

# 移動エントロピーによる動的ネットワーク化を用いた SNSと商品購買の相互関係の分析

## An experimental analysis of mutual relations between twitter and rakuten-ranking by transfer entropy

天神雄貴 尾崎知伸\*  
Yuuki Tenjin Tomonobu Ozaki

日本大学 文理学部  
College of Humanities and Sciences, Nihon University

**Abstract:** Recently, social media such as Twitter and Facebook play a key role to obtain a wide variety of information. In this study, we examine mutual influences between social media and buying behaviors by using dynamic networks constructed from Twitter and Rakuten Ranking based on transfer entropy.

### 1 はじめに

近年, Twitter<sup>1</sup> や Facebook<sup>2</sup> に代表されるソーシャルメディアは, 多くの人々に日常的に利用されており, 必要不可欠な情報インフラの一つとなっている. ソーシャルメディアは, Twitter におけるリツイートなどのように, 情報の拡散機能が豊富であり, テレビなどの放送メディアと比較して速報性や拡散性が高いという特徴を持つ. このことから, 最新の情報をソーシャルメディアから獲得する利用者も少なくない. これらの情報獲得行動を背景に, 企業などでは, 広告活動を行う際に, テレビコマーシャルや新聞だけでなく, ソーシャルメディアを利用するが増加している.

これらのことを踏まえ, 本研究では, 企業における効果的な販売戦略策定の一助となるような傾向を発見することを目的に, ソーシャルメディアでの投稿行動と商品購買行動の相互関係を分析する. 具体的には, Twitter での投稿であるツイートに含まれる単語と楽天市場<sup>3</sup> に現れる商品を分析の対象とする.

本論文の構成は以下のとおりである. 2章で分析手法の概要について述べる. 3章で実際の分析結果を示し, 考察を行う. 4章で関連研究について言及し, 最後に5章でまとめを行う.

### 2 動的ネットワーク化による分析

#### 2.1 分析手法の概要

本研究では, ソーシャルメディア上の投稿行動としての Twitter データと, 商品購買行動としての楽天ランキングデータを分析の対象とする. これらのデータを対象に, 以下に示す手順に従い, ソーシャルメディア上での投稿行動と購買行動の相互関係の分析を試みる.

まず各データから, 商品の売り上げ順位と単語の出現回数を, 一日を単位とした時系列データとして取り出す. 次に, 移動エントロピー [1] を用いて時系列データ間の情報伝播を計量し, その大きさに従い, 商品や単語を頂点とするネットワークを構築する. その後, ネットワーク上での中心性の高い要素を抽出することで, 影響力の強い商品や単語を特定する. さらに, 対象期間をずらしながらネットワーク構築を行うことで, 時間情報を伴う複数のネットワーク, すなわち動的ネットワークを構築し, 時間的な変化も考慮した分析を実現する. 分析の全体像を図 1 に示す.

#### 2.2 ネットワークの構築

本研究では, 楽天ランキングに登場する商品の集合  $I = \{i_1, i_2, \dots, i_n\}$  と, Twitter のツイート内に現れる単語の集合  $J = \{j_1, j_2, \dots, j_m\}$  を対象に, 相互影響力を考慮した二部グラフ

