

検索連動型広告の自動生成と集客効果の測定 — 飲食店ドメインを例題に

藤田 篤^{†1} 幾島 克洋^{†2} 佐藤 理史^{†2}

本論文では、飲食店ドメインの個々のコンテンツ（店舗）に対する検索連動型広告を自動生成するシステムについて述べる。このシステムは、ポータルサービス向けに記述された各店舗に固有のテキストデータから、飲食店ドメインにおける広告作成の知見に基づいて広告として有用な表現を抽出し、個々の店舗に対する複数の広告を生成する。自動生成した広告の適否を広告業務従事者が判定した結果、広告として使用するテキストを比較的高い被覆率・精度で生成できることが確認できた。さらに、自動生成した広告を実際の広告サービスを通じて約1ヵ月間運用した。その結果、我々のシステムが生成した広告の中には、テンプレートに基づく画一的な広告よりも高いクリック率を獲得できるものが、それほど多くはないもの含まれており、現在の広告サービス上で各店舗の集客効果を改善できる見込みがあることが確認できた。

Automatic Generation of Listing Ads and Assessment of Their Performance on Attracting Customers: A Case Study on Restaurant Domain

ATSUSHI FUJITA,^{†1} KATSUHIRO IKUSHIMA^{†2}
and SATOSHI SATO^{†2}

This paper describes a system that automatically generates content-specific listing ads by reusing textual data for promoting each content, taking restaurant domain as a platform of research. The textual data used in this research are primarily created for the use on a restaurant portal site on the Web. We therefore extracted and trimmed shop-specific expressions to make listing ads. Subjective evaluation confirmed that our system can generate listing ads at sufficiently high precision and coverage. A one-month experiment using Overture Sponsored Search revealed that the automatically generated ads do not necessarily have a higher click-through rate than template-based baseline ads, but they have a potential to improve the performance on attracting customers.

1. はじめに

インターネット上の広告（オンライン広告）にかけられる費用は急速に伸び続けている。文献5)によれば、日本における広告費の総額、およびテレビ、新聞などの従来のメディアを用いた広告費は減少しつつあるものの、オンライン広告に投じられたと推定される費用は媒体費、制作費ともに増え続けており、2009年には7,000億円を超え、今後さらに成長すると予想されている。オンライン広告には、バナー、テキスト、動画などが用いられる。なかでもとくに、**検索連動型広告**と呼ばれる、ウェブ検索の結果画面に表示されるテキスト広告が、Google AdWords*¹（以下、AdWords）や Overture Sponsored Search*²（以下、Overture）などの広告サービスの登場以降、多くのユーザによって利用されており（先述の文献によれば媒体費として1,934億円、PC向けが1,710億円、モバイル向けが224億円）、高い成長率を示している。

検索連動型広告とユーザの動線の間接的な関係を図1に示す。ユーザが検索エンジンに検索クエリを投じて検索を実行すると、検索結果画面にいくつかの広告が表示される。これらの広告は、検索エンジンが指定する広告サービスによって、検索クエリと登録済みの広告の**キーワード***³の照合、および適合した広告の品質の評価に基づいて選択される。表示される各広告は、**タイトル**、**説明文**、**表示用URL**の3つの要素からなる。表示された広告をユーザがクリックすると、広告の**ランディングURL**が指すウェブページ（**ランディングページ**）が表示される。検索連動型広告は、このように、ウェブ検索という近年頻繁に使用されるツールを通じて、より多くの消費者を広告主のウェブサイトに誘導する役割を担う。広告主は、より多くの消費者に広告対象の情報を効果的に伝達し、自らのウェブサイトを訪れるユーザに商品の購入、会員登録、メーリングリストへの登録などのアクション（コンバージョン）を起こさせるために広告を出す。

現在の広告サービスが採用している、クリック課金（Pay-Per-Click; PPC）と呼ばれるビジネスモデルを図2に示す。このモデルには、広告サービス、広告主、消費者の3者が

^{†1} 公立はこだて未来大学システム情報科学部

Faculty of Systems Information Science, Future University Hakodate

^{†2} 名古屋大学大学院工学研究科

Graduate School of Engineering, Nagoya University

*1 <http://adwords.google.com/>

*2 <http://listing.yahoo.co.jp/service/srch/>

*3 図1の例のように複数の文字列をスペースでつなげたもの。

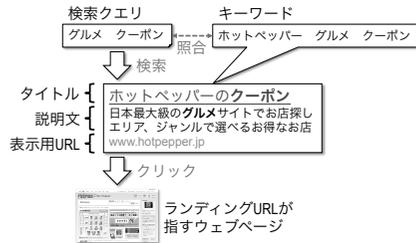


図 1 検索連動型広告とユーザーの動線
Fig. 1 Listing ad and users' flow.

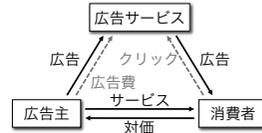


図 2 クリック課金広告のビジネスモデル
Fig. 2 Pay-Per-Click business model.

関わっている。広告主が広告費を支払うのは、ウェブ検索の検索結果画面に表示された広告が消費者（検索エンジンのユーザ）によってクリックされた場合のみである。検索連動型広告市場の成長は、広告の表示自体は無料という費用対効果に優れた仕組みと、容易に広告を掲載・管理できる広告サービスのプラットフォームに支えられている。

広告主が検索連動型広告の費用対効果を改善するには、集客効果の改善と費用の削減の両方が必要である。集客効果の改善に関しては、個々のコンテンツに固有の特徴をいかにして訴求するかが鍵となる。ただし、近年では、通信販売サイトをはじめとして多様なコンテンツを有するサイトが存在しており、数百万種類のコンテンツを有するサイトも珍しくない。個々のコンテンツに対して人間が個別に広告を作成することは人的・金銭的コストの面で現実的ではないため、費用の削減には、計算機による支援あるいは自動化が不可欠である。

我々は、広告主あるいは広告主と広告サービスを仲介する代理店の立場で、多様なコンテンツの各々をより適切に宣伝するような検索連動型広告を安価に作成するための技術について研究している。個々のコンテンツに対する広告を出そうとする場合、広告のランディングページが存在し、そこには、各コンテンツに固有の情報を記述したテキストデータが含まれると仮定できる。そこで、本論文では、

個々のコンテンツに関するテキストデータから、各コンテンツに固有の、集客効果の高い検索連動型広告を自動生成できるか

について論じる。先行研究では、検索連動型広告のうちキーワードの生成しか扱っておらず、個々のコンテンツに固有のタイトル、説明文を自動生成する技術は存在しない。そこで本論文では、広告のタイトルおよび説明文を自動生成する手法を提案し、その技術が人間による広告作成の支援技術となりうるか、あるいは完全な自動化が可能かを見極める。

ただし、ドメインによって消費者の分布や有効な宣伝手段が異なるため、広告の自動生成

の際も、それらを考慮する必要があると考えられる。そこで我々は、例題として飲食店ドメインを取り上げ、広告代理店の協力の下、様々な店舗に関するテキストデータおよび当該ドメインの広告事業に関する知見に基づいて、広告生成システムを作成した。

本論文の構成は次のとおりである。2章では、広告サービスおよび広告主の視点から検索連動型広告に関する研究を紹介する。3章で飲食店ドメインのポータルサービスにおける個々のコンテンツ（店舗）に関する一般的な情報および検索連動型広告を用いる際の課題、4章で我々が広告の生成に用いたデータについて述べた後、5章で我々の広告生成システムについて述べる。6章では広告を自動生成できる範囲と精度に関する評価、7章では自動生成した広告の集客効果に関する評価について述べる。最後に、8章で本論文をまとめる。

2. 検索連動型広告と先行研究

2.1 広告サービスの仕組みと先行研究

PPC方式のビジネスモデルにおいて、広告主から広告サービスへの広告費の支払いは、検索エンジンのユーザが広告をクリックすることによって発生する。したがって、広告サービスの運営者が広告収入を増やすには、よりクリックされやすい広告を表示する必要がある。検索エンジンのユーザは、検索意図に適合する情報を求めてウェブ検索を実行する。したがって、検索意図との関連性が高い広告ほどクリックされやすいと考えられる。

広告 a のクリックされやすさ（Click-Through Rate; CTR）は、次式で表される。

$$CTR(a) = \frac{Click(a)}{Imp(a)} \tag{1}$$

ここで、 $Imp(a)$ は広告 a が表示された回数、 $Click(a)$ は広告 a がクリックされた回数を表す。AdWords や Overture は、各広告の CTR を参照して広告の表示順位を決定する^{*1}。これは、CTR が高い広告を、検索クエリに対する関連性をユーザに対して主張できた広告と見なし、そのような広告を掲載することでさらなるクリックをねらっているものと解釈できる。この仕組みにより、CTR が相対的に低い広告は次第に表示されにくく、高い広告は表示されやすくなる。さらに、これらの広告サービスでは、同じキーワードを持つ複数の広告を1つの**広告グループ**にまとめる機能を広告主に対して提供している。広告サービスは、1つの広告グループ内の広告を1回の検索に対してたかだか1つしか表示しない。

