

商品紹介文の作成支援における生成文評価機能の開発

阿部 涼介^{1,a)} 横山 想一郎² 山下 倫央² 川村 秀憲²

概要: 近年の EC 市場の拡大に伴い、EC サイトなどに掲載する商品紹介文の必要性が高まっている。この課題に対して、著者らはテンプレートを用いた商品紹介文生成機能を有する商品紹介文の作成支援システムを開発してきた。開発システムの課題として、大量の生成文の中からシステム利用者に提示する生成文を選択するための評価機能が不足していることが挙げられる。本稿では、構文情報を用いた文の読みやすさなどの複数の評価指標に基づく生成文評価機能を開発した。

1. はじめに

現代の情報社会において、携帯型デバイスの所持の一般化、さらに他者とのコミュニケーションが容易に取れるインターネット・SNS の普及 [1] に伴い、電子商取引 (EC・Electronic commerce) の市場規模が拡大し続けており、売上は約 19 兆円にも及んでいる [2]。EC 市場は多くの人から一般的な商品の購買手段として利用されていることがわかる。

電子商取引 (EC) では、商品の販売担当者が商品を掲載・販売する EC サイトを作成し、顧客は EC サイトを利用し商品を購入する。例として、EC サイトの一つであるオムニ 7[3] では現在約 528 万種類の商品が扱われている。EC サイトを利用する顧客はこの商品情報を参考にどの商品を購入するか意思決定を行う。ここでいう商品情報とは商品名と共に掲載される、商品の画像、紹介文、生産地や価格等を示す。EC サイト上では、商品の触感や匂い、大きさなどの生の情報を自分の目や、商品に直接触れて確認したりすることが不可能である。そのために EC サイト上に掲載する商品情報は顧客の購買の意思決定において重要な情報である。

顧客の購買の意思決定には、商品の販売担当者によって作成される商品情報の質が重要になる。ここでいう質は、商品の生産地、商品名など一意に定まる情報はなく、紹介文や商品画像、他商品との類似点など商品の販売担当者の表現方法に依存する情報である。商品情報の質を高める

ことは商品の売り上げにつながるが、掲載する商品情報の組み合わせの試行錯誤や商品情報の取捨選択が必要となり、作成コストは大きい。例えば紹介文であれば、商品のどの特徴を含みどのような文章を作成するか、顧客にとって読みやすい紹介文となっているかなど。1 つの紹介文を作るとしても商品の販売担当者の負担は大きいものとなる。

EC サイトを利用する顧客にとって商品情報の中でも紹介文は、(1) 味覚や嗅覚等の情報 (2) 見た目だけでは確認出来ない要素 (ex. カーテンにおける断熱性、PC における機能等) (3) 他商品との類似・相違点の列挙などの商品の特徴に関する情報を多く含む。そのために商品の紹介文は商品情報の中でも特に重要となる。商品情報の中でも、紹介文は多くの場合、商品の販売担当者的手によって作成される。さらに、EC 市場の拡大に伴い、EC サイト上で扱われる商品数が増加し、商品の販売担当者が商品紹介文を作成するためにかかる負担も増加している。

また、商品紹介文には、EC サイトに掲載する商品の見出し文や、商品の詳細としての紹介文などが含まれる。更にはセールス、期間限定商品のお勧めのための特別な広告、メールマガジン SNS 等の他メディアに掲載するための紹介文なども存在している。そのために商品の販売担当者的商品紹介文の作成を支援する技術に高い需要がある。

本研究では、商品紹介文の作成支援システムを開発することで、紹介文作成にかかる商品の販売担当者的負担を削減することを目的とする。

EC サイトの商品紹介文において、複数のメディアに対応した複数の紹介文を、商品の数だけ商品の販売担当者的が作成するのは負担が非常に大きい。そこで本研究では開発したシステムにより、商品の特徴となるデータとテンプレートがあれば、商品の販売担当者的の複数の紹介文作成支援を行う。手始めとして、商品紹介文を作成しやすく、生

¹ 北海道大学 大学院情報科学院
Graduate School of Information Science and Technology,
Hokkaido University, Sapporo, Hokkaido, Japan

² 北海道大学 大学院情報科学研究院
Faculty of Information Science and Technology, Hokkaido
University, Sapporo, Hokkaido, Japan

a) aber_23@ist.hokudai.ac.jp



図 1 EC サイト (omni7) における商品掲載例

成した商品紹介文の質を評価しやすい条件を持つ商品として日本酒を対象商品とする。

本研究で提案する商品紹介文作成支援システムは、1つの商品に対し複数の紹介文候補を提示する。商品の販売担当者作業は、紹介文に必要な条件（商品の特徴・文の長さ）の指定と生成された紹介文候補から紹介文を選択するのみである。商品の販売担当者の負担となる、選択した特徴を含む紹介文の生成は、商品の販売担当者の選択に従い本システム内で処理する。

本稿では、商品の販売担当者にとって負担となる、顧客にとって読みやすい紹介文となっているかをシステム内のモジュールで評価する。モジュールで評価した紹介文は、顧客にとって読みやすい紹介文が商品の販売担当者に紹介文候補を提示される。

本稿では、第2章で関連研究について示す。第3章で著者らが開発してきたシステムについて示し、第4章で開発した評価モジュールについて示す。最後に第5章で本稿のまとめについて示す。

