オンラインレビューに基づいた観光地のキャッチコピー自動生成

逆瀬川　陽介†　　牛尼　剛聡‡

†九州大学芸術工学部　〒815-8540 福岡県福岡市南区塩原 4-9-1
‡九州大学芸術工学研究院　〒815-8540 福岡県福岡市南区塩原 4-9-1
E-mail:　†sakasegawa.yosuke.203@s.kyushu-u.ac.jp,　‡ushiama@design.kyushu-u.ac.jp

あらまし　近年，観光に関する情報源として，Web 上の観光情報サイトに投稿されたユーザレビューを利用することが一般的になった．しかし，地域の特徴を把握するためにユーザが膨大な数のレビューを確認することは時間がかかり，ユーザは類似したレビューを見返すことになり，必要な情報が効率的に取得できない場合もある．そうした背景の下，観光地の情報をユーザに短くわかりやすく伝えることが期待される．本研究では，ユーザが興味のある観光地に対して，そこに投稿されたレビューから，日本人にとってリズミカルで覚えやすく印象に残りやすい五七五形式のキャッチコピーで観光地の情報を表現する自動生成手法を提案する．提案手法では，まず，各地の観光地のレビューから五七五の音韻の文字列を抽出し，それらを学習データとして深層学習による文章生成の手法である seqGAN によって五七五形式のキャッチコピーを生成する．そして，観光地ごとの特徴的な単語や seqGAN の識別器に基づいてスコア付けを行い，スコアの高いキャッチコピーを提示する．

キーワード　観光情報，レビュー，キャッチコピー，自動生成，575 形式

1. はじめに

近年，観光情報を入手するために，「じゃらん」[1]などの Web 上の観光情報サイトが広く利用されている．観光情報サイトに掲載されているレビューや口コミからは，観光地を実際に訪れたユーザの体験や感想を知ることができ，ユーザが自分の好みに合う観光地を選定するのに役立っている．

しかし，膨大な観光地のレビューや口コミから自分の欲しい情報を取得するには時間がかかることや，調べた情報を覚えられずに何度も見返すことになることが問題である．そこで，レビューや口コミそのものではなく，そこに記載されている特徴が短く印象に残りやすい文章として表現されていれば，欲しい情報を探す時間や思い出す手間を省くことができると期待される．

そこで本研究では，短い文章で観光地の情報が伝わり，かつユーザの印象に残るような文章の提供を目指し，観光地の情報を短くまとめて伝えるキャッチコピーを自動生成する手法を提案する．キャッチコピーの形式には，印象に残りやすくするために，日本で昔から親しまれてきた俳句や川柳といった五七五の音韻的読みやすさを有するようにする．例えば，福岡県にある長浜屋台のレビューから図 1 のようなキャッチコピーが生成されることを想定する．

図 1　本研究によるキャッチコピー生成の概念図

観光地の情報を短くまとめることは，観光マップといった観光地の情報を地図やイラストを使って発信する方法とも親和性が高いため，観光マップとキャッチコピーを組み合わせた観光地の情報発信にも応用可能であると考えられる．

2. 関連研究

2.1 五七五の形式の有用性

五七五の形式を用いるメリットについて，安倍ら[2]は音韻的な読みやすさと記憶の残りやすさという二点を指摘している．

一つ目の音韻的な読みやすさについて述べる．五七五の形式には五七調・七五調と呼ばれる五音節と七音節を基調とした音数の規則がある．五音と七音は，二音一単位の「率拍」という概念に基づくとリズミカルでスムーズであるために心地よいと考えられる．渡辺ら[3]は五七調・七五調のリズムが現代人に対してもリズミカルに感じるか疑問を呈し，実証実験を行った．五七調，七五調が他の調子と比べてリズミカルに感じるか，四八調，六六調，八四調を比較対象として実験を行ったところ，「七五調，五七調，六六調，四八調，八四調の順にリズムが良いと評定される傾向にある」と述べている．また，「五音七音がリズミカルと感じるのは，単になじみのある型の持つ心地よさというだけでなく，五音七音という音の構成がリズム知覚に影響している」と結論づけている．

