

大規模電子商取引データからの効果的な訴求表現の発見

Discovering Effective Selling Phrases from Large-Scale E-commerce Data

櫻井 祐子*¹ 白井 康之*² 鶴間 浩二*² 小山 聡*³
 Yuko Sakurai Yasuyuki Shirai Koji Tsuruma Satoshi Oyama

*¹九州大学大学院システム情報科学研究所

Faculty of Information Science and Electrical Engineering, Kyushu University

*² (独) 科学技術振興機構 ERATO 湊離散構造処理系プロジェクト
JST-ERATO MINATO Discrete Structure Manipulation System Project*³北海道大学大学院情報科学研究科

Graduate School of Information Science and Technology, Hokkaido University

This paper proposes an approach to help forming effective selling phrases for goods or services on E-commerce sites. Given terms describing concrete attributes or functions of a product/service, terms describing feelings or atmosphere are discovered from large-scale E-commerce data. Combinations of the two kinds of terms can be efficiently discovered by introducing an emerging pattern mining algorithm. Preliminary experiments are conducted on a real dataset provided by Rakuten, Inc.

1. はじめに

楽天や Yahoo などの電子商取引サイトにおいては、類似した商品やサービスが複数の供給者から提供されることが多い。数多くの商品やサービスの中から消費者に選択してもらうために、供給者はサービスや商品の特徴を適切にアピール(訴求)することが必要となる。アピールする手段としては、写真や商品説明文などがあるが、特に商品やサービスの名前(タイトル)は消費者が最初に目にするものであり、ここで注意を引くことが、購入に繋げる上で重要である。特に、旅行やホテルの宿泊プランなどのタイトルは、名前の付け方の自由度が大きいため、同じ商品やサービスでもタイトルの付け方によっては販売数に大きな差が出てくることが考えられる。一方で、電子商取引サイトでは商品やサービスのタイトルは表示スペースの都合から文字数が制限されていることが多い。そのため、限られた文字数で適切にアピールするタイトル名を作成することは、供給者にとって困難な問題となっている。

表1に楽天トラベルにおけるプランタイトルの例を示す。限られた文字数の中で、商品の属性を表す名詞や、消費者の感覚に訴える形容詞を組み合わせてプランタイトルが作成されていることが分かる。本稿では、商品やサービスに対する効果的な訴求表現の作成を支援する方法について検討する。具体例として、宿泊予約サイトにおけるプランタイトルの作成支援を扱う。特に形容詞や副詞で表わされる感覚的な表現に着目し、提供者の入力した語に対して、修正候補の語を提示する。

関連研究としては、検索連動型広告における入札キーワードの選択を支援する研究がある[Bartz 06]。高橋ら[Takahashi 10]は料理レシピや旅行ツアーなど、恣意的に名前が付けられる Web 上に記載されたエンティティに対して、その名前と内容が一致しているかどうかを定量的に評価する方式を提案している。高橋らの研究は情報の信憑性の観点から消費者を支援する研究と考えることができ、本研究とは相補的である。

表 1: プランタイトルの例

楽天限定<ポイント10倍>まる得プラン
<人気>夕朝食ともに「部屋食で満足」旬の瀬戸内料理プラン
<特別企画>3/1~4 限定!最大40%引 とくつくプラン
パパママ応援企画<3歳以下・添い寝のお子様・赤ちゃん無料>赤ちゃん大歓迎
22~30歳限定 レイトチェックアウト ゆっくり朝寝坊プラン 朝食後たっぷり 貸切露天風呂
女性ひとり旅 レディースプラン~癒~
<仕事帰りにホット一息。こりゃいい湯!>露天感覚の大浴場で湯ったり 温泉旅館気分

2. 顕在パターンマイニングを用いたプランタイトル候補語の発見

顕在パターンマイニングは頻出パターンマイニングをクラスラベル付きのデータ用に拡張したものであり、あるクラスでのサポート(出現確率)が他のクラスでのサポートに比べて大きく増加するようなパターンを発見する[Dong 99a]。発見されたパターンは、クラス分類の際の特徴などにも用いられる[Dong 99b]。本研究ではプランタイトルに現れる単語をアイテムとみなして、ユーザによる評価が高くなるような単語の組み合わせを抽出し、プランタイトルに用いる語の候補として提示することを考える。

一般に、顕在パターンマイニングで発見されるパターンの数は膨大であるが、必ずしもそれらの多くがプランタイトルの作成に有用とは限らない。例えば、「温泉」という語を含むプランタイトルの評価が高かったとしても、温泉を持たないホテルのプランタイトルに使うことはできない。商品やサービスの属性を表す語は、その語を使うことができるか否かはプランタイトル作成以前に決定していることが多く、プランタイトル作成者自身も選択に迷うことは少ないであろう。一方、「ゆったり」や「のんびり」といった感性や雰囲気を表す語は、比較的プランタイトル作成者の自由度が大きいため、選択の支援が必

連絡先: 櫻井 祐子, 九州大学大学院システム情報科学研究所, 〒819-0395 福岡市西区元岡 744, ysakurai@inf.kyushu-u.ac.jp