2. 関連研究

2.1 ユーザの支援システム

ユーザの支援システムに関しては、将棋における局面把握支援 [4]、音楽における作曲の支援 [5] といった先行研究が行われている。また、文章の作成支援としては、ユーザの創造的な文章生成の支援 [6]、ユーザのエッセイ作成支援 [7] 等の研究が挙げられる。これらの研究は創作的な活動を対象としており、人間の活動の支援が目的となっている。文章の作成支援においては、1つのコンテンツに対し、1つの文章を生成することを目的としている。しかし、商品の紹介文の作成支援の際には、複数メディアに対応した複数の文章を生成する必要がある。本研究では先行研究とは異なり、1つのコンテンツ（商品）に対して複数メディアに対応した異なる内容の紹介文を生成可能となっている。

2.2 文章生成

文章生成における先行研究は2つに大別できる。深層学

習を用いた言語モデルで行う研究、ルールベースで文章生成を行う研究という区分である。

2.2.1 深層学習

まず深層学習を用いた言語モデルで行う研究においては、文章自動生成の手法として LSTM [8]、Transformer [9] などを用いて大量のテキストデータや画像を教師データとして文章の自動生成を行っている。LSTM を用いた観光案内における説明文の自動生成 [10] や、BERT [11] と強化学習を用いたテキスト要約 [12]、深層学習を用いた広告スローガンの生成 [13]、Transformer を用いた俳句生成 [14] 等の先行研究が行われている。深層学習を用いた言語モデルで行う研究は、画像データから文章生成が可能であること、大量のテキストデータを基に存在しない単語の意味の予測を行ったりと柔軟な表現が可能手法となっている。ただし、文法的な間違いや、単語の意味の取り違え等が起きてしまう可能性があり、質のある文章を生成することを保証出来ない。

2.2.2 ルールベース

ルールベースで文章生成を行う研究としては、単語の重要度に基づく文章要約 [15]、バイト符号化アルゴリズムによるテキストデータ圧縮 [16] 等が行われている。また、ルールベースの文章生成にはテンプレート文を利用しているものが存在する。テンプレート文を利用している文章生成では、予め用意もしくはシステム内で生成したテンプレートに、必要な特徴を与えて文章を生成する。関連研究としては、テンプレートを用いたロボットの挨拶文生成 [17]、アパレル EC サイト向けの紹介文生成 [18]、商品属性と商品説明をもとにテンプレート及び紹介文を生成 [19]、競輪におけるテンプレートを用いた記事生成 [20] などの研究が存在する。テンプレートを用いた文章生成では、一定以上の文章の質の保証及び、膨大な文章データの用意が不要等のメリットがある。ただし、テンプレートの数次第では、テンプレートの使い回しによる類似した文言の文章が生成されやすい。

本研究では、商品の販売担当者の紹介文作成の負担を削減することを目的としている。そのために、紹介文として質が一定以上であることが保証される、テンプレート文を用いた文章生成の手法を適用する。

2.3 文章の評価

文章自体の評価に関する研究も盛んに行われている。文章の読みやすさを文章の構造に基づいて評価 [21]、[22] や、文章の一貫性に関する評価 [23]、[24]、文章の内容自体を評価 [25] を行う研究が挙げられる。本稿では商品の販売担当者にとって負担となる文章の読みやすさを文章の構造に基づいて評価する手法を取り入れていく。

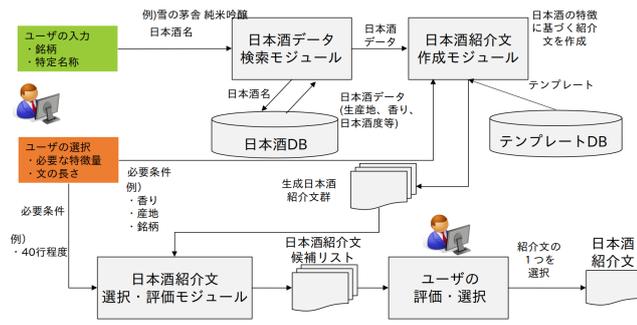


図 2 日本酒紹介文作成支援システムの概要

3. 日本酒紹介文作成システム

本稿では、著者らが開発してきた日本酒紹介文作成システム [26], [27], [28] について示す。日本酒紹介文作成支援システムの概要を図2に示す。著者らのシステムでは、以下に示すプロセスでユーザ（商品の販売担当者）の紹介文作成を支援する。

(1) ユーザによる紹介文を生成したい日本酒の種類を選択
日本酒DBに含まれる銘柄と特定名称をUIを通して入力する。銘柄と特定名称を組み合わせた「八海山 吟醸」のような文字列を、本稿では「日本酒名」と表記する。

(2) システムによる紹介文の生成

日本酒データ検索モジュールから得る「日本酒データ」とテンプレートDBから取得した「テンプレート群」を日本酒紹介文作成モジュールの入力となる。入力とユーザの指定した条件を元に日本酒紹介文作成モジュールが日本酒紹介文群を出力する。

(3) 日本酒紹介文選択・評価モジュールによる紹介文の評価

入力として渡された、生成日本酒紹介文群に対して、日本酒紹介文選択・評価モジュールによる紹介文候補の絞り込みを行い紹介文リストを出力する。本稿では、新たな評価機能を追加した。(4) ユーザによる紹介文候補の選択

