

クローラおよび機械学習を用いた観光ポータルサイトの提案

—小規模宿泊施設の ICT プロモーションのために—

Proposal of the Tourism Portal Site Using a Crawling and Machine Learning
—For the Purpose of ICT Promotion Strategy of Small-Scale Accommodations—尾崎剛 (Takeshi OZAKI)¹・倉田紀子 (Noriko KURATA)¹・土屋健 (Takeshi TSUCHIYA)²・
三代沢正 (Tadashi MIYOSAWA)³・山田哲靖 (Tetsuyasu YAMADA)³・広瀬啓雄 (Hiroo HIROSE)³¹諏訪東京理科大学経営情報学部 助教・²諏訪東京理科大学経営情報学部 講師
³諏訪東京理科大学経営情報学部 教授

[Abstract]

There is a disparity in a tourist's destination among Japan. This study is conducted for the purpose of construction of the portal site in order to resolve the problems with Information and Communication Technology (ICT) promotion of small-size accommodations. An analysis was performed on the public relations activities via the Internet, which was targeted accommodations within Suwa area in 2015. The results were revealed the current condition of the ICT promotion of the subjects. In addition, these showed a statistical difference by accommodations in recognition to ICT promotion. Furthermore, this study proposed a system of automated portal site using a crawling and machine learning in order to resolve the ICT promotion problems.

[キーワード]

観光ポータルサイト、地域情報化、機械学習、web クローリング、小規模宿泊施設

1. はじめに

2000年代におけるパーソナルコンピュータの低価格化および普及、さらに ICT (Information and Communication Technology) インフラが整備されたことにより、インターネット上の様々なサービスやアプリケーションを比較的容易に利用できるようになってきた。2014年における調査では、日本におけるインターネット人口普及率は82.8%であり、インターネット利用者のうち毎日1回以上インターネットを利用する割合が74.1%となっている[1]。また、近年利用者数が著しく伸びているアプリケーションがソーシャルメディアである。ソーシャルメディアとは、インターネットを利用した情報発信やユーザ同士のやりとりが可能な双方向のメディアである。Facebook や Twitter などの SNS (Social Networking Service) やブログは、ソーシャルメディアの代表例であり、世界的に利用者が増加している[1]。このようなインターネットの普及が、観光産業にも大きな影響を与えている。

現在、観光情報を発信するトラベルサイトや旅館・ホテル・民宿等(以下、宿泊施設)の予約が可能なウェブサイトは多数存在する。また、多くの宿泊施設では自社のウェブサイトや SNS、ブログなどを有しており、それらを通じて観光情報を発信している。財団法人日本交通公社は、日本人観光客が旅行を計画する際の観光地情報を収集する手段を調査し、『旅行年報 2015』として発行した[2]。図1は、旅行計画時の主な情報源を男女別にまとめたものである。男女ともに「ネットの検索サイト」と回答した割合が最も高くなっている。また、「宿泊施設のホームページ」や「ネットの旅行専門サイト」、「観光施設のホームページ」など、インターネットを利用しての情報収集が多くなっていることが分かる。ガイドブックやパンフレット、旅行雑誌といった紙媒体よりも多くの人がインターネットを利用して観光情報を収集していることが分かる。

『旅行年報 2015』では、宿泊施設の稼働率も調査された(表1)。日本全国の年間客室稼働率は、57.4%である。従業員数に着目すると、従業員数が10~29人では65.9%、30~99人では68.4%、100人以上では72.6%であり、従業員数が多いほど客室稼働率が高い。従業員数が10人未満のデータは明らかになっていないが、全体との比較により50%未満であることが予想できる。また、宿泊施設のタイプ別に見ると、ビジネスホテルおよびシティホテルの稼働率がそれぞれ72.1%、77.3%と高く、旅館の稼働率が35.2%と低い。さらに、首都圏のビジネスホテルの稼働率は、東京84.2%、大阪83.2%、京都81.2%と、他の地域と比較して高いことが明らかにされている。このように、宿泊施設の規模やタイプ、立地場所によって格差が生じていることは明らかである。