次に，記憶の残りやすさについて述べる．越場ら[4]は日本人に馴染みのある七五調のリズムを語呂合わせ

に使うことでより効果的な暗記学習が可能になると仮定し実験を行なった．歴史の暗記問題にて，七五調の語呂合わせを使い暗記する場合と語呂合わせを使用しない場合で比較し実験したところ，語呂合わせを使用した場合の方が，平均点が高かったことから，暗記学習に対して七五調の語呂合わせを使うことが効果的であると述べている．

以上より，五七五形式はリズミカルで心地よく，また記憶にも残りやすいという特徴があるため，五七五形式を使った観光地のキャッチコピーは，観光地のレビューを直接読むよりも読みやすく，また観光地の特徴を覚えやすいと考えられる．

2.2 俳句の自動生成手法

近年，機械学習を用いて俳句を自動生成する研究は活発に行われている．ここでは，提案手法に関連する代表的な研究について述べる．

太田[5]は seq2seq を使った深層学習による俳句の自動生成を行い，季語や拍数の素性や季節の制限を取り入れた生成手法を提案している．米田ら[6]は俳句データで学習した LSTM を用いて俳句候補を出力し，入力画像と適合する俳句をマッチングさせる手法を提案している．Wu ら[7]は，RNN，Stacked LSTM，Recurrent-Convolutional NN，seqGAN という 4 種類のディープニューラルネットワークを使い，それぞれのモデルで俳句を文字レベルで学習し生成させ，パープレキシティ(perplexity)に基づいて生成された結果を比較するという実験を行なった．廣田ら[8]は seqGAN を用いた俳句の自動生成を行い，識別器と生成器にそれぞれ異なるデータセットを用いることで性質の違う俳句が生成されることを示した．

本研究では，廣田ら[8]が示したデータセットを変えることによる性質の異なる俳句の生成手法に基づき，seqGAN を用いた五七五形式のキャッチコピーの自動生成を試みる．データセットをそれぞれの観光地によって変えることで，その観光地に合った特徴を持ったキャッチコピーを生成することができると考えられる．

3. アプローチ

本研究では，ユーザが観光地の大量の情報を読む手間や思い出す手間を省くために，観光地の情報を短くまとめて伝えるキャッチコピーを自動で生成する手法を開発することを目的とする．

提案手法では，観光地のキャッチコピー生成には入力データとして観光地のレビューや特徴的な単語を入力し，その観光地の特徴を持ったキャッチコピーを生成する．本手法の応用例として，ユーザは興味のある観光地を入力し，その観光地のキャッチコピーが掲載された観光マップを得られるようなシステムが考えられる．本システムの概要図を図 2 に示す．

図 2　システムの概要図

本論文で提案する手法の流れを以下に示す．

1. 観光情報サイトに掲載されている観光地のレビューを収集する．
2. それぞれの観光地に対するレビューを繋ぎ合わせた文章を 1 文書として，別の観光地の文書との比較によって，tf-idf に基づいて観光地に対する単語の特徴度をスコア付けする．tf-idf の詳細については第 4 章で述べる．
3. 観光地ごとにレビューを形態素解析し，レビューから五七五の音韻となっている文字列を抽出する．これを観光地五七五と呼ぶ．
4. 観光地五七五を学習データとして，seqGAN を用いてキャッチコピーを生成する．
5. 上記の生成結果を形態素解析し，五七五の音韻制約を満足していないキャッチコピーを除外する．
6. 生成したキャッチコピーに対してスコアを付ける．スコアを付け手法として以下の 3 種類を考える．
   (1) キャッチコピーの単語ごとに tf-idf を求めたものを合計した値をスコアとする方法
   (2) 識別器が本物だと判断した確率をスコアとする方法
   (3) 上記(1)のスコアと(2)のスコアをそれぞれ正規化した値の調和平均をスコアとする方法
7. スコアのより高いキャッチコピーを出力し，観光マップに掲載しユーザに提示する．

