

# EC サイト掲載商品の紹介文作成支援

阿部 涼介<sup>1,a)</sup> 横山 想一郎<sup>1</sup> 山下 倫央<sup>1</sup> 川村 秀憲<sup>1</sup>

概要：近年の EC サイト市場の拡大に伴い、掲載される商品数は非常に多くなっている。そのため EC サイト担当者の掲載商品の紹介文作成の負担が大きい。本稿では、担当者を想定利用者として、説明文作成の負担を軽減しつつ、より多くの商品に対して高品質な商品紹介文を提供するために商品紹介文作成支援システムの開発を行う。対象の商品として国内で豊富な種類が取り扱われている日本酒を取り上げる。本システムを利用して作成された商品紹介文や担当者の作成した商品紹介文を評価するために、日本酒 3 品の紹介文が掲載されたメールをメーリングリストの登録者 34 万人に配布した。メールに掲載されている商品掲載ページへ遷移するための URL のクリック率に基づく評価を行い、作成支援された紹介文のクリック率が有意に高いことが示された。

## Support for Making Introduction Sentences for Items on E-Commerce Websites

### 1. はじめに

近年、インターネット及びスマートフォンなどの携帯型デバイスの普及に伴って、日本国内の 2018 年度の BtoC における EC 市場規模は約 18 兆円、市場規模が対前年比 8.12 % 増と EC 市場は拡大している。ネットや実店舗を含む全ての商取引の中の EC サイトなどの電子取引の比率を EC 化率として、各産業における E コマースの利用状況を示している。EC 市場の物販系分野においては、EC 化率の伸び率が最も高い分野は「食品、飲料、酒類」である。物流拠点が整備されて、配送サービスが充実することにより、今後も EC 化率が伸びることが期待されている。

EC 市場において、一般顧客を対象とした EC サイトでは、商品名、価格、商品画像、商品紹介文等からなる商品情報が掲載されており、顧客は商品情報を閲覧して商品の購入に関する意思決定を行うことが多い。EC サイト上では、実際に商品を手に取って確認したり、店員に商品情報を確認できないため、商品の特徴を伝えるために商品情報は大きな重要性を持っている。

EC サイトに掲載される商品数の増加に伴って、必要とされる商品紹介文の数が増加するだけでなく、類似商品と

の相違点を明確にするために商品の特徴を記載するといった商品紹介文の質も求められるようになってきた。また、商品紹介文は商品を購入するための EC サイトだけではなく、メールマガジンや SNS などにも掲載されるようになってきている。そのため、ある一つの商品に対して、複数の掲載メディアに対してそれぞれ商品紹介文が必要となってきている。これまでは EC サイトの販売担当者が商品の製造元から情報を集めて、商品の紹介文を作成してきた。しかし、必要とされる商品紹介文の量と質がさらに求められているため、商品紹介文の作成が担当者の大きな負担となることが懸念される。

このような現状を踏まえて、本研究では商品紹介文の作成支援の対象商品として日本酒を選択する。商品紹介文を利用するユーザーのニーズを確認しながら商品紹介文を生成する紹介文作成支援システムの開発を行う。

本稿では、2 章で関連研究について述べる。3 章では、今回紹介文を作成する商品の説明を行う。4 章では、今回作成したシステムの概要の説明した後、各モジュールとユーザーインタフェースの説明を行う。5 章では、支援システムを利用して作成した商品紹介文について本研究で行った実験と得られた結果について述べる。

### 2. 関連研究及び対象商品の選択

本章では、文章自動生成において現在使われている手法

<sup>1</sup> 北海道大学  
Hokkaido University, Sapporo, Hokkaido 060-0814, Japan  
<sup>a)</sup> aber\_23@complex.ist.hokudai.ac.jp

についての説明する。

## 2.1 文章の自動生成の手法

従来行われてきた文章自動生成の方法としては主に

- (1) マルコフ連鎖を用いた文章生成
- (2) あるテキストの自動要約, 圧縮による文章生成 [1]
- (3) 深層学習を用いた文章自動生成
- (4) テンプレート文を用いた文章自動生成 [2]

