

推薦論文

口コミから美味しい料理店を手早く探すシステム

市村 哲^{1,a)}

受付日 2020年2月29日, 採録日 2020年8月7日

概要: グルメサイトを利用する際、他の人が書き込んだ感想（通称、口コミ）を参考にすることが多い。実際に食したであろう人たちの料理の感想は非常に有用であり、店舗から提供される写真や文面以外の多くの情報を知ることができる。しかしながら、グルメサイトに登録されている飲食店は非常に多く、かつ、飲食店について書き込まれた口コミ情報も大量であることが普通である。著者らは、美味しい料理が食べられる料理店を手早く探すことができる Web サービス「食探」を開発した。ユーザがどのような料理が食べたいかに関する自分の希望や質問を日本語自然文で入力すると、入力文に類似した口コミが多い飲食店を抽出して、さらに、レビュー文章から抽出した料理に関する要約文を作成して表示する。本論文では、料理に関する要約文を作成する方法、および、手早く料理情報を見つけられる要約文とはどのようなものかをユーザ実験を通して検証した実験結果について主に報告する。提案手法で生成した要約文は、従来手法で生成したものに比べ満足度が高いことなどから有効性が認められた。

キーワード: 推薦, 機械学習, 自然言語処理, 口コミ解析

Searching for Delicious Restaurants Using Review

SATOSHI ICHIMURA^{1,a)}

Received: February 29, 2020, Accepted: August 7, 2020

Abstract: When using gourmet site, we often refer to impressions written by others (user reviews for short). The impression of the cooking of the people who actually would have eaten is very useful and it is possible to know a lot of information other than photographs and texts provided from the restaurant. However, there are a lot of restaurants registered in the gourmet site, and it is also common that a large amount of reviews is written on each restaurant. Focusing on this problem, we have developed a web service “Syoku-Tan”, food explorer in English, that can allow you to quickly find a restaurant where you can eat delicious food. When a user inputs his or her wishes and questions about what kind of dish they want to eat in Japanese natural sentences, the system detects user reviews similar to the user’s input sentence, and find restaurants suitable to the user. Furthermore, instead of displaying a long review sentence, the system creates a short summary extracted from the long review sentence, and displays it on the screen. This paper mainly reports the results of an experiment on what kind of summary can allow you to quickly find a delicious restaurant. As a result of the experiments, it was found that the summary generated by the proposed method was more satisfactory than that of the conventional method.

Keywords: recommendation, machine learning, NLP, analysis of user reviews

1. はじめに

インターネットで食事に関する情報を収集したり発信したりすることが一般的になっている。たとえば飲食店を検索することができるサイト（通称、グルメサイト）では、

ユーザはまず場所や料理ジャンルなどの条件を入力して料理店を検索するが、それと同時に検索された料理店について他の人が書き込んだ感想（通称、口コミ）を参考にすることが多い。実際に食したであろう人たちの料理の感想

¹ 大妻女子大学社会情報学部情報デザイン専攻
Chiyoda, Tokyo, 102-8357 Japan

^{a)} ichimura@otsuma.ac.jp

本論文の内容は 2019 年 7 月のマルチメディア、分散、協調とモバイル (DICOMO2019) シンポジウムで報告され、グループウェアとネットワークサービス研究会主査により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である。

は非常に有用であり、店舗から提供される写真や文面以外の多くの情報を知ることができる。また飲食店を利用した後、その店の感想を口コミとして書き込むことも一般的になっている。料理についての質問を書き込むと他の人から有益な回答を得られることが多い。その回答もグルメサイトの口コミ情報として活用されることとなり、そうした口コミを利用して実際に行く店を決めるユーザは多い。

しかしながら、グルメサイトに登録されている飲食店は非常に多く、かつ、飲食店について書き込まれた口コミ情報も大量であることが普通である。実際、書き込みが多い店では数百件～数千件もの書き込み数になる場合もある。数件程度の書き込みであれば書き込みを1つ1つ見ていくこともできるが、数百件ともなると読むには膨大な時間がかかってしまう。そのため目についた口コミをいくつか読むだけで店やその店の料理の良し悪しを判断せざるを得ないことが多い。また口コミには多種多様な情報が含まれており、行きたい店を見つけるのに不要な情報も多い。自分が知りたい内容にたどり着くにはなおさら多くの時間がかかってしまう。

以上の問題に着目し、グルメサイトに登録された口コミ情報を利用し、効率的かつ効果的に希望する飲食店または口コミ情報を見つけることができるシステムを構築することを目的とした。そして特に本研究においては、美味しい料理が食べられる料理店を手早く探すことができるようにすることを目標に定めた。その店の料理が美味しいかどうかは、特に口コミ情報に頼る場合が多いと思われる。店員に面と向かっていうことが少ない料理の味付けや素材の鮮度に関する感想が口コミ文章のなかに含まれていることもしばしばあり、料理の味については店から提供される情報より、口コミ情報の方が信頼できる可能性が高い。

