

# テキストマイニングを用いた都市農村交流ニーズの把握

—岡山県吉永町ふるさと村の八塔寺山荘の落書き帳を対象として—

**Applying the Text mining method to the Analysis of Visitors' needs on Rural-urban exchange  
For the Messages in Doodling notes of Hattouji Villa, Yoshinaga town, Okayama prefecture**

石川 修\* 星野 敏\*\*

Osamu ISHIKAWA\* Satoshi HOSHINO\*\*

(\*神戸大学大学院自然科学研究科 \*\*神戸大学農学部)

(\*Graduate School of Science and Technology, Kobe University \*\*Faculty of Agriculture, Kobe University)

## I はじめに

本稿では、岡山県和気郡吉永町の「八塔寺ふるさと村」内にある町営宿泊施設「八塔寺山荘」を家族や友人達と共に訪れた利用者が書き残した落書き帳の記述内容を分析し、訪問者が見た山荘の長所・短所及び利用者のニーズの把握を試みる。

これまでこうした形式の落書き帳分析はほとんどされていない。その理由は、この種の落書き帳は訪問者が感想を自由に記入するために設置されているものであり、用意された設問に答えるアンケート方式ではない。記入される内容は訪問者の意思に任されており、時には施設と直接関係のないことも書き込まれる。このような不定型な質的情報から有益な情報を抽出することは非常に困難であると考えられてきたからである。

ところで近年、テキストマイニング手法の研究が進みつつある。この手法は、自由記述形式のテキストデータを対象としてその中に隠された情報を抽出することを目的とした手法である。

本稿では、事例宿泊施設に設置された落書き帳のメッセージにテキストマイニング手法を適用し、同施設（および周辺地区）に関わる来訪者のニーズを明らかにする。加えて、テキストマイニング手法の可能性について考察する。

具体的には、II節にてテキストマイニング手法の理論及びデータ可視化の意義と技術の概要について述べる。III節では分析対象となったふるさと村と落書き帳の詳細を述べる。IV節では本題であ

るテキストマイニング手法を用いて記述内容分析を行い、V節では結果からの考察を行う。

## II テキストマイニングと可視化技術

### 1. テキストマイニング<sup>1)</sup>

落書き帳のような自由記述形式のテキストデータには膨大で不均一な情報が残されるので、人間の目で一つ一つサンプリングし分析を実施するのは困難である。テキストデータの分析手法としてはKJ法が広く知られている。複数のデータから新しい仮説の発見や枠組みの形成のための、カードを用いたデータ整理法である。しかしKJ法では、人がテキストを読んで内容を理解し、頭の中でマッチングを行い分類しなければならない。一度に処理できるカードの数には限界があり、膨大な量のテキストをカードとして処理することには不向きである。また、KJ法は分析者の主観と熟練度に依存するため、分析者による「ばらつき」が避けられない。

そこで本稿では、テキストマイニング手法に注目する。同手法の特徴は、経験や熟練度の浅い分析者でも大量のフリー形式データを短時間で的確に分析できる点にある。スイスのRajmanはテキストマイニングとは「構造化されていないテキストデータからの情報抽出に関する技術」であり、その性能を向上させるために自然言語処理技術の発達が必要であると述べている<sup>2)</sup>。

テキストマイニングを根本から支える自然言語

処理技術において形態素解析と構文解析は不可欠な文法技術である。つまり形態素解析により分析対象となる文書データを言語学的に意味をなす最小の単位である形態素（一般的な単語よりも細かい単位）にまで分割する。例えば「あの服を買いたい。」という文は、

あの	(連体詞)
服	(名詞)
を	(格助詞)
買い	(動詞－「買う」連用形)
たい	(助動詞)
。	(句点)

と分ける。

そして構文解析によって語句間の関係を認識し文章のタイプ(疑問文であるか命令文であるか等)の判別を行い、語と語の係受け関係を調べる。「あの服を買いたい。」は『あの服を一買いたい。』という係受け関係に識別される。

この形態素解析と構文解析がテキストマイニングの基礎となっている。これにより文章内での重要語句の抽出(出現頻度の認識)、語句の意味的役割の認識、同義語の抽出等が可能となる。