の4つが挙げられる.[3] 以下ではその4つの手法の概要とメリット, デメリットについて述べていく。

## 2.2 マルコフ連鎖による文章自動生成

マルコフ連鎖とは, 確率分布の一つであるマルコフ過程のうち遷移確率が過去のものに依存せずに, 現在のもののみ依存する系列である.[4] 関連研究としてマルコフ連鎖モデルによって作られた文章と自然な文章の区別を行う研究 [5] が挙げられるように文章の不自然さが目立つ手法である。ある文章「私は 太郎 です。」という文章において, 「私は」という状態の次に「です」という状態が来る確率が  $P(\text{です} | \text{私は})$  で表されるとき,  $P(\text{私は} | \text{私は}), P(\text{太郎} | \text{私は}), P(\text{です} | \text{私は}), P(。 | \text{私は})$  の4つの確率のうち一番高い確率は  $P(\text{太郎} | \text{私は})$  である。このようにしてある単語につながりやすい単語の確率を求めて, この確率に従って生成された文章の良し悪しが決定する。メリットとして単語のつながりはおかしくなりにくく大量の文章データがあればバリエーションも豊富になるが, デメリットとしては文としてつながりがおかしくなりやすいということである。

## 2.3 テキストの自動要約, 圧縮による文章自動生成

テキストの自動要約, 圧縮というのは, ある大きなテキストが合った時にその文章の要約, 圧縮を行い文章を生成するもので, 基本的に元となる文章が存在する.[6] メリットとして文章としての完成度が高いものが出るが, あくまでも要約, 圧縮であるために, 今回のような紹介文を生成する際には, 要約, 圧縮するためのある程度の大きさの文章が必要となる。関連研究としては単語の重要度に基づくテキスト要約 [7] や, PPM 法を利用した日本語テキストの圧縮 [8] などが存在する。

## 2.4 深層学習にを用いた文章自動生成

深層学習を用いた文章自動生成では, 主に深層学習として LSTM[9], RNN[10], CN N[11], GAN[12] などが用いて画像や大量のテキストデータから文章の自動生成を行っている。関連研究では, CNN, RNN, LSTM を用いた感性的な画像説明文生成 [13] や, LSTM を用いた観光案内の説明文の生成 [14] が行われている。メリットとしては画像から文章生成が可能であっ

たり, 大量のテキストデータから未知の言葉の予測を立てることが出来柔軟な表現が可能であるが, デメリットとしては画像や大量のテキストデータが用意できたとしても, 正確さの高い文章を生成することを保証出来ないということである。

## 2.5 テンプレート文を用いた文章自動生成

テンプレート文を用いた文章自動生成では, 予めもしくは生成したテンプレートに, 必要な特徴量を入力することで文章を生成する。関連研究ではテンプレートを用いた新聞記事からの製品情報の抽出 [15] や, アパレル EC サイトにおける説明文の自動生成 [16] が行われている。メリットとしては他の手法では, 保証しきれない文章の正確さが非常に高い文章になるが, テンプレートの使い回しによる似たような言い回しの文章が増えてしまうということである。

上記の4手法の中から商品の紹介文の自動生成方法を選択する。本稿の目的は実際に商品紹介文を作成している担当者の支援が目的である。

そのため, システムが生成する文章は文法的に問題がないことが求められ, また, 商品によっては大量のテキストデータを集めることが難しい商品が存在している。そのため, 本稿ではテンプレート文を使用した紹介文生成を行う。商品によっては画像データがなかったり, 画像から得られる特徴量だけでは十分に商品の説明をすることが困難であるため, 本稿では商品紹介文に掲載する商品の特徴は事前に収集した商品データ群から抽出する。

また先行研究では生成された文章の質の評価は, 文章の各分野における専門家によってされていた。なので本研究では生成された紹介文の評価が顧客から収集したデータを用いて可能かどうか検証を行う。

## 3. 対象商品の選択理由

本研究で商品紹介文作成支援システムの対象商品として日本酒を選択する。日本酒を対象として選択した理由として, 下記の4つが挙げられる。

### (1) EC サイトで掲載される期間が長い

一定期間商品が販売されるために, 一度紹介文を生成すると一定期間実際に掲載されて, 商品に対するサイト利用者の情報収集が行いやすい。

### (2) 画像から特徴量の抽出が難しい(香り, 味など)

日本酒などの飲料や加工済み食品などは見た目よりも味や味覚などの画像からは得ることが出来ない情報が重要であることが多く, そのような情報は画像のみからは抽出が不可能である。