このような格差を生じさせる原因の一つが、ICTを利用した広報活動 (ICTプロモーション) であると考えら

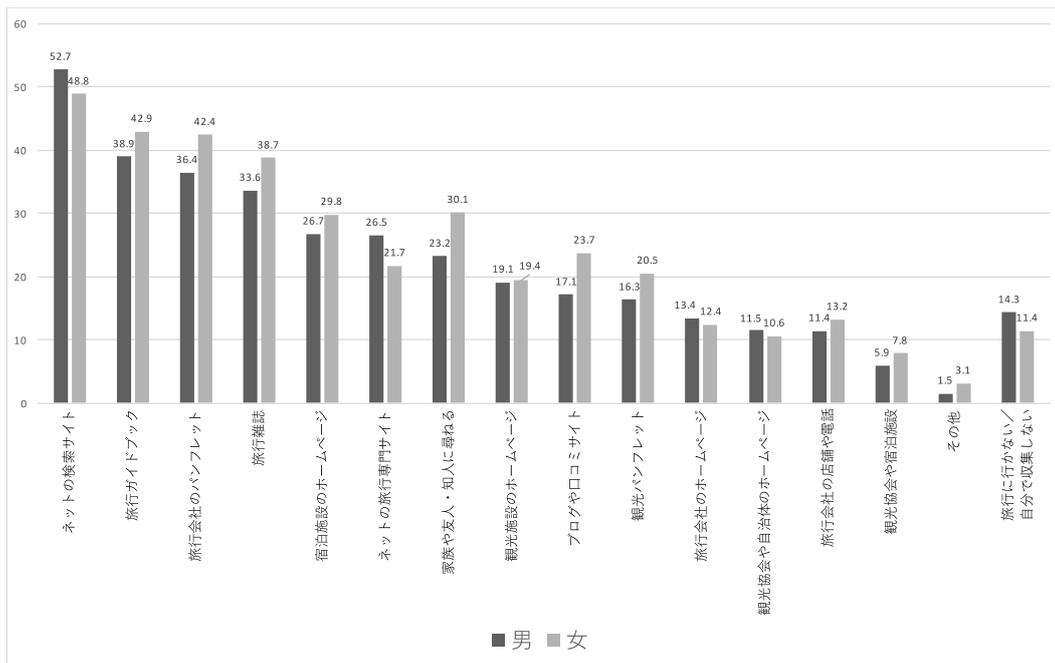


図1. 旅行の計画を立てる際の主な情報収集源 (〔2〕より著者が作成)

れる。地域ごとの観光協会や個別の宿泊施設がウェブサイトを開設したとしても、それら単独では掲載する情報量や更新頻度に限界があり、利用者にとって利便性の高いウェブサイトとはなりえず、十分な広報の効果を得ることは困難である。この困難への対応策として注目されているのが、地域の観光情報を収集したポータルサイト（観光ポータルサイト）である。ポータルサイトとは、インターネット上の様々なタイプのサービスやコンテンツにつなぐ役割を持ったウェブサイトである。観光ポータルサイトを用いることにより、利用者に詳細な密度の高い情報を提供できることが指摘されている[3]。宿泊施設が個別に開設したウェブサイト等に掲載された観光情報を、観光ポータルサイトに集約し、情報量や更新頻度の少ない小規模宿泊施設を集合体として扱うことにより、大規模宿泊施設や都市圏の観光地に匹敵する ICT プロモーションを展開することが可能となる。

観光ポータルサイトに関する先行研究として、田村は、観光情報を含む地域ポータルサイトの構築および運用に関する研究を行い、ポータルサイトの管理における最も重要な点は情報を収集するシステムであると主張した[4]。情報収集の自動化に関する研究が進められ、大槻らは地方自治体のウェブサイトおよびそこにリンクされた地域の情報を発信するウェブサイトを対象とするウェブディレクトリを自動的に構築するシステムを提案した[5]。遠藤らは、地域ウェブサイトやブログから、観光情報を抽出し融合する手法を提案した[6]。さらに石野らは、観光ブログのエントリから自動的に観光情報を収集するシステムを提案した[7]。これらポータルサイトでは、情

表1. 2014年客室稼働率 (〔2〕より著者が作成)

		客室数 稼働率 (%)
全体		57.4
従業者数別	10～29人	65.9
	30～99人	68.4
	100人以上	72.6
宿泊施設 タイプ別	旅館	35.2
	リゾートホテル	54.0
	ビジネスホテル	72.1
	シティホテル	77.3
会社・団体の宿泊所		28.8

報を収集するだけでなく、適切なカテゴリに分類し内容を更新する過程が必要であるが、従来の研究では情報収集に関する提案が中心であり、収集した情報の分類と更新については人的資源を割かなければならない問題が残されていた。

本研究は、小規模宿泊施設が展開する ICT プロモーションの効果を向上させるためのポータルサイトの構築を目的とする。このために、地方観光地における宿泊施設のインターネット利用状況に関する調査を、諏訪圏の宿泊施設を対象に行ない、実態を分析した。長野県は、宿泊施設の数が多い県であり、その 86.6%が従業員数 10 人未満の小規模宿泊施設となっている[8]。長野県の中央部に位置する諏訪圏は、岡谷市、諏訪市、茅野市、下諏訪町、富士見町、および原村で構成され、上諏訪温泉や蓼科高原などに加え、諏訪大社や御柱祭といった有数の観光資源を有する地方観光地である。この諏訪圏の宿泊施設を対象とすることで、地方の宿泊施設が ICT プロモーションをどのように捉え、実施しているのかを明らかにする。さらに、地方の宿泊施設が SNS やブログなどで発信した情報を、自動的に収集するだけでなく、分類および更新のプロセスをも自動化されたポータルサイトのシステムを提案する。

2. 諏訪圏における実態調査

2. 1. 調査概要

- (1) 調査対象：諏訪湖周辺の宿泊施設（上諏訪温泉 31 件、下諏訪温泉 10 件）41 件中 30 件
- (2) 収集方法：インタビューによる聞き取り調査
- (3) 調査期間：2015 年 1 月から同年 2 月まで

インタビューでは (i) 広報活動におけるウェブサイトの必要性、(ii) 広報活動における SNS やブログの必要性、(iii) 予約サイトやブログでの口コミが集客に与える影響の 3 点について、考えの強度を 4 件法で尋ね、さらにその理由について調査した。これらに加え、平均宿泊費（平日）、従業員数、客室数、ICT プロモーションにかかるコストの割合、現在取り組んでいるインターネットを利用した広報の方法についても調査を行った。調査を

表 2. インターネットを利用した広報の方法

ウェブサイトの利用			SNS、ブログの利用		
自社のウェブサイト	トラベルサイト	検索サイトでの広告	自社の SNS	ブログ	SNS での広告
20	20	7	8	8	3