本手法での観光地五七五の抽出からキャッチコピー出力までの流れを図 3 に示す．

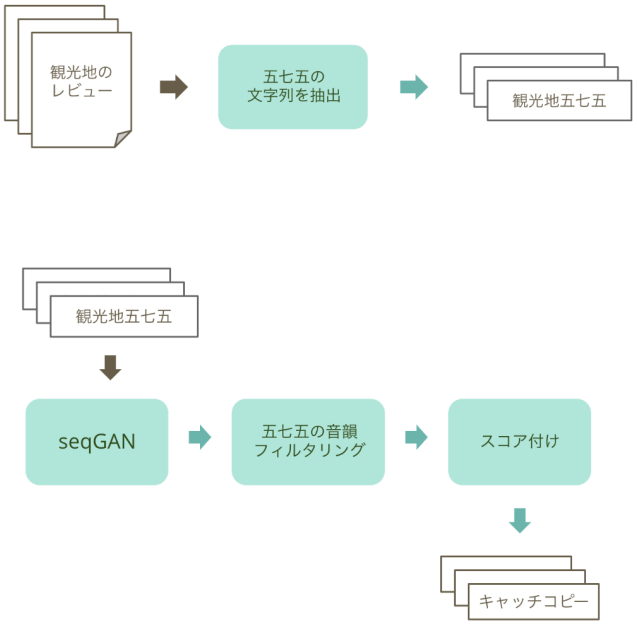

図 3　出力までの流れ

4. 提案手法

4.1 観光地五七五の抽出

キャッチコピーの生成に用いる教師データとして，観光地のレビューから五七五の音韻となっている文字列（観光地五七五）を抽出する．観光情報サイトからスクレイピングにより取得した観光地のレビューに対して形態素解析を行い，観光地のレビューを単語系列に変換する．そして，その系列中に含まれる部分系列に対して，形態素解析器から得られるヨミの情報を利用して，五七五の音韻で並んでおり，かつ五七五の各先頭の単語の品詞が自立語のものを抽出する．これは，五七五の各先頭の品詞を自立語に限定することで自然なキャッチコピーが生成されると考えたためである．表 1 に博多駅を例にレビューと抽出した観光地五七五を示す．

表 1 博多駅でのレビューと観光地五七五

レビュー	観光地五七五
これほど，日本の大都市で交通網の利便性が高い大都市はないです	大都市で 交通網の 利便性
飲食店は駅周辺も含めて常に混んでいるので注意です．	周辺も 含めて常に 混んでいる

4.2 seqGAN によるキャッチコピー生成

4.2.1 seqGAN の概要

GAN は画像生成を目的として提案されたが，Yu ら[10]は GAN を系列データに拡張した seqGAN を提案した．seqGAN は GAN と同様に生成器と識別器を用いて学習させていく GAN の一種である．生成器は LSTM[11]を用いて文章生成を行い，さらに生成器の報酬関数が最大となるように LSTM のパラメータ θ を強化学習によって更新する．識別器は，一般の GAN と同様に CNN を用いて訓練データと生成器が生成したデータを正しく識別するように学習を行う．図 4 にて seqGAN の概要を示す．

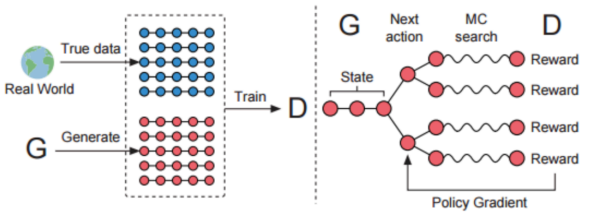

図 4　seqGAN の概念図([10]から引用)

4.2.2 本研究における seqGAN の概要

本研究では，観光地のレビューに含まれる五七五の音韻を有する文字列である観光地五七五を正例のデータとして，観光地五七五に近いキャッチコピーを生成するように学習を行う．本研究にて使用する seqGAN の概要を図 5 に示す．

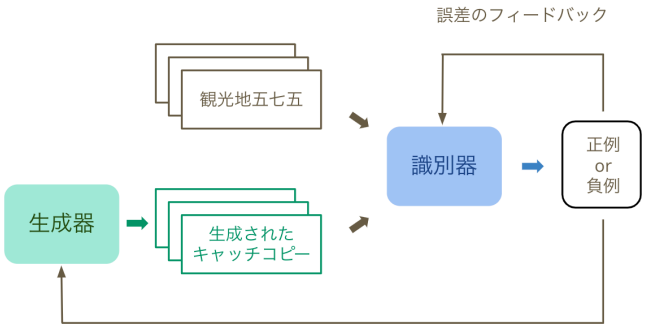

図 5 本研究の seqGAN の概要

4.2.3 データセットを分けた seqGAN の学習

廣田ら[8]は，seqGAN を用いた俳句生成の研究にて，seqGAN の生成器と識別器に与えるデータセットを変えることで，生成される俳句の性質を変えることができることを示した．これをもとに，本研究においても生成器と識別器に与えるデータセットを変えて生成を試みる．