表 2: 「食事」に関連する名詞への修飾語

サポート ランキング	パイキング	部屋食	蟹
1	たっぷり	ゆったり	一杯
2	おいしい	ゆっくり	甘い
3	どっさり	のんびり	堪能
4	早い	堪能	まるごと
5	うれしい	特別	極
6	堪能	新鮮	ゆったり
7	断然	たっぷり	ぷりぷり
8	楽しい	やっぱり	でっかい
9	広々	もちろん	美味しい
10	どうぞ	ひとり	たまらない

表 3: 「風呂」に関連する名詞への修飾語

サポート ランキング	温泉	湯	貸切
1	のんびり	のんびり	のんびり
2	ゆっくり	特別	ゆったり
3	特別	ゆっくり	特別
4	ひとり	白い	ゆっくり
5	たっぷり	つるつる	ひとり
6	嬉しい	ぶらり	極
7	ちょっと	ふんだんに	丸ごと
8	遅い	ちょっと	もちろん
9	広々	たっぷり	なんと
10	ゆったり	のんびり	しっとり

要であると考えられる。そこで、本研究では、以下の2種類の語の組み合わせによる顕在パターンを発見することを試みる。

1. 商品やサービスの具体的な属性や機能を表す語（一般に名詞で表現される）
2. 商品やサービスの感性や雰囲気を表す語（一般に形容詞や副詞で表現される）

抽出するパターンの形に制約を付けることで、計算量の増大を避けると同時に、抽出されたパターンの中での利用可能 (actionable) なパターンの割合を増やすことが可能となる。

3. 実験

本研究ではホテル予約サイト楽天トラベルのデータを用いた [楽天 10]。楽天トラベルのレビューには利用形態として、「ビジネス」「レジャー」などの区分が与えられている。今回の実験では一件以上「レジャー」のレビューが付与されたホテルのプラントタイトルを用いた。楽天トラベルデータのプラントタイトルを茶筌^{*1}を用いて形態素解析し、名詞の中から「食事」および「風呂」に関する属性や機能を表す語 100 語、50 語をそれぞれ人手で選択し、これらの語および「形容詞」、「副詞」、「名詞-形容動詞語幹」をアイテムとした。「食事」および「風呂」のそれぞれの評価値の平均が 4 以上のプラントタイトルを正例、4 未満のプラントタイトルを負例とし、IGVcaep^{*2}を用いて顕在パターンの発見を行った。表 2 および表 3 に顕在パターンで、食事に関連する名詞として「パイキング」「部屋食」「蟹」、風呂に関連する名詞として「温泉」「湯」「貸切」と共に現れる修飾語をサポートの大きな順に示す。

4. おわりに

本稿では、電子商取引で扱う商品やサービスに対する訴求表現の作成を支援する方式について述べた。本研究の貢献は以下のとおりである。

- 商品・サービスの訴求表現作成支援という新しい問題を提案した。
- 大規模電子商取引データに対する顕在パターンマイニングを用いた訴求表現作成支援の枠組みを考案した。

*1 <http://chasen.naist.jp>

*2 <http://kgmod.jp/mcmd/index.php?igvcaep.rb>

- 商品・サービスの属性を表す名詞と、それらを修飾する形容詞・副詞の組み合わせからなるパターンに着目することで、計算量の増加を避けつつ、有用なパターンを効率良く発見することを可能とした。

今後の課題としては、属性を表す名詞の自動抽出や、概念階層を用いて類似する名詞をまとめることで、有用なパターンの抽出数を増やすことが考えられる。

謝辞

本分析においては、楽天株式会社から研究目的のために公開されたデータを利用している [楽天 10]。大阪府立大学森田裕之教授、関西学院大学羽室行信准教授には顕在パターンマイニングプログラム IGVcaep の提供およびご助言を頂いた。ここに記して感謝する。

参考文献

- [Bartz 06] Bartz, K., Murthi, V., and Sebastian, S.: Logistic Regression and Collaborative Filtering for Sponsored Search Term Recommendation, in *Proceedings of the 2nd Workshop on Sponsored Search Auctions* (2006)
- [Dong 99a] Dong, G. and Li, J.: Efficient Mining of Emerging Patterns : Discovering Trends and Differences, in *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 43–52 (1999)
- [Dong 99b] Dong, G., Zhang, X., Wong, L., and Li, J.: CAEP: Classification by Aggregating Emerging Patterns, in *Proceedings of the 2nd International Conference on Discovery Science*, pp. 30–42 (1999)
- [Takahashi 10] Takahashi, R., Oyama, S., Ohshima, H., and Tanaka, K.: Evaluating Truthfulness of Modifiers Attached to Web Entity Names, in *Proceedings of the 11th International Conference on Web-Age Information Management (WAIM 2010)*, pp. 429–440 (2010)
- [楽天 10] 楽天株式会社: 楽天データ公開, <http://rit.rakuten.co.jp/rdr/index.html> (2010)