

Twitter ユーザの購買行動推定手法

岩田 麻佑[†] 服部 元[†] 池田 和史[†] 滝嶋 康弘[†][†] (株) KDDI 研究所

1 はじめに

近年, Twitter¹ などの SNS 上に, 特定の商品やサービスに関する自身の購買行動の状態 (〇〇欲しい, ×買った, など) を投稿するユーザが増えている. 企業は, これらの投稿から, 特定の商材に対して言及したユーザを発見し, 商材に興味を持つユーザへの広告配信, 購入者候補や購入者の意見の把握, 競合他社商材についての分析を行うことができる. しかし, 当該商材に対するユーザの購買行動プロセスのフェーズ (当該商材を知っているだけなのか, 興味があるのか, 購入したのか) を把握しなければ, 広告配信や意見分析などの十分な効果が得られない.

そこで本研究では, 任意の商材に関する Twitter 上の投稿から, そのユーザの購買行動プロセスのフェーズを推定する手法を検討する. 購買行動プロセスとして, 経済学における購買心理状態の遷移モデルである AIDMA[1] を利用する. AIDMA は, ユーザの購買行動に関する心理状態は, Attention (注目) → Interest (興味) → Desire (欲望) → Memory (記憶) → Action (行動) と変化するとしたモデルで, 認知フェーズ (Attention) → 感情フェーズ (Interest, Desire, Memory) → 行動フェーズ (Action) の 3 つに大きく分類される [1]. 本稿では, 同一ユーザの投稿を時系列に追跡し, 各フェーズに出現する語を検出することで, 各投稿を認知, 感情, 行動のフェーズのいずれかに分類する.

2 関連研究

SNS 上の投稿からユーザの属性を推定する研究は数多く行われている. 文献 [2] では, 各属性に特徴的に出現する単語を利用し, Twitter 上のユーザのプロフィール (性別, 年齢等) を推定する手法を提案している. しかし, 購買行動プロセスのように属性が時間に伴い変化することは想定していない. また, 購買行動プロセスに関連して, 文献 [3] では, 商品レビューから「～したい」というような購買意図を抽出する手法を提案している. この研究で対象とされている「～したい」「～欲しい」というような購買意図は, 購買行動プロセス

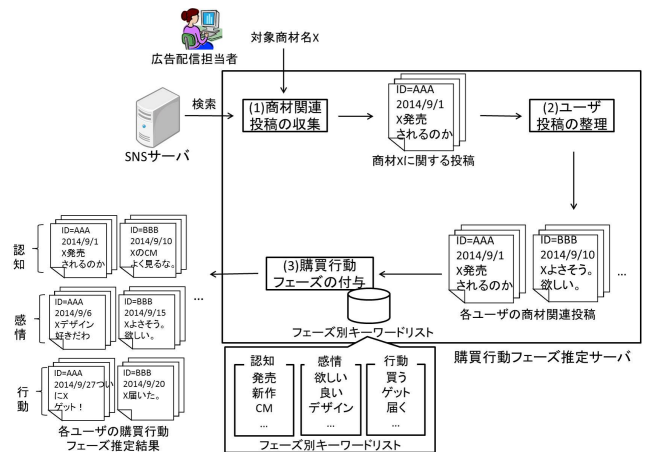


図 1: 購買行動フェーズ推定システムの概要

の感情フェーズに含まれるといえるが, 本研究で対象としている認知や行動フェーズは考慮されていない.

3 購買行動フェーズ推定手法

図 1 に購買行動フェーズ推定システムの概要を示す. 広告配信担当者が商材名を入力すると, システムは (1) 商材関連投稿の収集, (2) ユーザ投稿の整理, (3) 購買行動フェーズの付与の手順で, 各投稿のフェーズを決定する.

3.1 商材関連投稿の収集

Twitter 社が提供する streaming API を利用して, 当該商材の名称を検索キーとして, 当該商材に関する投稿を検索し, 収集する.

3.2 ユーザ投稿の整理

収集した商材関連投稿を, ユーザ ID に基づいて分類し, 投稿時刻に基づいて並べ, ユーザごとに時系列になるように投稿を整理する.