### (3) 日本酒は趣向品であり質が重視される

ティッシュなどの日用品であれば, 質よりも量や値段などが重要となり紹介文の重要度は高くないときが

あるが、日本酒などの趣向品では、基本的に質が重要視される傾向があり、質を表すために紹介文の質も重要となる。

(4) 日本の特産品で国内での種類が豊富であり得られる特徴量も豊富

日本酒は日本各地で多くの種類が生産されており、さらに味、香り、飲み口といった情報の他に、日本酒度、アルコール度などの数値的データも収集することが出来るために、テンプレートを作る際の特徴量候補が多く、様々なテンプレート文を用意できる。

4. 日本酒紹介文生成支援システム

本章では、日本酒紹介文支援システムの説明を行う。まずシステムの概要の説明を行う。次にデータベースについての説明、各モジュールの説明、最後にUIの説明という順番で行う。

4.1 システム概要

まずはじめにシステムの概要から説明する。

今回の日本酒紹介文支援システムでは

- (1) ユーザーによる日本酒の種類を選択  
 例) 八海山 吟醸
  - (2) システムによる日本酒紹介文の生成  
 内容: 選択された日本酒に対して二文からなる紹介文の生成
  - (3) ユーザーによる紹介文の選択及びシステムの評価による紹介文の評価  
 内容: 生成された紹介文にシステムによる評価とユーザーの選択を交えて紹介文候補の選択
- という流れのもとシステムを動かしていく。ユーザーの操作に関してはUI上でを行い、紹介文の生成・評価は日本酒紹介文支援システム上で行う。

次に日本酒紹介文支援システムの概要の説明を行う。

図1に表されているように、今回の支援システムは2つのデータベースと3つのモジュールで構成されている。支援システムの説明は、まず各DBの説明を行い、その後各モジュールの説明という順序で行う。

4.2 データベース

2つあるデータベースのうち、まず先に日本酒情報データベースの説明から行う。日本酒データベースは日本酒ごとに日本酒データカラムに基づいたデータ収集を行い、集まった日本酒データによって構成されている。表1にある日本酒データカラムは、インターネットで取得した日本酒のデータを元に制作された。制作されたデータカラム数は25種類存在し、こちらの方でおおよそ6グループに分けることが出来た。具体的なカラムの内容を表1を元に説明し

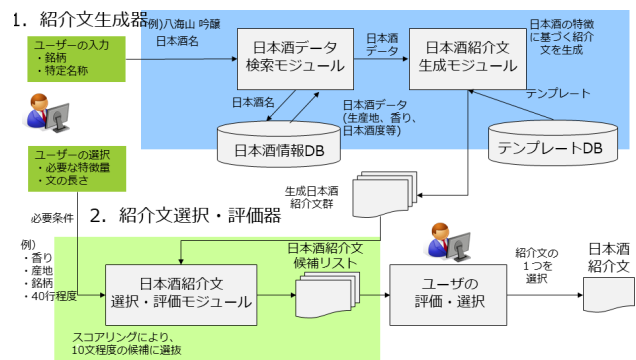


図1 日本酒紹介文生成支援システムの概要

表1 日本酒データカラム

商品情報	温度	製造関連	日本酒に関するデータ
名称	適した温度	蔵元	適した季節
銘柄	燗向け	地域	適した対象
特定名称	常温向け	使用米	適したシーン
説明	冷向け		適した料理
授賞			適した容器
			適したタイミング

数値データ	感覚データ
日本酒度	香り
アミノ酸度	味
酸度	飲み口
アルコール度数	

ていく。

日本酒データカラム説明

- (1) 商品情報  
 商品の名前、説明などのカラム。「名称」は商品としての名前を指しており、「銘柄」は日本酒の蔵元ごとのブランドであり、特定名称は日本酒を製作する際に原料や製造方法によって決定する名前である。「説明」は日本酒の説明文(紹介文)を保持するために用意したカラムである。「授賞」はその日本酒が過去にどのような賞を受賞しているかを表している。
- (2) 温度  
 温度に関わるカラム。「適した温度」をもとに「燗向け」、「常温向け」、「冷向け」のどの飲み方が適しているか判断している。
- (3) 製造関連  
 製造元などのカラム。「蔵元」は日本酒の製造元のこと、「地域」は日本酒が作られている蔵元がある場所(基本的に都道府県)を表している。「使用米」は日本酒を作る際に使われている米の種類のことである。
- (4) 日本酒に適した条件  
 日本酒を飲む際に、適している季節や料理などの情報表している。こちらのカラムは日本酒の専門家たちによって日本酒を飲む際にふさわしいと判断された時期や料理などをまとめたものである。「適した季節」は

