

# グラフ研磨による考慮集合の予測と推薦システム

中原 孝信

キーワード：グラフ研磨，推薦システム，考慮集合

本稿は、専修大学商学部にて提出した市川 雄太さんによる 2015 年度卒業論文および [1] をもとに加筆修正したものです。

## 1. はじめに

楽天市場や Amazon のような EC サイトでは 100 万種類以上の商品を取り扱っているため、顧客が好む商品をうまく推薦し販売につなげることが売上増加には必要です。伝統的な推薦方法は、協調フィルタリング [2] と呼ばれる方法で、推薦対象者と好みの似ている人たちを探し、その人たちが購入した商品の中から満足度の高い商品を薦めるという方法です。この方法は実用的でわかりやすい方法ですが、過去に購入したという事実だけに注目しているため、推薦対象者が過去に「迷ったけど買わなかった商品」、「知らなかった商品」など、買わなかったという情報は推薦時には利用されません。

そこで本稿では、考慮集合と呼ばれる購入の際に候補になった商品、つまり購入候補になったが実際には購入しなかった商品を推定し、その考慮集合を利用した推薦方法を用いた結果を紹介します。

## 2. 考慮集合の予測と推薦方法

考慮集合を予測するに際して、グラフ研磨 [3] を利用します。顧客が購買した商品を節点にし、同時購買されやすい商品間に枝を張ることで無向グラフを構築します (図 1)。このグラフを類似度グラフと呼び、枝で接続された商品は同時購入されやすい商品とみなします。

次にこの類似度グラフに対してグラフ研磨を適用します。グラフ研磨は、直感的には枝密度の濃い部分グラフはより濃く、薄い部分グラフはより薄くすること

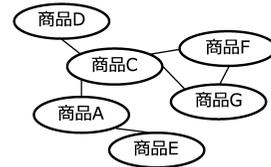


図 1 類似度グラフ

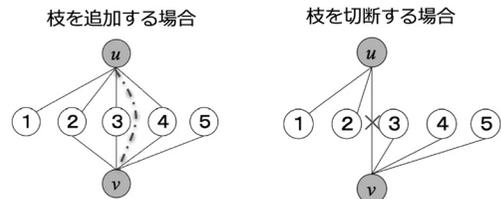


図 2 グラフ研磨 (枝追加・削除)

で、できる限り構造を明確化しておこうというものです。研磨の方法はシンプルで、すべての頂点ペアについて、その類似度が指定した閾値以上であれば枝を接続し、そうでなければ枝を接続しないというルールに従って、新たなグラフを再構成します。類似度は、グラフ上での二つの節点  $u, v$  の Jaccard 係数として式 (1) のとおり定義できます。

$$jc(u, v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|} \quad (1)$$

ここで  $N(u)$  は節点  $u$  に直接接続のある節点集合を表しており、図 2 の左では  $\{1, 2, 3, 4\}$  です。この類似度を用いてグラフを再構成すると、図 2 の左のように共通節点の多い節点間に枝が張られ (図 2 の点線)、図の右のように少ない節点間の枝は切断されます。そして、新たに構成されたグラフを入力として同様の研磨手法を繰り返して適用し、グラフの構成に変化がなくなるか、もしくはユーザの指定した最大繰り返し回数に達すれば終了します。この方法で、最終的に得られたグラフを研磨グラフと呼びます。

グラフ研磨によって頂点ペアに追加された枝は、頂点ペアで共通する共起関係を多くもっていることを示

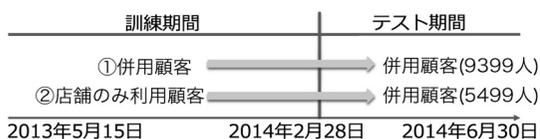


図3 実験の枠組み

しており (たとえば図2の左では  $\{2,3,4\}$  が共通集合), 関係性が類似していることを表しています. したがって, ある顧客が商品  $u$  を購入していれば,  $v$  はそのとき購入候補となった商品として考えることができるため, これらを考慮集合として捉えています. 商品の推薦は研磨グラフを利用して, 各顧客の購入商品と一致する節点から接続される節点の次数が多い順に  $N$  アイテムを推薦します. たとえば図1で, 商品 C, G を購入していれば最初に商品 F が推薦されます.