検索連動型広告サービスにおいて、検索クエリと登録済みの広告のキーワードを照合し、

*1 広告サービスの運営者にとっては、新しく登録された広告の CTR の予測も重要な課題である^{13),14)}。

適合する広告を順位付けして表示する処理は、情報検索の一種と見なすことができる。このため、照合および順位付けの技術は、それらを担う検索エンジンおよび広告サービスの運営者によって活発に研究されており²⁾、情報検索で培われた潜在意味解析や疑似関連性フィードバックなどの技術を用いることで改善できることが報告されている。また、ユーザが検索クエリとして用いた文字列が広告のタイトルや説明文に含まれる場合、広告サービスは、それらの文字列を太字で表示し、検索クエリに対する関連性の高さを強調する。

2.2 広告主の増収戦略と先行研究

広告主が広告を出す目的の一つは、広告を通じて自らのウェブサイトにはユーザを導き、コンバージョンを発生させ、収入を得ることである。PPC方式のビジネスモデルにおいては、ユーザのクリックによって支出が生じる。また、広告の作成にも費用がかかる。簡単のため、コンバージョンとしては、検索連動型広告を通じて発生する分のみを考えると、広告主所有の広告集合 A をある期間だけ掲載したときの利益は、次式で表される。

$$Profit(A) = \sum_{a \in A} [Conv(a) \cdot m(a) - Click(a) \cdot CPC(a) - p(a)]. \quad (2)$$

ただし、 $Conv(a)$ は広告 a を経由して発生したコンバージョンの回数、 $m(a)$ は広告 a によって得られた収入の平均、 $CPC(a)$ は広告 a のクリック 1 回あたりの広告費（平均クリック単価）、 $p(a)$ は広告の制作費を表す。広告代理店の利益も、 $m(a)$ 、 $CPC(a)$ 、 $p(a)$ の実体が異なるものの、式 (2) に準じる。ここで、広告 a のコンバージョン率は次式で表される。

$$CVR(a) = \frac{Conv(a)}{Click(a)}.$$

この式および式 (1) を変形して式 (2) に代入し、整理すると、次式が得られる。

$$Profit(A) = \sum_{a \in A} [Imp(a) \cdot CTR(a) \cdot \{CVR(a) \cdot m(a) - CPC(a)\} - p(a)]. \quad (3)$$

広告主および広告代理店にとっては、式 (3) の値を最大化することが目標となる。より具体的な戦略として、次の 5 つが考えられる。

- 広告の表示回数の向上： $Imp(a)$
- 広告のクリックされやすさの改善： $CTR(a)$
- クリック 1 回あたりの利益の向上： $\{CVR(a) \cdot m(a) - CPC(a)\}$
- 広告の制作費の削減： $p(a)$
- 広告集合全体の利益の向上：式 (3) 全体

2.2.1 広告の表示回数の向上

式 (3) においては検索連動型広告を通じて発生するコンバージョンのみを考えたが、実際には、検索連動型広告をクリックしなかったユーザが別の機会にコンバージョンを起こす可能性がある。このような意味でも、広告の表示回数の向上は重要である。

広告をより頻繁に検索結果画面に表示させるには、ユーザの検索クエリと直接照合されるキーワードの工夫が最も有効である。キーワードはユーザの検索意図をとらえるうえでも重要な役割を担う。キーワードが曖昧な表現や漠然とした表現の場合、広告が表示されたとしても、ユーザは検索結果も広告をクリックせず、検索結果を絞り込むために AND 検索を実行するだろう。検索クエリとしてよく用いられる表現をキーワードとすると、競合する広告が多くなるため、個々の広告は表示されにくくなり、広告のクリック単価も高くなる。逆に、検索クエリとしてあまり用いられない表現をキーワードとすると、広告があまり表示されないことになる。むやみにキーワードを増やすだけでは、ランディングページの内容に興味のないユーザを引き込んでしまい、コンバージョン率を下げってしまう恐れもある。したがって、コンテンツとの関連性が高い表現をキーワードとすることが望ましい。

与えたキーワードに基づいて表示回数を増やすような別のキーワードを発見する手法として、たとえば、意味的に類似する表現の発見^{1),8)}、上位語・下位語や共起語の利用^{3),4)}、検索クエリとして用いられやすい表現への変換¹²⁾ などの手法が提案されている。

2.2.2 広告の CTR の改善

AdWords や Overture などの広告サービスは、検索クエリとキーワードの照合に基づいて広告を表示する。ただし、キーワードは検索結果画面には表示されないため、検索クエリと広告との関連性を訴求し、ユーザのクリックを動機付けられるか否かは、広告として表示されるタイトルと説明文にかかっている⁷⁾。タイトルがユーザの検索クエリを含んでいる場合、2.1 節で述べたように、それらは広告サービスによって自動的に強調表示される。したがって、キーワードをタイトルや説明文に含めることによりユーザに対する求心力も向上すると期待できる。しかしながら、コンテンツに応じて広告のタイトルや説明文を自動的に生成する研究はなされていない。現在検索結果画面に見られる広告のタイトルや説明文の多くは、定型文や、せいぜいテンプレートにキーワードを埋め込んだものである。これらは、安価に作成できるが、画一的であるため、個々のコンテンツの情報を消費者に対して効果的に伝達できるとはいいがたい。

通常の検索結果と同様、検索連動型広告についても、表示順位が上位のものほどクリックされやすい⁷⁾。すなわち、表示順位を上げることが CTR の向上につながる。ただし、広告

の表示順位は、広告サービスによって、少なくとも、検索クエリに合致するキーワードを持つ他の広告の CTR, および広告の各要素を参照して定められる。参照方法や競合する広告の CTR を観測したり、内容を統制して表示順位を観測したりすることは不可能であるため、工学的研究において、広告の表示順位を直接的な評価指標とすることは適切ではない。

2.2.3 クリック 1 回あたりの利益の向上

クリック 1 回あたりの利益を向上させるには、

- ランディングページにおけるコンバージョン率 $CVR(a)$ の向上
- コンバージョンの単価 $m(a)$ を上げる (たとえば、高価な商品を販売する)
- クリックあたりのコスト $CPC(a)$ の上限値を下げる

ということが考えられる。しかし、いずれも、広告そのものの改良とは関係がない。

2.2.4 広告の制作費の削減

広告のタイトルや説明文に定型文やテンプレートを用いる場合、制作費は限りなく小さい。一方、膨大なコンテンツの各々に対して個別の広告を作成するには、人的・金銭的コストの面で、計算機による支援あるいは自動化が不可欠である。

2.2.5 広告集合全体の利益の向上

広告主は、広告サービスによる CTR に基づく広告の選択という仕組みと上述の広告グループを利用して、複数の広告を効果的に使用することができる。そして、広告グループ内で相対的に CTR が低い広告を削除することによって、CTR が高い広告の表示回数、クリック回数、ひいては利益を増やすことができる。ただし、1 つの広告グループに登録できる広告の数には、AdWords の場合 50 個、Overture の場合 20 個という上限がある。また、CTR に基づく広告の淘汰はそれほど急速に行われるわけではないため、コンテンツやウェブサイトで、および検索クエリとの関連性が低い広告を登録してしまうと、それが繰り返し表示され、広告対象のコンテンツやウェブサイトのイメージを傷つけてしまう恐れがある。

したがって、各広告の CTR の変動や安定性を分析して短い期間で有望な広告、集客効果の低い広告を見定めるとともに、CTR と相関のある指標を発見し、広告の登録前に取捨選択を行うことが考えられる。しかしながら、広告主あるいは広告主と広告サービスを仲介する代理店の立場での適切な広告運用の方法論は確立されていない。

3. 飲食店ドメインのポータルサイトと検索連動型広告

検索連動型広告の自動生成技術を研究するためのドメインとして、本論文では、多様なコンテンツに関するテキストデータが存在する飲食店ドメインを取り上げる。

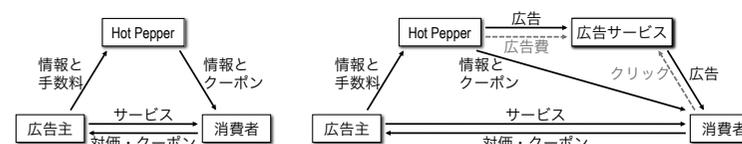


図 3 Hot Pepper のビジネスモデル (左: 基本モデル, 右: クリック課金広告による拡張モデル)
Fig. 3 Hot Pepper's business models (left: basic model, right: Pay-Per-Click extension).

