

地域に詳しいブロガーの発見

三浦 惇貴[†] 廣田 雅春^{††} 野澤 浩樹^{†††} 横山 昌平^{††††}

[†] 静岡大学大学院情報学研究科 〒432-8011 静岡県浜松市中区城北 3-5-1

^{††} 大分工業高等専門学校情報工学科 〒870-0152 大分県大分市牧 1666

^{†††} 株式会社シーポイント 〒432-8002 静岡県浜松市中区富塚町 1933-1 佐鳴湖パークタウンサウス 2F

^{††††} 静岡大学大学院情報学研究科 〒432-8011 静岡県浜松市中区城北 3-5-1

E-mail: [†]gs14043@s.inf.shizuoka.ac.jp, ^{††}m-hirota@oita-ct.ac.jp, ^{†††}hiro@c-point.co.jp,

^{††††}yokoyama@inf.shizuoka.ac.jp

あらまし 我々は、記事内で言及している場所 (Venue) を発見し、各 Venue に詳しく言及しているブロガーを地域 (町・字単位) ごとにランキング形式で Venue に関する記事とともに提示するシステムを実装した。実際に、浜松市を中心とした地域ブログサイト [はまぞう] の記事を利用したシステムを実装し、地域密着型 SNS サービス [LINK] に導入した。はまぞうは、地域の多くの Venue が公式ブログを開設しているだけでなく、一般ブロガーがそれら Venue に対する口コミ記事を多数投稿していることを特徴とする。本システムにより、Venue に関する情報を取得したいユーザにとっては、Venue の公式ブログだけでなく、Venue とその周辺に関する信頼できる口コミ記事に効率的にアクセスすることが可能となる。また、一般ブロガーにとっては、Venue についての口コミ記事を投稿するモチベーションとなると期待される。

キーワード Hyperlocal, Local Expert, 忘却の概念, ユーザ評価, Foursquare

1. はじめに

SNS におけるレストランやショップなどの店舗に関する口コミを分析する際、それがインターネットという地球規模のメディア上で発信されている情報であるにも関わらず、その情報はローカルな狭い範囲においてのみ有用な情報である事を考慮する必要がある。つまり、SNS ユーザの生活圏・行動範囲は限られており、旅行や出張などを除けば、生活圏の中で体験した情報を発信しているに過ぎない。しかし、この事を地域を軸として考えると、地域ごとにその場所に詳しいユーザが存在している事を意味している。また、口コミを検索するユーザにとっても、検索対象の地域に詳しいユーザを知る事により、情報の鮮度・確度を推し量る事ができるようになる。

地域情報サイトは、地域性が強く、その地域に関する情報が多いため、一般のサイトと異なり、地域に特化した情報を得ることが可能である。また、ブログはマイクロブログと比較して、ユーザが内容を慎重に考えるため、記事の内容が詳しく、文章が整形されている傾向にある。そのため、ブログを利用することで、少ない記事数でより詳しいユーザを容易に発見することが可能である。よって、地域に詳しいユーザを発見する情報源として、地域性・内容の詳しさ・容易性を持つ地域ブログサイトが最適であると考えられる。

本研究では、浜松市を中心とした地域ブログサイト [はまぞう]^(注1) のブログ記事内で言及している場所 (Venue) を発見する。また、ブログ記事中の口コミや画像に基づいて算出した各

Venue に対する詳しさのスコアを用いて、各 Venue に対して詳しく言及しているブロガーを地域 (町・字単位 (e.g. “浜松市中区城北”)) ごとにランキング形式で Venue に関するブログ記事とともに提示するシステムを実装する。ランキングを作成する際、投稿日時の新しいブログ記事から得られる情報は、投稿日時の古いものと比べて価値の高い情報であると考えられる。そこで、本システムでは、情報の鮮度を考慮する (新しいブログ記事ほどランキングに影響を持たせる) ため、時間とともに情報価値が減っていくことを忘却の概念を利用してモデル化し、ランキングの作成の際に利用する。さらに、実際に、地域密着型 SNS サービス [LINK]^(注2) に対して、本システムを導入する。

はまぞうは、2005 年 10 月にサービスを開始したブログサイトであり、ブロガーのほとんどが浜松地域に住んでおり、地域の多くの Venue が公式ブログを開設しているだけでなく、ブロガーがそれら Venue に対する口コミ記事を多数投稿していることを特徴とする。2015 年 12 月時点の総ブログ数は 32,795 (内、Venue の公式ブログは 2,461)、総ブログ記事数は 6,565,776、月間 PV 数は約 1,700 万、1 日の平均投稿数は 3,140 となっている。

我々はこれまで、はまぞうのブログ記事を利用して、ブログ記事の内容が良い口コミ・悪い口コミ関係なく、各 Venue に対して詳しく言及しているブロガーを Venue ごとにランキング形式で Venue に関するブログ記事とともに提示するシステムを実装し、そのシステムを LINK に導入してきた [1]。しかしながら、口コミサイトとは異なり、Venue の公式ブログの記事

