

情報量に注目した推薦のための商品間関係性の分類法

Discrimination method of the relation between merchandise pair based on the Information theory

園田 隆志*¹
Takashi Sonoda

出雲 英剛*¹
Hidetaka Izumo

佐藤 政寛*¹
Masahiro Sato

*¹ 富士ゼロックス株式会社 研究技術開発本部
Research & Technology Group, Fuji Xerox Co., Ltd.

Recommender system provides relevant items to users. Collaborative filtering is the most widely used technique in recommender systems. This technique is useful in the area of mass merchandiser market. However, user's purchased item and non-purchased item has contained substitution across other items. In this presentation, we propose discrimination method of the relation between merchandise pair based on the information theory.

1. はじめに

近年、インターネットを通じた通信販売やビデオレンタルなどの業種において推薦技術が広く利用されている。利用者の購買記録から類似の商品を購入した利用者を抽出し、推薦する商品を提示する協調フィルタリングに代表される、さまざまな手法が提案され[Ricci 2015]、多くの業種で成功を収めている。

特に音楽やビデオ配信では、興味に応じ購入することがあるために推薦により、ついで買いが起り売上の増加が期待できる。しかしながら、食料品小売業など日常的に利用される業種では、利用者の購入額はほぼ一定であるために、推薦による対象商品の購入率は増加しても、他の代替性のある類似商品の購入率は減少し、売上の増加には結びつかないという課題がある。[松村 16]では商品の欠品時に発生する代替できる商品購入行動をモデル化しており、[高橋 92]では牛乳におけるブランド間の代替関係を POS データから調査している。

このように代替性を持つ商品間においては、嗜好とは無関係に購入率を変動させてしまうため、協調フィルタリングの購入率予測精度が低下する要因となっている。

また、一方で精肉や野菜などの料理素材と、それを調理するための調味料のように、ある商品の利用に不可欠な相補性を持つ商品も存在する。このような商品では一方の推薦が他方の購入を促すことが期待できる。

そこで、これらの商品間の関係を見出すことができれば、売上の増加につながることを期待できると考え、本発表では、これらの商品間の関係を求める方法を考察する。

2. 食品小売業の購買傾向

ここで使ったデータは、食品小売業に協力いただき収集を行った。対象期間は 2014 年の 7 月から翌 2015 年の 6 月までの 1 年間である。この期間の利用者は 20,909 人、販売商品数は 5,690 品目、全販売数は 11,127,263 個である。図 1 に年間購入数が平均的(購入数約 700 個)な利用者 30 人の 1 回あたりの購入数のヒストグラムを示す。1 回の購入数は 13 個をピークとしてその両側では購入件数は減少している。1 回の買物における購入数に変化はあるものの、多くの場合は一定であることが推測される。

3. 情報量による商品間関係性の分類

ある 2 つの商品に何等かの関係がある場合、一方の商品の選択は他方の商品の選択に影響を及ぼす。情報論的に考えると一方の商品の選択は、他方の商品の選択にける情報量を変化させると考えることができる。そこで、商品の基本的な情報量を求めることで、商品間の関係性を知ることができると考える。

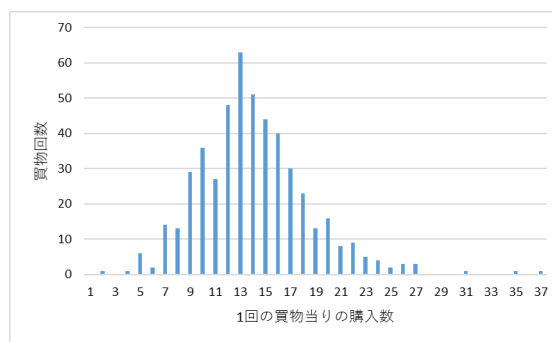


図 1. 1 回の買物当りの購入数

3.1 条件付きエントロピー

2 つの商品の選択が独立でなければ、一方の商品の選択が他方の商品の選択に関する不確度度は減少すると推測される。このように、一方の事象 A が何であるかが分かっているという条件のもとでの、他方の事象 B の不確度度は、次の条件付エントロピーで表される[Cover 2006]。

$$H(B|A) = - \sum_{i,j=0}^1 p(A_i) p(B_j | A_i) \log p(B_j | A_i) \quad (1)$$

ここで、添え字の 0 と 1 はそれぞれ商品を選択した/しないを表し、 $p(B_j | A_i)$ は条件付確率を表す。同様に、一方の事象 B が何であるかが分かっているという条件のもとでの、他方の事 A の不確度度は、 $H(A|B)$ で表すことができる。

3.2 相互情報量

さらに、一方の商品 A が選択されたかが分かると、商品 B についても情報を得ることができる。これは相互情報量と呼ばれる次の式で表すことができる。

$$I(A, B) = \sum_{i,j=0}^1 p(A_i, B_j) \log \frac{p(B_j | A_i)}{p(A_i)p(B_j)} \quad (2)$$

この情報量が高い商品間は、商品 A を選択されたとき、商品 B の選択の予測が高いことが推測される。

3.3 購買確率の計算

式(1)と式(2)により 2 つの商品の関係性は求めることができると考えられるが、これらの式を計算するためには、各確率値を求めておく必要がある。たとえば、確率 $P(A_i) (i=0,1)$ は、一定期間のお客様への販売数を来店回数で割ることで求められる。このとき、訪問回数のうち、商品 A が販売されていない期間の訪問回数は除いておく必要がある。さらに、本研究では 2 つの商品間の関係を考えているために、比較される商品 B も同時に販売されている期間を母数として考えなくてはならない。同様に、同時確率、条件付確率も、それぞれ商品間に対して計算を行う。