以上の目標に沿い、Web サービス「食探（しょくたん）」を開発した。食探は、口コミデータセットを所有するグルメサイト、グルメサイトの口コミデータセットを利用できる提携サイト、または、オープンデータとして公開されているグルメサイトの口コミデータセット（例：Yelp Open Dataset [1]）を利用するサイトにおいて運用されることを想定した Web サービスである。食探が導入されたサイトを利用する際、ユーザは通常のグルメサイトに入力するような地域や料理ジャンルなどの基本条件を指定することに加え、どのような料理が食べたいかに関する、自分の希望や質問を日本語自然文で入力する。すると食探は、あらかじめ機械学習しておいた口コミデータセットから、ユーザの入力文に類似した口コミ文章を検索するとともに、そのような口コミが多い飲食店を抽出する。

なお食探の検索機能については、口コミデータセット内の全文章のベクトル分散表現を作成しておき、ユーザが入力した日本語自然文のベクトルと近い文章を多数有する店を検索する手法であり、いくつかの関連研究 [2], [3] にも備

わっている機能である。よって本論文では、検索された飲食店について書き込まれた膨大な口コミ情報から料理に係る口コミ文章を優先的に抽出し、かつ、長い口コミ文章をそのまま表示する代わりに、料理に関する評価文を要約して表示する手法について提案する。そして、料理に関する要約文を作成する方法、および、手早く料理情報を見つけられる要約文とはどのようなものかをユーザ実験を通して検証した実験結果について主に報告するものである。評価実験では、口コミ文章をすべて表示する従来手法と、口コミ文章を要約して表示する3提案手法（要約表示、補足表示、抽出文表示）を比較した。

2. 関連研究

グルメサイトには大量の口コミ情報が掲載されているが、それがゆえに必要な情報にたどり着くのに時間がかかる問題がある。この問題はインターネット上の多種サービスにおいて発生しているが、多くの場合ユーザに検索機能を提供することでこの問題に対処している。

大山ら [2] は、ゲームレビューサイトに登録されている大量のレビュー文章を Word2Vec [27] を用いて機械学習し、ユーザがキーワードを入力した際に、類似度が高いゲームを検索して出力するシステムを提案している。Word2Vec は大量のテキストデータを解析して単語をベクトル化する手法である。単語をベクトル化することで、ベクトル間の距離を比較して、その単語に近い単語、すなわち類似した単語を出力することや、単語と単語の類似度を求めることができる。一般的に Word2Vec は入力層、中間層、出力層からなるニューラルネットワークで構成されている。

栗原ら [3] は、映画レビューサイトに登録されている大量のレビュー文書を自然言語検索するために、Word2Vec と Doc2Vec [25], [26] をあわせて用いる手法を提案している。Word2Vec が入力として単語列を受け取るのに対し、Doc2Vec は単語列と文書 ID をあわせて受け取るため、単語のベクトル化だけでなく任意の長さの文章のベクトル化が可能である。栗原らは映画ごとにレビュー文章をまとめて Doc2Vec に入力して機械学習するとともに、Word2Vec を用いて質問文に類語を追加してクエリ拡張している。これによりクエリ文に記述した語句が入っていないような場合でも、意味的に近い口コミ文を検索できるようになったと報告している。Doc2Vec も入力層、中間層、出力層からなるニューラルネットワークで構成されている。

谷本ら [4]、平山ら [5] および中野ら [6] は、通販サイトにおける大量の投稿レビューを自然言語処理によって解析し、各ユーザの関心事（商品のどの機能を重視するか）ごとに整理して表示する方法を提案している。係り元（ユーザの意見）および係り先（商品の機能）のペアを抽出し、どの意見がどの機能に対する意見なのかをグラフ表示する機能や、ユーザの意見がポジティブなものかネガティブな

ものかを日本語評価極性辞書を用いて解析しその統計情報をグラフ表示する機能を実装している。また紀本ら [7] は映画レビューサイトを対象として本手法を適用している。

同様に Suresh ら [8] は、グルメサイトにおける大量の投稿レビューを自然言語処理によって解析し、各ユーザの関心事（食事のおいしさ、店の雰囲気、価格帯など）に応じて検索結果の優先度を変更して表示する手法を提案している。英語版の形態素解析ツール NLTK (Natural Language Toolkit) [9] を用いて各単語の品詞、および、評価極性（レビュー文章がポジティブなものネガティブなものか）を解析し、検索時に利用している。

加えて Choi ら [10] は、グルメサイトに登録されたレビュー文章を対象とした検索機能を提案している。システムはレビュー文章中の文の係り受け構造に着目して係り元（形容詞：美味しい、安い、広い）および係り先（名詞：ハンバーガー、価格、駐車場など）のペアを抽出して記憶する。そうしてたとえば、ユーザが「美味しいハンバーガー」というクエリを実行すると、ハンバーガーという名詞と美味しいという形容詞を含むレビュー文章を多数有する店を検索して表示する。クエリとの類似度、レビューの評価、レビューの評価、レビューの新しさなどに応じて検索結果の表示優先度は変化する。また、検索された店について、その店がどの名詞に対する肯定的なレビューが多いのかをグラフィカルに表示する機能を備えている。