紹介文リストを元にユーザが使用したい紹介文を1つ選択する。

本節では、今回紹介文を生成する対象商品となった「日本酒」の選択理由について示した後に、システムのデータベース、及びモジュールについて示していく。

3.1 商品の選択理由

本研究の対象商品として日本酒を選択した理由は、紹介文が重要となる商品という観点から、主に3つ存在する。

(1) 五感の中でも「味覚」「嗅覚」等の情報を含む商品
ECサイト上では、画像や動画、音によって視覚や聴覚に関する情報を得ることが出来るが、味覚や球殻などの情報は文字として表すしかなく紹介文の需要が高くなるため。

(2) 商品情報となる特徴が多い商品

紹介文に含むことが可能な特徴が多ければ多いほど、紹介文に含む特徴の組み合わせが膨大になり、商品の販売担当者がどの特徴の組み合わせを選択するか考える必要が出てくるため。

(3) 購買頻度が週1～数カ月に一度程度である商品

購買頻度が毎日～週に数回あるような商品は、消耗品の中でも値段が重視される傾向があり、逆購買頻度が年に1度のような商品は、商品を必要とする顧客が少なく紹介文を見る機会が少ないため、データの収集等が行いづらく研究を進めていく上で支障が出ると思ったため。

したがって、本研究では商品カテゴリとして日本酒を対象として、日本酒の特徴とあらかじめ用意したテンプレートを入力とする、商品紹介文の作成支援システムを作成する。

3.2 日本酒について

本節では、本研究の対象商品となった日本酒について、名称に関する「銘柄」「特定名称」について説明を行う。日本酒の名称は「雪の茅舎 純米吟醸酒」といった形で形成されたおり、「雪の茅舎」が「銘柄」を示し、「純米吟醸酒」が「特定名称」を示している。

3.2.1 銘柄

日本酒におけるブランドを示す。日本酒の製造元である蔵元は、基本的に複数の「銘柄」を持っている。日本酒は蔵元がある地域の米と水を原料として、「杜氏（とうじ）」と呼ばれる代々日本酒製造の技術を引き継いできた職人によって製造される。そのために銘柄ごとの味わいに差が生じる。

3.2.2 特定名称

「特定名称」とは日本酒を製造する際に使用される原料や製造方法により定まる名称である。使用原料には「米」、「米麴」、「醸造アルコール」がある。製造方法に関しては「精米歩合」、「麴米使用割合」等がある。

3.3 データベース

データベースは「日本酒データベース」と「テンプレートデータベース」に2つがある。各データベースの説明の前に、双方に関連する「日本酒データカラム」について説明を行う。

3.3.1 日本酒データカラム

表1にある日本酒データカラムとは、ネットサイト酒仙人 [29] に掲載されている、日本酒の情報をもとに制作したカラムである。制作したデータカラム数は25種類であり、さらに大別として6グループに分類した。(1) 商品情報：商品の名前や説明等が分類されるカラムである。(2) 温度：日本酒に関する温度の情報が分類されるカラムである。(3) 製造関連：日本酒の製造に関わる情報が分類されるカラム

表 1 日本酒データカラム

カラム名	特徴
商品情報	名称, 銘柄, 特定名称, 授賞
温度	適した温度, 燗向け, 常温向け, 冷向け
製造関連	蔵元, 生産地, 使用米
適した状況	季節, 対象, シーン, 料理, 容器, タイミング
数値	アミノ酸度, 酸度, アルコール度数
感覚	味, 香り, 飲み口

である。(4) 適した条件：日本酒を飲む際に、適している季節や料理などの情報が分類されるカラムである。これらのカラムは日本酒の専門家たちが、日本酒を飲む際にふさわしいと判断した時期や料理などを要素としてもつ。(5) 数値データ：日本酒において「日本酒度」、「アミノ酸度」等の数値が分類されるカラムである。(6) 感覚データ：日本酒を飲む際に感じる事が出来る感覚が分類されるカラムである。

3.3.2 日本酒データベース

節 3.3.1 で説明した日本酒カラムを用いて、ネットサイト酒仙人 [29] と複数のサイトのデータより日本酒データベースを作成した。データによっては欠損があるものが存在しているが、本稿に掲載している日本酒に関してはデータの欠損はない。

3.3.3 テンプレートデータベース

テンプレートデータベースは日本酒紹介文を生成するために用意したテンプレート文によって構成されている。本研究においてテンプレート文は著者の手で酒仙人 [29]、オムニ 7[3] 等の複数の日本酒紹介文掲載サイトを参考に作成した。テンプレート文は全て 1 文となっており、1 文以上の紹介文を生成する際には文同士の組み合わせにより生成する。テンプレート文の参考にした日本酒紹介文は主に

- 紹介文内に日本酒カラムに該当する日本酒の特徴が含まれている点
- 汎用性が高く、一つの紹介文を参考にして複数のテンプレートを作れる点

の 2 点に注目して著者が収集して、テンプレートの作成を行った。テンプレートの作成例を表 2 で示す。1 文の紹介文が与えられた際に、日本酒データカラムに該当する日本酒の特徴部分をテンプレート化させる。このようにしてテンプレートを作成した結果、約 300 種類のテンプレートを用意した。ただし、作成したテンプレートの中には、1 つのテンプレート内の特徴を変更・削除した形のものを含んでいる。例を表 3 に示す。表 3 の例では、冒頭のテンプレート部分が「香り」「味」「飲み口」の三種類で異なるのみで、他の内容は同一の内容である。そのために全てのテンプレートが全く異なる内容のテンプレートではない。テンプレートは ID で管理しており、類似しているテンプレートの判別が出来るように ID を割り振っている。