\*連絡先: 日本大学 文理学部 情報科学科  
〒156-8550 東京都世田谷区桜上水 3-25-40  
tozaki@chs.nihon-u.ac.jp

<sup>1</sup><https://twitter.com/>

<sup>2</sup><https://ja-jp.facebook.com/>

<sup>3</sup><http://ranking.rakuten.co.jp/>

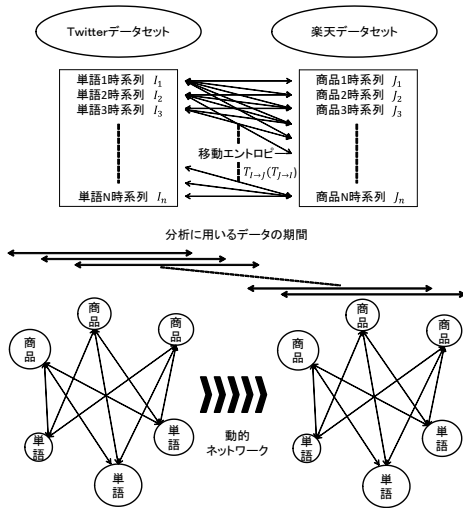


図 1: 分析の全体像

$$G = \langle I' \cup J', E \rangle \text{ where}$$

$$E = \left\{ (x, y) \mid \begin{array}{l} x \in I \wedge y \in J \wedge T_{x \rightarrow y} \geq \theta \vee \\ x \in J \wedge y \in I \wedge T_{x \rightarrow y} \geq \theta \end{array} \right\},$$

$$I' = \{x \in I \mid (x, y) \in E \vee (y, x) \in E\} \text{ and}$$

$$J' = \{y \in J \mid (x, y) \in E \vee (y, x) \in E\}$$

を構築する。ここで  $T_{x \rightarrow y}$  は、 $x$  から  $y$  への移動エントロピーであり、 $\theta$  は閾値である。

移動エントロピー [1] とは、2つの離散的な時系列間の情報伝播の大きさを定量化する尺度である。2つの離散時系列データ  $x = \{x_1, x_2, \dots, x_i, \dots\}$  と  $y = \{y_1, y_2, \dots, y_j, \dots\}$  に対し、 $y$  から  $x$  への移動エントロピー  $T_{y \rightarrow x}$  は、以下のように定義される。

$$T_{y \rightarrow x} = \sum p(x_n, x_{n-1}^{(n-k)}, y_{n-1}^{(n-l)}) \log \frac{p(x_n \mid x_{n-1}^{(n-k)}, y_{n-1}^{(n-l)})}{p(x_n \mid x_{n-1}^{(n-k)})}$$

ここで  $x_n^{(n-k)} = (x_n, x_{n-1}, \dots, x_{n-k+1})$  は、時刻  $(n-k+1) \sim n$  における  $x$  の状態を含むベクトルである。

本研究では、楽天ランキング中の商品  $x \in I$  の一日毎の売り上げ順位の系列  $\{x_1, x_2, \dots\}$  と、同期間でのツイート中の単語  $y \in J$  の一日毎の出現回数の系列  $\{y_1, y_2, \dots\}$  に対して相互の移動エントロピー  $T_{y \rightarrow x}$  と  $T_{x \rightarrow y}$  を算出し、ネットワークを構築する。

移動エントロピーは、本来、離散系列を対象としているので、データセットとして獲得した商品  $x$  の売り上げ順位の系列と、単語  $y$  の出現回数系列を、何らかの方法で離散化する必要がある。今回の分析では、前

日との変化の大きさに着目し、これらの時系列データを5つの離散値  $\{-2, -1, 0, +1, +2\}$  からなる離散系列  $\{f(x_1), f(x_2), \dots\}$  と  $\{g(y_1), g(y_2), \dots\}$  へと変換する。ここで、“-2”は負方向への大きな変化，“-1”は負方向への小さな変化，“0”は変化なし，“+1”は正方向への小さな変化，“+2”は正方向への大きな変化に対応する。なお、売り上げ順位の系列においては、値そのものの減少は順位の向上を意味し、正方向への変化である点に注意が必要である。

離散化された系列を用いて、商品  $x$  と単語  $y$  の全組み合わせに対して移動エントロピーを算出し、その値が閾値  $\theta$  以上の組み合わせに対して辺を付与し、ネットワークを構築する。またこのとき、辺数の増大を避けるため、新たなパラメタ  $K$  を導入し、付与する辺を全体の上位  $K\%$  以内の値を持つものに限定する。

上述の方法を用いて、重複を許した一定期間ごとにネットワークを構築することで、時間情報を伴った複数のネットワークを構築する。また得られたネットワークを期間順に並べることで、動的なネットワークを獲得し、時間的変化を伴う分析を行う。