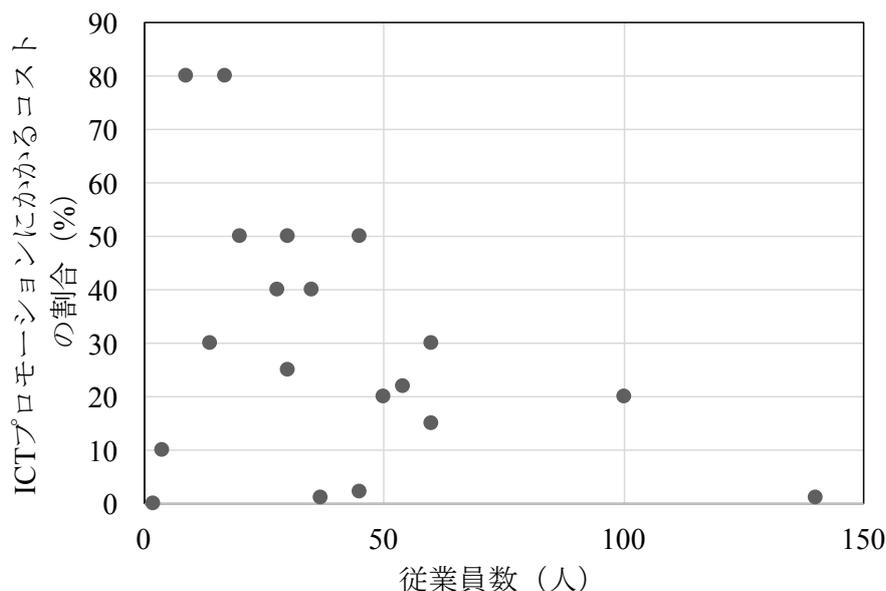


図 2. 従業員数と ICT プロモーションにかかるコストの割合の関係

表3. インターネットを利用した広報に関する調査結果

	強く感じる	感じる	あまり感じない	全く感じない
(i) ウェブサイトでの広報の必要性	16	8	1	0
(ii) SNS・ブログでの広報の必要性	5	11	7	1
(iii) 口コミの集客への影響	16	8	0	1

行った30件のうち25件から回答を得た。

2. 2. 結果および考察

表2では、諏訪圏の宿泊施設がインターネット上の広報で利用している方法を示した。最も利用されている広報の方法は、自社のウェブページおよびトラベルサイトで、それぞれ20件であった。一方、SNSやブログでの広報はそれぞれ8件とあまり行われていなかった。インターネット上での広告掲載の利用数は少なかったが、SNSでの広告は検索サイトでの広告の約半数であった。

図2は、従業員数に対するICTプロモーションにかかるコストの割合を表したものである。相関係数は-0.374であった。これは、従業員数の少ない宿泊施設ほどICTプロモーションにかかるコストの割合が高いことを意味する。

表3では、インタビューによる質問(i)～(iii)の4件法による回答の集計結果を示した。質問(ii)では未回答が1件あった。この結果から、諏訪圏の宿泊施設は、ウェブサイトにおける広報の必要性や、口コミが集客に与える影響を感じていることを示された。一方で、SNSやブログにおける広報の必要性はそれほど強く感じないことが分かった。

次に階層的クラスタ分析を用いて質問(i)～(iii)への回答理由を分析した。まず、回答を品詞に分解し、名詞、動詞、形容詞、副詞を抽出した。抽出した単語をWard法によりクラスタ分けした。この分析のためにKH coderを使用し、形態要素解析には茶釜を利用した[9-11]。図3～図5では、2回以上出現した単語の階層的クラスタ分析の結果を示した。横軸は抽出された単語、縦軸はJaccard係数を表す。破線で囲まれた単語は同一クラスタに属していることを意味する。

