

人気に対して中立な推薦のための類似度関数の評価

Evaluation of Similarity Functions for Popularity Neutral Recommendation

石岡 卓也

Takuya Ishioka

クリプトン・フューチャー・メディア株式会社

Crypton Future Media, INC.

This article shows which similarity function is appropriate for a neutral recommendation about popularity. Information neutral recommender systems are mainly researched about model base recommendation. However, many practical recommender systems made with similarity functions, because they can make a ordered list of items by similarity. We'd like to construct a popularity neutral recommendation using similarity function. Three similarity functions are evaluated as popularity neutral recommender system by using Calders & Verwer's discrimination score (CV score). This study shows jaccard coefficient makes better results than cooccurrence and simpson coefficient for popularity neutral recommendation.

1. はじめに

コンテンツの人気には偏りが生じやすい。コンテンツの販売数や閲覧数は、べき分布や対数正規分布のような裾野の広い分布になることが知られている。これは一度人気の出たコンテンツはさまざまな場面で言及され、その言及によってさらに多くの閲覧者を集めるといったメカニズムが働いているからだと考えられる。そうして非常に人気の出る作品が存在する一方で、作品が十分に高い品質を持っていても人気が出ない場合があることが知られている [Salganic 06]。

このような人気の偏りは、クリエイターにとっては「良い作品を作っても認められないことがある」という問題として、ユーザーにとっては「人気とは無関係に良い作品を見つけることが困難である」という問題として現れる。これは双方にとって望ましくないが、このような問題に対する解決策はあまり研究されていない。

推薦技術は 1990 年代から研究されていたが、2007～2009 年に実施された “Netflix Prize” で大きな注目を浴びてさらに進展した。実システムへの応用も盛んに行われ、売上を増加させることを目的として多くのサービスで導入されている。

コンテンツの推薦では、コンテンツとユーザーの関係を学習する協調フィルタリングを用いることで、ユーザーの好みに合致するコンテンツを推薦することが行われている。だが協調フィルタリングを利用した推薦手法では、既に人気のあるコンテンツが推薦されやすい傾向があると指摘されている [Celma 08]。

ある観点に対して中立な推薦に関しては、既にモデルベースの手法が提案されている [神鷹 14]。一方で、推薦に利用しやすい特性を持つアイテムベースの推薦手法も多くの実システムで利用されている [Linden 03]。アイテムベースの推薦であっても可能な限り中立性を保つために、本報告では類似度関数を用いた場合に推薦されるコンテンツについて検討する。中立性の指標としては Calders & Verwer の discrimination score (CV スコア)[Calders 10, 神鷹 13] を基にした特徴量を提案し

て利用した。また、piapro^{*1} の実データによって結果を確認し、Delicious^{*2} から取り出されたデータを用いて手法の妥当性を確認した。

2. 推薦と中立性

情報推薦において、特定の視点からの偏りがないように公正性に配慮した手法が提案されている。本研究では人気に着目しているが、他にも何らかの情報に対して偏りのない判断をしたい場面は多い。公正性に着目したデータマイニングとして [Pedreschi 08] が、公正な推薦の研究として [神鷹 14] がある。これらは、公正さが求められる情報であるセンシティブ情報と、分類される結果が統計的に独立であることを評価の基準としている。

中立性を定量的に扱うために、elift [Pedreschi 08], CV スコア [Calders 10] のような評価基準や、評価値予測関数に中立性項を導入するアルゴリズム [神鷹 14] などが提案されている。

2.1 中立性の評価

本研究では、中立性の指標として CV スコアを基にした評価値を用いる。

情報推薦は、アイテム集合を推薦されるものとそれ以外に分類する問題として考えることができる。分類結果として得られるクラスを $C = \{C_-, C_+\}$ 、センシティブ情報を $S = \{S_-, S_+\}$ としたとき、CV スコアの値 D は

$$D = P(C_+|S_+) - P(C_+|S_-) \quad (1)$$

と定義される。ここで P は確率である。意味としてはセンシティブ情報である人気 S_+ であるようなメジャーコンテンツが、推薦対象 C_+ になる確率と、マイナーコンテンツ S_- が推薦対象 C_+ となる確率の差を表す。この式はセンシティブ情報が $+, -$ どちらであっても推薦対象として選ばれる確率が変わらない場合に 0 となり、センシティブ情報が推薦結果に影響を