### 3 楽天ランキングの分析

#### 3.1 データセットの概要と実験設定

本研究では、2014年9月1日から11月30日の3か月間に収集した、楽天市場ランキングとツイートデータを分析の対象とする。楽天市場ランキングに含まれる商品とその売り上げ順位は、楽天ウェブサービスの楽天市場ランキング API<sup>4</sup> を用いて取得する。このAPIでは、ファッションや生活用品など、ジャンルごとの売り上げ順位1位から1000位までの商品情報を獲得することが可能である。今回の分析では、メンズファッションのジャンルを対象とした。

一方、同期間でのツイートデータの取得には、Twitter Streaming API<sup>5</sup> を用いる。取得したツイートに対して形態素解析を適用することで、名詞と形容詞を抽出し、出現回数の系列を作成する。また、単語より大きな粒度での相互関係の分析を目的に、感情語辞典 [2] を用いて作成した10の感情語グループ {喜 (268語), 怒 (217語), 哀 (246語), 恐 (163語), 恥 (65語), 好 (217語), 厭 (516語), 昂 (159語), 安 (109語), 驚 (135語)} に対しても、一日を単位とした出現回数系列を作成した。

実験では、収集した3か月間のデータから、10日間ずつ期間をずらしながら、30日間を一つの単位とするネットワークを7つ構築した。なお、辺の付与に関するパラメタ値は  $\theta = 1.2$ ,  $K = 5$  とし、移動エントロ

<sup>4</sup><https://webservice.rakuten.co.jp/api/ichibaitemranking/>

<sup>5</sup><https://dev.twitter.com/streaming/overview>

表 1: 楽天商品-ツイート単語間ネットワークのサイズ

期間	$V_x$	$V_y$	$E_{x \rightarrow y}$	$E_{y \rightarrow x}$
9/1-9/30	236	2793	79	19,506
9/11-10/10	268	3116	479	21,470
9/21-10/20	292	3366	958	22,289
10/1-10/31	297	3073	340	24,877
10/11-11/10	295	2974	145	24,799
10/21-11/20	282	3112	491	20,534
11/1-11/30	295	3170	355	21,102

$V_x$ : 楽天商品ノード数  
 $V_y$ : ツイート単語ノード数  
 $E_{x \rightarrow y}$ : 商品から単語へのエッジ数  
 $E_{y \rightarrow x}$ : 単語から商品へのエッジ数

表 2: 楽天商品-感情語グループ間ネットワークのサイズ

期間	$V_x$	$V_y$	$E_{x \rightarrow y}$	$E_{y \rightarrow x}$
9/1-9/30	115	8	180	160
9/11-10/10	129	9	195	185
9/21-10/20	122	9	210	205
10/1-10/31	143	9	215	205
10/11-11/10	146	9	215	195
10/21-11/20	154	5	210	190
11/1-11/30	161	5	195	180

$V_x$ : 楽天商品ノード数  
 $V_y$ : 感情語グループ数  
 $E_{x \rightarrow y}$ : 商品から感情語へのエッジ数  
 $E_{y \rightarrow x}$ : 感情語から商品へのエッジ数

ピーの値が 1.2 以上かつ上位 5% となる組み合わせに対して辺を付与した。

### 3.2 結果と考察

各期間で生成されたネットワークの大きさを表 1 と表 2 に示す。また、各ネットワークの頂点集合間の Jaccard 類似度を表 3 に示す。

表 1 と表 3 より、頂点数には大きな差はないが、ネットワークに現れる頂点集合自体は変化していることが読み取れる。また、商品から単語へのリンク数  $E_{x \rightarrow y}$  に着目すると、値のばらつきが大きく、期間によって影響を与える商品群や、その商品の購買行動が与える影響の範囲が異なること示された。

表 4 と表 5 に、各ネットワークにおける PageRank[3] 上位 3 位以内のノードと、最大移動エントロピーを持つ辺の一覧を示す。詳細な値は示さないが、全期間において、一般的に  $T_{x \rightarrow y} > T_{y \rightarrow x}$ 、すなわち、商品  $x$  から単語  $y$  への移動エントロピー値  $T_{x \rightarrow y}$  の方が、単語