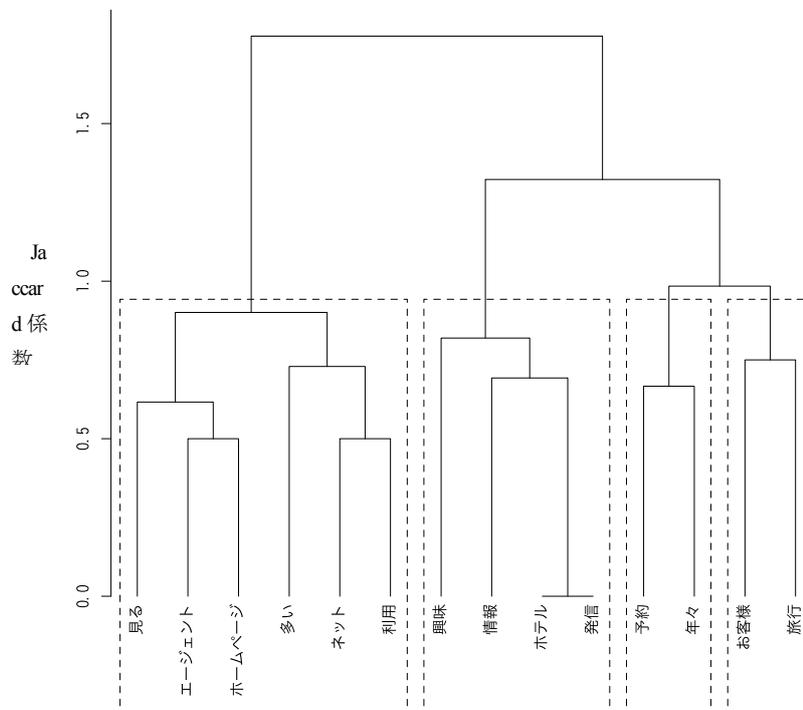


図3. ウェブサイトでの広報の必要性

図3は、ウェブサイト上の広報の必要性を感じる理由についての分析結果である。宿泊施設はウェブサイトでの広報について、

- 多くのユーザがウェブサイトを見ている
- ウェブ予約は年々増加している
- 宿泊施設の情報を送信することにより、ユーザに興味を持たせることが可能

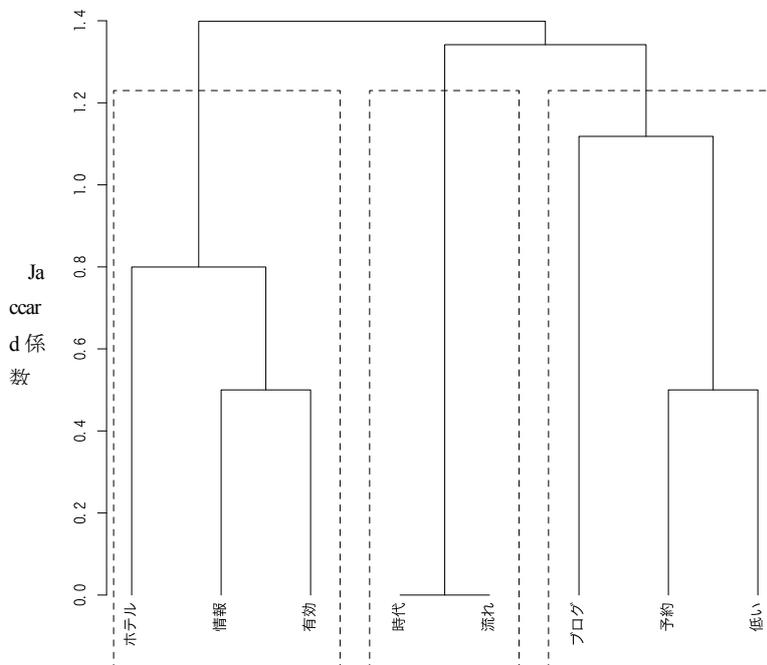


図4. SNSおよびブログでの広報の必要性

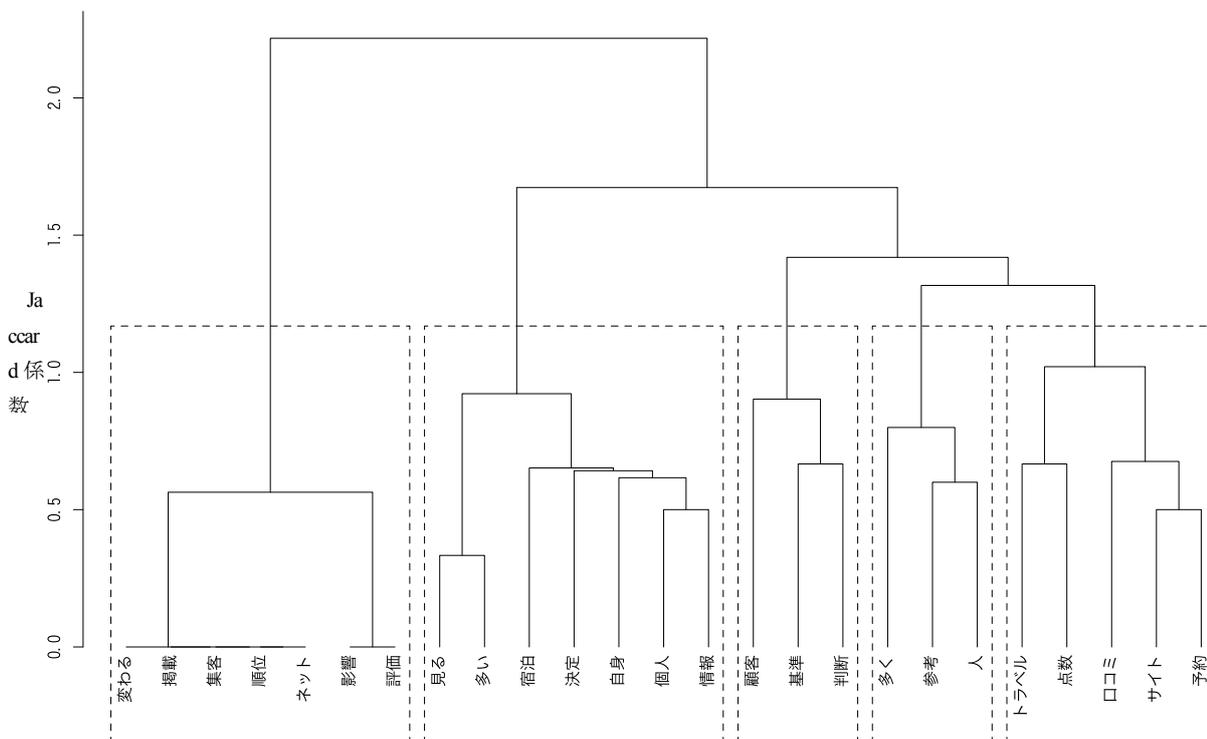


図5. 口コミの集客への影響

と認識し、重要性を高く感じていることが分かった。

図4は、SNS およびブログで広報の必要性を感じる理由についての分析結果である。SNS およびブログでの広報について、

- SNS やブログは、宿泊施設の情報を送信するのに有効である
- SNS やブログの使用は時代の流れ
- 予約への影響は低い

と、効果があることや時代の流れであることは認識している一方で、予約への影響が低いと認識しているため、重要性をそれほど高く感じていないことが分かった。

図5では、口コミの集客への影響を感じる理由についての分析結果を示した。口コミの集客への影響について宿泊施設は、

- 集客に強い影響を与え、集客数はインターネット上の順位に依存する
- 宿泊を決める時に見る人が多い
- 旅行サイトの口コミと点数は、予約に影響を与える

表4. ウェブサイトでの広報の必要性

	平均宿泊費		従業員数		客室数	
	1万円未満	1万円以上	30人未満	30人以上	30部屋未満	30部屋以上
平均値	3.308	3.917	3.267	3.900	3.364	3.786
標準偏差	0.630	0.289	0.704	0.316	0.674	0.426
データ数	13	12	15	10	11	14
t 値	-3.069		-2.659		-1.912	
P 値	0.003**		0.007**		0.034**	

*:p<.05 **:p<.01

表5. SNS・ブログでの広報の必要性

	平均宿泊費		従業員数		客室数	
	1万円未満	1万円以上	30人未満	30人以上	30部屋未満	30部屋以上
平均値	2.417	3.250	2.714	3.000	2.700	2.929
標準偏差	0.669	0.754	0.914	0.667	0.823	0.829
データ数	12	12	14	10	10	14
t 値	-2.865		-0.840		-0.668	
P 値	0.004**		0.205		0.256	

*:p<.05 **:p<.01

表6. 口コミの集客への影響

	平均宿泊費		従業員数		客室数	
	1万円未満	1万円以上	30人未満	30人以上	30部屋未満	30部屋以上
平均値	3.385	3.833	3.333	4.000	3.455	3.714
標準偏差	0.650	0.389	0.617	0.000	0.522	0.611
データ数	13	12	15	10	11	14
t 値	-2.070		-3.391		-1.123	
P 値	0.025**		0.001**		0.137	

*:p<.05 **:p<.01

- 多くの人が参考にする
- 顧客の判断基準となっている

と認識していることが分かった。