さらに Yatani ら [12] も、グルメサイトに登録されたレビュー文章を対象とした検索機能を提案している。システムはレビュー文章中の文の係り受け構造に着目して係り元（形容詞）および係り先（名詞）のペアを抽出して記憶する。そうして各ペアの出現をカウントし、各ペアの感情分析（分析には SentiWordNet [13] を使用）を実行した後、検索結果として形容詞と名詞のペアをクラウド表示する。これによってレビューの概要を提示しようとしている。各ペアの表示サイズは、ペアの発生数に応じて変化し、また、ペアの色は、単語ペアの感情分析結果に応じて変化するようにになっている。

以上、関連研究が提供している口コミ検索機能について述べた。食探の検索機能は、口コミデータセット内の全文章のベクトル分散表現を作成しておき、ユーザが入力した日本語自然文の検索文字列のベクトルと近い文章を多数有する店を検索する手法であり、大山らのシステム [2] および栗原らのシステム [3] が採用している機能と同一であって本論文で新たに提案するものではない。したがって本論文は、検索された飲食店について書き込まれた膨大な口コミ情報を要約してユーザに提示する手法について提案を行うものであり、上述した関連研究の検索機能と競合するものではない。なお、Choi ら [10] の検索機能にみられるレビューの評価、レビューの評価、レビューの新しさなどを考慮して検索結果の表示優先度を変化させるなどの工夫

は食探の検索機能に追加できる可能性がある。

次に、長い文章を短い文章に要約することによって、読むのにかかる時間を短くしようとする研究について述べる。

Erkan ら [11] は、文章中の文や単語の関係性を元に文章を要約する LexRank 法を提案している。文章のなかで、多くの文に類似する文は重要文であり、かつ、重要文に類似した文は重要文であるという考え方に基づいて重要箇所を抽出し提示する。投稿レビュー以外の一般的な文章に対しても汎用的に使用できる文章要約手法であるが、投稿レビュー文章に特化した要約手法に比べ希望どおりの要約が得られない可能性が高いという問題がある。

Nguyen ら [14] は、マイクロレビューサイトに登録されている短文レビュー（Foursquare に投稿されたある店舗に関連付けられた 200 字投稿など）の要約のために、他サイトに登録されている長文レビュー（Yelp に投稿された当該店舗に関連付けられた長文投稿など）を用いて要約を作成する手法を提案している。最小記述長（MDL: Minimum Description Length）基準を導入し、MDL コストが最小になるように長文レビューから最良の短文を抽出し、それを元に要約文を作成している。他の口コミサイトの情報を取得して文章を要約する手法は興味深いが、本論文の提案内容とは異なるほか、他サイトの口コミデータセットが利用できる場合に限られるという問題がある。

3. 提案

グルメサイトに登録された口コミ情報を利用し、自分の希望する料理店を手早く探すことができる Web サービス「食探」を提案する。本システムは、インターネットに公開されているライブドアグルメのデータセット [16] を利用して構築されている。このデータセットは学術研究用に公開されているものであり、飲食店 21 万 4 千件、口コミ数 20 万 6 千件の情報が入っている。2011.4.20 時点のスナップショットデータであり、すべてをダウンロードして利用することができる。

ユーザがどのような料理が食べたいかに関する希望や質問を日本語文で入力すると、システムはユーザの入力文に類似した口コミ文章が多い飲食店を検索する。ユーザの入力文に類似した口コミ文章が多い飲食店を検索する際は Dov2Vec による文章のベクトル化技術を利用している。得られる結果は検索された料理店に関するすべての口コミの全文表示である。ここまでの段階は、関連研究で述べた文献 [2] や [3] と同様の処理であることから従来システムと呼ぶこととする。

食探は、従来システムに追加して、料理に関係する口コミ情報を優先的に選んだうえで要約してユーザに表示することが特徴であり、口コミ文章から抽出した料理に関する評価文を要約文で作成する機能を有している。美味しい料理店を手早く探せるかどうかという観点で従来システムと

比較して本システムの優位性が示されれば、本システムは有効であるといえる。

本システムは Python [29] で実装されたサーバプログラムと、JavaScript で実装されたクライアントプログラムから構成されている。

3.1 プロトタイプ構築

本システムの機能について述べる。主な機能は、類似口コミ文の検索、係り受け文の作成、料理情報の抽出、評価文の抽出である。

(1) 類似口コミ文の検索：データセット内の全口コミ文章を Doc2Vec に入力してベクトル分散表現をあらかじめ作成しておく。料理店を検索する際にユーザが入力した入力文のベクトル分散表現と近いベクトルを持つ口コミ文章を検索するとともに、そのような口コミが多い飲食店を抽出する。

具体的には、1つの飲食店について書き込まれた多数の口コミを1つの文章にまとめて文書 ID を付与し、すべての飲食店についての文章（飲食店ごとにまとめられた口コミ文章）と文書 ID（飲食店 ID に相当）を Doc2Vec に入力する。本実装では Python のライブラリである gensim [28] に含まれる Doc2Vec を用い、Doc2Vec のモデルは PV-DM (Paragraph Vector with Distributed Memory)、ベクトルは 300 次元、入力に与える単語列の長さである window は 5 として利用した。