日本酒を飲むのに適している季節を主に春夏秋冬で表している。「適した対象」は日本酒を飲むことをおすすめする人のことである。「適したシーン」というのは日本酒を飲むのに適した場所、集まりなどの情報のことである。具体的には花見、パーティー、宴会などである。「適した料理」というのは日本酒と食べるのにふさわしい料理の情報のことである。「適した容器」というのは日本酒を注ぐのに適している酒器の情報のことである。「適したタイミング」というのは日本酒を飲むのにふさわしい時期のことである。主に 食前、食中、食後 の3つである。

#### (5) 数値データ

日本酒度、アミノ酸度などの数値に関わるカラム。「日本酒度」は日本酒の甘さや辛さを示す指標であり、日本酒内にどの程度糖分が含まれているかによって決まる値である。「アミノ酸度」、「酸度」、「アルコール度」はそれぞれ日本酒内のアミノ酸の量、酸の量、アルコールの量を表している。「アミノ酸度」が高ければ濃厚で芳醇、コクが深い味になり、低ければ淡麗でスッキリとした味の日本酒になる。「酸度」が高ければ辛口に感じ、低ければ甘口に感じやすくなる。[17]

#### (6) 感覚データ

香り、味、飲み口の日本酒では欠かせない感覚に関わるカラム。上で述べた数値データと密接に関わりがあるカラムである。製造方法や使用米なども関係がありこのカラムは日本酒カラムの中でも重要なカラムである。次にテンプレートデータベースの説明を行う。

テンプレートデータベースは日本酒紹介文を製作するために用意したテンプレートによって構成されている。今回テンプレートはインターネットで収集したデータを参考にして作成した。テンプレートの元となった紹介文は主に

- 紹介文内に今回のシステムで使用する特徴量が含まれているかどうか
- 汎用性が高く一つの紹介文から複数のテンプレートが作れるかどうか

の2点に注目して手作業で選びだした。結果として総数 234 種類のテンプレートを用意した。今回制作したテンプレートの中には、あるテンプレート内の特徴を変更、削除したものを新たなテンプレートとして加えているため全く異なるテンプレートだけではない

例) テンプレート作成

既存の紹介文: 富山の銘酒、満寿泉の純米大吟醸酒です。

日本酒データカラムに含まれている特徴をテンプレート化

生成されたテンプレート: [地域] の銘酒、[銘柄] の [特定名称] です。

表 2 生成された紹介文の文字数の分布 (日本酒: 雪の茅舎)

文字数	個数
10 以上 ~20 未満	458
20 以上 ~30 未満	23,660
30 以上 ~40 未満	76,614
40 以上 ~50 未満	243,780
50 以上 ~60 未満	296,830
60 以上 ~70 未満	361,802
70 以上 ~80 未満	173,738
80 以上 ~90 未満	82,672
90 以上 ~100 未満	30,854
100 以上 ~110 未満	5,182
110 以上 ~120 未満	368
合計	1,295,958

#### 4.3 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では、ユーザーの入力を受けて決まった日本酒を日本酒情報 DB 上から探しだして、取得した日本酒データを日本酒紹介文生成モジュールに渡し、データが存在しなかった場合はここでエラーを返す。

#### 4.4 日本酒紹介文生成モジュール

日本酒紹介文生成モジュールでは、まず日本酒検索モジュールから受け取った日本酒データと、テンプレートデータベースから受け取った日本酒データに含まれている日本酒データカラムを含むテンプレートを全て呼び出す。次に呼び出されたテンプレートに日本酒データに含まれている特徴を埋め込み 1 文の紹介文を生成。次に作られた 1 文の紹介文を総当たりで組み合わせることで 2 文の紹介文を生成。この際に同じテンプレートやあるテンプレートを元にして作られた類似テンプレートとは組み合わせないようにプログラミングしている。また本研究では紹介文は基本的に 2 文で扱うため、3 文以上の紹介文は生成しないこととする。実際に本日本酒紹介文生成モジュールで 2 文からなる紹介文群生成を行うと、日本酒データカラムで扱っているデータを全て持っている日本酒で 120 万程度の紹介文が生成された。例として雪の茅舎/純米吟醸で紹介文を生成したときの文字数の分布を表 2 に示した。