宿泊施設の属性による違いを明確にするために、(1) 平均宿泊費 (平日)、(2) 従業員数、および (3) 客室数の3点から分析を行った。これらの属性は、それぞれ宿泊施設のランク、人的資源にかかるコスト、および収容可能客数を反映している。これらの属性による違いを明確にするために、宿泊施設を2つのグループに分け、回答を強く同意する=4、同意する=3、あまり同意しない=2、全く同意しない=1と数値に変換し、t検定による比較を行った。2つのグループでデータ数が同程度となるように、平均宿泊費は1万円、従業員数は30人、客室数は30部屋を基準としてグループ分けを行った。それぞれの基準値を上回るグループを上位群、下回るグループを下位群と呼ぶこととする。上位群と下位群を比較したところ、全ての属性において上位群の平均値が、下位群の平均値より高いことが分かった。

表4は、ウェブサイト上での広報の必要性についてのt検定の結果を表している。(1) ~ (3) 全ての属性において、上位群と下位群に統計的有意差 (1%有意水準) があることが分かった。表5は、SNS やブログで広報の必要性についてのt検定の結果である。統計的有意差 (1%有意水準) があったのは、平均宿泊費のみであった。表6は、旅行サイトおよびブログの口コミの影響についてのt検定の結果であり、平均宿泊費と従業員数の二つの属性で上位群と下位群に統計的有意差 (1%有意水準) が見られた。特に宿泊費に関しては、全ての属性で上位群と下位群に有意差が見られた。これは、宿泊費が高くとも宿泊をしたいと思わせるだけの魅力を、インターネットを通じて伝えることができると宿泊施設が感じているからであると考えられる。従業員数と客室数に関しては、コスト面から多数の集客が必要であり、インターネットを利用した広報を行うことで多数の顧客へ宿泊施設の魅力を伝えることができると感じているからであると考えられる。

調査の結果、諏訪圏の宿泊施設は、ICTプロモーションの重要性と影響の強さを認識していることが分かった。また、自社のウェブサイトや旅行サイトは利用するものの、SNS やブログの利用は少ないことが明らかにされた。ICTプロモーションに関する認識に宿泊施設の属性によって統計的な有意差があった。宿泊費が高い、従業員数が多い、客室数が多いなど、集客やコスト面で不利な点があり、魅力を顧客にアピールする必要性が高い宿泊施設の方が、ICTプロモーションの重要性を強く認識していることが明らかにされた。

3. 自動化されたポータルサイトの提案

前節で述べたように、諏訪圏の宿泊施設は、ウェブサイトやSNSなどを用いたICTプロモーションを行っている。しかし、小規模宿泊施設ではICTプロモーションにかかるコストの割合が高いため、さらにそのコストを増やすことは現実的ではない。このような状況では、自動化を多岐に搭載したポータルサイトの利用が低廉であり最適である。そこで本研究では、クローリングおよび機械学習を用いて自動化されたポータルサイトのシステムを提案する。

クローリングとは、自動的にウェブサイトを巡回し、情報を収集することを指し、このプログラムはクローラと呼ばれる。クローラには、RSS (RDF Site Summary/Rich Site Summary/Really Simple Syndication) の情報やHTML (Hyper Text Markup Language) の分析に基づいて収集するタイプなどがある。RSSには、ウェブサイトなどの更新情報の概要が含まれており、SNS やブログ、ニュースフィードなどで広く使用されている。また、HTML型のクローラは、多くの検索エンジンで使用されている。