3.1 飲食店ドメインにおける一般的な情報

国内の飲食店ドメインのポータルサイトには、『Hot Pepper Foomoo』*1, 『ぐるなび』*2, 『食べログ』*3 などがある。いずれのサイトにおいても、個々の店舗に関するウェブページが存在し、トップページ、メニュー、口コミ、写真、特集やキャンペーンなどの複数のページから構成されている。これらのポータルサイトに共通の情報は、次の 3 種類に大別できる。

店舗の属性情報: 店舗名、店舗のコンタクト情報 (住所、電話番号など)、店舗の設備に関する情報 (収容人数、禁煙席の有無、個室の有無など) など。

宣伝用の情報: 店の雰囲気やメニュー、特集やキャンペーンに関する情報。メディアとしてはテキスト、写真、クーポン券など。

消費者によって提供される情報: いわゆる『口コミ』の情報。店舗へのアクセス、雰囲気、メニューなどに関する評点や、より個別的な評価に関するテキスト。

店舗の属性情報は、客観的事実を伝えるものである。したがって、ポータルサイト間でも共通であると考えられる。これに対して、宣伝用の情報はポータルサイトごとに作成されている。消費者によって提供される情報も、ポータルサイトごとに形式や掲載条件が異なる。

3.2 対象サイト

我々は、飲食店ドメインのポータルサイトの一つである『Hot Pepper Foomoo』(以下、単に Hot Pepper) を、運営者の協力のもと、研究の題材とした。

Hot Pepper は、各店舗に関する様々な情報を収集し、月刊のフリーペーパー、ウェブ上のポータルサイト、検索連動型広告などの媒体を通じてそれらを提供することによって、消費者を店舗へと導く (図 3 左)。消費者は、Hot Pepper のポータルサイトを通じて座席を予約したり、フリーペーパーやポータルサイトから入手できるクーポン券を店舗に持参したりす

*1 <http://www.hotpepper.jp/>

*2 <http://www.gnavi.co.jp/>

*3 <http://r.tabelog.com/>

ることにより、割引、無料サービス、特別メニューなどの特別なサービスを受けることができる。

Hot Pepper における宣伝用のテキストデータは、Hot Pepper の営業担当者と広告主である飲食店職員によって記述されている。営業担当者は店舗の業務内容を調査し、キャッチコピーだけでなく、キャンペーンおよびそのキャッチコピーを提案したりもする。その結果、店舗の様々な側面の特徴をもれなく伝えるようなテキストデータが蓄積されている。

PPC 方式のビジネスモデルにおける参加者と Hot Pepper の関係は、図 3 右によって表される。ただし、上述のテキストデータの本来の用途はポータルサイトにおける宣伝であり、広告サービスの基準（文字数や使用可能な文字など）を必ずしも満たしていないため、そのまま広告に使用することはできない。検索連動型広告はわずか数十文字であるため、各店舗に関するテキストデータの作成時に広告サービス向けのテキストも作成しておくことが可能であるように見えるかもしれない。しかしながら、数万件の店舗の各々に対して、広告サービスの基準を満たし、かつ有用そうなテキストを、広告サービスが許す限り多数作成するには、新たに無視できない規模の人的コストを要する。既存のテキストデータを人手で加工する場合でも、少なからずコストは小さいものの、それでも十分小さくはない。このため、Hot Pepper では、テンプレートに基づく広告を用いている。

Hot Pepper の運営者らは、ポータルサイトにおけるフリーワード検索のクエリログ、およびユーザのページ上の移動状況の分析に基づいて、ユーザの検索意図にはいくつかの典型例があることを明らかにした。たとえば、よく知られているチェーン店の最寄りの支店を探す、様々な銘柄の日本酒を出す店舗を探す、誕生日に訪れた来店客にケーキなどのサービスをする店舗を探す、などである。Hot Pepper は、ポータルサイトにおける検索意図がこの分野に関する人々の典型的な検索意図を反映していると仮定し、表 1 に示すような**特集サイト**を 200 種類（2009 年 9 月時点）運営している。各特集サイトは、Hot Pepper に登録されている店舗のうち、ユーザの典型的な検索意図に対応する条件を満たす店舗のみで構成されている。ユーザは特集サイトにおいて、検索クエリや検索条件を試行錯誤することなく、また、より少ないページの移動回数で、要件に見合う店舗を見つけることができる。

4. 広告生成に利用したデータ

Hot Pepper の特集サイトの各店舗に対する広告を生成するために、各店舗に関するテキストデータ、各特集サイトに関する情報、およびポータルサイトにおける検索クエリログの 3 種類のデータを使用した。

表 1 特集サイトの例

Table 1 Examples of special feature sites.

| 特集タイプ | 特集名 | URL |
|---------|-------|-------------------------|
| チェーン店 | 牛角 | gyukaku-navi.net |
| | 坐・和民 | zawatami-navi.net |
| | サンマルク | sanmaruku-navi.net |
| 店舗種別 | 居酒屋 | izakaya-jyouthou.net |
| | スイーツ | sui-tu-navi.net |
| | バー | bar-navi.net |
| 食べ物 | 日本酒 | nihonsyu-navi.net |
| | カレー | kare-navi.net |
| | ホルモン | horumon-navi.net |
| 状況 | 誕生日 | tanjyoubi-navi.net |
| | 二次会 | nijikai-navi.net |
| | デート | date-jyouthou.net |
| 設備・サービス | ソファ | sofa-navi.net |
| | ダーツ | darts-navi.net |
| | 食べ放題 | tabehoudai-jyouthou.net |

4.1 各店舗に関するテキストデータ

各店舗に関するテキストデータ（以下、**店舗データ**）のうち、5 章で述べる広告生成システムで使用した分*1を表 2 に示す。ただし、表中の「*」を付した 5 種類の要素は複数の店舗に共通に用いられているものである。店舗データは、元々は Hot Pepper のポータルサイトにおける宣伝のために記述されたものであり、例 (1) に示すように、印象的な、興味を引く、あるいは消費者にとって有益な情報を含む表現を多く含む。

- (1) a. 味・触感・香と 3 拍子揃ったうまい肉
 b. 黒毛和牛の飲放付コースが¥4515 → ¥3000
 c. 全席掘りごたつだから楽しんでね☆
 d. 毎月 29 日は何かか起こる！

ただし、これらのテキストは、多くの場合、広告サービスにおける長さの制約を満たしていない。また、店舗の属性以外のテキストには、「★」や「♪」などの、広告に使用できない文字が含まれる。したがって、これらを広告に使用するには、適切な加工を施す必要がある。

*1 公式の API (<http://webservice.recruit.co.jp/hotpepper/>) を通じて取得可能な情報に加えて非公開の情報も一部使用。

表 2 個々の店舗に関するテキストデータ
Table 2 Textual data for a shop.