また、ここで言うマイニング(mining: 採鉱)とは Web 検索システム等で行うことのできる一般的な情報検索とは別の意味をなす。探索者にとっては未知であってもその著者には既知である情報を探る、ということではなく複数のテキストの断片と探索者の知識の組合せから新たな仮説を創り出すことがマイニングの本質である。

## 2. 可視化技術<sup>3)</sup>

上記の操作によって大量のテキストデータの構文解析を行っただけでは役に立たない。膨大なデータの内容を要領よく把握し、それが意味するところを的確に把握することが必要である。そのためには、本来人手に負えない膨大な分析結果を視覚的に表現し、人間のデータ認識能力を最大限に駆使できるようにすることが重要となる。分析対象に対する事前知識が十分でない場合、こういった可視化によるアプローチはデータから新たな知見及び仮説を探索するのに非常に有効である。

今回用いたテキストマイニング可視化技術は「スプリング埋込み」と呼ばれる手法である。抽

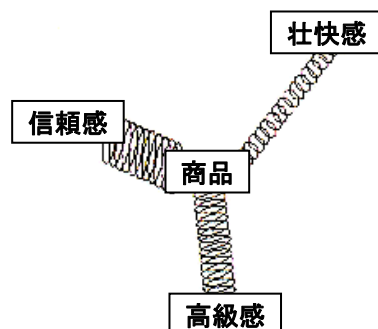


図1 スプリング埋込みの例<sup>4)</sup>

出されたキーワード間の関連性をバネ(スプリング)にみたて、全てのキーワードは互いに引っ張られながら、全てのバネの引力が釣り合う仮想的な物理モデルの安定状態に配置されレイアウトは決定される。その結果、関連性の強いキーワード同士は近くに、関連しないものは離れて配置される。直感的に情報全体の概観を把握することが可能となる技術である。

## III 分析対象概要

### 1. 八塔寺ふるさと村概要

1974年に岡山県下ではじめて誕生した「八塔寺ふるさと村」は吉永町の北部、岡山県と兵庫県との県境をなす標高539mの八塔寺山の南麓に位置し、一時は山



図2 八塔寺山荘外観

岳仏教の中心地として栄えた地域である。広大な田園地帯の中に10数軒の茅葺き民家や寺院が点在し、のどかな農村風景が広がっている。八塔寺山頂付近からは瀬戸内海を一望することができ、景観に優れており今村昌平監督の映画「黒い雨」のロケ地としても知られている<sup>5)</sup>。

八塔寺山荘はふるさと村内に1993年から設営された訪問者向けの町営宿泊施設である。村内の遊歩道脇に設営され、敷地面積が広く(総面積は247.52m<sup>2</sup>)井戸や「いろり」のある昔ながらの民家を再現している。村内に点在している茅葺き民家を象徴する建物と言える。利用は人数の多少に関わらず、一度に1グループとなっている(宿泊定員は15人まで)。宿泊費は基本料金5,000円+2,500円×宿泊人数。

## 2. 落書き帳

落書き帳は山荘開始当初の1993年4月から現在(2003年8月)まで山荘内に設置されており、八塔寺山荘に宿泊した利用者が自由に山荘及びふるさと村への感想や意見、そこでの出来事をノート10数冊分には書き記していったものである。

落書き帳の特性として「完全なる自由回答」及び「自由に閲覧が可能」が挙げられるが、その特性から考えられる利点・欠点を推察してみる。

**特性1**：設問に対し答えるという形式ではなく、記入者が自由に何でも書き込める

- 記入者の山荘に対する生の感想や意見が本音で汲み取れる(利点)
- アンケートをされる訳ではないので、書き込まなかった訪問者の負担にならない(利点)
- 画一的な答えが得られる訳ではなく、事後でのデータ整理が面倒(欠点)
- 山荘に関係しない記述もなされ、受動的であり、質問者が欲しい情報を得にくい(欠点)
- 質問者が思いも寄らなかった意外な答えを得ることができる(利点)

**特性2**：山荘に置かれてあるので、先の訪問者が書いた内容を後の訪問者が読むことができる

- 訪問者同士のコミュニケーションや情報の共有が可能(利点)
- 先の記述に影響されて、純粋な回答が得られない(欠点)