表 3: 各期間の頂点集合間の Jaccard 類似度

	9/11-10/10	9/21-10/20	10/1-10/31	10/11-11/10	10/21-11/20	11/1-11/30
楽天商品ノード						
9/1-9/30	0.71	0.55	0.46	0.39	0.35	0.32
9/11-10/10		0.76	0.61	0.49	0.42	0.36
9/21-10/20			0.76	0.57	0.50	0.40
10/1-10/31				0.71	0.56	0.44
10/11-11/10					0.75	0.56
10/21-11/20						0.70
ツイート単語ノード						
9/1-9/30	0.52	0.48	0.40	0.39	0.39	0.37
9/11-10/10		0.55	0.46	0.43	0.43	0.41
9/21-10/20			0.54	0.48	0.46	0.44
10/1-10/31				0.56	0.49	0.43
10/11-11/10					0.55	0.44
10/21-11/20						0.51

$y$  から商品  $x$  への移動エントロピー値  $T_{y \rightarrow x}$  より大きい傾向が確認された。このことから、ツイート行動が商品購買行動に影響を与えるというよりは、商品購買行動がツイート行動に影響を与えている可能性が示唆された。また、感情語グループ  $g$  を対象とすると、各方向での移動エントロピー値  $T_{x \rightarrow g}$  と  $T_{g \rightarrow x}$  に大きな差はなく、また全体的に小さな値となっている。このことは、感情語という大きな単位での分析が、必ずしも有効に機能するわけではないことを示していると考えられる。

図 2 に、期間 9 月 1 日～9 月 30 日のデータから生成されたネットワークの一部を示す。この図では、各商品の購買行動がツイートへと与える影響を直感的に示すとともに、相互に関連の強い単語群を視覚的に捉えることを助けるため、商品ノードを中心付近に、また単語ノードを周囲に配置するとともに、商品ノードの大きさを出次数（商品から単語へのリンク数）に比例させている。

## 4 関連研究

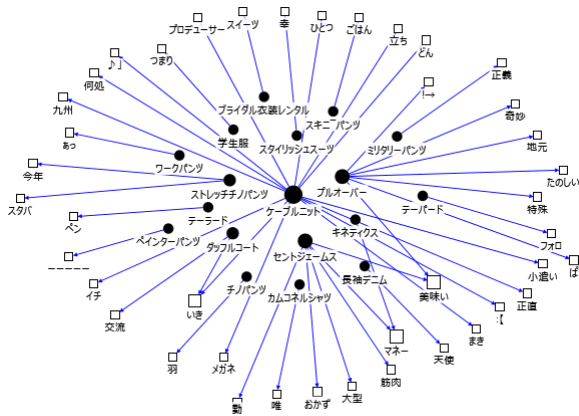
鳥海ら [4] は、移動エントロピーの定義を拡張した拡張移動エントロピーを用い、東京証券取引所のデータの各銘柄間のリターン情報から構築した個別銘柄間の定常的な影響度ネットワークを構築した。構築したネットワークをもとに分析を行い、結果として、拡張移動エントロピーを用いることで、刻み幅の小さい数秒単位での細かい情報伝播を捉えることに成功した [4]。さ

表 4: PageRank 上位 3 位以内のノード

期間	商品-単語間ネットワーク			商品-感情語グループ間ネットワーク		
	1 位	2 位	3 位	1 位	2 位	3 位
9/1 -9/30	かわいかっ	エア	おお	驚	怒	ラルフ ローレン
9/11-10/10	roshell ケーブルネット	improves チノパンツ	大人カジユ アルシャツ	恥	驚	スタイリッシュ スーツ
9/21-10/20	カーディガン	improves チノパンツ	デイリー	恥	Dickies ワークパンツ	カーディガン
10/1-10/31	improves チノパンツ	岡山	理	恥	驚	怒
10/11-11/10	いとこ	たこ焼き	神谷	恥	ジップパーカー	テーパード
10/21-11/20	improves チノパンツ	おいしかっ	スキニ丕 パンツ	恥	驚	キネティクス コロンビア
11/1-11/30	roshell ケーブルネット	期待	ダッフル コート	驚	恥	ダウンジャケット