表 2 紹介文とテンプレートの例

1 文の紹介文	富山の銘酒、満寿泉の純米大吟醸酒です。
テンプレート	「生産地」の銘酒、「銘柄」の「特定名称」です。

表 3 類似しているテンプレートの例

テンプレート 1	「香り」をお楽しみいただけます。
テンプレート 2	「味」をお楽しみいただけます。
テンプレート 3	「飲み口」をお楽しみいただけます。

表 4 モジュール比較 (雪の茅舎)

	処理時間	紹介文数	データサイズ
総当り (2 文)	33.56 [sec]	118 万個	11.1GB
木構造 (2 文)	20.48 [sec]	416 個	0.3MB

3.4 日本酒データ検索モジュール

日本酒データ検索モジュールは、ユーザーの入力として受け取った日本酒の名称を日本酒 DB から検索して、取得した日本酒データを日本酒紹介文作成モジュールに渡す。入力に対応するデータが存在しなかった場合はここでエラーを返す。

3.5 日本酒紹介文作成モジュール

本モジュールでは、文の数 (2 文, 3 文) が指定された際に、生成可能な全ての紹介文を生成せずに、指定された日本酒の特徴を含む紹介文のみを生成する。生成可能な全ての紹介文を生成した場合と比較して、PC の使用メモリとシステムの実行時間が削減されている。実際に生成可能な全ての紹介文を生成した場合 (総当り) と本稿で提案しているモジュール (木構造) で紹介文を生成した際の、PC の処理時間及び生成される紹介文数、データサイズについて表 4 に示す。表 4 にあるように、処理時間の短縮、および生成される紹介文数削減によるデータサイズの縮小を確認できる。

紹介文生成の流れとしては、

- (1) 特徴組リスト生成器：テンプレート文をテンプレート文が含む日本酒の特徴組リストに変換。(2) 木構造生成・枝刈り器：特徴組リストをもとに木構造を生成後、ユーザーの指定条件 (紹介文数・日本酒の特徴) を元に木構造の枝刈り。(3) 紹介文生成器：枝刈りした木構造を元に紹介文を生成。

という流れである。

本節では、各生成器におけるアルゴリズムについて示す。

3.5.1 特徴組生成器

特徴組リストでは、テンプレートが持つ特徴をリストにし、あるテンプレートが組み合わせられたときに含まれる特徴組のリストを用意する。表 5, 6 にテンプレートの特徴組に変換する例を示す。表 5 にあるテンプレート 1 とテンプレート 4 は異なるテンプレートであるが、特徴組に変換すると 2 つとも表 6 の特徴組 1 となる。このようにして、テンプレート 300 種程度からテンプレートの特徴組の

表 5 テンプレートの例

テンプレート 1	「香り」と「味」を感じられる「特定名称」。
テンプレート 2	「地域」を生産地とする銘酒「銘柄」です。
テンプレート 3	「香り」と「飲み口」を感じられる「銘柄」。
テンプレート 4	「特定名称」は「香り」と「味」が特徴のお酒。

表 6 特徴組リストの例

特徴組 1	「香り」「味」「特定名称」。
特徴組 2	「地域」「銘柄」
特徴組 3	「香り」「飲み口」「銘柄」

リストを生成する、日本酒にもよるが特徴組リストのサイズは 100 程度になる。テンプレートの増加と共に特徴組リストのサイズが大きくなるが、テンプレートの種類を変数 *type*、テンプレートの特徴組リストを変数 *list* と置くと、 $type \geq list$ が常に成立する。そのために特徴組リストを組み合わせる本手法は、テンプレートのまま組み合わせるより必ず少なくなる。

3.5.2 木構造生成・枝刈り器

本節では、3.5.1 節で生成した特徴組リストを元に、木構造の生成及びユーザの指定条件による枝刈りを行う木構造生成・枝刈り器について説明する。表 6 の特徴組リストを元に生成される木構造の例を図 3 に示す。木構造のノードの深さが生成される紹介文の「文数」、各ノードの要素がテンプレートが持つ特徴組となっている。ここでいう「文数」とは、「。」で終わる文を 1 文と数えている。そのため文数が 3 の紹介文とは紹介文内に「。」が 3 つ出てくる文章を表す。このようにしてあるノードまでの特徴組の組み合わせにより、紹介文が生成される際に含まれる全ての特徴と、紹介文の文数を表現した。例として図 3 の木構造における左下の「香り」「飲み口」「銘柄」を要素としてもつノード(対象ノードとする)に至った際に生成される紹介文を示す。まず深さ優先探索により深さ 3 の対象ノードを探索する。そして対象ノードに到達するまでの経路とした各ノード、深さ 1 における「香り」「味」「特定名称」のノードと深さ 2 における「地域」「銘柄」のノード、の要素の和集合である「香り」「味」「飲み口」「特定名称」「地域」「銘柄」の特徴組を得る。この特徴組は経路とした枝の要素を元に紹介文を生成した際に、紹介文に含まれる特徴となっている。このようにして生成される木構造に対して、ユーザの指定した条件を元に木構造の枝刈りを行う。枝刈りによって木構造に含まれるノードは、ユーザの指定条件に一致するもののみとなる。