以上のような利点・欠点が挙げられる。欠点である「分析方法が面倒」を解消するためにも、ま

た最大の利点と言える「思いも寄らない意外な答え」を得るためにもテキストマイニングを適用することは非常に有効であると言える。

## IV 記述内容の整理とその集計結果

### 1. 書き込み内容の整理

書き込み内容は、宿泊日や利用人数、利用目的、感想、意見を正確に書き込んでいるもの以外にも、分析サンプルとして利用することが不可能な記入者不明の子供が描いたような絵やただ自分の名前を書き記しているだけのもも多数ある。そこで分析の際には、山荘について全く触れていない内容のものは除外する。抽出した書き込み内容を、

- ① 訪問年月日
- ② 訪問者の居住地(都道府県別)
- ③ 同伴者(「家族・親戚」「友人」「カップル・夫婦」「合宿」「研修」「一人旅」と6分類)
- ④ グループの規模(1~5人を「小グループ」、6~10人を「中グループ」、11人以上を「大グループ」と3分類)
- ⑤ 訪問者詳細
- ⑥ 記入者性別
- ⑦ 記入者年齢(「高年者」「青・壮年者」「若年者」と3分類)
- ⑧ コメント内容
- ⑨ 備考

と読み取れる内容を振り分け整理した。コメント内容の分析を行う前に、これら振り分けた項目毎の集計結果から予備的分析を行う。

表1 書き込み整理項目表(一部)

訪問年月日	出身地	訪問形態	人数	訪問者詳細	記入者性別	記入者年齢	コメント内容	備考
1993/4/4		家族・親戚	小グループ	3歳2ヶ月の女の子も	男性	青・壮年	山荘のパンフレットもないので、どんな所か不安だったが、最新の電化製品や設備が色々揃っていてビックリ。静かでいいところ。子供も気に入った。次回は大勢で来たい。	はじめての訪問者
1993/4/5-12	静岡	友人	小グループ		男性	青・壮年	初日はとても寒い。フロ場に鏡があれば便利。自炊が大変。	
~中略~								
2002/11/3-4	大阪枚方				男性	高年	田舎を求めて。心と体の休息になった。包丁が切れないので研いで欲しい。こういう落書き帳スタイルはどうかと思う。	
2003/3/28-29		合宿	大グループ	作陽Trb.Ens				3年ぶり
2003/8/13-14							楽しかった。かもうどん美味しかった。	
2003/8/16-17					女性	若年	はじめて八塔寺に来た。ちょっと怖い。いろりははじめて。	

## 2. 集計結果の予備的分析

総サンプル数は合計で 146 あった<sup>注1)</sup>。年度別で見ると 1999 年度からサンプル数が激減している (2003 年は 8 月まで)。訪問者の全てが書き込みを残している訳ではないゆえ、山荘開始の 1993 年から 6 年ほど経ち、訪問者に何かしら感想を書き残させる程の魅力が山荘に薄れてきている、と考えられる。

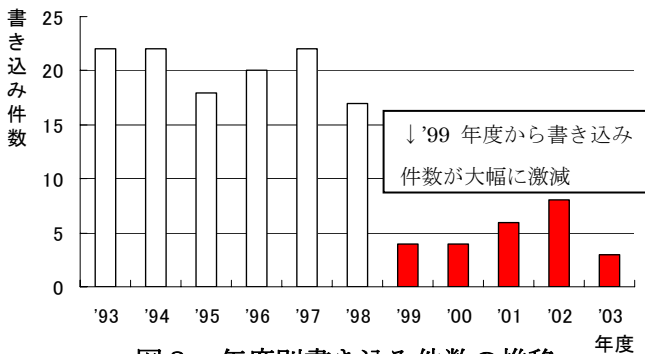


図3 年度別書き込み件数の推移

その他にも重要な要因として、この頃から単なるマンガや絵などの数が増え、目立ち始めている。2002 年 11 月には『こういう落書き帳スタイルはやめにして、別のものを考えてみてはどうか』(男性/高年者) という書き込みが残されていた。山荘及びふるさと村に関わらない書き込みが多数あると、感想や意見を書き残す気がそがれてしまう訪問者の意見として解釈できる<sup>注2)</sup>。

146 サンプル中で居住地が読み取れたのは 54 組である。岡山県 (13 組) ではなく兵庫県 (18 組) がトップに来ていることが興味深い。兵庫県に隣接している吉永町の地理的特徴と、岡山県と兵庫県が協力して共に備前地域の都市農村交流を推進してきた成果として理解できる<sup>注3)</sup>。

