

地域の認知と評価の集合知としての SALoT マップ

一場所の特徴的イメージを表す単語や文章の自動抽出

大森 宏 1, 羽生和紀 2, 山下雅子 3

1 東京大学大学院農学生命科学研究科, 2 日本大学文理学部, 3 東京有明医療大学

omori@lbm.ab.a.u-tokyo.ac.jp

概要 SALoT マップとは, ある地域において, スナップショット(写真)(S)と撮影した人の属性(A), 位置(Lo), その時感じたことや印象などのコメントのテキスト文(T)からなる 4 つ組の SALoT データを統合して作成される印象の集合知マップである. SALoT データは, 視覚的類似度から分類された見た目グループと撮影地点分布から分類された場所グループ, 属性による属性グループに分けられる. 形態素解析で単語を抽出し, コメントとグループ文集合の対応分析の結果が SALoT マップである. SALoT マップからグループを特徴づける単語やコメントの得点が算出される. TF-IDF による単語やコメント得点と総合して重要単語や重要文を抽出する. 地元大学生による埼玉県川越市内の気に入った景観を撮影する調査を行い, 177 名の学生から 242 件の SALoT データが集まった. 8 つの見た目グループと 7 つの場所グループが得られた. 場所グループのイメージを特徴づける重要語や重要文を抽出し, それらに基づき代表景観が抽出された.

キーワード 集合知, テキストマイニング, 対応分析, VisualRank

1 はじめに

ある地域の景観特性を明らかにすることは, 地域観光の振興や地元の住民が地域に対する愛着や満足を増進させるために重要である. 最近では地域の魅力や評判をブログや twitter などの SNS から抽出する研究が急速に進んでいる[1,2,3]. しかしながら, ウェブからの情報抽出では, 投稿者の性別や年齢層などの属性情報の取得は難しく, また, 投稿者の自由な意見や感想を受動的に取得する「受け身」の調査である.

一方, 従来からあるアンケート調査は, 被験者母集団の属性を定め, 聞きたい項目や内容は調査主催者が主体的に決められる「攻め」の調査であるが, コストの問題がある. 地域の認知と評価を写真投影法[4]を利用して行う参加型景観調査[5]を情報通信技術 (ICT) を利用して低コストで行い, 集合知を用いて計量的に扱う SALoT(Snapshot-Attribute-Location-Text)マップ法[6,7]を開発した. これは, ある指示を与え, 調査対象地域の景観写真を撮ってもらう. 撮影した写真(スナップショット)(S)とそれを撮った人の属性 (A), 撮影した位置(緯度・経度)(Lo), その時の印象や撮影理由などのコメントのテキスト文 (T) の 4 つの情報を 1 つのセットにした SALoT データを集め, それらを統合してその地域が与える印象の集合知を視覚化して具体的に表現するものである.

本稿では SALoT マップからの集合知の自動抽出法の改良点を述べる. SALoT マップから, あるグループを

特徴づける単語やコメントの得点を算出し, 重要語や重要文の自動抽出を行った. また, VisualRank[8]の補正項にコメント得点を用い, グループの代表景観の自動抽出も行った.

埼玉県川越市で地元大学生(東京国際大学)177 名による景観調査の結果得られた川越市の印章の集合知について報告する.

2 川越 SALoT マップの作成

2.1 多人数からの SALoT データの収集

調査参加者に何らかの指示を与え, 自身の所有する撮影機材を携行して対象となった場所を自由に散策させる. そこで, 指示に合っ目に止まった物や景色のスナップショットを何カ所かで撮ってもらう. このとき, 観測者の属性, 位置情報, コメントを記録してもらい, 電子メール等で収集する.

川越市内の「気に入った」景観を 2, 3 カ所撮影し, その理由やその時感じたことなどを記載してもらう調査を行った. 2010 年の調査では 49 名の学生(日本人男子 21 名, 日本人女子 23 名, 不明・外国人 5 名)から 64 件, 2011 年では 128 名の学生(男子 63 名, 女子 58 名, 不明・外国人 7 名)から 178 件, 計 242 件の SALoT データを収集した. その内訳は, 男子のものは 104, 女子のは 126, 不明・外国人のは 12 であった.

SALoT データの具体例は図 1 である. この例では地番から Google MAP で緯度, 経度情報に変換した.