本研究では，生成器に与えるデータセットとして，

キャッチコピーの生成対象の観光地以外も含む全ての観光地のレビューから抽出した観光地五七五を用いる．これは，生成器に与える語彙数を増やすことで，多様な表現のキャッチコピーが生成されると考えられるためである．一方で，識別器に与えるデータセットの観光地五七五はキャッチコピー生成対象の観光地のレビューのみから抽出した観光地五七五を用いる．GAN の性質上，生成器は，識別器を騙すように学習を進めるため，識別器に与えられた正例のデータセットと出来るだけ類似した文を生成できるように学習を行う．したがって，識別器に対象の観光地のデータセットを用いることで，生成器は対象の観光地の性質を持ったキャッチコピーが生成されると考えられる．

4.3 五七五の音韻フィルタリング

seqGAN によって生成されたキャッチコピーに対して，五七五の音韻の制約に基づいたフィルタリング処理を行う．生成されたキャッチコピーに対して形態素解析を行い，レビューから観光地五七五を抽出した際と同様に五七五の音韻になっているか，五七五の各先頭の単語の品詞が自立語であるかという基準でフィルタリングを行う．

4.4 観光地の特徴抽出

本研究では，観光地の特徴を抽出するにあたって，観光地のレビューに対して tf-idf を用いて特徴的な単語を抽出する．tf-idf とは，文書内に出現する単語について，文書の出現頻度(tf)と逆文書頻度(idf)からその単語の重要度を算出する手法である．ある文書d_jに出現する単語t_iについて考える場合，出現回数をfとするとき，tf 値を式(1)により求める．

$$tf(t_i, d_j) = \frac{f(t_i \in d_j)}{\sum_{t_k \in d_j} f(t_k \in d_j)} \quad (1)$$

また，ある文書集合における単語t_iについて考える場合，総文書数をN，$df(t_i)$を単語t_iが出現する文書数とするとき，idf 値を式(2)により求める．

$$idf(t_i) = \log\left(\frac{N}{df(t_i)+1}\right) \quad (2)$$

そして，上記で算出した tf 値と idf 値を掛け合わせることで，式(3)に示す式に基づいて単語ごとの重要度を計算する．

$$tfidf(t_i, d_j) = tf(t_i, d_j) \cdot idf(t_i) \quad (3)$$

例として図 6 に福岡エリアのレビューから，特徴的な単語の重み付けを行う流れを示す．この例では，福岡エリアのレビューと他の観光地のレビューを比較して単語ごとに tf-idf を求めている．また，福岡エリアのレビューから抽出した単語と tf-idf の表を見ると，福岡エリアのレビューには頻繁に出現するが他の観光地のレビューにはあまり出現しない単語の tf-idf の値が高くなっていることが示されている．そして，それらの単語が福岡エリアの特徴を表した単語だと考えることができる．

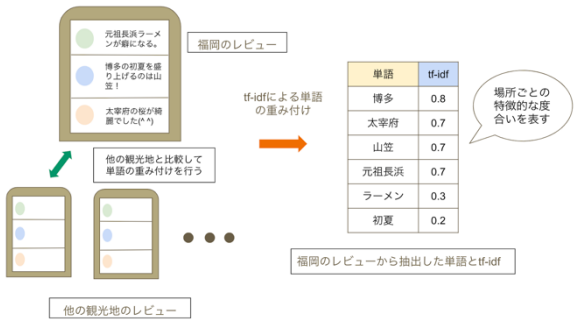

図 6　福岡のレビューを例にした，tf-idf による観光地の特徴的な単語の抽出

4.5 生成したキャッチコピーに対するスコア付け

4.5.1 単語ごとの tf-idf によるスコア付け

後処理をしたキャッチコピーに対して，tf-idf による観光地ごとの単語の重みを使い，キャッチコピーのスコア付けを行う．生成されたキャッチコピーに対して形態素解析を行い，含まれている名詞を抽出し，それぞれの名詞に対する tf-idf の値の合計を算出する．ここで得られた合計値を，キャッチコピーに対する観光地の特徴度として，スコア付けを行う．具体的には，キャッチコピーc_iに対するスコアを$S(c_i)$とし，対象の観光地のレビューを一つの文書としてまとめたものである文書d_jに出現する単語t_iに対する tf-idf の値を$tfidf(t_i, d_j)$とすると，$S_{tfidf}(c_i)$を式(4)のように定義する．