Doc2Vec に入力する文章が日本語の場合は、分かち書き表記された文章を入力する必要があるため、そのため MeCab [18] の分かち書き機能を利用した。MeCab を用いる際、分かち書きされるべきでない固有名詞などを MeCab 標準辞書に多数追加した mecab-ipadic-NEologd [21] を利用した。また、ストップワード除去処理のために、Web 検索研究支援プログラミングライブラリ SlothLib [15] を用いたワード除去処理などを施した。

(2) 係り受け文の作成：(1) で検索された口コミ文章を日本語係り受け解析器 CaboCha [19], [20] を用いて短い係り受け文に要約する。CaboCha は、日本語形態素解析エンジン MeCab [18] の開発者によって開発されたオープンソースソフトウェアであり、日本語文内の語句の係り受け関係を解析する機能を提供している。MeCab と CaboCha はそれぞれの Python 用インタフェースをサーバにインストールして用いた。

たとえば「太郎は花子が読んでいる本を次郎に渡した」という文を CaboCha に入力すると、「太郎は渡した」、「花子が読んでいる」、「読んでいる本を」、「本を渡した」、「次郎に渡した」という係り受け文（修飾関係の解析結果）が出力される。CaboCha に入力する文章

は句点から句点までのような単位文である必要があるため、各口コミ文章を複数の単位文に分割する前処理を行っている。単位文は、句点や「!」、「?」記号、および、改行を終点として文章を分割して得られるものである。

(3) 料理情報の抽出：(2) で作成された係り受け文の中から、料理に関する口コミ情報を抽出してユーザに提示する。具体的には、料理名が含まれている係り受け文、または、料理を評価する用語が含まれている係り受け文を抽出するようにした。

料理名は各種公開データベースソース [23], [24] から取得した。また料理を評価する用語については、早川ら [22] の、おいしさを評価する用語（官能評価のために用語選定された 445 語）の研究成果を利用した。たとえば料理名としては「ハンバーグ」、「唐揚げ」、「アイスクリーム」、「パンチェッタ」、「チンドロン」というように家庭で食べることが多いものから普通はレストランで食べるものまでを含んでいる。また料理を評価する用語には「ふわふわ」、「ふんわり」、「パリッ」、「パリパリ」、「サクサク」、「粒状」、「乳状」、「水飴状」などの用語が含まれている。

(4) 評価文の抽出：(3) で抽出された係り受け文のなかには、料理名は入っているがその料理に対する感想や評価が書かれていない文が多数混じっている。そこで、日本語評価極性辞書 [17] に登録されている語句（評価語）を含む係り受け文のみを抽出してユーザに提示するようにした。ここで評価とは「～が良かった」、「～が悪かった」というような個人の感想である。

日本語評価極性辞書には評価に用いられる単語とその極性（ネガティブ語/ポジティブ語）の 5,278 組が登録されているが、ポジティブ極性を持つ評価語を含む係り受け文に限定してユーザに提示するようにした。たとえば日本語評価極性辞書に登録されているポジティブ評価語には「おいしい」、「うまい」、「合う」、「充分だ」、「十分だ」、「十二分だ」、「上々だ」、「上級だ」、「出来だ」などが含まれている。

なお上記料理情報の抽出と評価文の抽出処理に関し、日本語には表記ゆれが多数存在するため動詞や形容詞などの活用語については MeCab を用いて原型を取得して比較に用いるようにした。たとえば、美味しかった、美味しくなる、美味しければ、はすべて原型の「美味しい」に変換する。加えて、漢字/ひらがな/カタカナ、全角半角の違いを吸収するために、MeCab を利用して語句の読みを取得して全角カタカナで登録するようにした。たとえば、「美味しい」、「おいしい」は全角カタカナの「オイシイ」に変換されて比較される。

システムの実際の出力例を用いて口コミ要約の動作を具体的に説明する。(2) の係り受け文の作成において「美味

入力: 鹿肉のキノコ添えソテーという肉料理を食べたい 送信 形式: 6:食品名&評価語解析&口コミ色分け 件数: 5 店番号:

ここは本当に色々アラカルトでお腹一杯食べて、飲んで、楽しいお店です。この日も満席。大分前から予約しないと、入ることが出来なくなっていました。寂しいような、嬉しいような。この日も赤松シェフの「厚い背中」から旨い物作っているオーラが出ています。で、今回は前菜に食べたいものが沢山あったので前菜ばかりをオーダーしてみました。というか、フランス料理の醍醐味って前菜だと思っているんですけど、どうでしょうか。メインに来ると、ああもう終わりだ...とちょっと寂しくなっていますが、デザートで復活。アミューズ・ブーシュはいつも季節感あふれるもの。これだけでもお金取れますよね。今回は2色メロンのマリネ、アニス風味です。完熟したオレンジと緑のメロンを両方一度に楽しめる楽しさ。たっぷりと入ったマリネ液でアニス風味のお酒が効いていてちょっとピリッとしています。メロンのまったりとした風味とアニスのスーっとする甘い清涼感が微妙なバランス。そして、自家製パン。この日は丁度良い焼き加減でした。クラストはパリッとして

図 1 システム利用画面（全文表示）

Fig. 1 Usage screen (full text mode).