属性 (A) : 日本人女子
位置情報 (Lo) : 川越市宮下町2-11-3 北緯 : 35.927998, 東経 : 139.488537
コメント (T) : ちょうど氷川神社に行ったときに戌の日で妊婦さんが安産祈願に来ていたり、お宮参りに来ていたり、縁結びのおみくじがあったりしたので幸せなスポットだと思い選びました。



図 1 SALoT データ例



図 2 撮影地点分布と場所グループ

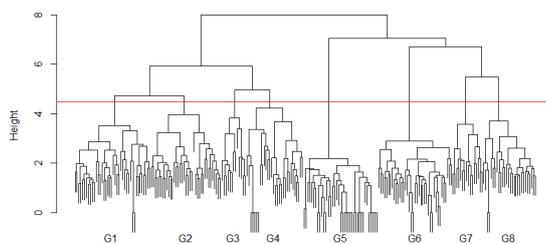


図 3 クラスタ分析から得られた 8 見た目グループ

表 1 属性と見た目, 場所グループとの 2 重分類

	G1	G2	G3	G4	G5	G6	G7	G8	計
男子	19	15	6	11	18	18	4	13	104
女子	21	19	7	17	19	19	11	13	126
計	40	34	13	28	37	37	15	26	230

	P1	P2	P3	P4	P5	P6	P7	計
男子	20	19	15	14	17	3	16	104
女子	19	29	6	17	12	22	21	126
計	39	48	21	31	29	25	37	230

2.2 属性の分類

興味ある属性で調査参加者を分類する. 今回の例では「日本人男子女子学生」とした.

2.3 位置情報の分類

撮影地点を電子地図上にプロットした (図 2), 川越中心部にある時の鐘付近は非常に狭い範囲であったが, 特異的な景観で, 多くの写真が集まったので独立した場所グループ (P1) とした. 時の鐘 (P1) から分布の連続性などを考慮して, 中心市街 (P2), 周縁部 (P3), 繁華街 (P4), 川越駅周辺 (P5), 郊外 (P6), 霞ヶ関 (P7) の場所グループに分けた.

表 1 の属性と場所グループの 2 重分類表の χ^2 検定は $p=0.002$ で, 男女で違いが認められ, P3 で男子, P6 では女子が有意に多く撮影していた. しかし, 図 2 の "M", "F" で示される男女の撮影地点平均は, ほぼ一致していて, 全体でみると偏りはみられなかった.

2.4 スナップショットの分類

242 枚のスナップショットを手動で似たものに分類するサイト[9]での分類課題を行ってもらった. 2 つのスナップショット間の視覚的類似度は, これらを同じグループに入れた人数の割合で測った. スナップショット間の類似度行列からクラスタ分析を行い, 図 3 のように 8 つの見た目グループに分けた.

表 1 の属性と見た目グループとの 2 重分類表の χ^2 検定は $p=0.868$ であり, 違いが認められなかった. これは, 男女でスナップショットの撮影内容に違いが無いことを示していた.

2.5 コメントのテキスト文の形態素解析

242 件の SALoT データのうち, 同一のコメントがいく

つか見られたので, これらを 1 つのもので代表させ, 235 のコメントを用いた.

コメントを RMeCab[10]を用いて形態素解析にかけて名詞, 形容詞, 動詞を抽出し, 整形を行って 1178 の形態素を抽出した. この中で「川越」や「喜多院」などの固有名詞, 「のどか」や「モダン」などの印象表現に関わる形態素を印象語とし, また, 「アパート」や「公園」など具体的な情報を表していた形態素を関連語とし, これら合わせて 907 を重要語として抽出した.

特定の個人の偏った印象表現を除くため, 2 つ以上のコメントで使用された 336 重要語のみを以後の解析に用いた. その結果, 固有名詞は 55 語, 印象語は 77 語, 関連語は 204 語になった. また, 組合せて使用される単語の重要性を増すため, 2 つ以上のコメントで同時に使用された重要語組合せ (「古い」「建物」など) 1725 通りも解析に用いた.