#### 4.5 日本酒紹介文選択・評価モジュール

日本酒紹介文選択・評価モジュールでは、まずユーザーの選択を元に紹介文の抽出を行う。本研究では紹介文に含まれてほしいカラムと文章の長さをユーザーの手で決めてもらう。紹介文に含まれていて欲しいカラムの設定は 2 パターンあり、指定されたカラムのみか指定されたカラム以外も含んでよいかである。後者の場合は次に説明する、カラムの組み合わせによる評価が加わる。

抽出された紹介文に対して次に紹介する評価基準による評価を加えて、紹介文候補を決定。最後にユーザーによる修正、

表 3 日本酒紹介文解析カラム

商品情報	温度	製造関連	適した状況
名称	適した温度	蔵元	適した季節
銘柄	燗向け	地域	適した対象
特定名称	常温向け	使用米	適したシーン
授賞	冷向け		

数値データ	感覚データ	販売データ	解析データ
日本酒度	香り	値段	ギフト
アミノ酸度	味	麹米	精米
酸度	飲み口	掛米	
アルコール度数		酵母種類	
		酵母	

追加を行ってもらい紹介文の完成となる。

本研究で使用した評価基準は

- 紹介文内の単語の重複率  

$$= (\text{重複単語数}) / (\text{出現単語数})$$
 という式で求められる。重複率が高いと文章内で同じ単語が何度も出て来ているということである。そのため文章量の割に紹介文内で使われている特徴が少なくなり質の高い紹介文とは言えなくなる。今回生成された紹介文内で重複率が高い文章は、似たようなテンプレートを使った文章だったり、紹介文内で使われている日本酒データカラムの特徴が同じものだったりしており、日本語が不自然な文章になっているものが多く見られた。実際に単語重複率を 0.1, 0.2, 0.4 の三段階に分けて紹介文の評価を行った結果 0.2 以下だと日本語がおかしな文章の数は少なかった。ただ、重複率が 0 に近い紹介文内には文章量が少ないために特徴量や単語が重複していないというものが含まれたおり、重複率が低い文章が全て紹介文として質の高い文章という結果ではなかった。先程述べたように単語重複率が 0.4 に近い紹介文内では日本酒の同じ特徴が使われていることがわかる。また単語重複率が 0.1 に近い紹介文の中には紹介文の長さが短いものが含まれている。よって本研究では単語重複率 0.2 以下の紹介文が紹介文候補をユーザーに表示する際に上位に出てくるように設定した。
- 紹介文内で頻出する日本酒データカラムの組み合わせでの評価  
 選択部分で少し述べたがこちらの評価は、紹介文内でよく出てくる組み合わせに対して高い評価を与えるというものである。  
 評価する組み合わせの選定方法は、インターネットに掲載されている日本酒紹介文 97 文に対してこちらで用意した解析カラムを用いて、頻出する組み合わせの調査を行った。使用した解析カラムというのは次に示す表 3 にある 30 種のカラムのことである。

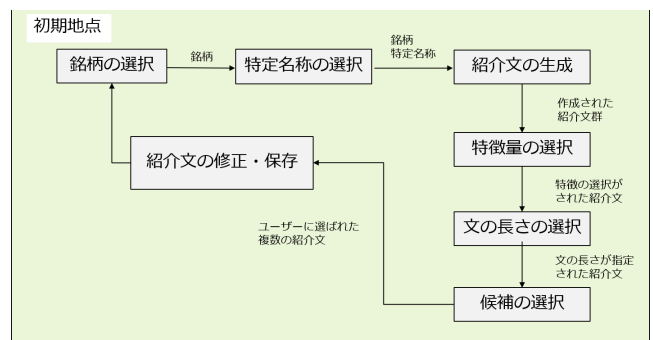


図 2 UI の画面遷移図

本研究ではこの 30 種の解析カラムに関して、インターネットから収集してきた日本酒紹介文内でのカラムに含まれる特徴同士の共起関係を求めた。その結果、収集したデータにおいて日本酒紹介文の頻出の組み合わせとして、味+香り (32/97)、味+特定名称 (44/97)、香り+特定名称 (25/97) が上位であった。本研究ではこの 3 つの組み合わせを含んでいる紹介文が紹介文候補をユーザーに表示する際に上位に出てくるように設定した。

