

研究指導 中澤 真 准教授

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

工藤 勇希

1. はじめに

近年のインターネットの普及により、誰もが大量の情報へ簡単にアクセスできるようになった。しかし、多くの情報を取り扱えるようになり便利になった反面、情報の肥大化によりその選択が困難になってしまうという問題も生じるようになった。このように、情報が多くなり過ぎてしまい、目的の情報が発見できなくなってしまう問題を「情報過多」と呼ぶ[1]。このため、多くの情報の中からユーザに代わって必要な情報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。情報推薦システムは、膨大なユーザの行動履歴を蓄積し、それらの情報を解析することでユーザにとって有益な情報を抽出してユーザに提示することができる。

また、企業側の視点でも情報推薦システムは大きなビジネスチャンスになりえる。実際に活用されている典型的な例としては、クロセリング¹を目的とした商品推薦が挙げられる。クロセリングにより顧客流出を最小限にし、かつ顧客価値を最大化することが可能になる[2][3]。このような、ビジネス的な可能性を実現させるためにも情報推薦システムの開発が必要とされている。

しかし、情報推薦システムを用いたサービスが数多く提供されるようになって、その推薦精度はまだ不十分であり、ユーザが満足できる情報を必ずしも提供できているわけではない。精度を高めるために多くの研究が取り組まれているが[4][5][6]、データの種類や数によっては十分な性能を発揮することができないケースも存在する。

そこで本研究では、情報推薦システムの要でもある精度向上に焦点を置き、ユーザにとって正確な情報を提供できることを目指す。具体的には、既存の推薦方法の式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項を過重和の計算から除外をすることで、精度の向上を目指す。

2. 現在の情報推薦システムの概要と問題点

まず、一般的な情報推薦システムの概要を説明

し、その問題点について説明をする。

2.1 情報推薦システムの説明と活用例

情報推薦システムとは、ユーザにとって有用と思われる対象、情報、または商品などを選び出し、それらをユーザの目的に合わせた形で提示するシステムである。推薦の方法はさまざまであるが、現在の多くの情報推薦システムでは、その推薦アルゴリズムに協調フィルタリングという手法が用いられている。この手法は図1に示したようにいくつかのタイプに分類することが可能である[7][8]。

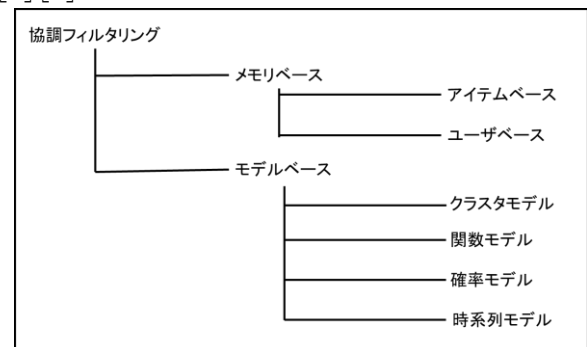


図1: 協調フィルタリングの分類

データベースの協調フィルタリングでは、アイテムベースとユーザベースの2種類にわけることが可能である。ユーザベースは、ユーザがアイテム(商品など)に対して行う評価を利用してユーザ同士の類似度を求める点に特徴がある。この値に基づいて似ているユーザ同士を抽出し、片方のユーザが高い評価をしているアイテムを、もうひとりのユーザに推薦するという手法である。一方のアイテムベースではアイテム同士の類似度を求める部分が先の手法と異なる点である。ユーザが高く評価したアイテムと類似度が高い別のアイテムをこのユーザに推薦するという方法である。現在、情報推薦システムで最も利用されている手法がこのアイテムベースである。

以上のような情報推薦システムは、Amazon²の商品推薦やGnosy³のニュース記事の推薦など、多様なサービスで活用されている。情報推薦システムの利点は、ユーザにとって満足度の高い情報をユーザの代わりに探してくれることである。これに

¹ ある商品を買ってもらった客に、同時に別の商品を薦めて複数の商品を買ってもらうこと

² Amazon.com [http://www.amazon.co.jp/]

³ Gnosy [http://gnosy.com/]

より、ユーザ自身が大量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、情報と深い関わりを持つようになった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。

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、最も重要なものが精度であり、ユーザの満足度に大きな影響を与える要因となる。例えば、オンラインショッピングのサイトにおいて、あるユーザが高く評価をされると思われる商品を推薦したときに、実際にユーザがその商品を高く評価した場合は精度が高いといえる。情報推薦システムの精度は、予測対象となる全アイテム数を N 、 i 番目のアイテムの実際の評価値を y_i 、 i 番目のアイテムの評価の予測値を \hat{y}_i と置くと一般的に以下の式で表される絶対平均誤差(MAE)で定義されることが多い。