3.5.3 紹介文生成器

本節では、3.5.2 節で生成された枝刈りをされた木構造をもとに紹介文生成を行っていく。まず、枝刈りをされた木構造の各ノードの要素を、要素となっている特徴のみを含むテンプレートに変換する。次にテンプレートを元に 1 文の文章を生成する。この際、ある日本酒が持つ 1 つのカ

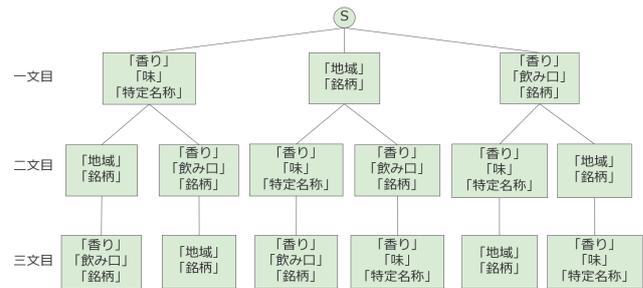


図 3 生成される木構造の例

ラムの要素は複数存在する場合がありますので、1 つのテンプレートに該当する日本酒の特徴を埋め込むと複数の文章が生成される。そして各ノードで生成された紹介文を、木構造の深さの順を保ちつつ組み合わせていくことで、ユーザが指定した文数と特徴の条件が満たされた紹介文の生成が行われる。

3.6 日本酒紹介文選択・評価モジュール

本研究では、節 3.5 で生成された紹介文に対してシステム内で評価を行った。そして評価を元にユーザに紹介文の選択を行ってもらい、評価機能としては、(1) 単語重複率、(2) 文章の複雑さ、(3) 単語の共起頻度、(4) 特異な表現の使用頻度、の 4 種類が存在する。本節では (1) 単語重複率、(2) 文章の複雑さ、(3) 単語の共起頻度、(4) 特異な表現の使用頻度、については 4 章で説明する。

3.6.1 文章の読みやすさに関する評価機能：単語重複率

紹介文に対して、紹介文内に出現する単語の重複率が低いほう(重複している単語が少ない)が良いという基準で評価を行う。単語重複率、紹介文内の重複単語数、紹介文内の出現単語数のそれぞれを、変数 dup , $w_{duplicate}$, $w_{appearance}$ とすると $dup = w_{duplicate} / w_{appearance}$ と表される。本研究では、テンプレートで生成した文章を組み合わせると複数分の紹介文を生成するために、似たようなテンプレートの組み合わせや、同一の特徴を含む紹介文が生成される可能性がある。それらの紹介文を単語重複率による評価によってユーザに提示しないようにした。表 7 は単語重複率が「0.1」「0.2」「0.4」付近であった紹介文の例である。単語重複率が 0.1 未満の紹介文は、単語の重複がほとんどなく読みやすい紹介文になっていることを確認した。

3.7 UI の概要

本研究では商品の販売担当者に実際に UI を通して日本酒紹介文作成システムを操作してもらい、紹介文作成を支援する。そのために、商品の販売担当者に負担がかからないように簡単な操作で紹介文作成を行ってもらうために UI を設計した。

UI では、(1) 日本酒の銘柄の選択。(2) 日本酒の特定名

表 7 生成紹介文の単語重複率 (日本酒：北雪)

単語重複率	紹介文
0.1 未満	原料米に山田錦を使用し、丁寧に仕込んだ大吟醸です。和洋のシーンでよく飲まれたり、伊勢海老の磯焼きと合わせたりして飲む方が増えています。
	ワイン通でも楽しめる北雪。和洋のシーンでも楽しめて、ローストビーフなどとの相性は抜群。
0.1 以上 0.2 未満	日本酒通でも楽しめて、鮑のバターソテーなどとの相性は抜群。フォーマルな雰囲気でも楽しめて、伊勢海老の磯焼きなどとの相性は抜群。
	原料米に山田錦を使用し、丁寧に仕込んだ大吟醸。山田錦で丁寧に仕込んだ、なめらかでキメの細かい味を有田焼でどうぞ。
0.2 以上 0.4 未満	華やかでエレガントな花の香りで、なめらかでキメの細かい味を醸し出します。北雪は華やかでエレガントな花の香りとなめらかでキメの細かい味が特徴のお酒。
	ワイン通でも楽しめて、鮑のバターソテーなどとの相性は抜群。フォーマルな雰囲気でも楽しめて、鮑のバターソテーなどとの相性は抜群です。

称の選択. (3) 紹介文に必要な特徴の選択. (4) 選択された日本酒の紹介文の生成. (5) 紹介文の長さの選択. (6) 紹介文候補から紹介文の選択. (7) 紹介文の修正・保存. という手順で紹介文作成の支援を行う. 基本的には, 商品の販売担当者には UI 上でボタンによる操作を行うが, 最初の (1) 日本酒の銘柄の選択. と最後の (6) 紹介文の修正・保存. は商品の販売担当者に入力してもらう.