$$S_{tfidf}(c_i) = \sum_{t \in c_i} \mathrm{tfidf}(t_i, d_j) \quad (4)$$

例として，図 7 に生成された福岡エリアのキャッチコピーに対して tf-idf をもとにスコア付けを行う流れを示す．

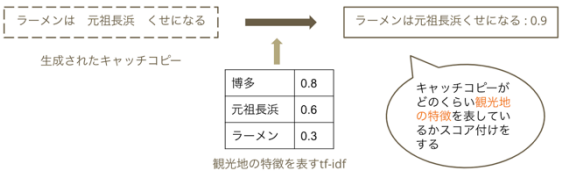

図 7　キャッチコピーに対するスコア付け

4.5.2 seqGAN の識別器を使ったスコア付け

本研究で利用する seqGAN では，生成器が生成したキャッチコピーに対して，識別器はそれが本物であるかの確率を算出する．この時，識別器が高い確率を出すと，生成器は識別器を騙すことができており，より精度の高い自然なキャッチコピーを生成できていると考えることができる．したがって，生成されたキャッチコピーに識別器が算出した確率をスコアとすることで，キャッチコピーに対する自然な文章度合いを数値化することができる．ここで，キャッチコピーc_iの文書の自然さを表すスコアを以下の式で定義する．

$$s_{natural}(c_i) = D(c_i) \tag{5}$$

ここで，$D(c_i)$はc_iを識別機に与えた際に返す確率の値である．

4.5.3 tf-idf と識別器の両方を使ったスコア付け

上記の生成したキャッチコピーに対する tf-idf を使ったスコアと識別器の確率を使ったスコアに対してそれぞれ正規化を行い，２つのスコアの調和平均を最終的なスコアとする．これよって，観光地の特徴と文章の自然度の両方を反映させたスコア$S_{total}(c_i)$を算出することができる．$S_{total}(c_i)$は以下の式で定義される．

$$S_{total}(c_i) = \frac{2}{\left(\frac{1}{S_{tfidf}(c_i)} + \frac{1}{S_{natural}(c_i)}\right)} \tag{6}$$

5. 実験

5.1 実験手法

5.1.1 データセット

本研究では，観光地のレビューを入手するにあたって，日本における代表的な観光情報サイトの一つであるである「じゃらん」に投稿されたレビューを使用した．観光地には，福岡の代表的な観光地である博多駅，マリンワールド海の中道，太宰府天満宮，大濠公園，福岡 PayPay ドーム，海の中道海浜公園の 6 箇所を対象に，それぞれの観光地から 1000 件のレビューを取得し，それらに基づいて観光地五七五を抽出した．各観光地にて抽出できた観光地五七五の数を表 2 に示す．

表 2　観光地ごとの観光地五七五の総数

観光地	観光地五七五の数
博多駅	332
マリンワールド海の中道	327
太宰府天満宮	387
大濠公園	312
福岡 PayPay ドーム	309
海の中道海浜公園	214

5.1.2 tf-idf による観光地の特徴抽出

tf-idf を計算する際には，オープンソースの日本語形態素解析エンジンである MeCab[12]を用いた．MeCab の辞書には，固有表現や新語などに対応できる mecab-ipadic-NEologd[13]を用いた．

観光地の例として博多駅において，tf-idf によって得られた特徴的な単語と，その tf-idf の値の上位 10 件を表 3 に示す．

表 3　博多駅の特徴的な単語と tf-idf

単語	tf-idf
新幹線	0.018
お土産	0.018
阪急	0.014
駅ビル	0.013
買い物	0.009
デパート	0.009
アミュプラザ	0.009
博多口	0.006
地下鉄	0.005
マルイ	0.005

5.2 キャッチコピーの生成

5.2.1 キャッチコピーの生成条件

キャッチコピー生成の際には，生成器と識別器に別々のデータセットを用いた．生成器には全ての観光地の観光地五七五を与え，識別器には対象とする観光地の観光地五七五を与えて学習させた．