| 情報種別 | 例 |
|--|--|
| 店舗の属性 | |
| 店舗 ID | J0000abcde |
| 店舗名 | 焼肉 XXX YYY 店 |
| エリア名 (大) * | 愛知 |
| エリア名 (中) * | 名古屋駅周辺 |
| エリア名 (小) * | 名古屋駅西口 |
| ポータルサイトにおける店舗トップページ中のテキストのうち写真見出し、キャンペーン情報以外 | |
| ジャンルキャッチ * | 食べ放題&飲み放題の焼肉 |
| 店舗キャッチ † | 当日 OK ★焼肉食べ放題! 名駅徒歩 30 秒! |
| 店舗コピー † | ★満足な宴会は焼肉で決まり★ 【焼肉】食べ放題 & 【ドリンク】飲み放題=クーポン利用で 3000・4000 円・5000 円がございます。 コラーゲン付! 国産牛しゃぶしゃぶの【食べ放題】&ドリンク【飲み放題】もクーポン利用で 3500 円です。 贅沢な食材で至福の時間を満喫してください。 全席掘りごたつで、カップルはもちろん宴会にピッタリの空間をご用意!! 又、創業 30 年、味自慢の XXX では、カルビが絶対オススメ!! 是非、味・食感・香と 3 拍子揃ったうまい肉をお召し上がり下さい!! 皆様の御来店を、スタッフ一同心よりお待ちしております |
| 交通アクセス | 名駅 西口 (新幹線乗り場) を左にスグ! |
| コース名 | (a) 当日 OK ★焼肉A【食放】+【飲放】¥3990 → ¥3000 ※+¥500 でタン+サラダをご提供! |
| (最大 5 種類) | (b) [1] 当日 OK ★焼肉B【食放】【飲放】¥5040 → ¥4000[2] 黒毛和牛★納得コース【飲放】¥4515 → ¥3000 |
| | (c) [1] 当日 OK ★焼肉C【食放】【飲放】¥6090 → ¥5000[2] 黒毛和牛★極上コース【飲放】¥5565 → ¥4000 |
| | (d) 当日 OK ★黒毛和牛のしゃぶしゃぶ or すきやき【食放】【飲放】¥4515 → ¥3500 |
| 写真の見出し | |
| 店舗トップ † | 名駅新幹線口、徒歩 30 秒★ 2 名〜当日 OK の焼肉食放&飲放充実!! 毎月 29 日は何かが起こる! |
| お薦めメニュー (最大 3 種類) | (a) 2 名〜当日 OK の新登場メニュー★コラーゲンボール入り♪すき焼きの食べ放題でスタミナバッチリ♪♪ (b) 2 名〜当日 OK ★全 10 品のお気軽コースは、学生さん・サラリーマンにも大人気! (c) 2 名〜当日 OK ★ちょっと贅沢に会社宴会にも…「牛タン」も食べ放題♪ |
| 雰囲気 (最大 3 種類) | (a) 全席掘りごたつだから楽しんだね☆ (b) カップルで落ち着いた一時を♪ (c) なんてたって店内がキレイだから気持ちいいよね!! |
| キャンペーン情報 | |
| キャンペーンキャッチ † (最大 6 種類) | (a) 名駅徒歩 30 秒★ガッツリ焼肉食べ放題♪飲み放題♪ナント 3000 円〜 (b) スタミナ◎黒毛和牛のすきやき食べ放題&飲み放題が 3500 円! (c) 毎月【肉の日= 29 日】は XXX の日★毎月 29 日は、何かが起こる…!! (d) 名古屋駅新幹線口徒歩 30 秒★掘りごたつで宴会、最大 60 名迄 OK! (e) 黒毛和牛の飲放付コースが 4515 → 3000 円 ★ 5565 → 4000 円もご用意! (f) 名駅 30 秒★ワイワイ食べ放題&飲み放題で宴会! 最大 60 名迄! |
| キャンペーンカテゴリ名 * (最大 6 種類) | (a) 食べ放題プランのあるお店 (b) お手頃! 4000 円以下の飲み放題付コース (c) おいしいお肉が食べたい! (d) 50 人以上の宴会が可能なお店 (e) 3000 円以下で楽しめるコース料理 (f) 20 人以上の宴会が可能なお店 |

表 2 中の「†」を付した 4 種類の要素 (店舗キャッチ^{*1}, 店舗コピー^{*2}, 店舗トップの写真の見出し, キャンペーンキャッチ) は、消費者に伝えるべき最重要項目として、Hot Pepper の各店舗のトップページに掲載されているものである。そこで、これらを**基本テキスト**と呼

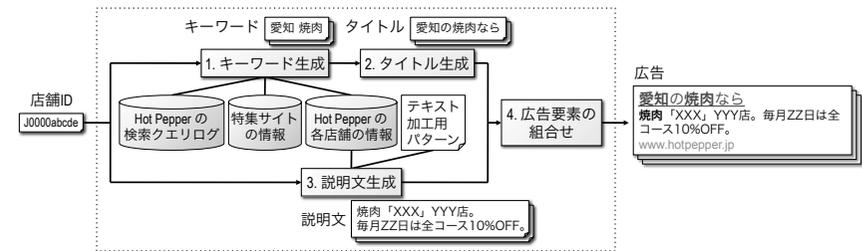


図 4 広告生成システムの概要
Fig. 4 Overview of our ad generator.

び、広告の説明文の生成元テキストとする。その他の要素は、広告の他の要素を生成する際に参照したり、説明文の生成元テキストの代替物として使用したりする。

4.2 各特集サイトに関する情報

各特集サイトに対して事前に付与した次の 2 種類の情報を用いる (表 1 を参照のこと)。

特集名: 特集サイトの URL の元になった日本語表現

特集タイプ: 特集名の分類であり、チェーン店、店舗種別、食べ物、状況、設備・サービスの 5 種類のいずれか

4.3 ポータルサイトにおける検索クエリログ

Hot Pepper のポータルサイトにおける検索クエリログも使用する。ただし、参照できるのは、検索クエリの文字列と、検索時に指定された検索対象タイプ^{*3}の組の情報のみである。各検索クエリの検索回数や検索結果数などは参照できない。

5. 検索連動型広告生成システム

各店舗に関する様々な情報に基づいて複数の広告を生成し、広告グループを用いてそれらを取捨選択しながら運用することによって、広告の自動生成の実現可能性および広告として有用な情報を明らかにすることをねらい、我々は、入力された店舗 ID に対して複数の広告を生成するシステムを作成した (図 4)。

広告の構成要素のうち、キーワードとタイトルは、飲食店ドメインにおける消費者の典型的な検索意図および検索連動型広告に関するこれまでの研究の知見を反映した、いくつかの

*1 店舗の特徴を簡潔に説明するテキスト。

*2 店舗の特徴を、記号や感情表現を用いつつ、冗長にならない程度に詳細に説明するテキスト。

*3 店舗名、駅・エリア、料理、フリーワードのいずれか。

表 3 地名の整形・抽出用のパターンの例
Table 3 Examples of patterns for trimming and extracting place names.

| 条件 | 整形結果 | 条件を満たすエリア名の例 |
|---------------------|---------------------|--------------|
| “XXX・YYY・ZZZ” | → “XXX” “YYY” “ZZZ” | 伊勢・志摩・鳥羽 |
| “XXX (YYY・ZZZ)” | → “XXX” “YYY” “ZZZ” | 田町 (三田・芝浦) |
| “XXX, YYY, その他 ZZZ” | → “XXX” “YYY” “ZZZ” | 茅野・上田・その他長野県 |
| “XXX 郊外” | → “XXX” | 岡山市郊外 |
| “XXX 周辺” | → “XXX” | 名古屋駅周辺 |

ヒューリスティクスに基づいて生成する。一方、説明文は、キーワードとタイトルの対とは独立に、各店舗に関するテキストデータを要約して生成する。

なお、広告サービスとして Overture を用いることを前提とし、本システムでは、タイトルおよび説明文の最大長 (各々15文字, 33文字), 使用可能, 不可能な文字などは, Overture の基準に沿うように調整した。

5.1 キーワードの生成

飲食店ドメインにおけるユーザの検索意図は、多くの場合、次のいずれかである。

- 条件に合う店舗を探す。
- 特定の店舗の詳細な情報を探す。

前者の場合、ユーザはまず地名を、そして随意的に店舗の種別名を検索クエリとして用いると考えられる。一方、後者の場合、店舗名のみで事足りるが、場所やチェーン店の支店を指定するために地名が用いられることもある。同様に、同じ名前を持つ異なる店舗の曖昧性を解消するために、店舗の種別名が用いられることもある。このような認識に基づいて、我々は、店舗の場所を表す地名と店舗の種別名の2つの文字列をスペースでつなげてキーワードとする。

所与の店舗のIDが入力されると、その店舗に対する地名のリストを次の手順で生成する。

- (1) 当該店舗の店舗データ (表2を参照のこと) から、店舗のあるエリア名 (大中小の3種類) を取り出す。たとえば、店舗のIDが「J0000abcde」の場合、「愛知」、「名古屋駅周辺」、「名古屋駅西口」の3つを取り出す。
- (2) 各エリア名を加工して、より簡潔な地名のリストを作成する。加工には、表3に示すようなパターンを用いる。今回はこのようなパターンを19種類作成した。
- (3) Hot Pepper のポータルサイトにおいて検索クエリとして用いられた地名のみを取り出す。

ステップ (3) で検索クエリログを参照するのは、ポータルサイトにおいて店舗を検索する

際に用いられる表現は同じ目的で検索エンジンを使用する際にも用いられる、という仮定に基づく。このようにして作成されたリストから1つ地名を取り出してキーワードとする。

店舗の種別名としては、特集サイトの特集名 (4.2節) を用いる。

5.2 タイトルの生成

キーワードの地名および店舗の種別名を用いてタイトルを生成する。具体的には、次のテンプレートをを用いる。

(2) 〈地名〉の〈店舗の種別名〉なら

これにより、「キーワードに地名を含める場合はタイトル、説明文にその地名そのものもしくはより具体的な地名を含めること」^{*1}、という Overture の基準も満たすことができる。

5.3 説明文の生成

Overture では、説明文の長さは最長33文字と定められている。ただし、長い表現を生成しようとする、文法的ではない表現を生成してしまいやすいと考えられる。そこで、説明文を次の2つの要素で構成する。

前半: 店舗名 (広告対象である店舗を消費者に宣伝するうえで最も効果的である)