### 3. 計算実験

計算実験で利用したデータは2013年5月15日から1年間のデータで, 約230万人の顧客による実店舗とECサイトの購買履歴データです. 店舗とECサイトの併用顧客は全体の6%で, 店舗のみ利用顧客が64%, ECサイトのみ利用顧客が30%でした. 提案する推薦方法によって, 全体の6割に及ぶ店舗のみの利用者にECサイトの利用を促し, 併用顧客にさせることが目的です. 最終的な推薦対象顧客は, 併用顧客と店舗のみの利用顧客の中で, 金額と購入回数の高い上位30%の顧客に限定し推薦アイテムを決定しました.

図3に実験の枠組みを示します. 推薦精度を確認するために, 2013年5月15日から8カ月間を訓練期間として研磨グラフを構築し推薦商品を決定しました. そして, 残りの4カ月をテスト期間として推薦した商品がその期間に購買されたかを調べました. 店舗だけを利用した顧客には, ネット商品が推薦できないため, 併用顧客の購買商品から推薦アイテムを決定しました.

### 4. 推薦結果

表1は推薦結果です. 提案手法の比較対象として類似度グラフを用いた方法, ユーザーベースの協調フィルタリング (User-CF), 一様ランダムにアイテムを選択した場合と比較しています. 評価尺度は正解率を表す precision で, それぞれ20アイテムを各顧客に推薦した結果を示しています. precision は提案手法が最も

表1 推薦結果

対象顧客	推薦方法	precision
併用—併用	提案手法	<b>0.0271</b>
併用—併用	類似度グラフ	0.0195
併用—併用	User-CF	0.0074
併用—併用	ランダム推薦	0.0008
店舗—併用	提案手法	<b>0.025</b>

高くほかの方法に比べて精度の高い推薦が行えていることが確認できます. また, 類似度グラフよりも高精度な推薦が行えていることから, 推薦におけるグラフ研磨の効果が示されています.

今回推薦対象となる商品種類数は53,000で, ランダムに推薦した場合には, precision が0.0008と低く, その値に比べると約33倍の正解率になっています. 表の一番下の結果は, 店舗のみ利用顧客を対象にした提案手法の結果で, ネット商品の推薦に関しても同程度の結果が得られました.

### 5. おわりに

本研究ではグラフ研磨を用いた考慮集合の予測と, 研磨グラフを用いた推薦方法を提案しました. 計算実験ではECサイトの購買履歴データを利用して提案手法の有効性を確認しました. 類似度グラフに対してグラフ研磨を適用した際の推薦精度は向上していることから, グラフ研磨を利用することで考慮集合として扱われた商品により推薦精度の向上につながったことが考えられます.

謝辞 本研究は, 市川 雄太氏, 田澤 有真氏 (専修大学) との共同研究です. 本研究の一部はJSPS 科研費 JP15K17146 の助成を受けたものです.

### 参考文献

- [1] 市川雄太, 田澤有真, 中原孝信, “データ研磨を用いた考慮集合の予測と推薦,” 日本オペレーションズ・リサーチ学会2016年春季研究発表会アブストラクト集, 1-F-11, 2016.
- [2] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, “GroupLens: An open architecture for collaborative filtering of netnews,” In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, pp. 175–186, 1994.
- [3] 宇野毅明, 中原孝信, 前川浩基, 羽室行信, “データ研磨によるクリーク列挙クラスタリング,” 情報処理学会アルゴリズム研究会報告書, 2014-AL-146(2), pp. 1–8, 2014.