2.6 分類情報の統合による拡大行列の作成

どの重要語とどの重要語組合せがどのコメントに使用されているかを示す $(336 + 1725) \times 235$ の (重要語 + 組合せ) \times コメント行列を作成した. 2 つの属性グループと 8 つの見た目グループ, 7 つの場所グループが作成されたので, それぞれのグループごとにコメント列をまとめてグループ列をつくる. 表 2 のように, グループ列をもとの (重要語 + 組合せ) \times コメント行列に並べて加えた, (重要語 + 組合せ) \times (コメント + グループ) 行列を拡大行列と呼ぶ. 拡大行列を作成することで分類情報が統合される.

2.7 TF-IDF による重要語得点とコメント得点

場所グループの分類により重要語 \times グループ行列が作成されたので, TF-IDF を用いて重要語の各グループ

表 2 (重要語+組合せ)×コメント行列と拡大行列

	コメント列					グループ列	
	コメント1	コメント2	コメント3	コメント4	コメント5	G1	G2
所属	G1	G2	G1	G1	G2		
重要語1	1	0	0	0	0	1	0
重要語2	0	0	1	0	0	1	0
重要語3	0	1	0	0	1	0	2
重要語4	1	0	1	1	0	3	0
重要語5	0	0	0	0	1	0	1
:	:	:	:	:	:	:	:
組合せ i, j	1	0	0	1	1	2	1
組合せ i, k	0	0	1	0	0	1	0
組合せ i, m	0	1	0	0	0	0	1
:	:	:	:	:	:	:	:

表 3 TF と P1 における TF-IDF

	tfidf(P1)	P1	P2	P3	P4	P5	P6	P7
コメント数		39	51	18	35	32	23	37
音風景百選	0.293	3	0	0	0	0	0	0
400年	0.488	5	0	0	0	0	0	0
イメージ	0.114	2	0	0	1	2	0	0
小江戸	0.324	7	8	0	3	4	0	0
時の鐘	1.066	28	2	0	1	1	4	0
川越	0.769	30	25	5	15	13	5	8

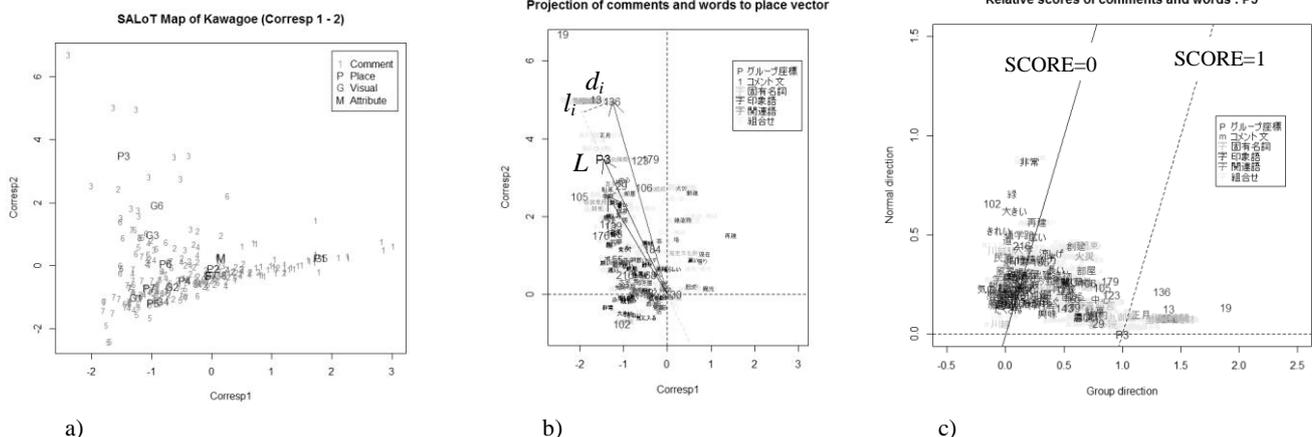


図 4 川越 SALoT マップ(第1-第2成分)とP3における重要語とコメントの相対得点分布

における得点を算出した。

TF(Term Frequency) は、ある場所グループでの重要語の頻度で、この重要語を用いたコメント数をグループ内コメント総数で割った値である。表 3 の例では P1 における「イメージ」は $TF=2/39$ となる。また、IDF(Inverse Document Frequency) は、ある重要語がどの程度場所グループで共通して使われているかを示す値で、 $IDF=\log(N/df)+1$ 、と定義される。ここで、 N はグループの総数 7 で、 df はこの重要語が使われたグループ数である。「イメージ」では $df=3$ である。TF-IDF はこの両者の積で定義される。表 3 に P1 での一部の重要語の TF-IDF 値を与えた。