後半: 自動生成した宣伝用テキスト

まず、店舗データ中の店舗名の文字列を加工して店舗名を特定し (5.3.1項)、店舗名が得られた場合のみ説明文の後半部分を生成する。後半の表現は、Overture の制約 (33文字) から店舗名の長さを減じたものを長さの上限とし、各店舗の基本テキスト (4.1節) に対して3種類の要約手法 (5.3.3項, 5.3.4項, 5.3.5項) を各々独立に適用して生成する。基本テキスト以外のいくつかのテキストデータに対しても、それらの有用性を調査するために、最も単純な手法 (5.3.3項) を適用する。なお、テキストデータの処理には、形態素解析器 MeCab^{*2} (辞書は IPADIC) および係り受け解析器 CaboCha^{*3}を用いた。

我々は、各店舗に固有の説明文の生成に主眼を置き、キーワードとは独立に説明文を生成する^{*4}。店舗固有のテキストから抽出できる各候補表現の良否については、機械学習に基づく手法^{6),10)}ではなく、各店舗に特徴的であること、および言語的に適格であることを基準として、文献11)で提案されたスコア関数 (5.3.2項) を用いて評価する。各店舗に対して

*1 http://listing.yahoo.co.jp/support/editorial/ss/05/compatibility_search.html

*2 <http://mecab.sourceforge.net/>

*3 <http://chasen.org/~taku/software/cabocho/>

*4 テキスト要約においては、クエリに応じた要約という問題設定⁹⁾も提案されている。広告生成においても、キーワードをクエリと見なして個別の説明文を生成することも考えられるが、今回は扱わない。

人間が作成した説明文も多様な説明文も存在せず、たとえば実際に広告サービスを用いてCTRを測定するなどしなければ広告の集客効果の大小を判断できないためである。

5.3.1 店舗名の抽出

店舗名は、検索結果画面に表示される際に1行に収まるように最大15文字とし、後半部分との境界を明示するために、末尾に句点を付加する。店舗名は飲食店ドメインの店舗データに必ず含まれるが、上記の制約よりも長い場合もある。そこで、各店舗が所属する特集サイトの特集タイプに応じて加工する。

当該店舗がチェーン店である場合、Hot Pepperの店舗データにおける店舗名は、チェーン名、支店名およびいくつかの修飾表現で構成され、次の例(3)に示す順に並べられている。

(3) AAA 居酒屋 ZZZ 料理 XXX YYY 店
 (修飾表現) (修飾表現) (チェーン名) (支店名)

伏字として用いている各アルファベットが実際の文字1文字に対応すると見なすと、例(3)の店舗名は、要素間のスペースを含めて全部で21文字からなる。「AAA 居酒屋」、「ZZZ 料理」のような修飾表現はチェーン店の特徴を伝えるうえで有用であると考えられるが、店舗名の要素としては、チェーン名「XXX」と支店名「YYY 店」の方が重要である。そこで、チェーン名に括弧を付けて強調し、修飾表現は、先頭から順に長さの制約を満たすまで削除する。上記の例の店舗名からは、例(4)に示す、スペースを含む14文字の文字列を得る。

(4) ZZZ 料理 「XXX」 YYY 店

チェーン店以外の特集タイプに関しては、いくつかの例を見た限り、店舗名が先頭に配置されることが多かった。そこで、チェーン店の場合とは逆に、要素を後ろから順に削除する。

5.3.2 スコア関数

ここでは、文献11)で提案された、任意の表現に対するスコア関数について述べる。我々は、後述の3種類の要約手法すべてにおいて、各候補表現の良否を定量化するために、このスコア関数を用いる。

店舗 x に対する説明文候補表現 s のスコア $S_s(x, s)$ を、次のように定義する。

$$S_s(x, s) = \sum_{(n_c, n_p) \in T(s)} S_r(x, n_c, n_p), \quad (4)$$

$$S_r(x, n_c, n_p) = S_e(n_c, n_p) \sum_{i \in \{c, p\}} S_n(x, n_i), \quad (5)$$

ここで、 $T(s)$ は s の依存構造を構成する文節対の集合である。 n_c, n_p は各々、係り側文節、受け側文節を意味する。 $S_r(x, n_c, n_p)$ は、1つの文節対のスコアである。 $S_e(n_c, n_p)$ および

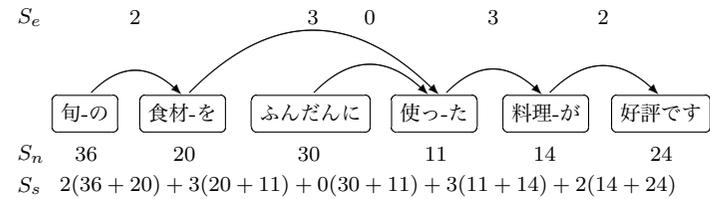


Fig. 5 Example of dependency tree of a sentence and its score.

表4 依存関係ごとのスコアの一覧
 Table 4 Score table for dependency edges.

| S_e | 依存関係の種類 | 例 ((n_c, n_p)) |
|-------|-------------------------------------|--|
| 3 | 動詞と格 サ変名詞と格 動詞の連体修飾 | “〈個室で, 楽しむ〉” “〈食材を, 使用〉” “〈くつろげる, 個室〉” |
| 2 | 形容詞の連体修飾 形容詞の述語用法 並列以外の名詞間の修飾 | “〈新鮮な, 魚〉” “〈気遣いが, うれしい〉” “〈旬の, 食材〉” |
| 1 | 並列名詞句 | “〈お造りと, 煮付け〉” |
| 0 | その他 | “〈ふんだんに, 使った〉” |

$S_n(x, n)$ は、各々、依存関係 (エッジ) および文節 (ノード) のスコアである。文の依存構造とそのスコアの例を図5に示す。

依存関係のスコア $S_e(n_c, n_p)$ は、各文節 n_c, n_p の種類に基づいて、表4に示すように定めた。述語とその項の間関係は、文の骨格をなす要素であるため、形容詞や副詞などの修飾要素よりも高いスコアを与える。

一方、文節のスコア $S_n(x, n)$ は次式で与える。

$$S_n(x, n) = \max_{w \in CW(n)} tf(x, w) \left[\log \left(\frac{N}{df(w)} \right) + 1 \right],$$

ここで、 $CW(n)$ は文節 n 中の内容語^{*1}の集合、 $tf(x, w)$ は店舗 x の基本テキストにおける語 w の出現回数、 $df(w)$ は全 N 店舗のうち基本テキストに語 w を含む店舗の数を表す。このように、情報検索で用いられる tf-idf に基づいて各内容語のスコアを定めることによつ

*1 各形態素の品詞情報に基づいて、非自立語を除く名詞、動詞-自立、形容詞-自立、副詞を対象とした。

表 5 説明文の刈り込みパターン
Table 5 Patterns for trimming descriptions.

| n_c | n_p | 表現数 |
|-------|--|-----|
| X の | { こだわり, カフェ, クレープ, コース, ブラン, メニュー, ランチ, ワイン, 宴会, 居酒屋, 魚, 空間, 鶏, 個室, 座敷, 焼酎, 食材, 数々, 鮮魚, 素材, 中, 店, 店内, 雰囲気, 味, 要望, 立地, 料理 } | 28 |
| X も | { OK, ある, おすすめ, できる, 可能, 楽しめる, 好評, 使える, 充実, 人気, 揃う, 多い, 豊富, 用意 } | 14 |
| X が | { おすすめ, オススメ, 円, 可能, 楽しめる, 自慢, 人気, 揃う, 登場, 豊富, 味わる } | 11 |
| X を | { いただく, 楽しむ, 楽しめる, 感じる, 使う, 提供, 用意 } | 7 |
| X に | { こだわる, どうぞ, 応える, 合う } | 4 |
| X まで | { OK, 用意 } | 2 |
| X から | { 選べる } | 1 |

て、各店舗に固有の語に高いスコアを与える。

5.3.3 手法 1：文と句の抽出

1つ目の手法では、与えられたテキストから、次の手順で説明文をただか1つ生成する。

- (1) テキストを切り分けて文のリスト S を得る。
- (2) S から長さの制限よりも長い文を取り除く。
- (3) S 中の、式 (4) のスコアが最も高い文を出力する。

この手法は、店舗データの有用性を調査するため、基本テキストだけでなく次の6種類の要素(表2を参照のこと)にも適用する。

- 交通アクセス(「XXX分」を含む場合は該当部分より後ろの文字列を削除)
- コース名(「XXXコース」を含む場合は該当部分より後ろの文字列を削除)
- お薦めメニューの写真の見出し
- 雰囲気の写真の見出し
- ジャンルキャッチ
- キャンペーンカテゴリ名