#### 4.6 UI の概要

本研究ではユーザー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。

図 2 の画面遷移図にあるように UI では

- (1) 日本酒の銘柄の選択
- (2) 日本酒の特定名称の確認
- (3) 選択された日本酒の紹介文の生成
- (4) 生成された紹介文から必要な特徴量の選択
- (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

を行ってもらう。

### 5. 生成紹介文の質の検証実験

本章では、今回提案した手法で生成された紹介文の質の検証実験を行いその結果について述べている。

#### 5.1 実験目的

本提案手法で生成された様々な特徴を含む複数の紹介文の質が、人間の手で作られている紹介文と同等の質、もしくはそれ以上の質を有しているのかを検証すること。また紹介文に使われている特徴の組み合わせパターンによって顧客の購買意欲がどの程度変化するのか検証することの 2 つである。

表 4 紹介文生成パターン

パターン 1	特定名称	香り	生産地	
パターン 2	特定名称	味	生産地	
パターン 3	特定名称	味	生産地	適した料理
パターン 4	特定名称	味	生産地	適したシーン
パターン 5	特定名称	味	使用米	
パターン 6	特定名称	味	飲み口	
パターン 7	特定名称	香り	生産地	味
パターン 8	味	香り	生産地	適した料理

## 5.2 実験方法 1

### ● 実験対象

EC サイト販売担当者二名（以後担当者）

### ● 実験内容

日本酒紹介文としての質が EC サイトの利用者に見せることの出来る質を有しているのか検証するための実験。

今回使用する日本酒は担当者と相談して決定した 3 種（雪の茅舎/純米吟醸, 大七/生もと純米, 龍力/大吟醸）を使用。またこの実験では次の実験内容である「紹介文に含まれる特徴と文字数の違いから EC サイトの利用者の購買意欲がどの程度変化するか」を検証するために、8 パターンの「特徴の組み合わせ・文字数」を条件とした紹介文を生成した。評価方法は生成された紹介文に対して担当者からの評価をもらうことで質の検証を行った。表 4 は今回紹介文を生成した特徴の組み合わせパターンである。また紹介文の長さは次の実験の際に配布するメールマガジンに載せるために 40 文字程度で生成した。

## 5.3 実験方法 1 の検証結果

実際に生成した雪の茅舎の紹介文に対して担当者から評価をもらったところ、誤解を生む表現やふさわしくない表現がところどころあるものの、文章としての完成度は高く少しの修正を加えれば実際に紹介文として使用できるという評価であった。また、担当者が修正したものを日本酒の蔵元 4 つにも評価して下さったところ、蔵元からも紹介文としての質は十分に確保されているという評価を得ることが出来た。

## 5.4 実験方法 2

### ● 実験対象

メルマガ会員約 33 万 5 千人。属性として性別、年代、購買実績が利用可能である。

### ● 実験内容

今回日本酒 3 種類に対して、先程の実験で生成した本システムが作った日本酒の紹介文 5 パターンと、人の手で制作した紹介文 1 パターンの系 6 パターンの紹介文を用意し、3 種類の日本酒に対して 6 パターンのうち

表 5 メールマガジ配布者の属性ごとの内訳

性別	年齢	購入実績	対象者数	クリック数	クリック率
男	20 代	なし	8812	11	0.125%
	30 代	なし	11599	22	0.190%
	40 代	なし	24404	86	0.352%
	50 代	なし	28736	196	0.682%
	60 代	なし	13371	196	1.466%
	70 代以上	なし	3544	85	2.398%
	20 代	あり	1966	3	0.150%
	30 代	あり	3972	11	0.277%
	40 代	あり	11072	73	0.659%
	50 代	あり	14240	157	1.103%
60 代	あり	5016	116	2.313%	
70 代以上	あり	733	32	4.366%	
女	20 代	なし	17953	24	0.134%
	30 代	なし	30567	56	0.183%
	40 代	なし	49351	224	0.454%
	50 代	なし	41700	238	0.571%
	60 代	なし	11631	91	0.782%
	70 代以上	なし	2208	19	0.861%
	20 代	あり	2934	6	0.204%
	30 代	あり	8728	18	0.206%
	40 代	あり	19368	92	0.475%
	50 代	あり	19063	169	0.887%
60 代	あり	5069	73	1.440%	
70 代以上	あり	763	14	1.835%	
合計	—	—	336800	2012	0.597%