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

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、この評価値の改善に取り組む。

3 先行研究の概要と問題点

先行研究[4][9]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 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。この予測値 $p(u)_i$ は、ユーザ 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。

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

ここで、 $R_{u,j}$ はユーザ u のアイテム j に対する評価値を、 $s_{i,j}$ はアイテム i と j の調整コサイン類似度を、 I はアイテムの集合とする。また、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。

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

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

しかし、後のセクションで検証結果を示すが、この手法では対象とする情報推薦のデータの種類によっては高い精度が得られない場合がある。その原因として、式(2)の右辺の分母と分子において、すべての評価済みアイテムを対象に和を求めていることが影響していると考えられる。例えばユーザが評価の判断に悩み、曖昧に評価してしまった場合、ユーザ自身もよくわからないような

値も足しこむことで、予測値が正しい値から離れてしまうことも想定される。そこで本研究では、この式(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項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る。

4. 提案手法

本研究の提案手法では、Badlur らの手法[4]と同様に、アイテム間類似度に基づく加重和を用いて、未評価アイテムの評価値を予測する。異なる点は加重和を求める対象範囲である。先行研究では式(2)の右辺の分母と分子において、すべての評価済みアイテムを対象に和を求めるとは、本研究では、閾値を設けることにより、和の計算で足しこむ項を制限して予測値を計算する。ここでは3種類の仮説を立てて、それに対応する制限の方法を考える。

一つ目は、未評価アイテム i と評価済みアイテム j の $|s_{i,j}| \geq \alpha$ の条件を満たす項のみを足しこみ、加重和を計算する手法である。これは予測対象のアイテムとの類似度が高い、あるいは逆にまったく似ていないという極端な評価済みアイテムの評価値のみで予測をする手法である。類似度が0に近い、すなわち似ているとも似ていないとも明確に区別できない評価済みアイテムの情報は、予測精度をかえって下げてしまうのではないかとこの仮説に基づいている。この手法を「手法1」とする。

二つ目は、 $|R_{u,j}| \geq \beta$ の条件を満たす項のみを足しこみ、加重和を計算する手法である。ユーザが評価した値において、好きでも嫌いでもない中間的な値は、ユーザが迷いながら入力した可能性が高く、その情報には曖昧さが多く含まれていると考えられる。このような曖昧な情報を用いることは、予測値の精度に悪影響を及ぼす可能性が高いため、これらを除外して計算することが精度向上につながるかと仮説を立てた。この手法を「手法2」とする。

三つ目は、 $|R_{u,j}| \leq \beta$ の条件を満たす項のみを足しこみ、加重和を計算する手法である。極端に高い評価、または低い評価は思い込みでつけてしまっている場合があると考えられるからである。例えばユーザが評価をする際、ある程度好きだったとしても、とても好きだと思いついてしまい、つい極端に高い点数をつけてしまう場合もあるであろう。しかも、極端な評価値は予測計算への影響が大きく、適正でない評価の場合はより強い悪影響が懸念される。そのため、これらを除外して計算することが精度向上につながるかと仮説を立てた。この手法を「手法3」とする。

5. シミュレーションによる検証

先に挙げた3つの仮説について、Badlurら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行う。以下にその詳細を述べる。

5.1 データ

本研究では、ベンチマークデータである Jester Joke データを用いた[10]。このデータは、24983人のユーザが100個のジョークに対して最低でも36個以上のジョークを評価している⁴。なお、評価値は-10.00点から10.00点までの範囲の値をとる。表1はそれらの一例である。

表1:データのサンプル (一部)⁵

	アイテム1	アイテム2	アイテム3	...	アイテム98	アイテム99	アイテム100
ユーザ1	-7.82	8.79	-9.66	...	99	99	99
ユーザ2	4.08	-0.29	6.36	...	0.34	-4.32	1.07
ユーザ3	99	99	99	...	99	99	99
ユーザ4	99	8.35	99	...	99	99	99
	⋮				⋮		
ユーザ24980	9.13	-8.16	8.59	...	-8.59	9.13	8.45
ユーザ24981	99	99	99	...	99	99	99
ユーザ24982	99	99	99	...	99	99	99
ユーザ24983	2.43	2.67	-3.98	...	99	99	99