5.3.4 手法 2：依存構造上の部分構造の抽出

手法1では長さの制限よりも長い文を捨ててしまうが、文の一部に広告として有用な表現を含む長い文が存在する場合もある。手法2では、文の刈り込みパターンの適用と部分構造の列挙によって、そのような表現の抽出を試みる。ここでいう刈り込みパターンとは、広告の説明文として有用そうな末尾表現を抽出するためのものである。今回は、後述の評価実験に用いていない150件の店舗データを参照して、表5に示すような、依存構造上の1つの文節対を表し、受け側文節(n_p)における表現を条件として持つパターンを、合計7種類(n_p の表現はのべ67種類)作成した。

具体的には、ある店舗の基本テキストが与えられると、次の手順で説明文をただか1つ

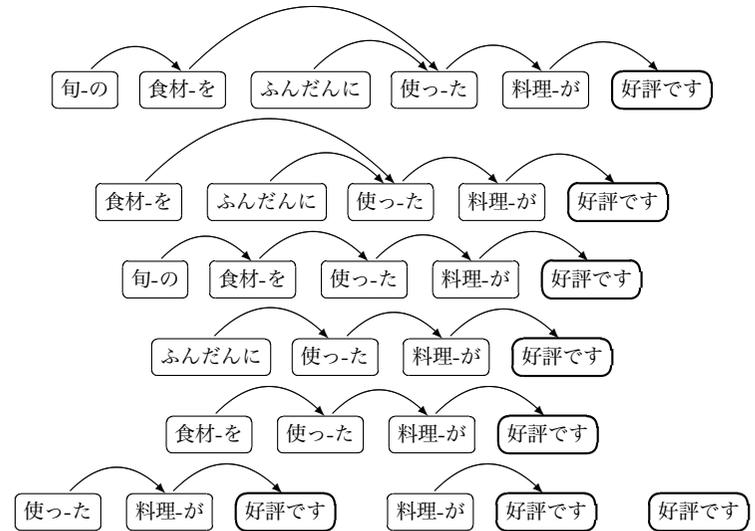


図 6 図 5 の依存構造の根にあたる文節を含む部分構造の一覧
Fig. 6 Dependency subtrees containing the root node of original dependency tree in Fig. 5.

生成する。

- (1) テキストを切り分けて文のリスト S を得る。
- (2) S 中の各文を構文解析して、依存構造のリスト DS_1 を得る。
- (3) DS_1 中の各依存構造と文の刈り込みパターンの各々を照合し、パターンに適合する場合は、パターンの受け側の文節(n_p)を根とする依存構造を DS_1 に追加する。
- (4) DS_1 中の各依存構造について、その構造の根にあたる文節(日本語の場合は末尾の文節)を含むあらゆる部分構造を列挙し¹⁰⁾、新たなリスト DS_2 を作成する。たとえば図5の依存構造からは、図6に示す8つの部分構造を抽出する。
- (5) DS_2 中の各依存構造について、対応する表層表現の長さを計算し、制限よりも長い場合は、その依存構造を取り除く。
- (6) DS_2 中の、式(4)のスコアが最も高い部分構造を選択し、その表層表現を出力する。

5.3.5 手法 3：文の再生成

手法2を適用するには、ドメインに応じて刈り込みパターンを人間が作成しなければならない。また、依存構造の根にあたる文節を含む候補しか生成できない。より挑戦的な方法

表 6 生成する説明文の一覧
Table 6 Inventory of generated descriptions.

| タイプ | 説明文の生成元テキスト | 手法 |
|-------|----------------|----|
| d_1 | | 1 |
| d_2 | 基本テキスト | 2 |
| d_3 | | 3 |
| d_4 | 交通アクセス | 1 |
| d_5 | コース名 | 1 |
| d_6 | お薦めメニューの写真の見出し | 1 |
| d_7 | 雰囲気の写真の見出し | 1 |
| d_8 | ジャンルキャッチ | 1 |
| d_9 | キャンペーンカテゴリ名 | 1 |

として、手法 3 では、基本テキスト中のすべての依存構造の断片から表現を生成しなおすことを試みる。

具体的には、ある店舗の基本テキストが与えられると、次の手順で説明文をただか 1 つ生成する。

- (1) テキストを切り分けて文のリスト S を得る。
- (2) S 中の各文を構文解析し、依存構造中のすべての文節対の集合 $E = \bigcup_{s \in S} T(s)$ を得る。
- (3) E の中から、式 (5) のスコアが最も高い文節対 e_{core} を選択し、それを E から取り除く。
- (4) 次のいずれかに該当する文節対を E から取り除く。
 - e_{core} とは異なる文から抽出された。
 - スコアが 0 である。
 - いずれかの文節に代名詞を含む。
 - 係り側文節に動詞、非自立語の名詞、副詞的な名詞のいずれかを含む。
 - e_{core} と文節を共有しない。
- (5) E 中の文節対 e を、式 (5) のスコアの低い順に、それを接続したとしても対応する表層表現の長さが制限を超えない限り、 e_{core} に接続する。
- (6) 生成した依存構造の表層表現を出力する。

ステップ (4) における条件付けは、非常に保守的に見えるが、非文法的な表現を生成しないために導入したものである。

5.4 広告要素の組合せ

n 種類のキーワード、タイトルの対と、最大 9 種類の説明文 (表 6) を網羅的に組み合わせ、最大 $9n$ 種類の広告を生成する。広告のその他の要素は次のように定める。

- 表示用 URL : Hot Pepper のポータルサイトの URL (www.hotpepper.jp) .
- ランディング URL : Hot Pepper のポータルサイトにおける対象店舗のページの URL (http://www.hotpepper.jp/str<対象店舗の ID>/) .

6. 広告生成能力の評価

本章では、システムの広告生成能力に関する評価について述べる。まず、広告の各構成要素の生成数を示し、その後、人間による広告の適否の評価に基づく説明文の精度および被覆率を示す。なお、全構成要素の組合せに関する評価は今後の課題とする。たとえば、タイトルと説明文が偶然同じ表現を含む可能性はあるが、今回はそれに言及しない。

6.1 広告の自動生成

まず、全 200 種類の特集サイトからランダムに 20 種類を選択し、各特集サイトから国内の 9 つの地域^{*1}の各々における最大 10 店舗を取り出した。地域によっては登録店舗数が 10 未満の場合もあるため、2009 年 10 月 27 日時点で抽出できた店舗の数は 575 であった。

次に、個々の店舗に対する広告の各要素を図 4 の手順で生成した。詳細は次のとおりである。

- 店舗データとして、2009 年 10 月 27 日時点の 25,815 店舗分 ($N = 25,815$) を使用。
- 検索クエリログとして、2009 年 8 月に収集された 482,700 行を使用。
- 各店舗に対して、5.1 節の方法で抽出した地名からランダムに 2 種類を選択。

6.2 各要素の生成数

575 店舗すべてに対して、2 種類の地名を抽出し、キーワードを生成することができた。地名および店舗の種別名を用いて生成したタイトルも、すべて Overture の文字数制限を満たしていた。一方、説明文を 1 つも生成できない店舗が 62 店舗 (11%) あった。説明文の前半部分として用いる店舗名を抽出する時点で、15 文字という制限を満たす文字列を抽出できないことが原因であった。この問題は、説明文の長さの制限を緩める、または説明文として用いる短い店舗名を店舗データに新たに含めることによって、回避することができる。

説明文の後半部分は、店舗名を抽出できた 513 店舗についてのみ生成した。各手法で生成された説明文の数を表 7 の「生成数」の列に示す。同じ店舗に対する d_1 , d_2 , d_3 が偶然同じ表現となる場合もあるため、説明文の異なり数は 3,775 であった。基本テキストからは、3 種類の手法のいずれを用いた場合でも、もれなく説明文を生成できた。一方、ジャン

*1 北海道、東北、関東、北陸甲信越、東海、関西、中国、四国、九州。

表 7 説明文生成の結果
Table 7 Results of generating descriptinos.

| 広告タイプ | 生成数 | 適格 | 不適格 A | 不適格 B | 精度 | 被覆率 |
|-------|-----|-----|-------|-------|------|------|
| d_1 | 513 | 496 | 16 | 1 | 0.97 | 0.86 |
| d_2 | 513 | 428 | 77 | 8 | 0.83 | 0.74 |
| d_3 | 513 | 400 | 88 | 25 | 0.78 | 0.70 |
| d_4 | 219 | 219 | 0 | 0 | 1.00 | 0.38 |
| d_5 | 290 | 234 | 56 | 0 | 0.81 | 0.41 |
| d_6 | 470 | 450 | 20 | 0 | 0.96 | 0.78 |
| d_7 | 487 | 479 | 8 | 0 | 0.98 | 0.83 |
| d_8 | 513 | 513 | 0 | 0 | 1.00 | 0.89 |
| d_9 | 469 | 469 | 0 | 0 | 1.00 | 0.82 |