4. 情報量による商品間関係性の分類

これまでに定義した情報量により商品間関係性を分類する。

4.1 同時性・代替性

相互情報量が高い商品間関係は、一方の商品 A を選択した場合、B の選択か否かが高い確率で定まることが予想される。高い確率で同時に選択される場合は同時性、一方の選択により他方の選択がなくなる場合を代替性と呼ぶことにする。

ただし、この指標では、同時性と代替性を分離することはできない。

4.2 同時性と代替性の分離

そこで、これらを分離するために 2 つの商品 i と j の間に、次の時間相関係数 $C(i, j)$ を導入する。

$$C(i, j) = \sum_{u \in U(i, j)} \frac{\sum_d I_{ui}(d) I_{uj}(d)}{\sqrt{\sum_{d'} I_{ui}^2(d') \sum_{d''} I_{uj}^2(d'')}} \quad (3)$$

ここで、 $I_{ui}(t)$ は、利用者 u が日 d に商品 i を購入した個数を、また、 $U(i, j)$ は、分析期間中に商品 i と商品 j を購入したことがある利用者の集合である。この相関係数の大きさによって代替性と同時性の分離が可能と考えられる。

4.3 相補性

条件付エントロピーは、一方の選択による他方の選択の情報量の変化であることから、非対称の関係性が成立する 2 つの条件付エントロピーの差が大きい場合、商品 A を選択するとき必ず商品 B が選択されるが、逆に、商品 B の選択は、商品 B の選択に影響しない場合想定される。これは、牛肉に対する焼肉のたれのように一方の選択は、他方の選択に影響するが、逆の場合には、影響がないことに相当する。

4.4 独立性

相互情報量の低い商品間は、一方の商品 B の選択は、他方の商品 A の選択とは無関係に行われる。この論文では、この関係性を独立性と呼ぶことにする。

5. 商品間関係性の検証

検証はそれぞれ上記の情報量で表される関係性を持つ商品対に対し、一方の商品データを削除した場合と削除無しの場合で、他方の購買の予測の変化を、アイテムベースの協調フィル

タリングによって求める。すなわち、商品対 A と B に上記の関係があるならば、A のデータの欠如により、B の購入予測の精度の低下が起これと考えられる。評価は、実際の購入数と予測数の間の平均二乗誤差で行った。各商品対の上位 200 品目に対し、ユーザ-アイテム行列の要素 1000 個を削除した場合と、削除しない場合について比較している。この結果を図 2 に示す。

ここで、青色の棒グラフは商品 A データを削除した商品 B の予測を、橙色の棒グラフは削除無しの予測である。また、“同時性”は相互エントロピーと時間相関ともに大きい商品間を、“代替性”は相互エントロピーが大きく、時間相関が小さい商品対を、“相補性 A|B”は条件付エントロピー $H(A|B)$ が大きく、 $H(B|A)$ が小さい商品対を表し、“相補性 B|A”はその逆の関係を表している。さらに、“独立性”は相互情報量が小さい商品対を表している。

この図を見ると分かるように、同時性・代替性については商品 A のデータが商品 B の予測に影響していることが分かる。相補性については、予測方向によって予測に変化が無い場合とある場合がある。これは、相補性の場合、一方の選択が、他方の選択に影響しない状況を表していると考えられる。さらに、独立性の商品間は、お互いに影響を与えないこと示している。

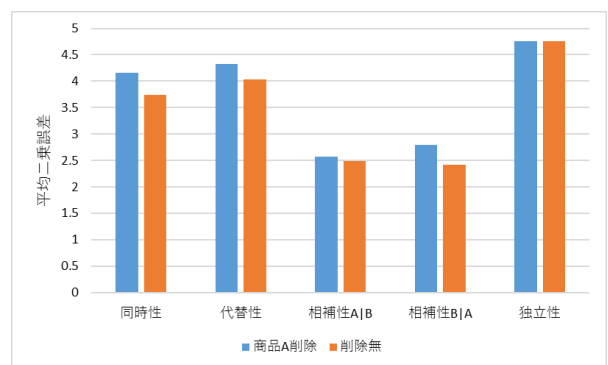


図 2. 商品間関係性に対する購入推定

6. さいごに

この発表では、スーパーマーケットなどに代表される食料品小売業の推薦のために、情報理論で定義される指標に注目し、商品間関係性の分類方法を提案した。食料品のように日常的に購入されるものは、予算の上限が定まっていることが多いが、この分類を推薦に利用することで、売上の向上が期待できると考えている。

参考文献

- [Cover 2006] T. M. Cover, J. A. Tomas, Elements of Information Theory 2nd edition, Wiley Interscience publication (2006)
- [Ricci 2015] F. Ricci, L. Rokach, B. Shapira(eds.), Recommender Systems Handbook 2nd edition, Springer Verlag, (2015).
- [高橋 92] 高橋 克也: 食品アイテム間の競合分析 -POS データを用いた実証分析-, 農総研季報第 15 号, 農林水産政策研究所, 1992
- [松村 16] 松村 直樹, 和泉 潔, 山田 健太: POS データに基づく欠品時の顧客行動を考慮した小売店舗の購買シミュレーション, 人工知能学会論文誌 vol. 31, 2016