4. 開発した生成文評価モジュール

本章では, 節 3.6 で説明した評価モジュールの新たな機能として開発した, 評価機能 (1) 文章の複雑さ. (2) 単語の共起頻度. (3) 特異な表現の使用頻度. について説明する.

4.1 文章の複雑さ

紹介文に対して, 係り受けの多さによる文章の複雑さを基準として評価を行う. 係り受けの多さによる文章の複雑さ [21] というのは, 文章を文節にわけて, 文節毎にその時点まで「係り先が未決定な文節」の係り先を確認していった際に, 「係り先が未決定な文節」の数が最大となっている値を複雑さの値としている. 以下に載せる例文 1 をもとに複雑さについて説明する.

例文 1)

花や果実のような甘い香りが特徴の秋田の日本酒です。

例文 1 を文節に分けると

文節 1)

花や / 果実のような / 甘い / 香りが / 特徴の / 秋田の / 日本酒です。

表 8 文節 1 における文節ごとの係り先と「係り先が未決定な文節の数」

文節番号	文節	係り先の文節番号	係り先が未決定な文節の数
1	花や	2	1
2	果実のような	4	1
3	甘い	4	2
4	香りが	5	1
5	特徴の	7	1
6	秋田の	7	2
7	日本酒です。	-	0

表 9 生成紹介文の複雑さの評価 (日本酒：雪の茅舎)

評価値	紹介文
0.772	雪の茅舎は適度な酸の後味が特徴のお酒。花や果実のような甘い香りが特徴の秋田のひと品です。
0.914	雪の茅舎は花や果実のような甘い香りと適度な酸の後味が特徴のお酒です。秋田生まれのひと品。
1.0	花や果実のような甘い香りと適度な酸の後味を感じられる雪の茅舎。秋田が誇る銘酒です。

という形になります. 先頭から順番に文節を確認した際の, 文節の係り先とその際に「係り先が未決定な文節の数」を表 8 に示す. 表 8 にあるように「係り先が未決定な文節の数は最大で「2」であることがわかる. そのために例文 1 の複雑さは「2」となる. 評価機能の実装には, かかり受け解析器として「CaboCha/南瓜」[30], [31] を使用した. 本研究では評価値として 0 1(1 に近い値ほど優れている) の範囲で値の表現を行うために以下のような計算式で評価値を計算した. まず, 複雑さの値は評価対象が 2 文である際にそのまま複雑さの値を足して評価とすると, 1 文目の複雑さの値 + 2 文目の複雑さの値において, $2+5=6$ と $3+4=7$ といった合計値が等しくなる場合の区別をつけることが出来ない. その区別を可能とするために, n 文で構成される文章において, 複雑さの評価値の合計を変数 c_{all} , x 文目の複雑さの評価値を変数 c_x とおくと, $c_{all} = \sum_{x=1}^n \frac{1}{c_x+2}$ という式で評価値の計算を行った. 分母において 2 を足しているのは複雑さは 1 以上の値であるが, 1 と 2 は逆数とした際に大きな値となり評価値の値が必要以上に大きくなる. そのために 2 を足すことで分母は必ず 3 以上の数になるようにした.

表 9 にシステムが生成した紹介文に対して評価を行った例を示す. 表 9 における評価値は, 生成された紹介文に対する評価値の中で最大値となるものを 1.0 となるように正規化している. 表 9 にあるように実際に評価が高い紹介文は, 係り受けによる文章の複雑さが少なく, 読みやすい文章になっていることを確認できる.

4.2 単語の共起頻度

紹介文に対して, 文章内の単語の共起頻度を元にした評

表 10 単語「香り」との共起頻度の高い単語

	単語	共起頻度		単語	共起頻度
1	味わい	192	8	苦味	100
2	口	180	9	後味	96
3	酸味	154	10	主体	95
4	原料	127	11	特徴	87
5	ふくよか	116	12	華やか	86
6	甘み	110	13	軽快	86
7	爽やか	108	14	旨味	85

表 11 生成紹介文の単語の共起頻度による評価 (日本酒：雪の茅舎)

評価値	紹介文
0.524	お花見などに美味しくいただける日本酒。花や果実のような甘い香りでさらりとした口当たり。
0.677	花や果実のような甘い香りと適度な酸の後味を感じられる雪の茅舎。秋田が誇る銘酒です。
1.0	花や果実のような甘い香りが特徴の秋田の日本酒です。適度な酸の後味が特徴の雪の茅舎。

評価を行う。共起頻度の高い単語組が含まれる文章は、その単語組を同時に含むことが一般的であり、また文章がわかりやすい表現となっていることを示している。そのために紹介文を読んだ人が商品のイメージをより明確に持てるかどうかの指標として使用する。

本研究では、教師データに MeCab[32] による形態素解析を行い、単語の共起頻度に関する辞書データを作成し、その辞書データを元に紹介文における単語の共起頻度による評価を行うことにした。辞書データを作成するにあたり、教師データとしてウェブサイト酒仙人 [29] に掲載されていた 705 種類の日本酒の紹介文 (合計 30574 文字) を使用した。教師データを選択する際には、データ数が多い商品に対する口コミを利用することも考えたが、本研究におけるユーザは商品の販売担当者であり、商品の紹介文を生成しているという点より紹介文を教師データとした。