*1 「piapro(ピアプロ)」とは、クリプトン・フューチャー・メディア株式会社が運営する、CGM 型コンテンツ投稿サイト。音楽・イラスト・歌詞など、ユーザー同士が得意な作品を投稿し、コラボレーションし合い、新たな作品を生み出すことができる“創作の場”。
http://piapro.jp/

*2 Delicious website, http://www.delicious.com

与えるほど大きな値となるため、できるだけ 0 に近づけるように推薦対象を決めることになる。

3. 情報推薦におけるクラスタリングと類似度関数

協調フィルタリングを用いて情報推薦する手法として、ユーザーとアイテムの関係を直接利用するメモリベース法と、データから構築したモデルを利用して推薦を行うモデルベース法が存在する。メモリベース法ではユーザー間やアイテム間の類似度を用いる方法が主流であり、モデルベース法ではクラスタモデルを用いることが多い。

3.1 モデルベースのクラスタリング

情報中立推薦に関しては、主にクラスタモデルを基に議論が進められてきた。クラスタモデルは数理的な基礎がしっかりしているため、予測の精度を定量的に示すことができる、尤度でモデルの良さを表現できるなど、扱いやすい特性を持っている。

推薦の手法としては、クラスタモデルで対象アイテムの分類をモデル化し、同じクラスに分類されたアイテムを推薦するといった手法がとられる。この手法では全アイテムの中から有望なアイテム集合を取り出すことができるが、集合のサイズはアイテム全体のサイズとクラス数、使用するアルゴリズムによって変動し、制御することは難しい。

3.2 メモリベースの類似度関数

一方で類似度を用いた方法では、順序つきリストを容易に生成できるため、「アイテムを類似度順に r 個表示する」といったことを容易に行うことができる。これは web サービスにおいて表示可能なアイテム数が限られていたり、レイアウトに応じてアイテム数を変更する必要がある場合には重要な意味を持つ。

同様なリストをクラスタモデルで作成する場合は、クラスタリング以外に順序を定義する何らかの基準を導入する必要がある。また、「推薦されたアイテムの中から一つを選んで閲覧することを繰り返す」という実際に起こりそうなユースケースにおいて、クラスタリング結果からアイテムを推薦すると、同じクラスから選ばれるアイテムが何度も表示されてしまうため閲覧の体験が単調になりやすい。一方で類似度を利用した場合は、アイテム a から見てアイテム b が最も類似したものだとしても、 b から見て a が最も類似したアイテムとは限らないため、ある程度変化のある閲覧体験を提供できるという利点がある。

4. 実験の設定

上述のことを踏まえて、類似度関数の中立性を評価する。類似度関数として扱うものを 4.1 に、中立性の評価方法を 4.2 に示す。データとしては中立な推薦を行いたいサイトである piapro と、妥当性検証用の Delicious のものを用いた。

4.1 対象とする類似度関数

以下の様な類似度関数を用いた。アイテム a を評価したユーザーの集合を I_a 、アイテム b を評価したユーザーの集合を I_b としたときの類似度は、以下のように表現される。

Cooccurrence(共起度)

$$Cooccurrence = |I_a \cap I_b| \quad (2)$$

Simpson 係数

$$Simpson = \frac{|I_a \cap I_b|}{\min(I_a, I_b)} \quad (3)$$

Jaccard 係数

$$Jaccard = \frac{|I_a \cap I_b|}{|I_a \cup I_b|} \quad (4)$$

いずれも集合同士の類似度としてよく用いられるものだが、中立性に関する議論は少ない。

4.2 多数の分類結果に対する CV スコア

類似度関数を利用して推薦を行う場合、対象アイテムと似たものから上位 r 件を推薦対象として利用することになる。これはアイテム集合を C_- と C_+ に分類することに相当する。この分類結果をもとに CV スコアを計算することになるが、類似度の値は閲覧中のアイテムと推薦対象のアイテムの対に定義されるので、閲覧中のアイテムが異なれば全く別の分類が行われることとなる。閲覧中のアイテム毎に CV スコアが存在するが、そのままでは数が多すぎて扱いにくいので、以下のように各アイテムが C_+ となる確率を平均し、その差を新たな中立性スコア D_m として利用する。