また、コメントの得点は、そこで使用された重要語の TF-IDF 得点の総和とした。

TF-IDF 得点の高い重要語は、そのグループで特徴的に使用された単語と考えられ、また、TF-IDF 得点の高いコメントはそのグループの特徴を良く表現している文章(文集合)と考えられる。しかし、コメント TF-IDF 得点は、多くの重要語を使用した文章ほど高い得点を得る傾向(平均相関 0.91)があり、必ずしもグループの特徴を体現した文章とは言えないと感じられた。

2.8 川越 SALoT マップ

TF-IDF とは別の重要語とコメントの評価方法も用いた。先ほど説明した拡大行列を対応分析にかけた結果が SALoT マップ (図 4(a)) 第 1-第 2 成分である。このマップに各グループ、コメント、重要語、重要語組合せが同時にマップされ、これらの関係が視覚化される。

3 SALoT マップからの情報抽出

SALoT マップは、調査参加者が対象地域に抱いた印象の集合知を数値化したものと考えられる。このマップから各グループの印象を具体的に抽出する方法を述べる。

3.1 場所グループからの情報抽出

図 4(a)の P3 付近を拡大したのが図 4(b)であり、P3 グループ座標とコメント座標、P3 で使用された重要語と組合せ座標も同時に記載した。P3 グループの位置ベクトル方向にあるコメントと重要語は、P3 固有の情報を持っていると考えられる。

いま、グループ位置ベクトルの長さを L とし、コメント i の位置ベクトルをグループ位置ベクトルへ射影した長さを li 、グループ位置ベクトル方向からの偏差の大きさを di とすると、 li が大きく di が小さなコメントは P3 固有の特徴をよく表現していると考えられる。同様なことが重要語に関しても言える。

場所グループや見た目グループ間の分離をよくするため、SALoT マップの 6 次元までを用いることにした。ここで、 li や di を L で割ると、コメントと重要語の射影と偏差の相対得点が得られる。P3 での分布が図 4(c) である。

3.2 コメントと重要語の相対得点

場所グループ k でのコメントや重要語の相対得点を、グループ位置ベクトルの長さを L_k として

表 4 各場所グループを特徴づける特徴語上位 10

P1(時の鐘)	P1(中心市街)	P3(周縁部)	P4(繁華街)	P5(川越駅周辺)	P6(郊外)	P7(霞が関)
時の鐘	菓子屋横丁	喜多院	クリアモール	川越駅	安産	川
鐘	りそな銀行	本丸御殿	川越	モニュメント	祈願	入間川
川越	街並み	化粧の間	通り	工夫	ごみ集積所	第2キャンパス
400年	蔵造り	家光	買い物	雑多	川越観音	釣り
シンボル	古い	客殿	本川越駅	普通	光	バーベキュー
4回	川越	指定	近年	噴水	具合	川越線
川越大火	建物	重要文化財	宣伝	形	沼	霞ヶ関
音	建築	春日局	民家	駅	場所	歩道橋
観光地	登録有形文化財	書院	現代	クリアモール	立派	テニス
音風景百選	川越支店	誕生の間	川越市駅	川越	小さい	好き