入力: 鹿肉のキノコ添えソテーという肉料理を食べたい 送信 形式: 0:係り受け&食品名&評価語解析 件数: 5 店番号:

307035 ビストロ ラ シブレット(フランス料理)

甘鯛の美味しさが

甘鯛を選びました。

甘鯛を選ぶ

鰻も美味しかったのですが、

うにも合います。

海老はよかったな。

大分落ち着いてきて、光っていました。

図 2 システム利用画面（要約表示）

Fig. 2 Usage screen (summary mode).

しい 魚料理でした」,「真鯛の カマは」,「ブログにて 公開しています」という係り受け文が含まれていた場合を考える。

まず(3)の料理情報の抽出において,「ブログにて 公開しています」は料理名も料理を評価する用語も含まないため候補から外れ,「美味しい 魚料理でした」,「真鯛の カマは」が残る。

次に(4)の評価文の抽出において,日本語評価極性辞書内の用語との比較を行うが,「真鯛の カマは」については「真鯛の」の部分も「カマは」の部分も日本語評価極性辞書内の用語とマッチングしない。店で出される真鯛のカマが美味しいのかどうかはこの係り受け文からは分からないため,美味しい料理が食べられる料理店を手早く探すという観点から除外すべきである。この処理により,「真鯛のカマは」が候補から外れて「美味しい 魚料理でした」が残り,これがユーザに提示されることとなる。

料理情報の抽出も,評価文の抽出も,実際は分かち書きされた活用語の原型を全角カタカナ読みに変換して実行するため,「おいしい さかな料理 でした」や「美味しい 魚料理 です」であっても候補から外れることはない。

3.2 プロトタイプの試用評価

プロトタイプを実装して大学生 10 名に実使用させ,使った感想をヒアリングした。

口コミ文を要約して表示する機能が本研究において提案

する部分である。そこで無作為に選んだ料理店 2 店について,図 1 のように口コミを全文表示した場合(従来システム)と,図 2 のように食探によって要約表示(提案システム)した場合とを比較させた。

なお全文表示時は平均 4,860 字,要約表示時は平均 398 字であり,要約機能によって文字数は約 8%に減少していた。また,表示されたレビュー文章をすべて読むのにかかった時間の平均は,全文表示時が 341.6 秒,要約表示時が 54.3 秒であり,要約機能によって時間が約 16%に短縮されていた。

得られた感想を以下に記す。

- (全文表示) 文字数が多いため本当に食べたい料理かを判断するのに時間がかかりました。
- (全文表示) 長い文章を読み進めて他に食べたい料理があるかを探すのは時間がかかりました。
- (要約表示) 料理についての部分が簡略表示されているので興味のある料理は早く見つけることができました。
- (要約表示) 全文よりも内容が頭に入ってきてやすく読みやすかったです。

以上のように,全文表示に比べて要約表示の方がおおむね良い感想が多かった。

3.3 プロトタイプの改良

しかしながら一方で,ユーザからは以下のような指摘があった。

- (要約表示) 指摘 1: 料理名が欠落している文があり、何の料理についてかわからず悩むことがあった。
- (要約表示) 指摘 2: 語尾が切れていたりして通常の日本語ではないため読んでいて違和感があった。

指摘 1 は、システムが生成した係り受け文は、具体的な料理名が欠落している場合がある点を指摘している。システムが抽出した係り受け文とは別の係り受け文に料理名が含まれているケースがあるからである。たとえば、(4)において、「チーズには焼き込みしっかりのパンが合いますね。」という文から「チーズには 合いますね。」という係り受け文が抽出された場合、この係り受け文だけでは「パン」についての書き込みであるということが分からない。

そこで、さらに文の係り受け構造を追加解析し、同じ文から生成された他の係り受け文の中に料理名があつて、その料理名が同じ係り受け関係にあった場合には料理名を補足表示するようにした。たとえば上記例では「チーズには合いますね。」と「パンが 合いますね。」がともに同じ文から生成される係り受け文であるが、「チーズには」も「パンが」も同じ「合いますね。」にかかっており、かつ、「パンが」は料理名を含んでいる。よって「パンが」という語句を補足情報として選定し、

「(パンが) チーズには 合いますね。」

という文に修正して提示するようにした。この機能を「補足表示」と呼ぶこととする。

指摘 2 は、システムが生成した係り受け文が不自然である点を指摘している。そこで、抽出された係り受け文と料理名の補足表示結果とを含む単位文を表示する機能を実装した。たとえば上記例であれば、「チーズには 合いますね。」という係り受け文の元の単位文「チーズには焼き込みしっかりのパンが合いますね。」をユーザに提示する。係り受け文を表示する場合と比較してユーザが読むべき文字数は増えるが、読みやすい自然文を提示することができるメリットがあると考えた。この機能を「抽出文表示」と呼ぶこととする。