表 5: 各期間における移動エントロピー最大の商品-単語・感情語対

期間	単語 → 商品	商品 → 単語	感情語 → 商品	商品 → 感情語
9/1 -9/30	兄ちゃん ↓ roshell ケーブルネット	結婚式衣装 モーニングレンタル ↓ 金額	驚 ↓ ラルフローレン	パンスーツ福袋 ↓ 驚
9/11-10/10	判断 ↓ roshell ケーブルネット	クルーネック ↓ 普段	驚 ↓ スタイリッシュ スーツ	無地パーカー ↓ 恥
9/21-10/20	月曜 ↓ improves チノパンツ	Dickies ワークパンツ ↓ ツッコミ	驚 ↓ カーディガン	improves チノパンツ ↓ 驚
10/1-10/31	立ち ↓ improves チノパンツ	SEANA ドルマンニット ↓ 当時	驚 ↓ SEANA ドルマンニット	チャンピオン パーカー ↓ 恥
10/11-11/10	作品 ↓ roshell ケーブルネット	roshell ケーブルネット ↓ ホテル	恥 ↓ テーパード	デザイン Y シャツ ↓ 恥
10/21-11/20	よく ↓ roshell ケーブルネット	roshell ケーブルネット ↓ 次第	驚 ↓ P コート	形態安定 長袖 Y シャツ ↓ 恥
11/1-11/30	金 ↓ roshell ケーブルネット	roshell ケーブルネット ↓ 期待	恐 ↓ roshell ケーブルネット	PROCLUB パーカー ↓ 驚



Created with NodeXL (<http://nodexl.codeplex.com>)

図 2: 期間 9 月 1 日～9 月 30 日に対するネットワーク (一部のみを抜粋)

らに、個別銘柄間の情報伝播を可視化することで、予測し難い銘柄間の関係性を示し、投資指標に還元することで、個人投資家にとって有益な情報を見出す方法を考察した [5]。そこでは、情報伝播の遅延や複数の銘柄間からの影響を考慮することが、投資指標を検討する上で有用である結論付けている。

笹原ら [6] は、Twitter のつぶやきに含まれる感情に関わる顔文字や形容詞の出現頻度時系列を対象に、有効移動エントロピー [7] を用いたネットワーク構築を行った。構築されたネットワークを東日本大震災の前後で比較することにより、震災による集合気分の変遷への影響を考察している。

## 5 まとめ

本研究では、ソーシャルメディアにおける投稿行動と商品購買行動の相互関係を分析する第一歩として、移動エントロピーとネットワーク分析手法を用い、楽天ランキングに現れる商品とツイート内の単語との相互関係の分析を試みた。結果として、いくつかの初歩的な知見が示唆された。

今後の課題としては、離散系列への効果的な変換手法の追求や、ユーザ属性との組み合わせなどの技術的な側面に加え、より深い分析を行うために、他メディアや他ジャンルを含めた広領域を対象とした長期間にわたる継続的な実験の実施があげられる。

## 参考文献

[1] T. Schreiber : Measuring Information Transfer, *Physical Review Letters*, Vol.85, No.2, pp. 461–464

(2000)

[2] 中村 明 : 『感情表現辞典』, 東京堂出版 (1993)

[3] S. Brin and L. Page : The Anatomy of a Large-scale Hypertextual Web Search Engine, *Computer Networks and ISDN Systems*, Vol.30, No.1–7, pp.107–117 (1998)

[4] 小村 和輝, 鳥海 不二夫, 大橋 弘忠 : 移動エントロピーを用いた銘柄間ネットワークダイナミクスの分析, 人工知能学会全国大会, 3L4-OS-26b-5in (2014)

[5] 小村 和輝, 鳥海 不二夫, 大橋 弘忠 : 移動エントロピーを用いた銘柄間影響度ネットワークによる投資指標の分析, 第 13 回 人工知能学会 金融情報学研究会, SIG-FIN-013-08 (2014)

[6] 笹原和俊 : SNS における集合気分のダイナミクス, 人工知能学会全国大会, 2J4-OS-16a-3 (2014)

[7] R. Marschinski and H. Kantz : Analysing the Information Flow between Financial Time Series, *The European Physical Journal B*, Vol.30, No.2, pp.275–281 (2002)