表 5 各場所グループを説明する重要文上位 2

P1	午前6時、正午、午後3時、午後6時に「鐘」の「音」が鳴り、街の風景、漂う香りなどを加えると小江戸「川越」の情緒をたっぷり味合わせてくれるので、この「時の鐘」の写真を撮りました。 そして、なぜ私が時の鐘を選んだかと言うと、時の鐘は1624年～44年に酒井忠勝が、建てたのが最初と言われているようで、現在あるものは1893年に起きた「川越大火」の翌年に再建されたものだそうです。
P2	「菓子屋横丁」通りは、「蔵造り」の「町並み」や大沢家住宅と言った、江戸の風情を感じさせる「建築物」が全国的に有名である。 「古い」し、特別だし、または「建物」の前は昔の人力車が止まって、昔の時代を思い出した。
P3	「客殿」や「書院」には「家光」「誕生の間」や「春日局」「化粧の間」と伝えられている部屋があり、これらは国の「重要文化財」に「指川越の名所をいくつかまわったが「喜多院」は広々としていて休憩場所もあり木もたくさん植えられて緑が生い茂っており涼しい場
P4	「川越」にある「クリアモール」などの賑やかな感じとは一変してとても静けさを感じられ心が安らぎ何か神秘的なものを感じられるこのあたりはあまり「現代」的な建物が建っておらず、昭和の空気漂う「通り」なのだが、この「民家」だけがずば抜けて古い。
P5	「普通」は駅前には人が溢れ、「雑多」なイメージが大きいですが、小江戸とも呼ばれる「川越」の駅前には人の心を落ち着かせるような「工夫」がなされていると思い、非常に気に入りました。 「川越駅」前にある「モニュメント」。
P6	これは聖地となっており、開運・「安産」等の「祈願」ができるそうです。 この「場所」を選んだ理由は「小さい」けれど「立派」な神社で魅力的だったからです。
P7	川越の代表的な「川」である「入間川」は土日になると家族が「バーベキュー」や水遊びなどをして賑わい、またその上を走る「川越線」、東武東上線、は川越らしさを引き立てていると思いました。 「テニス」が「好き」で、夏に毎日お世話になったコートが私のお気に入りの場所です。

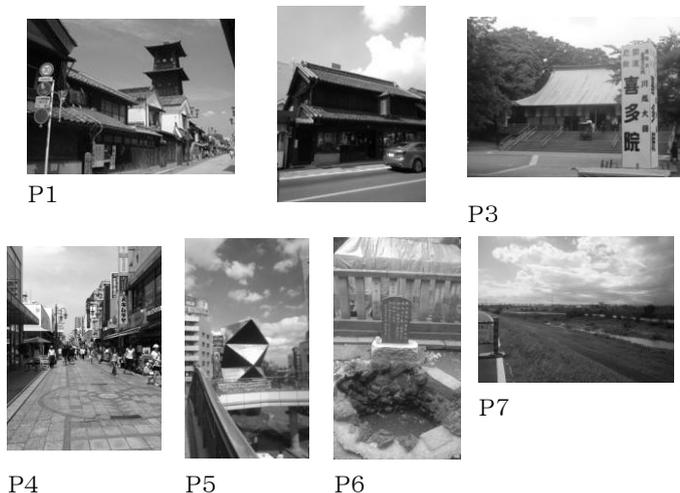


図 5 VisualRank で抽出された各場所グループの代表景観

$$SCORE_i = l_i/L_k - (1/2)d_i/L_k$$

と定義した。SCORE_iが 0 以下は、この場所にあまりふさわしくないコメントや重要語であり、1 以上のものはこの場所の固有な特徴をよく表現していると考えられる。d_i/L_kの係数は、コメント文全体をみて、これを 1/2 にすると、0 以下が約 30%、1 以上が約 20%となり、分布がちょうど良いと思われたからである。

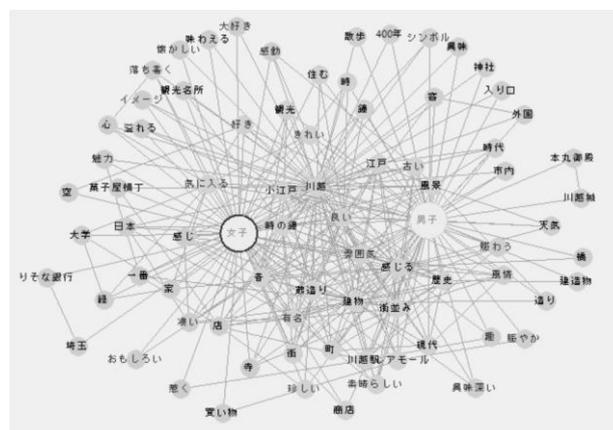


図 6 重要語の同時生起関係の男女比較

3.3 総合得点からの重要語抽出

場所グループで、重要語やコメントの TF-IDF 得点と相対得点が得られたので、この両者を足して 2 で割ることにより総合得点を求めた。具体的には両者を平均 0 標準偏差 1 に標準化してその平均をとった。場所グループごとで総合得点が高かった上位 10 の重要語を表 4 に

