微小なものであった。また、手法1に関しては、逆にMAEが大きくなってしまっている。

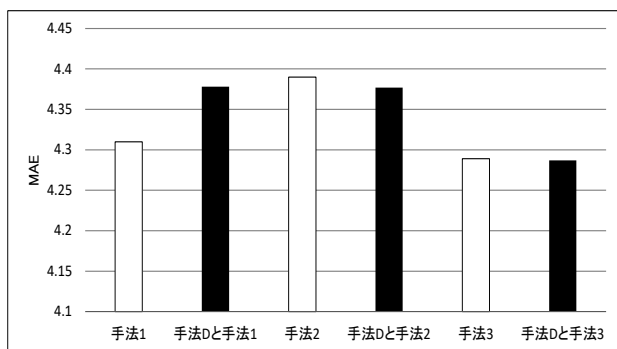


図5: 工藤の手法と手法Dの比較

6. 考察

以上のシミュレーション結果より、提案した4つの手法のうち2つについては微小ながら工藤の手法を改善できた。一方、手法2と手法3に関してはMAEが増加し、精度の向上どころか逆に精度が悪化してしまっている。そのため、これらについては閾値の設定に関して再検討を要すると考えられる。さらに、手法1と手法4について、MAEの低減が微小であったのは式(3)において各評価値から平均評価値を差し引いていることが影響していると思われる。具体的には、各評価値から平均評価値を差し引く計算は、すでに偏差や標準化といった偏りやズレの成分を加味しているのではないかということである。そのため、工藤の手法と比較してMAEがあまり大きく下がらず、手法によっては悪化してしまったのだと推測される。

また、一般的にサンプル数が少なければ少ないほど実際の値との誤差は大きくなる。本研究においても、その可能性は排除しきれない。そこで、今回のシミュレーション結果で最も結果がよかった手法Aと手法3を組み合わせたものと工藤の手法3について、ランダムに抽出するユーザーの数を増やし、同様にシミュレーションをした。その結果を図6に示す。なお、括弧内の数字は抽出したユーザーの数を表している。このグラフからわかるように、若干の変動を見せているものの、MAEはどのサンプル数のときにおいても本研究の手法のほうが低くなっている。したがって、本研究で提案した手法は予測精度の向上に有効である。

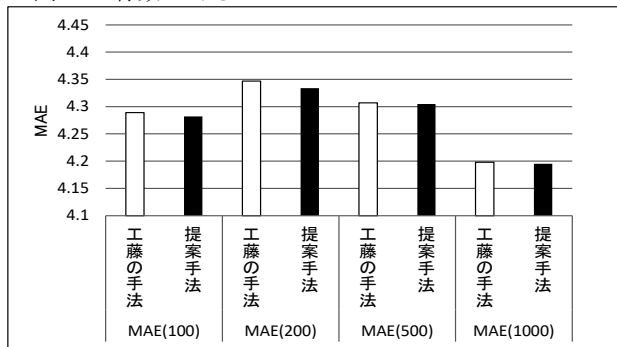


図6: 抽出するユーザーを増やした際のMAEの変化

7. おわりに

本研究では、偏差の値や標準化した値を用いることでユーザーごとの評価値の偏りを考慮した予測値を計算する手法を提案した。また、工藤の手法に修正を加えた予測値の算出方法も提案した。そして、提案した4つの手法のうち2つについては部分的で微小ながらも工藤の手法を改善できることを確認した。

今後の課題としては二点あげられ、一点目は使用したベンチマークデータがJester JokeだったからこそMAEの改善が見られた可能性がないとは一概にはいえない点である。そのため、MovieLensやBookCrossingなどのほかのベンチマークデータでもシミュレーションによる評価検証をする必要がある。二点目は本研究の手法でMAEの改善が見られなかった2つの手法についてである。本研究では類似度の算出と加重和の計算をする際に工藤の手法で最も結果がよかった閾値を設定していたが、閾値を変えることで誤差の改善が見られる可能性がある。また、本研究ではアイテムベース協調フィルタリングを用いたが、ユーザーベース協調フィルタリングでの評価検証によってMAEに何らかの変化が見られる場合も考えられる。したがって、今後は様々な手段で精度の向上に取り組んでいきたい。

参考文献

- [1] Badlur Sarwar, Geoge Karypis, Joseph Konstan, John Riedl, “Item-Based Collaborative Filtering Recommendation Algorithms”, Proceedings of the 10th international conference on World Wide Web, ACM, pp.285-295, 2001
- [2] 多田 美奈子, 菊池 浩明, “アイテム間類似度に基づくプライバシー保護協調フィルタリングの提案”, 情報処理学会論文誌, vol.51 no.9, pp.1554-1562, 2010
- [3] 伊藤 星, 吉見 哲哉, バンティヌラク スティエラ, 坂本 竜太, 柿崎 淑郎, 辻 秀一, “複合型アルゴリズムを用いた映画推薦システムの提案(一般)”, 情報処理学会研究報告.情報システムと社会環境研究報告, Vol.2008, no.81, pp.69-75, 2008
- [4] 大賀 さとみ, “Web上の児童被害防止におけるフィルタリングのありかた”, 2010年度卒業研究論文要旨集, pp.125-128, 2010
- [5] 工藤 勇希, “WeightedSum法を用いた閾値の設定に基づく情報推薦システムの精度向上”, 2014年度卒業研究論文要旨集, pp.121-124, 2014
- [6] 増永 良文, ソーシャルコンピューティング入門—新しいコンピューティングパラダイムへの道標—, サイエンス社, 2013
- [7] 奥 健太, 情報推薦システム入門: 講義スライド, <https://www.slideshare.net/KentaOku/ss-50762836>
- [8] Ken Goldberg, Anonymous Ratings Data from the Jester Online Joke Recommender System, Jester Joke Data Set <http://www.ieor.berkeley.edu/~goldberg/jester-data/>