事前学習にて，生成器では，LSTM を使った文章生成を 300 epoch 学習させ，その後，識別器に生成器が生成させたデータと訓練データの識別を 15 epoch 学習させた．敵対的学習時には 300 epoch の学習を行い，1 epoch に付き生成器は 5 回，識別器は 1 回の学習と

パラメータの更新を行った．生成器の学習の割合を増やした理由は，実験をしていく中で識別器の精度が上がり過ぎており生成器の学習が足りていないと考えたからである．最適化手法には，事前学習時，敵対的学習時共に Adam を用いた．識別器の損失関数には categorical_crossentropy を用いた．seqGAN の学習時の識別器の accuracy と loss（損失関数の値）のグラフを図 8,9 に示す．

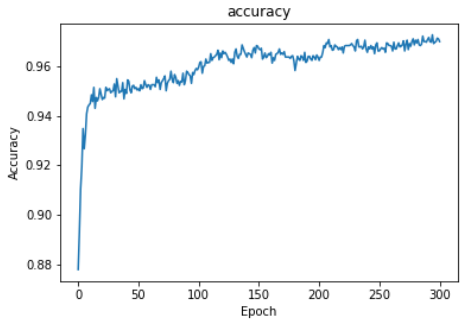

図 8 識別器の精度のグラフ

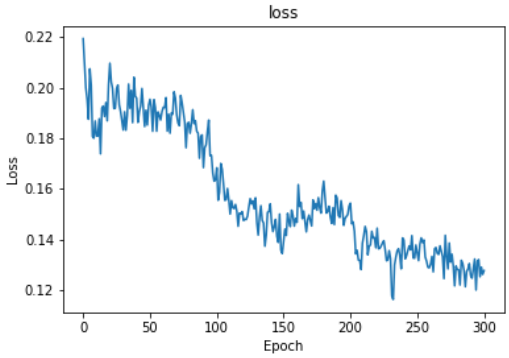

図 9 識別器の損失関数のグラフ

5.2.2 キャッチコピーの生成例

生成されたキャッチコピーとそのスコアを表 4 に示す．score_1 は単語ごとの tf-idf の合計，score_2 は識別器が本物だと判別した確率，score_3 は score_1 と score_2 をそれぞれ正規化した値の調和平均を示している．

表 4　博多駅のキャッチコピーの生成例とスコア

生成例	score_1	score_2	score_3
九州の 玄関口の お土産に	0.006	0.611	0.011
沢山の 商業施設 ビルがある	0.813	0.128	0.222

実際に生成されたキャッチコピーを使用し，地図上にマッピングした例を図 10 に示す．

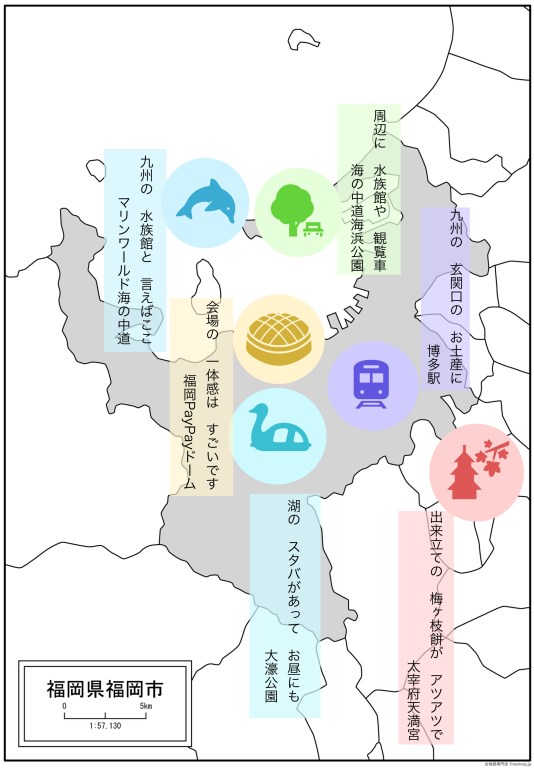

図 10 生成したキャッチコピーを地図にマッピングした例

5.3 評価実験

5.3.1 評価方法

本研究では，出力されたキャッチコピーに対するスコア付けの手法に対して， DCG および NDCG を用いて評価実験を実施した．DCG はランキングによる順位の再現度を表す評価指標であり，NDCG は得られた DCG を理想ランキングの DCG で割った値である．Järvelin ら[14]が定義した DCG は式(7)で表される．ここで，rel_iはランキング中のi番目のスコアを表す．