$$D_m = \frac{1}{N_+} \sum_{t \in T_+} P^t(C_+|S_+) - \frac{1}{N_-} \sum_{t \in T_-} P^t(C_+|S_-) \quad (5)$$

ここで、 T_+ は S_+ であるアイテム集合、同様に T_- は S_- のアイテム集合である。 t はアイテムを表し、 $P^t(C|S)$ はセンシティブ情報が S であるアイテム t がクラス C と分類される確率、 N_+ と N_- はそれぞれ S_+ と S_- を有するアイテム集合のサイズである。

4.3 対象とするデータ

4.3.1 piapro のブックマークデータ

piapro の 2015 年 10 月時点での音楽のブックマークデータを用い、音楽に対するブックマークがあるユーザーと、ブックマークされた楽曲の中から今回の対象データを抽出した。対象データのユーザー数は約 320,000、アイテム数は約 98,000、ブックマーク数は約 3,600,000 で、被ブックマーク数が最も大きい曲は約 11,000 名のユーザーにブックマークされていた。アイテムの人気指標として被ブックマーク数を用い、被ブックマーク数 100 未満を S_- 、100 以上を S_+ とした。推薦対象とするアイテム数 $r = 50$ として、類似度の高いものから 50 件を推薦した。

4.3.2 Delicious のブックマークデータ

比較のために、RecSys 2011 で用いられた小規模な Delicious のブックマークデータ [Cantador 11] を用いた。内容はユーザー 1,867 名、対象ブックマークは 104,799 種類で、最もブックマークの多い記事は 29 名のユーザーにブックマークされていた。ブックマークに付与されたタグについては考慮に入れないこととする。

5. 実験と結果

上記のような設定で類似度関数による推薦を行ったと仮定して、人気に対する中立性を調べた。

5.1 中立性の評価

piapro と Delicious の双方について、前述した 3 つの類似度関数による推薦の中立性スコア D_m を求めた。結果を表 1 に示す。被ブックマーク数が piapro と Delicious では大きく異なるため、 S_- と S_+ の閾値は piapro で 100, Delicious で 10 とした。

表 1: 類似度関数ごとの中立性スコア D_m

類似度関数	piapro (D_m)	delicious (D_m)
Cooccurrence	0.00356	0.00476
Simpson	0.00139	0.00373
Jaccard	0.00033	-0.00026

どちらのデータにも同様の傾向があり、Cooccurrence の D_m が最も大きく、Simpson 係数、Jaccard 係数の順で小さな値になっている。 D_m は 0 に近いほど中立であることを示すスコアなので、3 つの尺度の中では Jaccard 係数が最も中立に近い対象を推薦できている。

5.2 被ブックマーク数と推薦のされやすさ

前節では S_- と S_+ を分けるブックマーク数を 100, 10 としたが、公平性の観点からは閾値がどんな値であっても同等な確率で推薦されることが望ましい。そこで piapro のデータについてアイテムの被ブックマーク数 k を横軸に、推薦対象としての選ばれやすさ u_k を縦軸に示したのが図 1 である。 u_k は

$$u_k = \frac{f_k}{r|T_k|} \quad (6)$$

と定めた。 f_k は被ブックマーク数 k に該当するアイテムが推薦先として選ばれた回数。 $|T_k|$ は k に該当するアイテム集合のサイズで、 r はあるアイテムを閲覧した際に推薦対象として選ばれるアイテムの数である。 u_k が大きいほど推薦されやすく、小さいほど推薦されにくい。推薦対象が k とは独立に選択されているとき、 $u_k = 1.0$ となる。 u_k が被ブックマーク数に関わらず 1.0 に近いほど、人気に対して中立な推薦がなされていることになる。

