

ソーシャルメディアからの自動知識獲得

石 野 垂 耶*

1. はじめに

近年、Twitter, Facebook, レビューサイトなどのソーシャルメディアが急速な広がりを見せている。Twitter や Facebook には、今日の出来事、食事の内容や交通状況、レビューサイトには、購入した商品や鑑賞した映画の感想など、様々な情報が絶えず投稿されている。現在、これらのソーシャルメディアに投稿された情報を、ビジネスやサービスに役立てるための研究が活発に行われている。例えば、新商品の評判分析、株価の予測、交通トラブルの検出、商品推薦システムへの利用が挙げられる。このようなサービスを構築するためには、ソーシャルメディアに投稿された膨大な量の情報から、有益な情報を自動的に抽出や分類する技術が必要である。

本研究集会報告会では、まず、ソーシャルメディアに投稿されたテキストを処理するために使用される自然言語処理の基礎的な技術を紹介した。次に、ソーシャルメディアを活用した研究やサービスについて紹介し、使用されている技術について説明した。また、現在行っている研究事例として、Twitter からの震災時の行動経路の自動抽出 (Ishino et al., 2012)、商品レビューからの購買意図の自動抽出 (石野他, 2014) について説明した。本報告では、上記の研究事例について紹介する。

本報告の構成は以下の通りである。2節では Twitter からの震災時の行動経路の自動抽出、3節では商品レビューからの購買意図の自動抽出について説明し、4節でまとめる。

2. Twitter からの震災時の行動経路の自動抽出

2.1 研究の背景と目的

2011年3月11日14時46分に発生した東日本大震災では、日本で観測史上最大の Mw (モーメントマグニチュード) 9.0を記録し、岩手県から千葉県までの8県におよぶ広域で震度6弱以上の強い地震動が発生した。東北日本の太平洋側で10メートルを超す大津波が襲来、東京電力福島第一原子力発電所の事故にとまなう大規模な避難が行われるなど、我々が経験したことのない大規模かつ複合的な大災害となった。このような災害時には、被災者の避難経路や救援物資の配送に利用可能な経路の情報は非常に重要な情報である。

震災発生時以降、Twitter 上では、安否確認を求めるツイートや、被災地の交通情報を発信するツイートなど、震災に関連する様々な情報が発信された。自然言語処理研究者を中心としたプロジェクト ANPI NLP (Neubig et al., 2011) では、Twitter などから、個々に述べられている安否情報に注目し、Google 社の Google Person Finder¹⁾ 上のデータと照合しながら、最新の安否情報を整理する活動が行われた。本研究では、ANPI NLP で提供されている

* 広島経済大学経済学部助教

東日本大震災情報に関連する Twitter のデータから、機械学習を使用して、ユーザの行動経路を抽出する手法を提案する。震災時の行動経路のデータベースを作成することで、被災者や輸送物資を配送する援助者へ、利用可能な経路情報を提供できることが期待できる。また、得られた避難経路を地図上にマッピングする事により、集約してユーザに提示し、閲覧できるシステムを開発する。

2.2 震災時の行動経路提示システムの動作例

本研究で構築した行動経路提示システムについて、その動作例を紹介する。図1は避難経路提示システムの画面である。約50件の行動経路をマッピングしてある。

行動経路として、ユーザが移動元（図中①）から移動先（図中②）に移動したことがわかるようになっている。また避難経路には、移動手

段に応じたアイコンが表示されている。図中③の場合は、移動手段としてバスが用いられているため、バスのアイコンが表示されている。また、アイコンをクリックすると、詳細な情報（図中④）が表示できるようになっている。

本研究では、ANPI NLP で提供されている Twitter データから、ユーザの行動経路として、移動元、移動先、移動手段を抽出する手法を提案する。

2.3 行動経路の抽出手法

行動経路として、移動元、移動先、移動手段を機械学習により抽出するために、以下の3種類のタグを定義する。図2がタグを付与した例である。

- FROM: 移動元
- TO: 移動先
- METHOD: 移動手段



図1 避難経路提示システムの動作例

義弟の安否確認が取れました。<FROM> 石巻 </FROM> から <METHOD> 徒歩 </METHOD> で <TO> 仙台市内 </TO> の家まで帰って来たそうです。

図2 行動経路を含むツイートにタグを付与した例

機械学習として CRF を使用した。CRF は与えられた文に含まれる語を分類するために使用した。予備実験の結果から、窓幅は 6 と定めた。CRF には、単語、品詞、括弧 (「, 『』 など)、手掛かり語の有無を素性に与える。以下に使用した手掛かり語の一部を示す。

- ANPI NLP で提供されている駅名
- ANPI NLP で提供されている東北地方の地名
- “を出発” など FROM タグの後に使用される単語
- “へ避難” など TO タグの後に使用される単語
- “で行く” など METHOD タグの後に使用される単語