機械学習は、明示的にプログラムすることなくコンピュータを動作させるための方法である。データマイニングおよびパターン認識などの幅広い分野で使用されている。学習データの種類と必要な結果に応じて、教師あり学習、半教師あり学習、および教師なし学習のアルゴリズムが存在する。教師あり学習は、既知の情報 (正解付き学習データ) を使って、未知の情報の予測器や識別器を生成することを目的とする。一方、教師なし学習は、データに含まれるパターンを見つけ出し、データの性質の理解をすることを目的としている。半教師あり学習は、教師あり学習と教師なし学習の中間的手法で、代表的なものとして正解付き/なしの混在データに対する識別学習がある。この識別学習では、正解付きデータで識別器を生成し、正解なしデータを用いて識別器の性能向上を図ることができる[12]。

本研究で提案するポータルサイトのシステムでは、クローラは宿泊施設のSNSおよびブログから発信されるRSSから情報を収集する。仮にHTMLタイプのクローラを使用しても同様のシステムを作ることは可能である。しかし、RSSの方がHTMLに比べて文書構造が簡潔であるため、RSSを用いることとする。図6は、提案するシステムの概念図である。クローラは、登録したサイトのリストを受け取り、リストに登録された宿泊施設のSNS

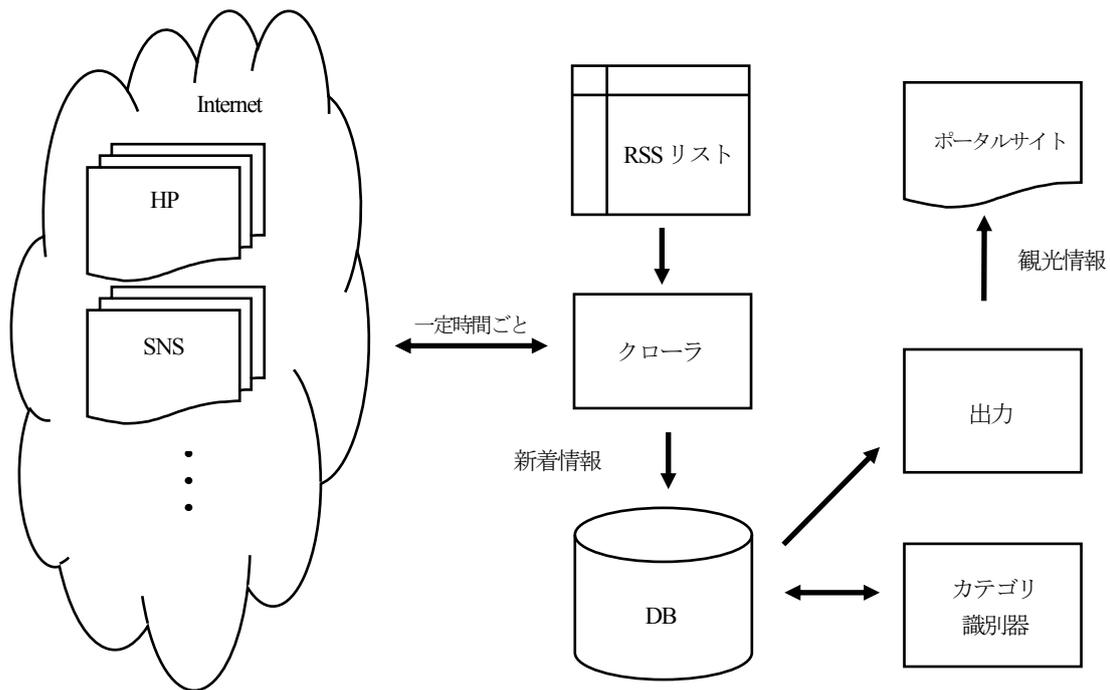


図6. システム概念図

およびブログから RSS を一定間隔で収集する。カテゴリ識別器は機械学習により生成され、収集した情報をカテゴリ分けし、その情報はポータルサイトの適切なカテゴリに出力される。

提案するポータルサイトのシステムは、次の3つの特徴がある。

(i) 宿泊施設が情報をアップロードする先を SNS またはブログとする点である。ポータルサイトの運用において多くの場合に問題となるのが、情報のポータルサイトへのアップロードである。ウェブサイトの更新は関連する専門知識が欠如している人には困難であり、外部に委託する場合は更新の度にコストがかかる。この問題を解決するためには、情報更新が簡単にできるシステムが必要である。そこで、情報更新のプラットフォームとして SNS およびブログに着目した。これらは誼訪圏の宿泊施設でもすでに導入している所があり、新規導入する場合であっても、ほとんどのサービスが無料で利用可能である。SNS やブログなどは、登録さえすればオンラインで記事を投稿することが可能である。このため、ウェブサイトの更新とは異なり頻繁な情報更新が行われることにより、情報の陳腐化を防ぐことができる。また、記事を投稿すると同時に RSS も作成されるため、SNS およびブログは提案するポータルサイトシステムの情報源として最適である。

(ii) 観光情報を収集するサイトを制限する点である。クローラは通常、収集されたデータに含まれるリンクをたどり、異なるウェブサイトの情報を収集する処理を繰り返す。しかし、通常の方法では観光情報以外の情報が入り込むことが明白であり、ICT プロモーションの観点から考えるとポータルサイトとしての効果を低下させることになる。このため、RSS を収集するウェブサイトなどを決め、情報源を制限することによって、ポータルサイトに掲載される情報の精度を高め、広報としての効果を向上させることが可能となる。