5.2 シミュレーション方法

本研究では、上記のデータを使って、Badlurらの手法、提案した3つの手法を加えたものそれぞれに対してシミュレーションと検証を行う。シミュレーション方法は、ランダムに一人のユーザ u^* を選び出し、そのユーザの評価済みアイテムの中から一つだけ評価値 R_{u^*i} を未評価へと置き換えて、予測値の計算をする。このとき、予測値と置き換え前の評価値 R_{u^*i} との差が、ユーザ 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、その平均を求めたものがユーザ u^* に対する予測誤差となる。シミュレーションでは、これをランダムに選び出した100人のユーザに対して計算し、その平均値を算出した値をその手法のMAEとしている。

次に、提案手法の閾値についての説明をする。手法1の条件 $|s_{i,j}| \geq \alpha$ を満たす項のみを足しこみ、加重和を計算する手法については、 $\alpha=0.25, 0.5, 0.075, 0.1$ の場合でシミュレーションする。次に手法2である $|R_{u,j}| \geq \beta$ を満たす項のみを足しこみ、加重和を計算する手法では、 $\beta=1.00, 2.00, 3.00, 4.00, 5.00$ の場合でシミュレーションする。最後に手法3の、 $|R_{u,j}| \leq \beta$ を満たす項のみを足しこみ、加重和を計算する手法は、 $\beta=9.00, 8.00, 7.00, 6.00, 5.00, 4.00$ の場合でシミュレーションする。

このように提案した3つの手法については閾値

を設けて、どの閾値のときにMAEが小さくなるかを検証する。

5.3 シミュレーション結果

初めに、Badlurらの手法をそのまま用いた場合のMAEは4.49であった。これ以降の手法に対しても同様にシミュレーションによってMAEを求めて、誤差の比較を行う。

まず、手法1の検証結果を図2に示す。 $\alpha=0$ は閾値を設けていないことと同義であるため、これはBadlurらの手法[4]を用いた場合と同じ結果となる。このグラフからわかるように、どの閾値でもBadlurらの手法のMAEより小さくなるという結果が得られた。したがって、手法1によって閾値を設けることで誤差を小さくすることが可能である。なお閾値を決定する際には誤差を最小限に抑えられるように適切な値を考慮する必要がある。

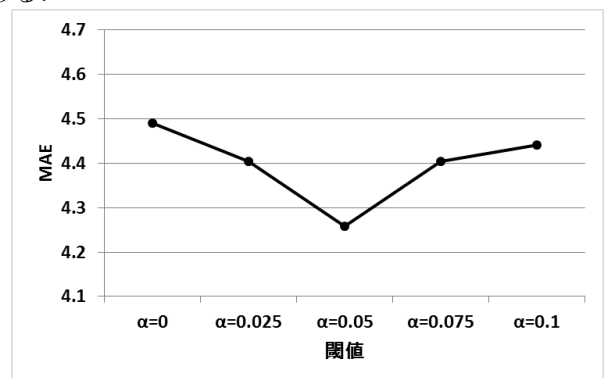


図2: Badlurらの手法と手法1の比較

次に、手法2についての検証結果を図3に示す。このグラフにおいても $\beta=0$ のケースはBadlurらの手法を用いた結果と同義である。図3より、 $\beta=1, \beta=2$ においては、誤差がわずかに改善されることが確認できたが、閾値を3以上の値とすると、かえって精度が悪くなることも明らかになった。しかも、手法2は手法1と比べると、誤差改善はわずかなものであるため、精度を高めるための有用な手法であるとはいえない。

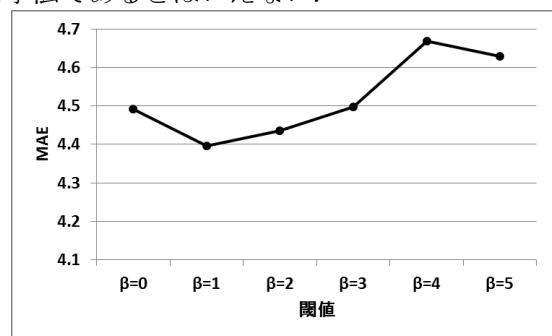


図3: Badlurらの手法と手法2の比較

⁴ 平均評価アイテム数は72.5

⁵ 「99」は未評価を意味する。

最後に手法3についての検証結果を図4に示す。グラフからも読み取れるように、 $\beta=10$ のときのBadlurらの手法よりも手法3のほうが、誤差が小さくなるのがわかる。しかし、閾値を大きい値に設定してしまうとBadlurらの手法よりも誤差が大きくなってしまうため、他の手法同様に、閾値は最適な値を考える必要がある。