$$DCG = rel_1 + \sum_{i=2}^{k} \frac{rel_i}{log_2 i} \tag{7}$$

また，理想的な DCG の値を ideal_DCG とすると，NDCG の値は式(8)で表される．

$$NDCG = \frac{DCG}{ideal\_DCG} \tag{8}$$

本研究の評価実験には，博多駅と太宰府天満宮の二つを対象にそれぞれ 20 件の生成したキャッチコピーに対して，以下の 5 つの項目で 5 段階評価を行った．

1. 印象に残りやすいか
2. 読みやすいか
3. 観光地の特徴を表しているか
4. 文章は自然か
5. 観光地に興味を持ったか

「読みやすいか」と「文章は自然か」の質問は，文章として違和感がなく生成できているかを調べるための質問である．「印象に残りやすいか」と「観光地の特徴を表しているか」と「観光地に興味を持ったか」の質問は，観光地のキャッチコピーとして情報を伝えることができており，魅力的なキャッチコピーとして成り立っているかを調べるための質問である．

評価実験に使用したキャッチコピーを表 5,6 に示す．

表 5 評価実験に使用した博多駅のキャッチコピー

今回は実家にあげるお土産屋
沢山の商業施設ビルがある
２．３杯飲み一階のお土産屋
お土産を買えるお店も多数あり
長崎のカステラとかをお土産に
お土産を買い忘れても大丈夫
お土産や飲食店が充実し
お土産を買うのにとても便利です
長崎のカステラとかをお土産屋
今回は実家にあげるお土産が
お土産を飲食店が充実し
九州の玄関口のお土産屋
お土産も多くなじみのお土産屋
中国や韓国からの旅行者の
お土産を新幹線で JR
九州の玄関口のお土産に
お土産や買い忘れます JR
スムーズに電車が利用出来ました
お土産や飲食店も充実し
今回はなんでもここで揃います

表 6 評価実験に使用した太宰府天満宮のキャッチコピー

出来立ての梅ヶ枝餅がアツアツで
焼きたての梅ヶ枝餅を食べながら
名物の梅ヶ枝餅を焼き立てで
死を遂げた天神さんの荒魂を
定番の梅ヶ枝餅を食べながら
現代の梅ヶ枝餅を食べながら
学問は全く関係ないのです
京都にも天満宮はありますが
朝早くお参りそして九国で
お守りや鉛筆なども充実し
素晴らしい神社さなあと思いまし
子連れにはもってこいだと思います
海外の観光客が多いので
交通の便もよく行きやすかった
してるって気持ちになれる初詣
知り合いが国家試験を受ける際
観光が楽しくなると思います
韓国の観光客が多かった
ドライブの途中一年半振りに
韓国や中国からの訪問者

5.3.2 評価結果

単語ごとの tf-idf の合計のスコアに基づいた出力の手法を手法 1，識別器の確率に基づいた出力の手法を手法 2，手法 1 と手法 2 のスコアをそれぞれ正規化した総合のスコアに基づいた出力の手法を手法 3 として，博多駅と太宰府天満宮でのキャッチコピーに対する各手法，アンケートの各項目の NDCG の値を表 7,8 に示す．

表 7　博多駅のキャッチコピーに対する NDCG

	手法 1	手法 2	手法 3
印象に残りやすいか	**0.958**	0.909	0.885
読みやすいか	**0.946**	0.927	0.924
観光地の特徴を表しているか	**0.953**	0.902	0.906
文章は自然か	0.869	**0.901**	0.880
観光地に興味を持ったか	**0.955**	0.903	0.890

表 8　太宰府天満宮のキャッチコピーに対する NDCG

	手法 1	手法 2	手法 3
印象に残りやすいか	**0.979**	0.881	0.960
読みやすいか	**0.979**	0.905	0.961
観光地の特徴を表しているか	**0.990**	0.846	0.959
文章は自然か	**0.976**	0.914	0.953
観光地に興味を持ったか	**0.900**	0.761	0.873