同伴者は家族・親戚や友人達と回答した訪問者のサンプル数が全体の約 8 割を占め、また、グループの規模で見ると 10 人以下のグループで山荘を利用するのが全体の 7 割以上と大多数である。稀に学校のクラブでの合宿等によって大人数で使用されている。

記入者は性別で見ると女性が多かった。落書き帳記入は女性の方が積極的であると考えられ、記述内容も女性と

思われる書き込みの方が男性よりも丁寧で長い文章であった。年齢別サンプル数比較は青・壮年者が約 6 割を占めており若年層のサンプル数は 1 割にも満たず、実際の訪問者の年齢層をほぼ正確に反映していると思われる。

## 3. キーワード集計

書き込みコメントにおいて、「ホテル」「いろり」「寒い」といった同一の内容を複数人が書き残していった要素が多数あることに着目する。それら同一コメントは山荘訪問者全員が共通して感じた魅力や特徴であると考え、キーワードとして抽出し、その数を集計した。その集計結果から考えられる仮説と、後のテキストマイニングツールを用いた分析結果を比較することでコメント全文に潜む新たな情報の抽出を試みる。

地域特性を象徴するキーワードは、のべ 141 個あった。図 4 にそれを示す。

キーワードとして上位に来たのは「自然」と「いろり」である。ここでは、5 番目に挙がっている「設備が整っている」という項目に注目する。1995 年 10 月に記された書き込みを抜粋すると『山荘ということで、設備の面を心配していたが、設備の良さにビックリした』(女性/青・壮年者) というものがある。山荘はいろりや茅葺屋根に代表される昔ながらの田舎的要素と、テレビや冷蔵庫、ラジカセ等といった現代的な電化製品も用意されて

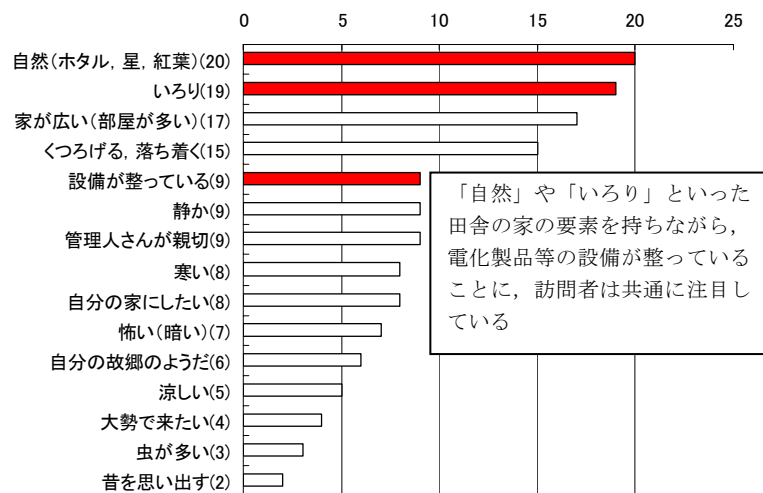


図4 キーワード数集計

いる。「典型的な田舎の日本家屋にて、都会と大差  
ない水準の生活を過ごせ、周囲はのどかで美しい  
自然に囲まれ、広い家で家族（または友人）との  
んびりできる」という点を、山荘の際立った特長  
として多数の訪問者が評価していると判断できる。

#### 4. マイニングツールによる可視化

本編では「コンセプトマッピング」という独自の  
スプリング埋込み型可視化技術を持つビジュアル  
テキストマイニングツール「トレンドサーチ(富士通ハイパーソフトテクノロジー社製)」を用いた。  
コメント内容と記入者を表現する要素（出身地、  
訪問形態、人数・・・）との間について順番に関  
連性を調べていった。すると年齢層を3分類する  
ことで新たに興味深い結果を導くことができた。  
その結果を図5に示す。

この図は「スケルトンマップ」と言われる表示  
方法で関連度の弱いスプリングを削除し、骨格と  
なる構造のみを残して作成している。このマップ  
では中心的なテーマとなる重要な単語が放射状の  
中心になる。高年者グループのコメント内容と分  
別できるものをAで、  
若年者をB、青・壮  
年者のグループをCの円  
で囲い、明確に分類し  
た。この図から