2.4 実験と考察

実験には、ANPI NLP で提供されている震災情報に関連するツイート 1,303 件に対し、FROM, TO, METHOD タグを手で付与したデータを使用した。タグは訓練用に 728 件、評価用に 208 件付与された。機械学習には CRF++²⁾ を用いた。精度と再現率を用いて評価を行った。

実験の結果を表 1 に示す。精度に比べ、再現

表 1 行動経路にタグを自動で付与した結果

タグ	精度 (%)	再現率 (%)
FROM	78.4	40.9
TO	76.3	59.2
METHOD	91.7	64.7
平均	78.2	53.4

率は低い結果となった。再現率低下の原因は、主に地名や施設名に関する手掛かり語の不足であった。今後発生する震災に備え、各地の知名の詳細な辞書を予め構築しておく必要がある。

3. 商品レビューからの購買意図の自動抽出

3.1 研究の背景と目的

あらゆる人には欲求があり、人々の活動はそれによって説明することができる。アメリカの心理学者 Maslow は、人が生まれつき持つ欲求 (潜在欲求) を 5 種類 (生理的欲求, 安全の欲求, 社会的欲求, 尊敬欲求, 自己実現の欲求) に分類し、それらの間には階層的な構造があるとした (Maslow, 1943)。人が何かを手に入れる (購買する) 際にも、これらの欲求が源にあると考えられ、マーケティングにおいては Maslow の理論をもとに様々な研究が行われている。代表的な考え方は、人が持つ何らかの潜在欲求が、外からの何らかの刺激により顕在化し、購買行動に結びつくという解釈である。例えば、「喉が渇く」と言う潜在欲求の一つである生理的欲求に対して、それを気づかせるような刺激が与えられると「喉を潤したい」という顕在欲求となり、何らかの飲料を手に入れるという行動に結びつく。

本研究では、「顕在化された欲求」を購買意図と呼ぶ。上記の例における購買意図は、「喉を潤したい」である。購買意図を具体的に抽出、分類することは、最適な販売手段や消費者との適切な関係を構築するために非常に重要な情報になる。

まずそのために、顕在化された欲求をどのようにして抽出するかを考える必要がある。これまでのマーケティングではアンケートという手段が主に使われていた。しかし、楽天³⁾ や Amazon⁴⁾ などを利用したオンラインショッピングが日常的となっている現代、こうした欲求

が明示的に記述されている資源として、商品レビューがある。商品レビューには、ユーザの購買意図、購買した手順、商品やショップに対する感想、他者への推薦等の記述がされている。本研究では、商品レビューを対象に購買意図の抽出を行う。

3.2 商品レビューデータ

楽天市場の約1,660万件の商品レビューデータから購買意図の抽出および分析を行う。楽天市場の商品レビューデータは、楽天技術研究所より公開されている⁵⁾。商品レビューには、投稿者に関する情報（ただし、氏名、性別、住所などの個人情報に含まれていない）、商品名、商品カテゴリ ID、購入した商品に対するレビューが登録されている。この商品レビューから語彙構文パターンを使って購買意図の抽出を行う。

3.3 購買意図の抽出手法

「顕在化された欲求」である購買意図は、実際の商品レビューには様々な表現で記述される。我々は、商品レビュー分析（安藤他、2014）で定義される、買い手に関する分類項目中の「購入理由」に分類される表現を購買意図とした。10,000件の商品レビューを分析した結果、購買意図は、表2中の（例1）の「～たくて購入」や、（例2）の「しようと思ひ購入」など、ある限られた数の語彙構文パターンで記述される傾向があることがわかった。

高頻度で出現した語彙構文パターンを使って、楽天市場の商品レビューの全体から約40万件の購買意図を含むと考えられる文を抽出し

表2 購買意図を含む商品レビューの例

（例1）	焼きたてのワッフルが食べたくて購入しました。
（例2）	自分でジェルネイルに挑戦しようと思ひ購入しました。

表3 語彙構文パターンの一例

- たくて購入／買／注文
- しようと思ひ購入／買／注文
- するために購入／買／注文

た。利用した語彙構文パターンの例を表3に示す。

次に、購買意図を含む文から、購買意図の抽出を行う。基本的に、購買意図は動詞句を主辞とし、その格要素を含めた格フレームのインスタンスとして表現されることが多いため、購買意図を格フレームで抽出する。実際には、語彙構文パターンを含む文を係り受け解析器 CaboCha⁶⁾ を用いて解析し、語彙構文パターン直前の動詞を含む文節と、その文節に直接かかる助詞を含んだ文節を抽出する。表2の（例1）からは、「ワッフルが／食べたくて」、（例2）からは「自分で／ジェルネイルに／挑戦しよう」が、購買意図として抽出される。この結果、約45,000種類の購買意図（約9,000種類の動詞句）を抽出できた。抽出された購買意図のうち頻度の高い5件の動詞句と出現する商品ドメイン⁷⁾を表4に示す。頻度は、動詞句中の動詞の原形