ルキャッチ (d_8) を除く他のテキスト要素からは、説明文を生成できない場合があった。とくに、交通アクセス (d_4) やコース名 (d_5) からは、説明文の生成を試みた 513 店舗のうち 43% (219/513)、57% (290/513) に対してしか説明文を生成できなかった。表 2 に示したように、広告としては使用できない記号を多く含んでいたことが原因であった。

6.3 説明文の精度と被覆率

自動生成された説明文の適否を、次の手順で評価した。

設問 1: 説明文は文法的か? → Yes: 設問 2 へ, No: 「不適格 A」

設問 2: 説明文は広告として適切か? → Yes: 「適格」, No: 「不適格 B」

これらの評価には、飲食店ドメインにおける 5 年間の広告作成の経験がある者 1 名が従事した。設問 2 においては、ユーザの誤解を招かないかどうかという点も評価させた。

評価の結果は表 7 の「適格」、「不適格 A」、「不適格 B」の列に示すとおりである。手法 2, 3 によって生成した d_2 , d_3 に不適格な説明文が多かった。不適格な説明文の例を (5) に示す。

- (5) a. 付き選べる鍋コース 4780 円最大 (不適格 A, d_3)
 b. 2500 円、飲み放題+500 円 (不適格 B (価格の対象が不明), d_2)
 c. まずはこれを食べてほしい。 (不適格 B (指示対象が不明), d_5)

適否の判定結果に基づいて次の式で算出した値を表 7 に示す。

$$\text{精度} = \frac{\text{適格な説明文の数}}{\text{生成した説明文の数}}$$

$$\text{被覆率} = \frac{\text{適格な説明文を生成できた店舗の数}}{\text{対象店舗数 (= 575)}}$$

不適格と判定された説明文の異なり数は 299 であったので、全体の精度は 92% (1-299/3,775) となった。また、説明文の後半部分の生成を試みた 513 店舗すべてに対して、1 つ以上の適格な説明文を生成できたので、全体の被覆率は 89% (513/575) となった。我々のシステムが、多くの店舗に対して、説明文として使用するテキストを生成できることが確認できた。なお、店舗あたりの適格な説明文の平均数は 6.8 種類 ((3,775-299)/513) となった。

被覆率は、各手法で n ベスト解を出力したり、店舗データ中の他の要素を用いたりすることで、向上させることができる。精度に関しては、今回の評価実験を通じて適否のラベル付きの説明文が収集できたため、これらを教師信号として分類器を構築することで改善できる可能性がある。

7. 自動生成した広告の集客効果の評価

本章では、自動生成した広告の集客効果に関する評価について述べる。

7.1 実験の設定

個々の広告の表示回数や CTR はキーワードによって著しく異なる。そこで、キーワードおよび対象店舗が同一である広告を、2.1 節で述べた広告グループを用いてひとまとめにする。同一の広告グループ内の広告は排他的にしか表示されず、ユーザによるクリックの状況に応じて徐々に表示回数に差は出るものの、短い期間ではその差はそれほど大きくならない。そこで、個々の広告の集客効果を定量的に示す指標として CTR を用いる。具体的には、個々の広告の CTR の大小を、同一の広告グループ内のベースライン広告と CTR を比較することで評価する。任意の 2 つの CTR の差の有意性は、2 群の比率差の検定 (連続性の補正あり、片側検定) によって判定する*1。有意水準は 5% とする。

6 章で対象とした 575 店舗のうち、20 種類の特集サイトからランダムに選択した 5 つのいずれかに属する全 105 店舗を対象とした。これらの店舗に対する広告の各要素は、各店舗に対するキーワードとタイトルの組が 2 種類ずつと、合計 709 種類の適格な説明文であった。自動生成した説明文に焦点を当てた評価を行うために、各店舗に対する説明文の後半の

*1 実験期間が長いほど広告の表示回数 (比率の分母) が増え、差が有意になりやすくなるため、CTR の差の有意性は評価のうえでは本質的ではない。また、広告グループ内の CTR が低い広告は次第に表示されにくくなるため、顕著に CTR が低い広告があったとしても、損害にはならない。

表 8 CTR に基づく広告の対比較結果
Table 8 Pairwise comparison of ads based on CTR.

| 広告対 | 対の数 | $CTR(a_i) > CTR(a_0)$ | | $CTR(a_i) < CTR(a_0)$ | | $CTR(a_i) = CTR(a_0)$ |
|-----------------|-----|-----------------------|------------|-----------------------|------------|-----------------------|
| | | すべて | $p < 0.05$ | すべて | $p < 0.05$ | |
| a_1 vs. a_0 | 210 | 73 (35%) | 7 (3%) | 102 (49%) | 15 (7%) | 35 (17%) |
| a_2 vs. a_0 | 158 | 53 (34%) | 4 (3%) | 76 (48%) | 11 (7%) | 29 (18%) |
| a_3 vs. a_0 | 164 | 73 (45%) | 5 (3%) | 64 (39%) | 13 (8%) | 27 (16%) |
| a_4 vs. a_0 | 80 | 26 (33%) | 1 (1%) | 40 (50%) | 5 (6%) | 14 (18%) |
| a_5 vs. a_0 | 94 | 31 (33%) | 3 (3%) | 47 (50%) | 6 (6%) | 16 (17%) |
| a_6 vs. a_0 | 184 | 79 (43%) | 3 (2%) | 80 (43%) | 10 (5%) | 25 (14%) |
| a_7 vs. a_0 | 196 | 71 (36%) | 5 (3%) | 95 (48%) | 14 (7%) | 30 (15%) |
| a_8 vs. a_0 | 210 | 79 (38%) | 5 (2%) | 93 (44%) | 18 (9%) | 38 (18%) |
| a_9 vs. a_0 | 196 | 76 (39%) | 3 (2%) | 91 (46%) | 9 (5%) | 29 (15%) |

表現のベースライン*1 d_0 を, (6) に示すテンプレート*2を用いて生成した.

(6) 〈店舗の種類名〉ナビでお店探し

合計 814 種類の説明文と 2 種類のキーワード, タイトルを組み合わせると 1,628 種類の広告を生成し, 対象店舗およびキーワードごとにまとめて合計 210 種類 (105 店舗, 2 種類のキーワード) の広告グループを作成した. そしてこれらを Overture に登録し, 1ヵ月強の期間 (2009 年 11 月 18 日から 12 月 26 日までの 39 日間), 各広告が表示された回数およびクリックされた回数を計測した. なお, クリック単価の上限値は一定とした.

7.2 実験結果

説明文の後半が d_i であるような広告を a_i と記す. 我々のシステムによって自動生成した広告とベースライン広告 a_0 の CTR の対比較の結果を表 8 に示す. 少数ではあるが, a_0 よりも顕著に高い CTR を持つ広告をいくらかは生成できていたことが確認できた. 6 章の評価では, a_3 は適格でない場合が多いことが明らかになったが, 適格な説明文については, a_1, a_2 よりも高い割合の広告グループ (量は同程度) において, a_0 よりも高い CTR を示した. また, a_3 のみ, a_0 よりも CTR が高い広告の数が, a_0 よりも CTR が低い広告の数

*1 店舗データを人手で加工したものを比較対象とすることも考えられるが, 次の 3 つの理由からそのような実験は行わなかった. (i) 人間が作成した広告が理想的であるとは限らず, 恣意性も排除できないため, 実験の位置付けが不明瞭になり, 公平性も損なわれる. (ii) 3.2 節で述べたように広告作成作業にかかるコストは相当の規模となるが, それ以上に利益を得ることが保証されていないため, 実験の実施が困難である. (iii) 本論文の目的は, テンプレートに基づく従来手法からの改善の可否を明らかにすることである.

*2 このテンプレートは, Hot Pepper のブランドイメージの保護のため新規に作成したものである. ただし, 店舗の種類名を含み, 店舗に固有の情報を含まないという点は, 実験の実施時に運用されていたものと同じである.