表 10 に単語「香り」との共起頻度の高い単語上位 14 単語を示す。また、表 11 にシステムが生成した紹介文に対して評価を行った例を示す。表 11 における評価値は、生成された紹介文に対する評価値の中で最大値となるものを 1.0 となるように正規化している。表 11 における評価値が 1.0 となっているものは表 10 において「香り」と共起頻度が高い「後味」「特徴」が含まれていることを確認できる。

4.3 特異な表現の使用頻度

紹介文に対して、文章内での特異な表現の使用頻度による評価を実装した。特異な表現は文章内にあると目立つ言葉であるので、文章を見た人の興味を惹くことが可能である。ただし、人によっては特異な表現は含まず一般的な表現を好む場合があるので、どちらがいいとは断定できない。そのために紹介文における特異な表現による評価を元にユーザが選択する形式をとる。

表 12 生成紹介文の特異な表現の使用頻度による評価 (日本酒：雪の茅舎)

評価値	紹介文
0.102	花や果実のような甘い香りが特徴の秋田の日本酒です。適度な酸の後味が特徴の雪の茅舎。
0.551	秋田産まれのひとつ品です。花や果実のような甘い香りと適度な酸の後味を感じられる雪の茅舎。
1.0	適度な酸の後味が特徴の秋田のひとつ品。初夏でも楽しめて、酔の物などとの相性は抜群です。

本研究では、4.2 節と同様に、単語の出現頻度による評価を行うことにした。教師データに関して 4.2 節と同様にウェブサイト酒仙人 [29] に掲載されていた 705 種類の日本酒の紹介文を使用した。

表 12 にシステムが生成した紹介文に対して評価を行った例を示す。表 12 における評価値は、生成された紹介文に対する評価値の中で最大値となるものを 1.0 となるように正規化している。表 12 における評価値が 1.0 となっているものは「抜群」「ひとつ品」等の特異な表現が含まれていることを確認できる。

5. まとめ

本稿では、紹介文の読みやすさの評価を行うために、新たに生成文評価モジュールの開発を行った。開発した評価モジュールは、日本酒紹介文作成システム内で生成される紹介文に対して、それぞれの評価指標において、読みやすい紹介文を高く評価していることを確認した。

今後の展望としては、他の評価基準として文章に含まれる特徴の組み合わせによる評価を追加したり、日本酒以外の商品の紹介文に対しても、評価を適切に行うことが出来るか検証を行うということがあげられる。

謝辞 本研究は、株式会社セブン&アイ・ホールディングスの協力を受け実施されたものです。情報の御提供ならびに種々の御助言を賜りました。ここに深甚なる謝意を表します。

参考文献

- [1] 経済産業省商務情報政策局情報経済課：令和 2 年度産業経済研究委託事業 (電子商取引に関する市場調査)。(参照 2022-01-27)。
- [2] 総務省情報通信政策研究所：令和 2 年度情報通信メディアの利用時間と情報行動に関する調査報告書。(参照 2022-01-27)。
- [3] Seven & i Holdings Co., L.: セブン&アイの総合通販サイト | オムニ 7。(参照 2022-01-27)。
- [4] 西原陽子, 高山玲央名, 菱田賢祐, 山西良典: コマの配置を用いた争点と玉の危険度の評価による将棋初心者の局面把握支援, 知能と情報, Vol. 30, No. 6, pp. 796-803 (オンライン), DOI: 10.3156/jssoft.30.6.796 (2018)。
- [5] 山下 峻, 藍 圭介, ベンジャミンエバンズ, 棟方 渚, 小野哲雄: 予測変換のアイデアを用いた作曲支援システムの提案, 研究報告音楽情報科学 (MUS), Vol. 2017-MUS-114,