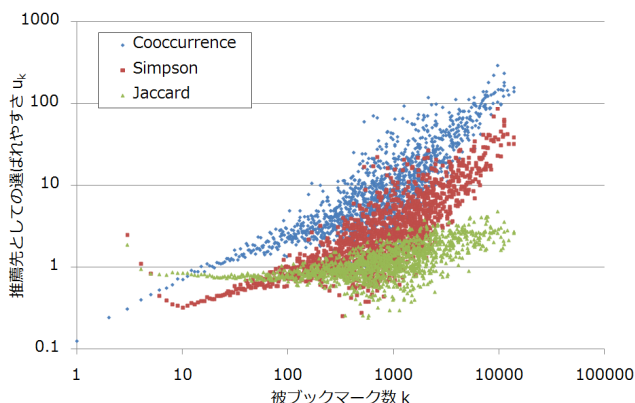


図 1: 被ブックマーク数と推薦のされやすさ

各距離関数について k と u_k の関係を図 1 に示した。被ブックマーク数が 10 以上の部分を見ると、 k が大きいほど u_k が大きくなる傾向がある。特に Cooccurrence に顕著で、 $k = 10,000$ 付近では $k = 10$ の 100 倍ほど推薦先として選ばれやすい。一

方で Jaccard 係数を利用した場合は $k = 10$ と $k = 10,000$ での u_k の差は 4 倍程であり、他と比べて人気に対して中立性の高い推薦ができていていることがわかる。

6. おわりに

本研究では Cooccurrence, Simpson 係数, Jaccard 係数の 3 つの類似度に関して、中立性スコア D_m と被推薦確率を確認することで評価を行った。結果として、この中では Jaccard 係数が最も中立に近い推薦を可能とする類似度であることがわかった。これらの結果をもとに、人気に対して中立で、実用的な推薦を実装して行きたい。

今回の報告では推薦対象の中立性を評価したが、精度に関する数値は扱わなかった。これは、類似度関数を用いた推薦では正解が定義されないため精度の議論ができないことに起因する。サービスとしての利用があれば CTR (Click-Through Rate) などによって推薦の良さを測ることができるため、実際の利用を通して議論を深めたい。また、今回の結果では他と比較して小さいとはいえ人気のあるアイテムが推薦されやすい傾向が残っている。この傾向を消すには、CV スコアを基に推薦対象を入れ替える、中立性に関する制約を導入した類似度を作成する、などの方法が考えられるため、今後の課題としたい。

参考文献

- [Salganic 06] Salganik, M. J., Dodds, P. S., and Watts, D. J.: Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market, *Science*, Vol. 311, (2006)
- [Celma 08] Celma, Ò. and Cano, P.: From hits to niches? or how popular artists can bias music recommendation and discovery, *Proc. of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*, ACM KDD, (2008).
- [神鷹 14] 神鷹敏弘, 赤穂昭太郎, 麻生英樹, 佐久間淳: 情報中立推薦システムの高速化, 2014 年度人工知能学会全国大会, JSAI 2014, (2014).
- [Linden 03] Linden, G., Smith, B. and York, J.: Amazon.com Recommendations: Item-to-Item Collaborative Filtering, *IEEE Internet Computing archive*, Volume 7 Issue 1, (2003).
- [Calders 10] Calders, T. and Verwer, S.: Three naive Bayes approaches for discrimination-free classification, *Data Mining and Knowledge Discovery*, (2010).
- [神鷹 13] 神鷹敏弘, 赤穂昭太郎, 麻生英樹, 佐久間淳: 情報中立推薦での中立性項の改良, 電子情報通信学会技術研究報告, IBISML 2013, (2013).
- [Pedreschi 08] Pedreschi, D., Ruggieri, S., and Turini, F.: Discrimination-aware Data Mining, *ACM KDD 08*, (2008).
- [Cantador 11] Cantador, I., Brusilovsky, P., and Kuflik, T.: 2nd Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011), *Proceedings of the 5th ACM conference on Recommender systems*, ACM, (2011).