表 9 a_0 よりも高い CTR を獲得した広告の数と CTR の最高値が改善された広告グループの数
Table 9 Ads and groups that gained a higher CTR.

| 集計単位 | 広告 | 広告グループ |
|------------|-------------|-----------|
| N | 1,418 | 210 |
| すべて | 532 (38.9%) | 168 (80%) |
| $p < 0.05$ | 35 (2.7%) | 21 (10%) |

を上回った. 他のテキスト要素から生成した広告も, a_4 と a_5 を除いて, a_1 や a_3 と同程度の量の広告グループにおいて, a_0 よりも高い CTR を獲得した. a_8 や a_9 の生成に用いたテキストは, 店舗の種類やキャンペーンの種類ごとに記述されたものであり, 特定の店舗に固有の表現ではないが, CTR を改善するうえで有用な場合もあることが明らかになった.

a_0 よりも CTR が高かった広告の数, そのような広告が 1 種類以上存在した広告グループの数を表 9 に示す. 我々のシステムによって自動生成した 1,418 種類の広告のうち 532 種類 (38.9%) が, 同じ広告グループの a_0 よりも高い CTR を獲得した. しかし, 顕著に高い CTR を獲得した広告は 35 種類のみであった. このように, 広告単位で見た場合, CTR を改善できた割合は十分に高いとはいえないが, 広告グループ単位で見た場合, 80% の広告グループに対して a_0 よりも CTR が高い広告を 1 つ以上生成できた. 広告サービスは CTR が高い広告を優先的に表示しようとする (2.1 節) ため, 各広告グループにおいて既存の広告よりも高い CTR を持つ広告が 1 つでも得られれば, 理論上は当該広告グループの CTR は改善される. ただし, 今回の設定とは異なり, 広告グループに登録できる数よりも多くの広告を生成したい場合や, 運用のための人的なコストをかけられない場合も考えられる. したがって, 高い CTR を持つであろう広告をあらかじめ自動的に選択するための技術の開発が今後の課題としてあげられる.

広告の表示回数や CTR は, 予想どおりキーワードによって著しく異なっていた (表示回数が最多のものと最少のものでは 200 倍, CTR は 0% のものもあるため算出不可). そこで, 広告グループ内の CTR の比に基づく CTR の改善率を算出した. 210 広告グループのうち a_0 が 1 度以上クリックされた ($CTR(a_0) > 0$) のは 157 件であった. これらの広告グループの各々において CTR が最も高い広告と a_0 の CTR の比を求めたところ, 平均は 1.85 (標準偏差 1.21, 最大 7.98, 最小 1.00 ($CTR(a_0)$ が広告グループ内で最大)) であった. すなわち, 各広告グループにおいて CTR が最も高い広告を十分に短い時間で特定できれば, CTR を約 2 倍に改善できると解釈できる. ただし, そのような判断のためには, CTR の変動や安定性に関する分析が不可欠である.

参考までに、 a_0 に対する CTR の比が最も高かった広告の説明文を例 (7) に示す。ただし、“X” は抽出された各店舗名の 1 文字に対応する。

- (7) a. XXXXXXXXXXXX。MED のディナーコースが女性に大好評！
(特集名：バイキング, d_1 , 30 文字)
- b. XXXXXXXX。刺身桶盛りは、豪快さに圧巻！ (特集名：日本酒, d_3 , 22 文字)

8. おわりに

本論文では、飲食店ドメインを対象として作成した、検索連動型広告の自動生成システムについて述べた。このシステムは、飲食店ドメインにおける検索連動型広告に関する知見およびポータルサイト向けに記述された各店舗に固有のテキストデータを活用して、個々の店舗に対する複数の広告を生成する。自動生成した広告の適否を広告業務従事者が判定した結果、我々のシステムが、多くの店舗に対して、広告として使用しうるテキストを生成できることが確認できた。さらに、自動生成した広告を実際の広告サービスを通じて運用し、広告の CTR を測定した結果、我々のシステムが、テンプレートに基づく画一的な広告よりも高い CTR を獲得できる広告を生成しうることが確認できた。以上より、個々のコンテンツに関する既存のテキストデータを利用して検索連動型広告を自動生成することはできると結論付ける。本論文で述べたシステムは、完全な自動化技術であるとはいえないが、人間による広告作成の支援技術となりうると考えられる。

検索連動型広告の自動生成という試みは先に例がないため、本論文での評価結果をふまえて、今後様々な方向での発展が期待される。まずは、適格な説明文、高い CTR を持つであろう広告の自動判別が挙げられる。実用面では、Hot Pepper の他の特集サイトの店舗に対する広告を生成することや、店舗データの更新に応じて広告を生成し、長期的に運用することも考えられる。人間が作成すべきパターンなどの資源を効率的に開発する方法の検討も重要な研究課題である。これには、開発したシステムを他のドメインに援用してみる必要がある。『口コミ』が非常に高い宣伝効果をもたらすことはよく知られているため、消費者がコンテンツに関して記述したテキスト (Consumer Generated Media; CGM あるいは User Generated Contents; UGC) の活用も興味深い。ただし、広告主や広告代理店のブランドイメージを損なわないよう、慎重に調査する必要がある。

謝辞 株式会社リクルートの上手亮氏、石山洸氏、玉地理氏、岩本宣武氏、横川睦氏、片岡亮氏に、データ提供、自動生成した説明文の適否の判定、CTR の計測実験において、お力添えをいただきました。本論文の査読者の方々には、論文を精読していただき、有益なご

指摘を賜りました。皆様に深く感謝いたします。

参考文献

- 1) Abhishek, V. and Hosanagar, K.: Keyword generation for search engine advertising using semantic similarity between terms, *Proc. 9th International Conference on Electronic Commerce (ICEC)*, pp.89–94 (2007).
- 2) Broder, A., Ciccolo, P. and Gabrilovich, E.: Online expansion of rare queries for sponsored search, *Proc. 18th International World Wide Web Conference (WWW)*, pp.511–520 (2009).
- 3) Chang, W., Pantel, P., Popescu, A.-M. and Gabrilovich, E.: Towards intent-driven bidterm suggestion, *Proc. 18th International World Wide Web Conference (WWW)*, pp.1093–1094 (2009).
- 4) Chen, Y., Xue, G.-R. and Yu, Y.: Advertising keyword suggestion based on concept hierarchy, *Proc. 1st ACM International Conference on Web Search and Data Mining (WSDM)*, pp.251–260 (2008).
- 5) 電通：2009 年日本の広告費 (2010).
<http://www.dentsu.co.jp/news/release/2010/pdf/2010020-0222.pdf>
- 6) Filippova, K. and Strube, M.: Dependency tree based sentence compression, *Proc. 5th International Natural Language Generation Conference (INLG)*, pp.25–32 (2008).
- 7) Jansen, B.J. and Resnick, M.: Examining searcher perceptions of and interactions with sponsored results, *Proc. Workshop on Sponsored Search Auctions, 6th ACM Conference on Electronic Commerce (EC)* (2005).
- 8) Joshi, A. and Motwani, R.: Keyword generation for search engine advertising, *Proc. 6th IEEE International Conference on Data Mining (ICDM) Workshop*, pp.490–496 (2006).
- 9) Mani, I. and Bloedorn, E.: Machine learning of generic and user-focused summarization, *Proc. 15th National Conference on Artificial Intelligence (AAAI)*, pp.821–826 (1998).
- 10) Nomoto, T.: A generic sentence trimmer with CRFs, *Proc. 46th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp.299–307 (2008).
- 11) 岡満美子, 小山剛弘, 上田良寛: 句表現要約の句合成手法, 情報処理学会研究報告, NL-129-15, pp.101–108 (1999).
- 12) Ravi, S., Broder, A., Gabrilovich, E., Josifovski, V., Pandey, S. and Pang, B.: Automatic generation of bid phrases for online advertising, *Proc. 3rd ACM International Conference on Web Search and Data Mining (WSDM)*, pp.201–210 (2010).
- 13) Richardson, M., Dominowska, E. and Ragno, R.: Predicting clicks: Estimating the click-through rate for new ads, *Proc. 16th International World Wide Web Confer-*

ence (WWW), pp.521–529 (2007).

- 14) Wang, X., Broder, A., Fontoura, M. and Josifovski, V.: A search-based method for forecasting ad impression in contextual advertising, *Proc. 18th International World Wide Web Conference (WWW)*, pp.491–450 (2009).

(平成 22 年 9 月 27 日受付)

(平成 23 年 3 月 7 日採録)



藤田 篤 (正会員)

1977 年生. 2005 年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了. 博士 (工学). 京都大学大学院情報学研究科産学官連携研究員, 名古屋大学大学院工学研究科助手, 同助教を経て, 2009 年より公立ほこだて未来大学システム情報科学部准教授. 現在に至る. 自然言語処理, 主に言い換え表現の生成と認識, テキスト生成の研究に従事. 言語処理学会, 人工知能学会, ACL 各会員.



幾島 克洋

1985 年生. 2010 年名古屋大学大学院工学研究科博士前期課程修了. 同年 (株) NTT コミュニケーションズ入社. 現在に至る. 映像配信の業務に従事.



佐藤 理史 (正会員)

1960 年生. 1988 年京都大学大学院工学研究科博士後期課程電気工学第二専攻研究指導認定退学. 京都大学工学部助手, 北陸先端科学技術大学院大学助教授, 京都大学大学院情報学研究科助教授を経て, 2005 年より名古屋大学大学院工学研究科教授. 工学博士. 自然言語処理, 人工知能, 情報の自動編集等の研究に従事.