まとめた。これをみると、各場所グループの特徴がよく抽出されているように思われた。

3.4 総合得点からの重要文抽出

場所グループをよく説明する重要文を抽出することを試みた。場所グループに含まれるコメントの総合得点が算出されているので、総合得点の最も高いコメントがその場所を説明している文章とすることもできる。しかし、総合得点が高いコメントでもその場所のある特定の特徴しか説明していなかったり、冗長な表現である場合もあったので別の方法を考えることにした。

コメントはいくつかの文の集まりからなる文章になっているので、コメントではなく、文を抽出することを考えた。コメントを読点等で文に分解したところ、235 のコメントは 619 の文に分けられた。コメントと同様に文に対しても(重用語+組合せ)×文全体からなる行列を作成した。

SALoT マップの重要語と組合せの得点から、それぞれの文の場所グループ相対得点が算出できる。また、重要語の TF-IDF 得点からそれぞれの文の場所グループ TF-IDF 得点が算出できるので、両者を統合した総合得点が算出できる。

重要文の抽出は文集合から文選択を行う文書要約 [11]と同様であると考えられる。すなわち、ある場所グループに含まれるコメントを分解した文集合から良い文を選び出していく問題だからである。文選択では各文にスコアを与え高いものから逐次的に選択していく方法がとられる。この場合問題となるのが冗長性である。すなわち、高いスコアの文は互いに類似していることが多く、スコア順に選んでしまうと同様な内容の文が選ばれてしまう。

この問題に対処するために、本稿では一旦抽出された文に含まれた重用語に対しては、(重用語+組合せ)×文全体からなる行列において、その重用語に対応する部分を 0 にして文のスコアを再計算する方法を用いた。このようにすると、次の文選択では今までの文選択で重要とみなされた単語の重みが 0 となっているので、今まで文選択に使用されていなかった単語を用いた文の総合得点が算出されるからである。

このような方法で各場所グループから抽出された重要文上位 2 つを表 5 に与えた。文内で括弧がついた単語は、その場所で上位 10 にランクされた重要語である。これをみると各場所がうまく説明されているように感じられた。

3.5 場所グループ代表景観の抽出

VisualRank を用いてグループ内スナップショットの基準化得点ベクトル(要素の和が 1) r を求めた。対角成分を 0 とした類似度行列の列を基準化したものを S とし、補正ベクトルとしてコメントの基準化総合得点 p を用い、

$$r_i = \alpha S r_{i-1} + (1 - \alpha)p$$

を一樣な初期ベクトル r_0 から逐次更新して収束させた。なお、 $\alpha = 0.9$ を用いた。

これは、グループ内で類似写真が多いほど得点が高くなるもので、 p の値で補正するものである。この補正は、高い総合得点を持つコメントを作成した調査者が撮影したスナップショットはその場所を代表するであろう、という考えを反映している。

基本的には r 値が最大のスナップショットを代表としたが、映りが悪い場合は次点のスナップショットを代表とした。各場所グループの代表景観を図 5 に与えた。

3.6 属性による印象の違い

表 1 に示されたように、撮影した写真の内容には男女間で違いが認められなかったが、図 4(a) の SALoT マップではコメント文の内容に男女間で多少の差があるように見えた。

男女間での使用重要語の違いを視覚化するため同時生起関係をグラフ化した(図 6)。男女それぞれで 3 つ以上のコメント文で使用された重要語を用いた。

中央部に女子と男子の頂点があり、この間に位置する単語は男女から枝が伸びているので、男女ともによく使用された単語を表す。一方、女子の左側が女子のみによく使用された単語、男子の右側が男子にのみよく使用された単語をあらわしている。

これを見ると、女子は「菓子屋横丁」、「買い物」など自分の行動に関連した単語を用い、「落ち着く」、「おもしろい」、「大好き」など自分の感情をストレートに表現していた。一方、男子は「400年」、「川越城」など歴史的なことに興味を示し、「シンボル」、「賑やか」など客観的な印象を記述する傾向があった

4 おわりに

地元大学生 177 名による川越市の景観調査を行った。収集された 242 の SALoT データは、学生の目に映った川越のイメージが投影されていると考えられた。SALoT マップ法で、川越の集合知的なイメージを具体的に抽出することができた。

コメントの中には、「統計学の調査課題により、改めて川越を知るきっかけになり良かった。」というもあり、調査を行うこと自体が自分の住む街の愛着を高める効果があることが示唆された。