表4 頻度の高い5件の動詞句

動詞句	頻度	主な商品ドメイン（頻度）
使用する	6,007	日用品雑貨・文房具・手芸 (859)／インテリア・寝具・収納 (675)
食べる	3,934	食品 (2,279)／スイーツ、キッチン用品・食器・調理器具 (596)
送料無料にする	3,689	日用品雑貨・文房具・手芸 (530)／バッグ・小物・ブランド雑貨 (461)
プレゼントする	3,522	日用品雑貨・文房具・手芸 (486)／バッグ・小物・ブランド雑貨 (454)
作る	3,493	日用品雑貨・文房具・手芸 (1,549)／キッチン用品・食器・調理器具 (524)

表5 抽出された購入意図の例

動詞句	格	頻度	主な商品ドメイン (頻度)
プレゼントする	友人に	250	日用品雑貨・文房具・手芸 (34)／おもちゃ・ホビー・ゲーム (23)
	母に	246	バッグ・小物・ブランド雑貨 (38)／靴 (24)
	子供に	60	キッズ・ベビー・マタニティ (13)／おもちゃ・ホビー・ゲーム (11)
	父に	56	ダイエット・健康 (12)／バッグ・小物・ブランド雑貨 (12)

で求めた。また、「プレゼントする」という購買意図と、その購買意図が出現する商品ドメインを表5に示す。

3.4 実験と考察

提案手法により抽出した購買意図の網羅性について評価を行う。実験には、1,000件の商品レビュー (2,970文) から、安藤ら (安藤他, 2014) の定義に従い購買意図を含むと判定された247文を評価対象とした。評価は人手で行った。例えば、評価対象に含まれていた「母のために購入」という購買意図は、自動で抽出した「母にプレゼントする」、もしくは「母に送る」という購買意図と同意であるといえる。このような場合に、評価対象の購買意図は、自動で抽出した購買意図によりカバーできていると評価する必要があるためである。結果、197文 (79.8%) の購買意図をカバーすることができていた。カバーできなかった購買意図を含む文には「他社製の3～5万円程度のギターの購入を考えていたのですが、それらと同程度の品質が望めると考え、購入を決めました。」のように、購買意図が複雑な例があった。

提案手法により自動で抽出した購買意図を使用し、商品検索システムを構築した。今後は、購買意図間の包含関係を自動的に発見することで、購買意図の構造化を行いたいと考えている。購買意図の構造化を行うことで、潜在的欲求に近い抽象的な購買意図から、具体的な購買意図、さらには商品の購買へと、人間の購買意図がどのように遷移するのかを明らかにすることができる。これにより潜在的欲求のみを持つ消費者に対する購買の促進が可能となる。

4. ま と め

本報告では、現在行っている研究事例として、まず、Twitterからの避難経路の自動抽出について説明した。今後の災害発生時にも適用できるように、地名や施設名の辞書構築を行いたいと考えている。

次に、商品レビューからの購買意図の自動抽出について説明した。今後は、購買意図を使用した商品検索システムの改良、購買意図の構造化を行う予定である。

注

- 1) <http://www.google.org/personfinder/japan/>
- 2) <http://www.chasen.org/~taku/software/CRF++>
- 3) <http://www.rakuten.co.jp/>
- 4) <http://www.amazon.co.jp/>
- 5) <http://rit.rakuten.co.jp/opendataj.html>
- 6) <http://chasen.org/~taku/software/cabocha/>
- 7) 商品ドメインとは楽天商品カテゴリ階層の第1層目のカテゴリであり、34種類存在する。

参 考 文 献

- Ishino, A., Nanba, H., and Takezawa, T. (2012) "Extracting Transportation Information and Traffic Problems from Tweets during a Disaster", Proceedings of the Second International Conference on Advances in Information Mining and Management.
- Maslow, A. H. (1943) "A Theory of Human Motivation", Journal of Psychological Review, Vol. 50(4), pp. 370-396.
- Neubig, G., Matsubayashi, Y., Hagiwara, M., and Murakami, K. (2011) "Safety Information

Mining —What can NLP do in a disaster—”,
Proceedings of the 5th International Joint
Conference on Natural Language Processing, pp.
965–973.

安藤まや, 関根 聡 (2014) “レビューには何が書か
れていて, 読み手は何を読んでいるのか?”, 言

語処理学会第20回年次大会.

石野亜耶, 村上浩司, 関根 聡 (2014) “商品レ
ビューからの購買意図の抽出とそれをういた商
品検索システムの構築”, 言語処理学会第20回年
次大会.