実際にある洋食店の口コミ文章から抽出文に選ばれた文の例を以下に記す。

- パリパリのクラストが香ばしく、バターもミルクィで美味しかったです。
- 牡蠣のフリットは香ばしい衣の中にミルクィな牡蠣がとても美味しく、単品としていただきたいほどでした。
- スズキには白子のソテーが添えられていて、香ばしい焼き具合の表面とクリーミーな中身の組み合わせが良く美味しかったです。

一方、抽出文に選ばなかった文の例を以下に記す。

- ソムリエとシェフがお店の外まで見送りに出て下さいましたが、なかなかタクシーが捕まらなくて待たせて

しまいました。

- 今回、ブーダンノワールをいただくことが出来なかったため、また別の機会に伺ってみたいと思っています。
- 同行者が持ち込んだスパークリングワインで乾杯です(写真はブログにて公開しています。よろしかったらご覧ください)。

4. 評価

試作したシステムを Web サーバ上に構築し、大学生 11 名に使用させる実験を行った。データベースには、飲食店 21 万 4 千件、口コミ数 20 万 6 千件の情報が入っている状態である。

システムの画面例を図 1、図 2 および図 3 に示す。ページの最上部に入力欄が設けられており、入力欄（日本語自然文で入力）、送信ボタン、出力形式指定（口コミ全文表示/口コミ要約表示）、出力件数指定、店番号入力欄（通常は空白）が配置されている。被験者にはシステムの使い方について以下の文面で説明した。

『美味しい料理が食べられる料理店を手早く探すためのシステムです。入力欄には、どのような料理が食べたいかに関する自分の希望や質問を日本語自然文で入力してください。たとえば、「鹿肉のキノコ添えソテーという肉料理を食べたい」と入力して送信ボタンを押します。すると件数欄で指定された件数のおすすめ店舗が出力されます。出力形式は、口コミ全文表示または口コミ要約表示が選べます。なお、店番号の欄に店番号をいれて送信ボタンを押すと、おすすめ店舗検索ではなく、その店の口コミが表示されます。この場合も出力形式は、口コミ全文表示または口コミ要約表示が選べます。』

被験者全員に特定の店番号を店番号欄に入力させ、全員に同じ店の口コミ文章を見せて実験を行った。本実験では料理店 2 店を実験対象とした。

実験に用いた表示形式は表 1（画面に表示された文字数を付記している）に示す 4 種類であり、口コミ文章をすべて表示する従来手法と、口コミ文章を要約して表示する 3 提案手法（要約表示、補足表示、抽出文表示）である。

被験者には、表示形式それぞれについて、表示された文章をすべて読むのにかった時間、満足度に対する 5 段階評価（1 が低く、5 が高い）、および自由文による感想を記入させた。被験者を 4 つのグループに分け、それぞれのグループで 4 つの表示形式を読む順序が異なるようにした。また流し見にならないよう、被験者には「読んだ結果、何を食べたくなったか回答してください」と指示した。

美味しい料理店を手早く探せるかどうかという観点で従来システムと比較して本システムの優位性が示されれば、本システムは有効であると判断することとした。

実験結果を表 2 に示す。

ユーザ満足度は、抽出文表示が最も高く、次いで係り受

入力: 鹿肉のキノコ添えソテーという肉料理を食べたい 送信 形式: 6:食品名&評価語解析&口コミ色分け ▼ 件数: 5 店番号:

この黒板メニューを見ているときが一番楽しいかもしれませんね。
 かなりシンプルで、牡蠣にエシャロット&赤ワインヴィネガーと生クリームをかけ、シブレットをちょこつとのせただけです。
 ヴィネガーと生クリームの味の効かせ加減がたまりません。
 季節ごとに最良の牡蠣を用意しており、いつ食べても楽しめます。
 また牡蠣だったので、合わせようと思いました。
 クスクスはポロポロとしていて、軽くヴィネグレットのような味とトマトコンカッセ、生玉ねぎ、セロリ、セルフィユなどの生
 もうちょっと強めに甘辛くてもよいかもしれません。
 なんだか珍しい野菜もあって、勉強になります。
 ドフィノワはチーズは入っていないと思われます。
 ほんのりと器にこすりつけたと思われる大蒜の風味が効いていて、柔らかな乳製品の味。
 トロトロとした白桃はしっかりと形を残していて、色々な風味のフルーツの優しい煮汁がしっかりとしみこんでいました。

図 3 システム利用画面（抽出文表示）

Fig. 3 Usage screen (sentence extraction mode).

表 1 実験対象

Table 1 Evaluated modes.

要約機能	表示文字数（平均）
全文表示（従来手法）	4,483
要約表示（係り受け文）	358
補足表示（係り受け文）	546
抽出文表示（単位文）	941

表 2 実験結果

Table 2 Evaluation results.