どれか 1 パターンの紹介文を載せてあるメールマガジンを配布した。メールマガジンからクリックすると EC サイトの日本酒販売のページに行くことが出来るリンクを用意し、配布人数内メールを受け取った人がどれくらいクリックするのか、またクリックした人の中で実際に購買にまで至った人は何名いるのかデータを収集した。今回の実験ではメール配布の対象者数に対して、配布されたメールからクリックして販売サイトに遷移した人の割合であるクリック率を元に、会員の商品に対する興味関心の度合いについて検証した。メルマガ会員の属性の内訳とクリック数、CTR は表 5.4 に記載した。今回得られたデータに対して比率の多重検定であったので、全体における検定と、二群の比率の差の検定としての 2 群の比率の差の場合のフィッシャーの正確確率検定 [18] と同等の結果が得られる Z 検定 [19] を用いた。

## 5.5 実験方法 2 の検証結果

表 4 のデータを元に検定を行った結果次のような結果を得ることが出来た。

- システム作の紹介文と人間作の紹介文での検定多重検定の結果今回パターン別のクリック率においては全体として有意な差はみられないという結果になっ



表 6 パターン別の検定

パターン No	Z	P	有意な差
01	2.762	0.575%	あり
02	1.678	9.344%	なし
03	1.018	30.852%	なし
04	2.571	1.013%	あり
05	2.406	1.613%	あり
システム合計	2.406	1.613%	あり

表 7 パターンごとの検定 (性別)

パターン No	Z	P	有意な差	クリック率が高い対象
00	5.369	0.000%	あり	男性
01	3.940	0.008%	あり	男性
02	5.046	0.000%	あり	男性
03	3.408	0.066%	あり	男性
04	4.340	0.001%	あり	男性
05	3.528	0.042%	あり	男性
システム合計	9.079	0.000%	あり	男性
パターン合計	10.445	0.000%	あり	男性

表 8 パターンごとの検定 (購買経験)

パターン No	Z	P	有意な差	クリック率が高い対象
00	3.392	0.069%	あり	購買経験あり
01	4.730	0.000%	あり	購買経験あり
02	5.622	0.000%	あり	購買経験あり
03	3.710	0.021%	あり	購買経験あり
04	4.131	0.004%	あり	購買経験あり
05	3.927	0.009%	あり	購買経験あり
システム合計	9.903	0.000%	あり	購買経験あり
パターン合計	10.450	0.000%	あり	購買経験あり

た。次に人間作の紹介文とシステム作の紹介文の合計で2群検定を行った。全体として差はないという結果打破あったがパターンごとの2群検定も行った。システム作の紹介文と人間作の紹介文の比較では表6に示されているように、人間作の紹介文よりシステム作の紹介文の方がよりよいクリック率であることがわかったが、やはり各パターン別に比較すると有意な差はあまり認められなかった。

● 属性別の紹介文の検定

性別、購買経験においては表7,8に示されているように双方においてクリック率に対して有意な差が認められ  
 - 男性は女性よりクリック率が高い  
 - 購買経験が[あり]の人は[なし]の人よりクリック率が高い  
 という結果であった。年代については母数の関係で有意な差がほとんど認められなかった。

まず今回の実験の際に収集したデータより、メルマガを用いた顧客データ(今回はクリック率)の収集は十分可能であるということ。またシステム作の紹介文においては人間

作の紹介文と比較して、システム作の紹介文全体でクリック率が高くなるという有意な差を確認することが出来た。なのでクリック率を用いた紹介文の評価は可能であるという結果になった。男女や購買経験の有るかどうかによって興味関心度が変化することも確認できた。したがって本研究で作られたシステム作の紹介文は人間作の紹介文と比較しても十分な質を持っていると考えられ、今後クリック率を用いた評価をシステムに組み込みフィードバックを行うシステムを作成出来る可能性が見受けられた。