今回は「気に入った景観」の調査であったが、「お薦めのスポットは」など指示内容を変えることにより、別の調査も可能である。データの取得は Web からでは無く、人手による調査からであったが、ICT の利用により低コストで行うことができた。

撮影された景観の内容には男女で差が認められなかったが、その印象の記述には違いが見られた。女子は自分と川越との関わりを記述するのに対し、男子は第三

者的な目で見た川越を記述していた。

今回ののは地元学生の調査だったため、大学が P7 で撮影されたスナップショットが多かったが、もし、旅行者が調査者であったならば、P7 で撮影した SALoT データは多くなかったと考えられる。このように調査参加者の属性は地域の認知と評価を分析する上で非常に重要である。

学生以外の旅行者などの別属性をもった人物に調査協力してもらうには、調査協力者に現地商店での割引クーポンを発行するなどの謝礼を準備すれば、比較的簡単に集めることができると考えられる。人数は数百人程度で十分と考えられるのでコストはこの場合でもそれほどかからないと思われる。

得られたテキストデータから、対応分析と TF-IDF を利用して場所グループにおける単語のスコアを求めた。このスコアを用いて、コメントや文のスコアが算出された。見た目グループにおける単語スコアも同様に算出できる。このスコアを用いて、グループ文集合から重要文の自動抽出を行い、冗長性の少ない文が抽出できた。しかし、抽出された文には重要でないと考えられる部分もあるので、係り受け解析から「枝狩り」により文短縮を行い、より簡潔で重要な印象のみを含む文にする改良が必要だと考えられる。

スナップショット間の類似度行列から VisualRank により場所の代表景観の抽出を行った。コメント総合得点で補正することにより、場所を代表する景観が抽出できたと考えられる。現在は、人手により視覚的類似度を計測しているので、正確ではあるがコスト的に問題がある。また、100 程度のスナップショットではこの方法で類似度は計測できるが、数千、数万のスナップショットでは不可能である。このため、画像解析による類似度と人手による類似度を組み合わせて行く必要がある。

謝辞

東京国際大学の統計学の基礎の受講生におかれましては、「川越市の気に入った景観を探す。」という課題に多くの学生が快くご協力くださり、自由闊達な調査データを提供していただいたことに感謝しております。

参考文献

- [1] 石野亜耶, 難波英嗣, 竹澤寿幸, 旅行ブログエントリーからの観光情報の自動抽出, 知能と情報, 22(6), 667-679, 2010.
- [2] Yin, Z., Cao, L., Han, J, Zhai, C., Huang, T., Geo-geographical topic discovery and comparison, In: WWW, 247-256, 2011.
- [3] 奥健太, 橋本拓也, 上野弘毅, 服部文夫, 地理情報推薦のための観光スポットと位置情報付きユーザ生成コンテンツの対応付け手法の提案, DEIM2013 論文集, 2013.
- [4] 野田正彰, 漂白される子供たち—その眼に映った都市へ, 205pp, 情報センター出版局, 1988.
- [5] 古賀誉章, 高明彦, 宗方淳, 小島隆矢, 平手小太郎, 安岡正人, キャプション評価法による市民参加型景観調査—都市景観の認知と評価に関する研究その1, 日本建築学会計画系論文集(517), 79-84, 1999.
- [6] 大森宏, 羽生和紀, 山下雅子, SALoT マップ: スナップショット・位置・テキストによる印象の集合知と街歩きマップ, 日本建築学会計画系論文集 78(638), 159-166, 2013.
- [7] 大森宏, 羽生和紀, 山下雅子, 埼玉県川越市の景観に対し地元大学生が抱いた印象の集合知としての SALoT マップ, ランドスケープ研究, 77 (5), 491-496, 2014.
- [8] Jing, Y. and Baluja, S., VisualRank: Applying pagerank to large-scale image search. IEEE trans. Pattern Analysis and Machine Intelligence, 30(11), 1870-1890, 2008.
- [9] 大森宏ホームページ, 川越景観グルーピングサイト, <http://lbn.ab.a.u-tokyo.ac.jp/~omori/landscape/groupingkawagoe.htm>.
- [10] 石田基広, R によるテキストマイニング入門, 173pp, 森北出版, 2008.
- [11] 高村大也, 奥村学, 最大被覆問題とその変種による文書要約モデル, 人工知能学会論文誌, 23(6)C, 505-513, 2008.