要約機能	平均（秒）	標準偏差（秒）	満足度
全文表示（従来手法）	287.0	139.2	2.1
要約表示（係り受け文）	46.5	13.5	3.4
補足表示（係り受け文）	50.0	19.1	3.4
抽出文表示（単位文）	60.2	14.4	4.4

け文表示の2つ、全文表示の順となった。また満足度に関しウィルコクソンの符号付き順位和検定を行ったところ、全文表示と要約表示、および、全文表示と補足表示の対比においては有意水準両側5%で有意差が認められ、全文表示と抽出文表示の対比においては有意水準両側1%で有意差が認められた。特に抽出文表示は、全文表示や係り受け文表示に比べて満足度が高いうえに、読むのにかかる時間は係り受け文表示の2つとそれほど差がなかった。

抽出文表示についてユーザからヒアリングした意見を以下に列挙する。

- 良：料理の感想が取り出されていて一番わかりやすかった。
- 良：美味しい料理の部分が抽出されていてすぐわかった。
- 良：構成が良いので食べたい料理が早く見つかった。
- 良：簡潔に文章がまとめられていてわかりやすい。
- 良：一番読みやすい。丁度よい文章量。
- 良：ちょうどよい。流し読みができる。
- 良：係り受け文より少し時間がかかったが、ちょうど

よい。

- 良：一番知りたいところわがかりやすい。ちゃんと読みたい時にも使える。

以上のように今回のヒアリングでは悪いコメントは1つも収集されなかった。係り受け文表示の2つに比べ抽出文表示の方が読む時間が若干長くかかったことを不満に思う被験者もいなかった。

一方、補足表示の満足度は、要約表示をかなり上回るのではないかと当初予想していたが、結果的に差は見られなかった。補足表示についてユーザからヒアリングした意見を以下に列挙する。結果的に悪い面についてのコメントの方が良い面についてのコメントより多い結果となった。

- 良：全文表示や要約表示よりはわかりやすくなった。
- 良：どの料理が美味しいのか（）内を見たらわかった。
- 良：文章は少し変だが理解はできた。
- 悪：短いので読む気になるが、何の料理かわからない場合もあった。
- 悪：文章の方がよい。
- 悪：理解しにくい場合があった。
- 悪：補足説明としてはまだ足りない。内容に合っていない場合がある。
- 悪：短くて読みやすいがところどころ料理が分からないケースもある。

以上のことから、食探が最も優先して提供すべきは抽出文表示機能であると結論づけた。また抽出文表示機能を提供する本システムには有効性が認められると結論づけた。

ここで考察として、前述した関連研究と比較しながら提案システムの位置づけについて述べる。

検索機能としてはあるが、関連研究のなかにも文の係り受け構造に着目して係り元と係り先のペアを抽出し、日本語または英語の評価極性辞書を用いて口コミの中の感情語を抽出して利用する例が存在している。たとえば、口コミ文章から感情語（形容詞）を抽出してユーザがどのようなこと（名詞）に関心を持っているかをグラフィカルに表

示する例 [4], [5], [6], [7] や、口コミ文章中の感情語（形容詞）を分析して、その分析結果に応じて検索結果（形容詞と名詞のペア）の表示色を変更する例 [12] がある。

ただしこれらの関連研究が作成する係り元と係り先のペアは単なる形容詞と名詞のペアであり、このまま食探の要約機能に採用したとすれば、改良する前の食探プロトタイプと同様、何の料理についての評価か分からないという問題や、通常の文章でないため理解しにくいという問題が残る。

食探の場合は、係り元と係り先のペアを抽出した後に、さらに文の係り受け構造を追加解析して料理名を表示（補足表示）したり、係り元と係り先のペアと料理名の補足表示結果とを含む自然文を提示（抽出文表示）したりして、この問題に対応しているところが関連研究と異なっている。

5. まとめ

本論文では、美味しい料理が食べられる料理店を手早く探すためのシステム食探を提案した。レビュー文章から抽出した料理に関する良い評価文を要約して表示する機能が特徴である。

評価実験の結果から、システムに実装した3つの要約機能の満足度はどれも従来システム（全文表示）を上回った。特に抽出文表示は、要約表示や補足表示に比べて満足度が高いうえに、読むのにかかる時間は要約表示とそれほど差がない。これらのことから、食探が最も優先して提供すべきは抽出文表示機能であると結論づけた。

