

# ラフ集合理論を用いた web 上の酒類に対する製品評価文構造の視覚化

## Visualization for Structure of Sentences on Liquor Products Reputation on the Web

竹之下 遼† 高見諭史‡ 原田利宣†

Ryo Takenoshita Satoshi Takami Toshinobu Harada

### 1. はじめに

消費者が製品を購入する際、Webで製品評価文を参考にすることが多い。このような認知から購入の決定という感性は、「明るい色」や「角ばった形」のような具体的な部位の知覚から、「スポーティーな」、「モダンな」というイメージ、「かわいい」、「買いたい」のような認識、態度といった、抽象度別に階層化されることが知られている。感性工学の分野ではこのような製品の感性評価を表す言葉を感性ワードと呼び、主に、形容詞・形容動詞で表される[1][2]。一方、従来のデータマイニングの問題点として、多くのデータから算出された結果の解釈・理解の難しさ、有益な情報の発見が困難であることが挙げられている。その問題点を解決する技術として、ビジュアル・データマイニングが注目されており、様々なツールが開発されている[3]。ビジュアル・データマイニングとは、分析結果を可視化することで、より直感的な理解を支援する技術のことである。その結果、今まで発見することが難しかった有益な情報を得ることができると考えられている。

そこで本研究では、Web上の製品評価文を元に、ラフ集合を用いて、消費者の感性の階層構造を可視化するシステムを開発する（以下、視覚化システム）。具体的には、酒類の評価文をケーススタディとして、固有表現抽出法を用いて、階層ごとの感性ワードを抽出する。製品のレビューページから、抽出された感性ワードを用いて、ラフ集合における決定表を生成する。その後、生成した決定表から決定ルールを算出し、その結果から、感性の階層構造を可視化する。

### 2. テキストマイニング手法について

視覚化システムに用いる手法の概要を以下に示す。

#### 2.1 ラフ集合とは

ラフ集合とは1982年にZ.Pawlakによって提唱され、認知構造や感性情報をモデル化する方法として、近年注目されている[4][5]。本研究では、ラフ集合理論における決定ルールを用いて、結論（ポジティブ評価/ネガティブ評価）に対する、感性ワード間の関連を視覚化する。決定ルールとは、If-Thenルール形式で表現できる決定表から導き出されるルールのことである。さらにその決定ルールの結論に対する寄与度の指標としてCovering Index（以

下C.I）を用いる。C.Iとはそのルールと同じ結論の対象数のうちで、ルールにあてはまる対象数の占める割合を示している。決定ルールを算出することで、結論に関係する重要な属性値の組み合わせを把握することができる。

#### 2.2 固有表現抽出とは

固有表現抽出とは自然言語処理技術の一つであり、固有名詞や固有物名を文書から抽出する技術である。

例として以下の文書の場合をあげる。

『4月24日、Apple Watchが日本で発売された』

「4月24日」、「Apple Watch」、「日本」をそれぞれ、時間、物名、地名として判別する。本研究では、形態素解析と構文解析を用いた固有表現抽出を行う。形態素解析とは、文書を意味のある単語に区切り、辞書を利用して品詞や内容を判別する自然言語処理技術である。構文解析とは、文書を定義された文法に従って解釈し、文の構造を明確にする技術である。本システムではYahooテキスト解析APIを利用して形態素解析と構文解析を行った。

### 3. 視覚化システムのアルゴリズム

視覚化システムのアルゴリズムについて説明する。本システムのフローを図1に示す。

#### 3.1 Web上のレビューの抽出と感性ワード辞書構築

感性ワード辞書構築のために、焼酎5種（「赤霧島」、「赤兎馬」、「いいちこ」、「富乃宝山」、「百年の孤独」）から、消費者が書いた評価文をAmazon、楽天のレビューページから1945件抽出し、その中に含まれる形容詞・形容動詞を感性ワードとして辞書に登録した。さらに、その中でも意味の近い語句は類義語として、一つにまとめた。その後、それらの感性ワードに対してレマ化を行った。レマ化とは、単語の頻度研究などに用いられる手法で、基本形と活用形の関係を一つの表記系にまとめる手法である[6]。これにより、活用形の個々の頻度は基本形頻度に集約される。レマ化された形容詞に対して、「計量データとして計測可能なもの」<色、距離、面積、時間など>、「官能評価で個人の差があまり無い物」<味覚、触覚、聴覚など>「それ以外のもの」の観点から、抽象度を3つの階層に分類し、これらの情報を付加して、感性ワード辞書の構築を行った。

† 和歌山大学システム工学部, Faculty of Systems Engineering, Wakayama University

‡ 和歌山大学大学院システム工学研究科, Graduate School of Systems Engineering, Wakayama University