(iii) 自動的に観光情報のカテゴリを識別する識別器を生成する点である。ポータルサイトへの新たな情報をアップロードする場合には、管理者または投稿者によってカテゴリが選択されることが一般的である。この手順に人的資源が必要であり、ポータルサイトの運用コスト増大につながる。この点、本研究では機械学習を導入しカテゴリ識別器を使用することでコストを軽減するだけでなく、投稿者はカテゴリを考慮することなく情報を投稿することができるため、情報更新の頻度向上が期待できる。

以上の特徴から、クローラと機械学習を用いるポータルサイトには次のような利点があると期待できる。

- SNS またはブログの活用による導入コストの低減
- 情報の陳腐化の防止
- 情報更新にかかる人的コストの削減

図7は、提案するポータルサイトシステムのフローチャートである。本システムでは、クローラが一定時間ごと

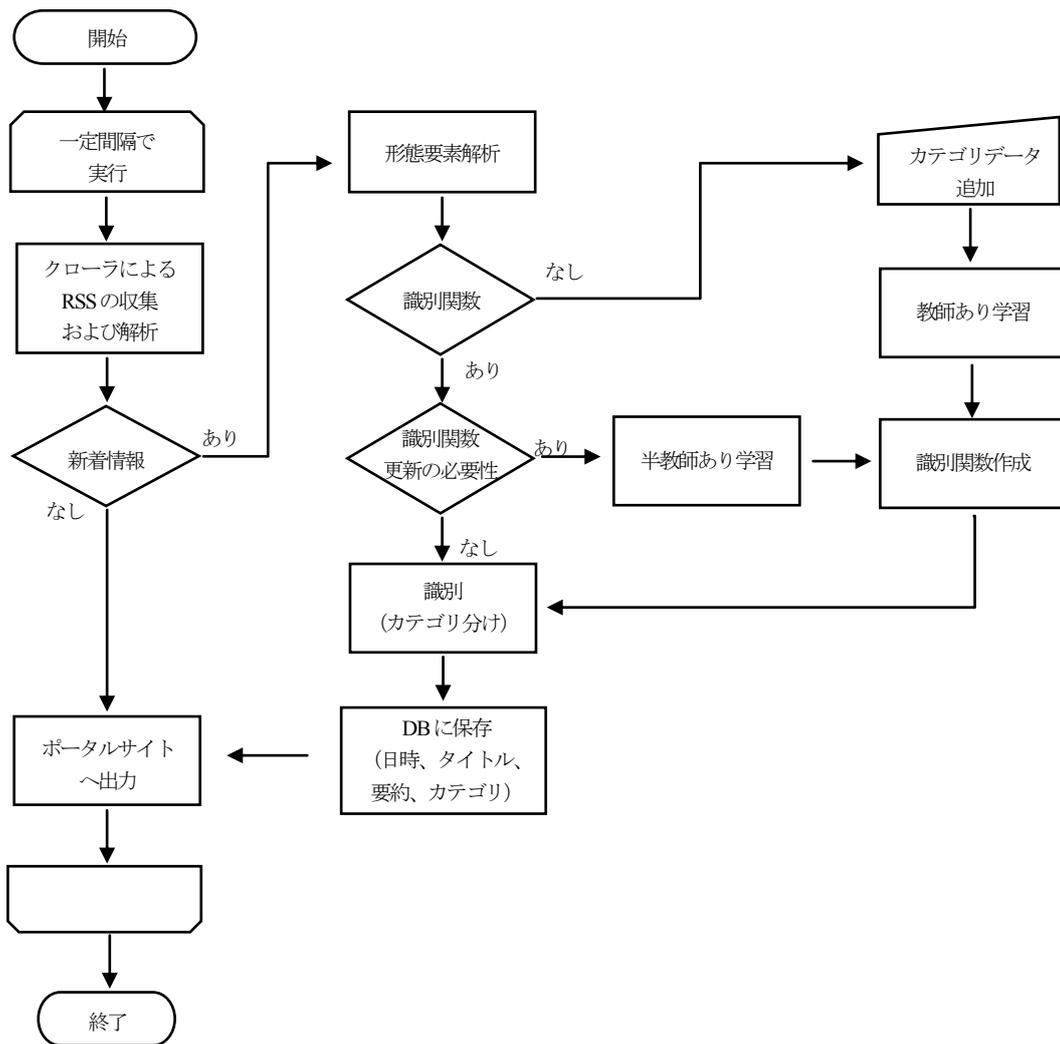


図7. ポータルサイトのフローチャート

とに実行され、RSSの収集および分析を行う。収集した情報の中に新着情報がなかった場合は、情報の更新をせずにポータルサイトへ観光情報を出力する。新着情報があった場合は、形態要素解析を行ない、カテゴリ分けを行った後にデータベースに保存し、ポータルサイトへ出力する。システム導入時にはカテゴリ識別器(識別関数)がないため、教師あり学習を用いて生成する。このために、システムの導入前に過去に書かれた SNS およびブログエントリをいくつかのカテゴリに分類する必要がある。このカテゴリデータを正解データとして扱い、コンピュータが教師あり学習によって適切なカテゴリに識別する方法を学習し、カテゴリ識別器を生成する。新着情報は、このカテゴリ識別器によって分類される。この識別器の生成はナイーブベイズ識別法が適していると考えられる。ナイーブベイズ識別法は迷惑メールの識別などに使われている識別法で、実装が容易でありモデルの計算時間が短い、新しいデータに対するモデルの更新がしやすい、手動でのチューニングが容易であるなどの利点がある。また、宿泊施設によりアップロードされる観光情報のカテゴリは、季節などの時間経過やイベント、広報戦略などによって変化する。この変化に対応するために、数ヶ月に一度または任意のタイミングで再び機会学習を行う必要がある。正解付きデータと正解なしデータを合わせて半教師あり学習を実行しカテゴリ識別器を更新することで、カテゴリ識別器の正解率および再現率を向上させることができる。この時に、新規カテゴリの追加を許可し正解データを追加することで、カテゴリの陳腐化を防ぐことも可能である。新規カテゴリの追加にあたっては、教師なし学習による階層的クラスタリングが有効である。クラスタ分析を行うことで戦略に合ったキーワードの候補を探すことができる他、潜在的な観光資源の発掘が可能となる。