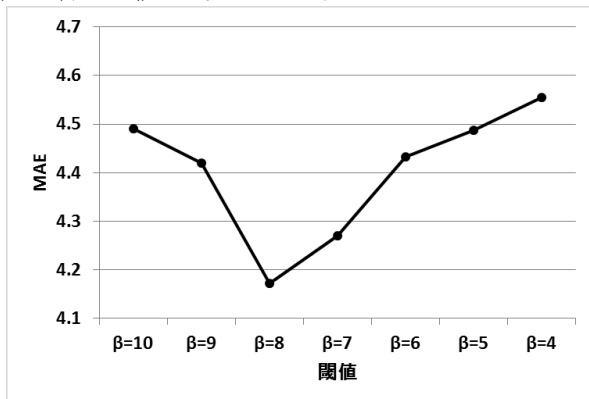


図4: Badlurらの手法と手法3の比較

6. シミュレーション結果の考察

以上のシミュレーションの結果より、提案した3つの手法が従来手法よりもMAEを小さくできることを確認した。また、それぞれの手法は閾値が適切な値でないと最適な結果が得られないため、閾値の設定を慎重に行う必要がある。今回の結果について、手法2の誤差の変化は手法1や手法3と比較して微小なものであった。したがって、手法2と手法3の結果より、「思い込みによって極端に高い点数や低い点数をつけてしまう」という考えの方が、「好きでも嫌いでもない中間的な値は、ユーザが迷いながら入力した可能性が高く、その情報には曖昧さが多く含まれている」という考え方よりも誤差を小さくさせる要因として大きく影響を与えたと考えられる。

また、今回の結果が過学習の解消によるものかどうかについても検討した。つまり、加重和の計算に用いている評価済みアイテムの個数が多すぎることで過学習となり精度を悪化させているのであれば、ランダムに選んだ項を除外して加重和を計算しても、精度の改善がみられるはずである。そこで、加重和に足しこむ評価済みアイテムを、全体の中からランダムに絞り込み、その割合を90, 80, 70, ..., 30%と変化させた場合のシミュレーションを実施した。その結果、足しこむ項を除外する割合が増加しても、誤差が小さくなる傾向はみられず、その変化にも規則性はみられなかった。したがって、加重和の計算に用いる評価済みアイテムを、提案手法のように適切な手段で取捨選択しない限り、予測精度は向上しないといえる。

7. 終わりに

本論文では、情報推薦システムの精度を向上させるために、閾値を設けてデータの取捨選択を行い、予測計算をする手法を提案した。さらに、従来の手法との比較をシミュレーションによって検証し、閾値を適切な値に設定することでMAEを従来よりも下げられることを確認した。

今回は、ExcelのVBAを用いたシミュレーションであったため、試行回数が十分でない面があった。今後は試行回数を増やしてシミュレーションを行う必要がある。また、今回の手法により一定の精度の改善は見られたが、ベンチマークデータにおける誤差はまだ4.2ほどあり、実際の評価値から多少離れていた。そのため、類似度の計算方法そのものを見直し、精度の向上に取り組む必要がある。

参考文献

- [1] 神畷 敏弘 “協調フィルタリングの課題:プライバシー, サクラ攻撃, 評価値のゆらぎ” 情報処理, vol.48, no.9, pp.966-971, 2007
- [2] Dietmar Jannach, Markus Zanker, Alexander Felfernig, Gerhard Friedrich, 情報推薦システム入門 -理論と実践-, 共立出版, 2012
- [3] Foster Provost, Tom Fawcett, 戦略的データサイエンス入門-ビジネスに活かすコンセプトとテクニック, オライリー・ジャパン, 2014
- [4] Badlur Sarwar, George 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
- [5] 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
- [6] 伊藤 壘, 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
- [7] 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
- [8] 神畷 敏弘, “推薦システムのアルゴリズム (1)”, 人工知能学会誌, vol.22, no.6, pp.826-837, 2007
- [9] 伊藤 寛明, 吉川 大弘, 古橋 武, “アソシエーションルールを用いた推薦システムにおける精度と意外性の向上”, 情報処理学会研究報告, Vol.2014-MSP-100 No.19, pp.1-6, 2014
- [10] Ken Goldberg, Anonymous Ratings Data from the Jester Online Joke Recommender System, Jester Joke Data Set <http://www.ieor.berkeley.edu/~goldberg/jester-data/>