- A) 高年者はこの山  
荘に「童心にかえ  
る」や「暖かい」、  
「田舎」の風景を  
求めている。
- B) 若年者（子供）は  
「ホテル」や「星」  
といった都会の  
生活では得られ  
ない自然の魅力  
を感じている。
- C) 青・壮年者は「美  
味しい」「感動」  
「親切」といった  
キーワードがあ  
り、ふれあいを感じ

じている。

以上のことが新たな情報として導かれる。

可視化の結果から、年齢層ごとに感じている山  
荘及びふるさと村への魅力の差異が明らかになっ  
た。年代によって感じる魅力が違うということは、  
どの年代が主体であってもそれぞれに楽しむ要素  
が村内には存在するということになる。もうホテル  
や虫採りに夢中になる年ではなくなった頃には、  
温泉や美味しい料理で農村の文化と触れ合うこと  
を楽しみ、また高年になればのんびりと昔を思い  
出すことができる。ふるさと村及び周辺地区は、  
来訪者の年代に合わせて次々と新たな魅力を提供  
することが可能であり、その可能性を一層広げること  
が重要となるであろう。

#### 5. 八塔寺地区の改善方策についての提言

上述の分析に加え、記入者全員に共通して表れた  
キーワードとして筆者が注目したものに「いろり」  
があり、若年層と青・壮年層にもいろりを評価  
しているコメントは多かった。「いろり」のある  
山荘は対象を老若男女問わない素晴らしい魅力で

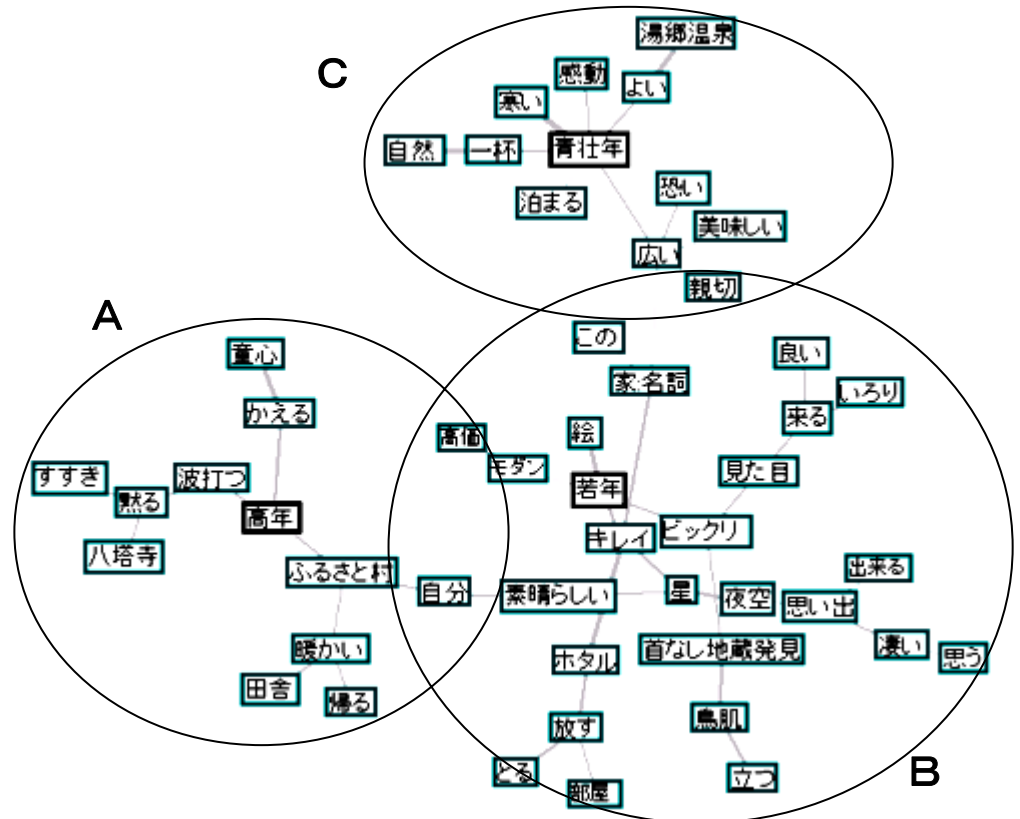


図5 コメント内容と記入者年齢層との相関図

あり、この点を踏まえて提言を行う。

- ◆ 幅広い年齢層に受け入れられ得るので、小学校にターゲットを当て、そこから家族連れをも引き込む。
- ◆ 「いろいろのある家」といったキャッチフレーズをつけて、いろいろを囲んだ食事・団欒の楽しさをアピールする。