4. おわりに

小規模宿泊施設が展開する ICT プロモーションの効果を向上させるためのポータルサイトの構築を目的に研究を行った。このために諏訪湖周辺の宿泊施設を対象に ICT プロモーションの実態調査を行った。その結果、諏訪圏の宿泊施設は ICT プロモーションの重要性とその影響を認識していることを明らかにした。インターネット上で広報においてウェブサイトを使用する宿泊施設の数が増える一方で、SNS やブログを使用している宿泊施設の数はそれぞれ 8 件にとどまっていた。また、宿泊施設の平均宿泊費、従業員数、および客室数の大小により 2 グループに分け、それぞれ t 検定を行ったところ、ICT プロモーションに対する認識に統計的有意差があることが分かった。集客にあたり魅力をアピールする必要性が高い宿泊施設の方が、ICT プロモーションの重要性を強く認識していることが明らかになった。

そこで、小規模宿泊施設での ICT プロモーションの問題を解決するために、クローラと機械学習を用いた観光ポータルサイトのシステムを提案した。提案するポータルサイトシステムは、低コストでの導入、運用コストの削減、情報の陳腐化の防止などの効果が期待でき、小規模宿泊施設における ICT プロモーションの問題に対して適していると考えられる。このポータルサイトを導入することによって、自動的に小規模宿泊施設が SNS またはブログにアップロードした観光情報を収集し、観光ポータルサイト内の適切なカテゴリに出力することができる。SNS およびブログは記事の投稿が簡単であり、複数の宿泊施設の SNS およびブログを登録することによって、観光ポータルサイトに掲載される観光情報の更新頻度が高くなる。これにより小規模宿泊施設の集合体による観光ポータルサイトが、利用者にとって有益な利便性の高い情報源となり、大規模宿泊施設や都市圏などの観光地のウェブサイトと匹敵する存在となり得る。都市圏と比較して小規模宿泊施設の多い地方でこの観光ポータルサイトを導入することで、観光客の増加や潜在的な観光資源の発掘といった地方観光振興に繋がることが期待される。

本研究は諏訪地域のみを対象として分析したが、同様の調査を他の地域でも行ない、本研究の成果の一般性を検証する予定である。また、本研究で提案したポータルサイトについて、著者らは構築を開始しており、クローラによって RSS を収集しポータルサイトの内容を更新するプロセスまでは構築した。今後は、Sebastiani のまとめた英語文書の分類の手法を参考に[13]、機械学習による日本語文書を分類するシステムを構築し、観光ポータルサイトを完成させたい。観光ポータルサイトを構築した後、諏訪圏およびその他の地域で実際に運用し、実用性を検証する予定である。

[参考文献]

- [1] 総務省『平成 27 年度版 情報通信白書』, 2015.
- [2] 公益財団法人日本交通公社『旅行年報 2015』, 2015.
- [3] 福島県ベンチャー・SOHO・テレワーカー共働機構『地域ポータルサイト導入手引き書』, 2008.
- [4] 田村信之『地域ポータルサイトについて』情報の科学と技術 Vol. 55 No. 2 pp. 76-80, 2005.
- [5] 大槻洋輔, 佐藤理史『地域情報ウェブディレクトリの自動編集』情報処理学会論文誌 Vol. 42 No. 9 pp. 2310-2318, 2001.
- [6] 遠藤雅樹, 中村信也, 奥秋清次, 大野成義『地域サイト及びブログからの観光情報抽出と融合の提案』情報処理学会研究報告 Vol. 2012-DBS-155 No. 6 pp. 1-6, 2012.
- [7] 石野亜耶, 難波英嗣, 竹澤寿幸『旅行ブログエントリからの観光情報の自動抽出』知能と情報 (日本知能情報ファジィ学会) Vol. 22 No. 6 pp. 667-679, 2010.
- [8] 観光庁『宿泊旅行統計調査 平成 28 年 2 月分 (第 2 次速報値)』, 2016.
- [9] 樋口耕一『テキスト型データの計量的分析 —2 つのアプローチの峻別と統合—』, 理論と方法 (数理社会学会) Vol. 19 No. 1 pp. 101-115, 2004.
- [10] 樋口耕一『社会調査のための計量テキスト分析 —内容分析の継承と発展を目指して—』, ナカニシヤ出版, 2014.
- [11] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 浅原正幸.『日本語形態素解析システム「茶筌」 Version 2.0 使用説明書』, NAIST-IS-TR99012. 1999.
- [12] 荒木雅弘『フリーソフトで始める機械学習入門』, 森北出版, 2014.
- [13] F. Sebastiani『Machine Learning in Automated Text Categorization』ACM Computing Survey, Vol. 34, No. 1 pp. 1-47, 2002.

(2016 年 7 月 7 日受理)