3.3 購買行動フェーズの付与

投稿時刻と投稿内容に基づいて, 各ユーザの投稿に対して購買行動フェーズを付与する. フェーズを付与するために, あらかじめフェーズ別キーワードリストを作成しておき, キーワードリストを利用して分類器により各投稿のフェーズを自動推定する. 購買行動フェーズは, 認知 → 感情 → 行動と時系列で遷移するため, 時系列でフェーズが遷移するようユーザごとに補正処

¹<http://twitter.com/>

理を行い、各投稿の最終的なフェーズを決定する。

以下では、フェーズ付与の詳細な方法について述べる。

1. フェーズ別キーワードリストの作成

人手により認知、感情、行動のいずれかのラベルが付与された学習用文書を形態素解析によって単語分割し、各フェーズに該当する文書に出現する単語 unigram と bigram をキーワードリストに登録する。bigram も利用するのは、「買いたい」「買った」などの「動詞+助動詞」の連続した単語が重要になると考えたからである。

2. 分類器による購買行動フェーズ推定

フェーズ推定には CNB(Complement Naive Bayes)[4] を用いる。CNB は、そのラベルに属さない文書を使って学習し、単純ベイズを用いた分類を行う。CNB を用いて学習用文書におけるキーワードの出現傾向を学習することで、未知の投稿内容に対してフェーズを推定することが可能となる。具体的には、3.3 節で説明した方法でフェーズごとに抽出したキーワードと、学習用文書それぞれにおける各キーワードの出現回数からなる行列を CNB の入力として与える。学習段階では、各フェーズのラベルを合わせて与えることで CNB を学習させる。CNB 分類器はフェーズごとに構築し、各分類器を用いて、未知の投稿内容が各フェーズに属する確率（尤度）を求め、尤度が最大のフェーズを採用する。

3. 時系列補正処理

フェーズが時系列で遷移し、逆戻りしないことを考慮し、付与されたフェーズを補正する。時系列で並んだ各ユーザの投稿 $S_1, S_2, S_3, \dots, S_n$ (n はユーザの投稿数) の中で、投稿時刻が最も早い感情フェーズの投稿 S_{idm} 、投稿時刻が最も早い行動フェーズの投稿 S_{ac} を各フェーズへの変化点とする。そして、 $S_1 \sim S_{idm-1}$ を認知フェーズ、 $S_{idm} \sim S_{ac-1}$ を感情フェーズ、 $S_{ac} \sim S_n$ を行動フェーズと補正する。

4 評価実験

4.1 実験環境と実験手順

単語を抽出する形態素解析器として MeCab² を用いた。実験データとして、2012年2月～2014年5月に発表された通信関係の新商品やサービス5商材に関する投稿15,006件を利用した。これらの投稿に人手でフェーズのラベルを付与し、キーワードごとに8-Cross validation で実験を行った。

²<http://mecab.sourceforge.net/>

表 1: 実験結果 (5 キーワードの平均)

		unigram	uni+bigram
Accuracy		0.71	0.73
認知	Precision	0.83	0.84
	Recall	0.77	0.79
感情	Precision	0.55	0.60
	Recall	0.25	0.26
行動	Precision	0.62	0.61
	Recall	0.71	0.71

unigram を利用する手法 (unigram) と unigram と bigram を利用する手法 (uni+bigram) において、評価指標として、Accuracy (正解数/商材名を含む投稿数)、各フェーズの Precision (フェーズ X の正解数/商材名を含む投稿の中でフェーズ X と推定した件数)、Recall (フェーズ X の正解数/商材名を含む投稿の中でフェーズ X が正解である件数) を算出した。

4.2 実験結果

表 1 に結果を示す。uni+bigram は全項目において unigram と同程度以上の値となった。特に、感情フェーズの Precision で 5% の向上が見られた。これは、bigram により「買う+た」「買う+たい」「申し込む+たい」などの特徴的な語を検出できたためである。

ただし、認知フェーズと比較すると、感情、行動フェーズの Precision や Recall は 0.7 に満たず低かった。これは、学習文書のうち 53% が認知フェーズに該当し、感情、行動フェーズが正解である投稿の多くが認知フェーズに分類されたためと考えられる。

5 まとめ

本稿では、Twitter 上の商材に関する投稿から、ユーザの購買行動フェーズを推定する手法の検討を行った。同一ユーザの投稿を時系列に追跡し、各フェーズに出現する語を検出することで、各投稿を、認知、感情、行動のいずれかのフェーズに分類することで、推定精度が 0.73 であることを確認した。今後は、感情や行動フェーズを精度良く推定するため、キーワードリストに登録する単語の抽出方法について検討する。

参考文献

- [1] Hall, S.R., Writing an Advertisement. An Analysis of the Methods and the Mental Processes That Play a Part in the Writing of Successful Advertising, General Books LLC, 2009.
- [2] 池田和史, 服部元, 松本一則, 小野智弘, 東野輝夫, マーケット分析のための Twitter 投稿者プロフィール推定手法, 情報処理学会論文誌, Vol.2, No.1, pp. 82-93, 2012.
- [3] 石野亜耶, 村上浩司, 関根聡, 商品レビューからの購買意図の抽出とそれを用いた商品検索システムの構築, 言語処理学会第 20 回年次大会, pp. 622-625, 2014.
- [4] Rennie, J.D.M., Shih, L., Teevan, J., and Karger, D.R., Tackling the Poor Assumptions of Naive Bayes Text Classification, ICML 2003, pp.616-623, 2003.