5.3.3 考察

生成されたキャッチコピーに対する NDCG の結果では,「印象に残りやすいか」,「観光地の特徴を表しているか」,「観光地に興味を持ったか」の 3 つの項目に関して単語ごとの tf-idf の合計に基づいた手法の NDCG の値が他の手法の値より上回っていた．これは，tf-idf に基づいて出力しているため観光地の特徴的な単語をキャッチコピーに含んでおり，観光地の魅力を表現できていたからだと考える.「読みやすいか」と「文章は自然か」の項目に関しては，手法 1 と手法 3 の値が手法 2 の値より上回っていることが多かったが，どの手法が適しているかを明示することはできなかった．しかし，NDCG の値が高い場合が多いため，全体的に読みやすく自然なキャッチコピーが生成できていると考えられる．

また，生成実験を通して，生成手法にて単語毎の tf-idf の合計のスコアを用いた場合,似たような内容のキャッチコピーが多数出力されており，一方で識別器の確率を用いた場合は，バラエティに富んだ内容のキャッチコピーが出力された．したがって，2 つの生成手法を上手くパラメータを調整しながら組み合わせることで理想的な内容のキャッチコピーを生成できると考えられる．

6. まとめ

本研究では，観光情報サイトのレビューを使用して seqGAN によるキャッチコピーの自動生成手法を提案した．また，生成されたキャッチコピーに対して，単語ごとの tf-idf の合計によるスコアに基づいて出力すると，印象的で観光地の特徴や魅力を持ったキャッチコピーが生成できることが示された．

今後の課題として，データセットに韻を踏んだ詩的な文章を用いることや，生成されたキャッチコピーの単語ごとの tf-idf と識別器の確率の両方を上手く組み合わせるなどで，理想的なキャッチコピーが生成できるよう改善していきたい．また，最終的に出力したキャッチコピーを観光マップとしてユーザに提示するシステムを開発し，それを用いた運用を行いながらさらなる評価・分析に取り組みたい．

謝辞

本研究は JSPS 科研費 19H04219 の助成を受けたものです．

参 考 文 献

[1] じゃらん HP, https://www.jalan.net/
[2] 安倍文紀, 寺田実, ”575 の音韻的読みやすさを付与した学術論文の要約文自動生成手法” DEIM Forum 2018 E3-1
[3] 渡部涼子, 小磯花絵, ”五七調・七五調のリズム知覚に関する予備的研究” 言語処理学会第 20 回年次大会発表論文集 2014
[4] 越場千絵, 鈴木克明, 藤原康宏, 市川尚, ”暗記学習のための語呂和歌作成支援システムの開発” 卒業論文 熊本大学 2004
[5] 太田 瑶子 ”深層学習による俳句の自動生成” 修士論文 奈良先端科学技術大学院大学 2018
[6] 米田航紀, 横山 想一郎, 山下 倫央, 川村 秀憲, ”LSTM を用いた俳句自動生成器の開発” 人工知能学会 2018
[7] Xianchao Wu, Momo Klyen, Kazushige Ito, Zhan Chen. “Haiku Generation Using Deep Neural Networks”, 2017.言語処理学会
[8] 廣田敦士, 岡 夏樹, 荒木 雅弘, 田中 一晶, ”学習データセットを分けた seqGAN による俳句生成” 言語処理学会第 24 回年次大会発表論文集 2018
[9] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza,B. Xu, D. Warde-Farley, S.Ozair, A. C. Courville, and Y. Bengio. “Generative Adversarial Nets.” InProceedings of NIPS,pp. 2672–2680, 2014.
[10] Lantao Yu,weinan Zhang, Jun Wang, Yong Yu. “SeqGAN: Sequence Gener-ative Adversarial Nets with Policy Gradient.” InAAAI 2017.
[11] Hochreiter, S. and Schmidhuber, J “ LongShort-Term Memory”, Neural Computation, Vol. 9, No. 8, pp.1735–1780, 1997
[12] MeCab http://taku910.github.io/mecab/
[13] “ mecab-ipadic-NEologd : Neologism dictionary for MeCab, ” https://github.com/neologd/mecab-ipadic-neologd
[14] K. Järvelin and J. Kekäläinen, “Cumulated Gain-based Evaluation ofIR Techniques,” ACM Transactions on Information Systems (TOIS),vol.20, no.4, pp.422–446, 2002.