謝辞 本研究は JSPS 科研費 16K00506 の助成を受けたものです。

参考文献

- [1] yelp Open Dataset (2020), available from <https://www.yelp.com/dataset>.
- [2] 大山浩暉, 竹川佳成, 平田圭二: レビュー文を考慮したゲーム推薦システムの実現に向けた単語の類似度調整の取り組み, 情報処理学会エンタテインメントコンピューティングシンポジウム 2017 論文集 (2017).
- [3] 栗原光祐, 莊司慶行, 藤田澄男, Martin, J.D.: Doc2Vec 手法による映画レビューサイトからのクエリと意味的に類似した評価表現の発見, DEIM2019, データ工学と情報マネジメントに関するフォーラム, C4-2 (2019).
- [4] 谷本融紀, 太田 学: 評価属性を考慮した評判情報の可視化, 情報処理学会 DBS 研究報告, Vol.151, No.4, pp.1–8 (2010).
- [5] 平山拓央, 湯本高行, 新居 学, 佐藤邦弘: 語の共起と極性に基づく商品レビュー閲覧支援システム, 情報処理学会 DBS 研究報告, Vol.155, No.3, pp.1–9 (2012).
- [6] 中野裕介, 湯本高行, 新居 学, 佐藤邦弘: 商品レビュー要約のための属性-意見ペア抽出, 情報処理学会 DBS 研究報告, Vol.160, No.15, pp.1–7 (2014).
- [7] 紀本雅大, 伊藤淳子, 宗森 純: 評価表現に着目した映画レビューからの評価情報抽出, 情報処理学会 DCC 研究会報告, Vol.21, No.43, pp.1–8 (2019).
- [8] Suresh, V., Roohi, S. and Eirinaki, M.: Aspect-based opinion mining and recommendations system for restaurant reviews, *Proc. ACM Conference On Recommender Systems (RecSys 2014)* (2014).
- [9] NLTK: Natural Language Toolkit (2020), available from <https://www.nltk.org/>.
- [10] Choi, J. et al.: Smith Search – Opinion-Based Restaurant Search Engine, *Proc. ACM Conference on World Wide Web (WWW 2015)* (2015).
- [11] Erkan, G. and Radev, D.R.: LexRank - Graph-based Lexical Centrality as Salience in Text Summarization, available from <http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume22/>.
- [12] Yatani, K., Novati, M., Trusty, A. and Truong, K.N.: Review Spotlight – A User Interface for Summarizing User-generated Reviews Using Adjective-Noun Word Pairs, *Proc. ACM Conference on Human Factors in Computing Systems (CHI 2011)* (2011).
- [13] SentiWordNet: A lexical resource for opinion mining (2020), available from <https://github.com/aesuli/sentiwordnet>.
- [14] Nguyen, T.-S., Lauw, H.W. and Tsaparas, P.: Review Synthesis for Micro-Review Summarization, *Proc. ACM Conference On Web Search and Data Mining (WSDM 2015)* (2015).
- [15] SlothLib, Web 検索研究支援プログラミングライブラリ, pp.223–227 (2017), 日本語ストップワード, 入手先 <http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt> (2019).
- [16] livedoor グルメの DataSet, 学術研究用アーカイブ (2011), available from <http://blog.livedoor.jp/techblog/archives/65836960.html> (2019).
- [17] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: 意見抽出のための評価表現の収集, 日本語評価極性辞書 (用言編), 自然言語処理, Vol.12, No.3, pp.203–222 (2005).
- [18] 工藤 拓: MeCab - Yet Another Part-of-Speech and Morphological Analyzer (2019), 入手先 <http://taku910.github.io/MeCab/>.
- [19] 工藤 拓: CaboCha - Yet Another Japanese Dependency Structure Analyzer (2019), 入手先 <http://taku910.github.io/CaboCha/>.
- [20] 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, 情報処理学会論文誌, Vol.43, No.6, pp.1834–1842 (2002).
- [21] mecab-ipadic-NEologd: Neologism dictionary for MeCab (2019), available from <https://github.com/neologd/mecab-ipadic-NEologd/blob/master/README.ja.md>.
- [22] 早川文代: おいしさを評価する用語, 官能評価のために用語選定された 445 語, 日本調理科学会誌, Vol.41, No.2, pp.148–153 (2008).
- [23] ルーラル電子図書館 日本の食生活全集 データベース (2019), 入手先 <http://lib.ruralnet.or.jp/yougo/dentou/a.phtml>.
- [24] 料理名一覧: 建帛社入手先 https://www.kenpakusha.co.jp/pdf/pdf_ee7.0_ryouri_ichiran.pdf (2019).
- [25] Lau, J.H. and Baldwin, T.: An Empirical Evaluation of Doc2Vec with Practical Insights into Document Embedding Generation (2016), available from <https://arxiv.org/abs/1607.05368>.
- [26] models.doc2vec - Doc2vec paragraph embeddings (2019), available from <https://radimrehurek.com/gensim/models/doc2vec.html>.
- [27] models.word2vec - Word2vec embeddings (2019), available from <https://radimrehurek.com/gensim/models/word2vec.html>.
- [28] gensim, Free Python Library (2019), available from

<https://radimrehurek.com/gensim/>).

[29] python, available from <https://www.python.org/> (2019).

推薦文

本論文は、DICOMO2019 の発表論文のなかで特に評価の高い論文であった。膨大な口コミ情報のなかから重要なコメントを絞り込む方式の研究であり、技術として優れており良い効果が得られているため、実用性のある論文として推薦する。

(グループウェアとネットワークサービス研究会主査
齊藤 典明)



市村 哲 (正会員)

1989 年慶應義塾大学理工学部計測工学科卒業。1994 年同大学大学院理工学研究科博士後期課程修了。博士(工学)。同年富士ゼロックス(株)入社。1997～1999 年富士ゼロックスパロアルト研究所(FXPAL)駐在。2002 東

京工科大学助教授。2015 年より大妻女子大学教授。グループウェア、生体情報活用、人工知能応用等の研究に従事。『IT TEXT 基礎 Web 技術』、『IT TEXT 応用 Web 技術』(オーム社)、『AI のキホン：人工知能のしくみと活用』(日本経済評論社)。DICOMO 2019 優秀論文賞受賞。