

ソーシャルメディア上の嗜好情報を利用する商品推薦方式の研究

1130396 山岡翔太 【清水研究室】

1 はじめに

EC 市場の規模は、2013 年に 10.2 兆円になると推測されている。また、多くのユーザが EC サイトを用いて商品を購入する傾向にある。

しかし、商品の多様化によって商品数が増加し、探していた商品が見つからないといったユーザが増加している。これに伴い、多くの情報から有用な情報を推薦するサービス「レコメンデーション」が注目されている。

レコメンデーションには、主にコンテンツベースフィルタリングと協調フィルタリングの二つの技術が使われている。コンテンツベースフィルタリングは、商品内容を利用した推薦手法である。この手法は、購入した商品から推測可能な推薦しかできず、ユーザの利点が少ない。それに比べ、本人も知らない意外性のある商品推薦が可能な協調フィルタリングがある。協調フィルタリングは、ユーザ情報、購買履歴といった行動履歴を用いて、類似ユーザを決定する。決定した類似ユーザの購買履歴から商品を推薦する手法である。しかし、新規ユーザに対して、商品推薦することができないコールドスタート問題が発生する。故に、行動履歴がないとユーザに商品を推薦することができない。

2 類似ユーザを決定する提案方式

本研究では、コールドスタート問題に対して、ソーシャルメディア上の嗜好情報を利用し、類似ユーザを決定する商品推薦手法を提案し、評価する。図 1 に提案手法の流れを示す。以下では、抽出単語における共起ネットワーク解析と集合の類似度を表す Jaccard 係数について述べる。

2.1 共起ネットワーク解析

共起ネットワークは、解析対象のデータの特徴を繋ぎ、単語同士の共起関係を明らかにする解析手法である [1]。本研究で抽出する単語は、ソーシャルメディアの代表的なサービス Twitter の投稿情報から名詞を取得する。取得した抽出単語から、出現頻度や出現パターンの似通った単語を関連づけ、複数の単語が集合したネットワークを構成する。

2.2 Jaccard 係数

Jaccard 係数を用いることで、二つの共起ネットワークの類似度の数値化を行う。Jaccard 係数は、集合の類似度を数値化する係数であり、web 上の文書から共起ネットワークを構成する際に用いられる [2]。Jaccard 係数の値が 1 に近いほど、類似度が大きい値を示す。Jaccard

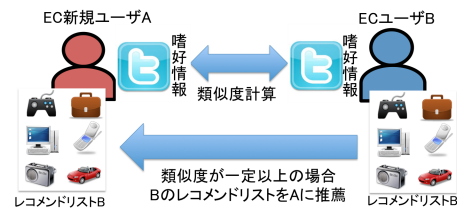


図 1 本研究の流れ

係数を式 (1) で示す。

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

本研究では、 A と B は Twitter から共起ネットワーク解析によって構成された集合が入る。

3 評価

本研究では、Twitter ユーザ 25 名を調査対象とした。まず、Jaccard 係数で数値化された集合の組み合わせを類似度の大きい順に並べた。その上位 30 組を評価の対象とした。評価方法として、協調フィルタリングが使われている既存 EC サイトの推薦商品を利用した。推薦されている商品が評価者に適合しているか評価する。コールドスタート問題は、先行研究において約 20 % 以上の適合率が必要という結果がある [3]。

評価結果より、上位 15 組目 (Jaccard 係数の数値 0.134) で平均適合率 20 % を下回ってしまう。この結果より、Jaccard 係数の値が 0.134 を上回っていれば、コールドスタート問題に対する有効性を示すことができた。

4 おわりに

本稿では、ソーシャルメディア上の嗜好情報を利用する商品推薦方式について述べた。Jaccard 係数の数値が 0.134 以上であれば、適合率が 20 % 以上となりコールドスタート問題に有効であることを示した。

今後の展望として、類似ユーザを決定する際の判断基準が増えることで推薦の精度が上がると考えられる。よって嗜好情報とは異なる推薦技術を併用することでより、有用な推薦が可能となると考える。

参考文献

- [1] 松尾豊, 大澤幸生, 石塚満, “Small World 構造に基づく文書からのキーワード抽出” <http://www.miv.t.u-tokyo.ac.jp/papers/matsuoIPSJ02.pdf> .
- [2] H. Kautz, B. Selman, and M. Shah “The Hidden Web”, AI magazine, Vol18, No2, pp.27-35(1997)
- [3] 和泉憲明, “共有コンテンツのアクセス履歴分析に基づく情報推薦エンジン,” <http://www.ai-gakkai.or.jp/conf/2007/data/pdf/100320.pdf> .