- No. 20, pp. 1–6 (2017).
- [6] Özbal, G., Pighin, D. and Strapparava, C.: BRAINSUP: Brainstorming Support for Creative Sentence Generation, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Sofia, Bulgaria, Association for Computational Linguistics, pp. 1446–1455 (2013).
- [7] Tsai, C.-T., Chen, J.-J., Yang, C.-Y. and Chang, J. S.: LinggleWrite: a Coaching System for Essay Writing, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Online, Association for Computational Linguistics, pp. 127–133 (online), DOI: 10.18653/v1/2020.acl-demos.17 (2020).
- [8] Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural computation*, Vol. 9, No. 8, pp. 1735–1780 (1997).
- [9] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I.: Attention Is All You Need (2017).
- [10] 生田和也, 品川政太朗, 吉野幸一郎, 鈴木 優, 中村 哲: 観光案内におけるニューラル言語モデルを用いた説明文の生成, 人工知能学会全国大会論文集, Vol. JSAI2018, pp. 2K103–2K103 (オンライン), DOI: 10.11517/pjsai.JSAI2018.0.2K103 (2018).
- [11] Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *CoRR*, Vol. abs/1810.04805 (2018).
- [12] Wang, Q., Liu, P., Zhu, Z., Yin, H., Zhang, Q. and Zhang, L.: A Text Abstraction Summary Model Based on BERT Word Embedding and Reinforcement Learning, *Applied Sciences*, Vol. 9, No. 21 (online), DOI: 10.3390/app9214701 (2019).
- [13] Znidarsic, M., Martinc, M., Repar, A. and Pollak, S.: BISLON: BISociative SLOgaN generation based on stylistic literary devices, *Proceedings of the Ninth International Conference on Computational Creativity, Salamanca, Spain, June 25-29, 2018* (Pachet, F., Jordanous, A. and León, C., eds.), Association for Computational Creativity (ACC), pp. 248–255 (2018).
- [14] 平田航大, 横山想一郎, 山下倫央, 川村秀憲: Transformerによる言語モデルを用いた俳句生成とその評価, 技術報告 2, 北海道大学大学院情報科学院, 北海道大学大学院情報科学研究所, 北海道大学大学院情報科学研究所, 北海道大学大学院情報科学研究所 (2021).
- [15] 緑川貴大, 林 篤司, 岩下志乃: 単語の重要度に基づいたニュース記事の自動要約手法, 日本知能情報ファジィ学会ファジィシステムシンポジウム講演論文集, Vol. 35, pp. 694–697 (オンライン), DOI: 10.14864/fss.35.0.694 (2019).
- [16] 晶 中野, 大城小泉, 敏泰松嶋: バイズ符号化アルゴリズムを用いたテキストデータ圧縮, 技術報告 5(2007-AL-110), 早稲田大学理工学部経営システムエ学科, 早稲田大学理工学部経営システムエ学科, 早稲田大学理工学部経営システムエ学科 (2007).
- [17] 吉村枝里子, 渡部広一, 河岡 司: 知能ロボットにおける挨拶文の自動生成方式, 人工知能学会全国大会論文集, Vol. JSAI04, pp. 140–140 (オンライン), DOI: 10.11517/pjsai.JSAI04.0.140.0 (2004).
- [18] 園田亜斗夢: アパレルECサイトにおける説明文自動生成, 経営情報学会 全国研究発表大会要旨集, Vol. 2018t10, pp. 125–127(オンライン), DOI: 10.11497/jasmin.2018t10.0\125 (2018).
- [19] Wang, J., Hou, Y., Liu, J., Cao, Y. and Lin, C.-Y.: A Statistical Framework for Product Description Generation, *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, Taipei, Taiwan, Asian Federation of Natural Language Processing, pp. 187–192 (2017).
- [20] 吉田拓海, 横山想一郎, 山下倫央, 川村秀憲: 競輪における予想記事生成のためのレース結果予測, 情報処理学会論文誌, Vol. 60, No. 10, pp. 1641–1652 (2019).
- [21] HIROSHI, Y., AKIRA, S. and RIN'ICHIRO, T.: Generating Candidates for Rewriting Based on an Indicator of Complex Dependency and Its Application to a Writing Tool, *IPPSJ journal*.
- [22] 尚子宮下, 憲一上條: Evaluation of Understandability : Grammatical structure and recall ratio, 全国大会講演論文集, Vol. 第44回, No. データ処理, pp. 311–312 (1992).
- [23] 庵 愛, 富永敦子, 竹川佳成, 平田圭二: 文章構造を取り入れた文間類似度グラフに基づいた一貫性に関する評価指標の提案, 人工知能学会全国大会論文集, Vol. JSAI2018, pp. 3G201–3G201 (オンライン), DOI: 10.11517/pjsai.JSAI2018.0.3G201 (2018).
- [24] 庵 愛, 竹川佳成, 平田圭二, 寺井あすか: 推敲支援に向けた文章の階層構造を考慮した一貫性に関する評価指標の提案, 日本教育工学会論文誌, Vol. 44, No. 4, pp. 513–525 (オンライン), DOI: 10.15077/jjet.44048 (2021).
- [25] 和代八木下, 純 宗森, 勝 首藤: KJ法文章のVA手法に基づく評価法の提案と実装, 技術報告 23(1996-GN-022), 大阪大学大学院基礎工学研究科情報数理系専攻, 大阪大学大学院基礎工学研究科情報数理系専攻, 大阪大学大学院基礎工学研究科情報数理系専攻 (1997).
- [26] 阿部涼介, 横山想一郎, 山下倫央, 川村秀憲: ECサイト掲載商品の紹介文作成支援, 行動変容と社会システム vol.06 (2020).
- [27] 吉田拓海, 阿部涼介, 横山想一郎, 山下倫央, 川村秀憲: ECサイト掲載商品の紹介文作成支援システムの開発と評価, 第25回デジタルコンテンツクリエーション研究会, Vol. 2020-DCC-25, No. 1 (2020).
- [28] 阿部涼介, 横山想一郎, 山下倫央, 川村秀憲: ECサイト掲載商品の紹介文作成支援システムにおける複数文組合せ機能の開発, 情報処理北海道シンポジウム 2020 講演論文集 (2020).
- [29] 日本酒サービス研究会・酒匠研究会連合会 (SSI) : 日本酒香味評価データベース酒仙人. (URL:http://www.sake-sennin.com/fbo/web/app.php/user/index 参照 2019-10-26).
- [30] 工藤 拓, 松本裕治: チャンキングの段階適用による日本語係り受け解析, Vol. 43, No. 6, pp. 1834–1842 (2002).
- [31] Taku Kudo, Y. M.: Japanese Dependency Analysis using Cascaded Chunking, *CoNLL 2002: Proceedings of the 6th Conference on Natural Language Learning 2002 (COLING 2002 Post-Conference Workshops)*, pp. 63–69 (2002).
- [32] KUDO, T.: MeCab : Yet Another Part-of-Speech and Morphological Analyzer, <http://mecab.sourceforge.net/> (2005).