参考文献

- [1] 奥村 学, 難波英嗣: テキスト自動要約に関する研究動向(巻頭言に代えて), 自然言語処理, Vol. 6, No. 6, pp. 1-26 (オンライン), DOI: 10.5715/jnlp.6.6.1 (1999).
- [2] MCKEOWN, K.: Generating summaries of multiple news articles, SIGIR '95, pp. 74-82 (online), available from (https://dl.acm.org/doi/10.1145/215206.215334) (1995).
- [3] 太田博三: 文章自動生成における主な手法と独自性に関する検討, 技術報告 23, 放送大学教養学部 (2017).
- [4] フリー百科事典『ウィキペディア (Wikipedia)』: マルコフ連鎖, https://ja.wikipedia.org/wiki/ (Accessed on 01/22/2020).
- [5] 鴨志田芳典, 菊池浩明: 文章合成の不自然さの評価と応用, 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. 26, pp. 245-245 (オンライン), DOI: 10.14864/fss.26.0.245.0 (2010).
- [6] Rush, A. M., Chopra, S. and Weston, J.: A Neural Attention Model for Abstractive Sentence Summarization, CoRR, Vol. abs/1509.00685 (online), available from (http://arxiv.org/abs/1509.00685) (2015).
- [7] 努 平尾, 木谷強: 単語の重要度に基づくテキストの要約, 情報処理学会研究報告データベースシステム (DBS), No. 34(1998-DBS-115) (1998).
- [8] NISHIDA, Y. and IMAI, H.: Japanese Text Compression Using PPM Algorithm, IEICE technical report. Information theory, Vol. 101, No. 177, pp. 19-24 (online), available from (https://ci.nii.ac.jp/naid/110003197757/en/) (2001).
- [9] Hochreiter, S., J. Schmidhuber: Long short-term memory, Neural computation, Vol. 9, No. 8, pp. 1735-1780 (online), available from (https://www.bioinf.jku.at/publications/older/2604.pdf) (1997).
- [10] Mikolov, T., Karafiat, M., Burget, L., Černocký, J. and Khudanpur, S.: Recurrent Neural Network Based Language Model, In INTERSPEECH-2010, (online), available from (https://www.isca-speech.org/archive/archive\_papers/interspeech.2010/i10.1045.pdf) (2010).
- [11] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems 25 (Pereira, F., Burges, C. J. C., Bottou, L. and Weinberger, K. Q., eds.), Curran Associates, Inc., pp. 1097-1105 (online), available from (http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf) (2012).
- [12] Goodfellow, I.: Generative Adversarial Networks, NIPS 2016 Tutorial, (online), available from (https://arxiv.org/pdf/1701.00160.pdf) (2016).

- [13] 三由裕也, 萩原将文: 感情推定を利用した感性的な画像説明文自動生成システム, 日本感性工学会論文誌, Vol. 18, No. 2, pp. 151–158 (オンライン), DOI: 10.5057/jjske.TJSKE-D-18-00071 (2019).
- [14] 生田和也, 品川政太郎, 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/pj-sai.JSAI2018.0\_2K103 (2018).
- [15] IDE, Y., FUJIYOSHI, M., NAGAI, H., NAKAMURA, T. and NOMURA, H.: Template-based Products Information Extraction from Newspaper Articles, *IPSJ SIG Notes*, Vol. 115, pp. 83–90 (online), available from <https://ci.nii.ac.jp/naid/110002934966/en/> (1996).
- [16] 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).
- [17] 沢の鶴株式会社: 日本酒の甘辛度が分かる! 「日本酒度」「酸度」「アミノ酸度」について, <https://www.sawanotsuru.co.jp/site/nihonshu-column/knowledge/nihonshudo-sando-aminosando/>. (Accessed on 1/29/2020).
- [18] (株) データサイエンス研究所: 検定の種類と選択方法, <https://www.statweb.jp/method/sentakuhouhou>. (Accessed on 1/30/2020).
- [19] 2020 統計学用語辞典: 二群の比率の差の検定, <https://www.weblio.jp/content/> (Accessed on 1/30/2020).