山荘を都会の住民に、特に若年層の小学生に向けてアピールをする。家庭はもちろん、学校にも、児童の林間学習の場や総合学習の時間として山荘を利用してもらおう(山荘の定員は15人と人数のキャパシティーに若干の限度があるが)。子供に十分なアピールができれば今度は親を含めた家族での来訪も期待できるのではなかろうか。

## V 考察

今回の経験を踏まえて、来訪者ニーズを把握する対象として落書き帳を有効に活用するための留意点を考察すると、以下の点を指摘できる。

- 不要な書き込みが増えてきた場合には、早めに新しいノートを追加・交換しておく。落書き帳で前の来訪者による評価を読むことで、読んだ訪問者が先入観をもってしまう。マンガ等のつまらない書き込みが増加すると、返って悪印象を与えることになるので、ある程度の日常管理が必要である。
- 利用者の書き込みに対し、山荘管理者がしっかり返信する。この返信で訪問者(都市住民)と管理人(農村住民)との意思の疎通が図れ、落書き帳に目を通した訪問者も「これは農村側にも確実に感想を伝えられるもの」との意識が生まれ、建設的な意見を書き残してもらうことにつながる。
- 各ページに負担にならない程度のプロフィール欄を設ける。落書き帳書き込みの自由な印象を失われないように配慮しつつ、事後に活用する際に必要なデータを得やすくなる

ようなプロフィールを予め印字しておく。

## VI おわりに

本稿では、従来、計量的・客観的処理が困難であったテキストデータの解析手法として、テキストマイニング手法を導入した。落書き帳への適用事例の一部を紹介したが、ある程度有益な情報の抽出に成功しており、本手法は「農村計画や地域分析の現場に十分定着しうる実践的分析手法」であると判断できる。

反面、コンセプトマップの意味を正しく理解するためには、分析者がテキストデータの「読み込み」(意味の把握)作業を丁寧に行う必要がある。

なお本研究は、文科省科学研究費補助金(基盤研究(B)(1)、課題名:中山間地域における農村環境整備の計画手法に関する研究、代表者:松本康夫岐阜大学教授、課題番号:15380162)による成果の一部であることを付記する。

### 【注釈】

- 注1)居住地は書き込んでいるが同伴者については書かれていないもの等が多数あるので、項目毎の合計数は総サンプル数合計の146と一致しない。
- 注2)書き込み件数減少の正確な理由について地元住民の方に聴き取り調査を行ったが、はっきりとした原因を明らかにすることはできなかった。
- 注3)兵庫県西播磨県民局・岡山県東備地方振興局・岡山県勝栄地方振興局が合同で作成した「兵庫岡山県際交流マップ」にも八塔時ふるさと村は記載されている。

### 【参考・引用文献】

- 1) 那須川哲哉, 河野浩之, 有村博紀(2001):『テキストマイニング基盤技術』, 人工知能学会誌, No.16, Vo.2, pp201~211
- 2) 市村由美, 長谷川隆明, 渡部勇, 佐藤光弘(2001):『テキストマイニング事例紹介』, 人工知能学会誌, No.16, Vo.2, pp192~199
- 3) 渡部勇(2001):『ビジュアルテキストマイニング』, 人工知能学会誌, No.16, Vo.2, pp226~232
- 4) 富士通ソフトウェア生産技術研究所:『トレンドサーチ操作ガイド』
- 5) 岡山県吉永町:「八塔寺山荘」パンフレット

In this paper, the authors applied the text mining technique to the messages in the doodling books which the visitors of the "Hattouji Sansou" cottage in Yoshinaga town, Okayama prefecture, wrote down during 1993 - 2003. We do not have paid enough attention to doodling book as object of analysis until now. By the text mining technique, we can grasp the visitors' needs toward this accommodation. This paper made clear the potential advantages that this accommodation has and also pointed out an improvement direction that the advantages of this cottage are made fully use of.