(注1) : <http://www.hamazo.tv/>

(注2) : <http://linkz.jp/index.html/>

と、一般ブロガーのロコミ記事が混在するブログというメディアにおいて、公式ブログとともに、悪いロコミ記事を並べて表示する事は、実ビジネス上現実的とは言えない。一方で、情報を得たいユーザにとっては良いロコミよりも、悪いロコミを参考にするため、より多くの影響を与えるという調査^(注3)もあり、単に、良いロコミだけを掲載して、悪いロコミを排除するという方法では、ユーザの満足度に悪影響を与える事となる。そのため、良いロコミ記事のみを扱ったサービスも重要であるが、悪いロコミ記事も扱ったサービスはより重要であると考えられる。以上を踏まえ、本研究では、各 Venue に対して詳しく言及しているブロガーを Venue ごとでなく、地域ごとにランキング形式で Venue に関するブログ記事とともに提示するシステムを実装する。

本システムにより、Venue に関する情報を取得したいユーザにとっては、Venue の公式ブログだけでなく、Venue とその周辺に関する信頼できるロコミ記事に効率的にアクセスすることが可能となる。また、ランキングのようなゲーム要素は、SNS におけるユーザの投稿に対するモチベーションの向上すると言われている [2]。そのため、地域に詳しいブロガーをランキング形式で提示することで、ブロガーにとっては、Venue についてのロコミ記事を投稿するモチベーションとなると期待される。さらに、既存のサービスに本システムを導入するため、他のサービスのように専用のサイトを立ち上げる必要がなく、加えて、システム側でブログ記事の内容をロコミとして解析するため、ブロガーはわざわざロコミ記事を投稿するのではなく、今まで通りブログ記事を投稿するだけで良い。よって、容易に新サービスを始めることが可能になる。

2. 関連研究

2.1 Venue の発見に関する研究

河野らの研究 [3] は、Venue を発見する際に、形態素解析とキーフレーズ解析を行い、形態素解析の結果から「地名」や「固有名詞」と判定された単語、キーフレーズ解析の結果から「地名を含んだキーフレーズ」を Venue としている。そのため、それらの条件を満たさない Venue の発見ができない。また、倉島らの研究 [4] は、形態素解析を用いて、地名・ランドマークと格助詞「へ/に」の共起に着目し、格助詞と共起している単語列を Venue としている。さらに、中嶋らの研究 [5] は、場所を表す格助詞が出現する文を係り受け解析し、その格助詞の直前の名詞を Venue としている。これらの研究は、形態素解析や、係り受け解析に基づいており、特定の品詞で表現された Venue 以外は発見できないという課題がある。そのため、本研究では、Foursquare から収集した Venue を形態素解析辞書に事前に登録することで対応する。Foursquare には、2013 年 10 月 7 日の時点で、浜松市役所から半径 16km の範囲に 10,918 件もの Venue が登録されているため、課題が解決できると考えられる。

2.2 レビュー分析に関する研究

平山らの研究 [6] は、商品の評価項目を自動で抽出する手法を提案している。平山らの手法では、レビュー文をに対して、係り受け解析を行い、「形容詞」、「名詞（形容動詞語幹）」、「名詞（ナイ形容詞語幹）」、「動詞（係り先の次の語が形容詞の場合のみ）」に係っている語を評価項目として抽出している。また、小西らの研究 [7] は、トピックモデルを用いて、評価項目を推定する手法を提案している。小西らの手法は、レビュー文に対して、周辺文を考慮するトピックモデルの distance dependent Chinese Restaurant Process(dd-CRP) を応用することで、評価項目を推定している。本研究では、ブログ記事を扱うため、レビュー以外の文も数多く現れる。そのため、これらの手法を適用した場合に、適切な評価項目が得られないと考えられる。そこで、本研究では、評価項目を恣意的に与えることで対応する。

2.3 ユーザランキングに関する研究

Bao らの研究 [8] は、Foursquare を用いて、詳しいユーザを抽出する手法を提案している。Bao らの手法は、Foursquare のチェックイン履歴とチェックイン回数を用いて、あるカテゴリにおける詳しいユーザを抽出している。Bao らの手法では、チェックインのみで詳しいユーザを抽出している。本研究では、ブログ記事を用いて、Venue に対するレビューも反映することで、チェックイン回数だけではなく、Venue に対する詳しさを考慮する。そうすることで、Venue に対する詳しさを考慮した地域に詳しいブロガーをブログ記事から抽出する。Venue に対するレビューも反映する理由は、Venue に対する詳しさの度合いを算出する際に、訪れた回数だけではなく、Venue に対する意見や感想も重要なためである。

稲垣らの研究 [9] は、体験熟度度に基づいてブロガーのランキングを作成する手法を提案している。“マニア”、“ファン”、“フリーク”といった単語の直前の名詞句 (e.g. “鉄道”、“コンピュータ”) をブログ記事から抽出し、名詞句の共起語と名詞句を含むブログ記事の回数を基にブロガーのランキングを作成している。本研究では、Venue に対して詳しく言及しているブロガーを発見するため、“マニア”、“ファン”、“フリーク”といった単語を利用して Venue を発見することは困難である。よって、Foursquare から収集した Venue を形態素解析辞書に事前に登録することで対応する。また、稲垣らの手法は、情報の鮮度を考慮していないため、古いブログ記事から得られる情報は、現在の情報と異なっている場合がある。よって、情報の鮮度を考慮するため、ランキング作成後に投稿されたブログ記事の重みを重視しつつランキングを更新する。

3. 提案システム

本研究で提案するシステムの構成図を図 1 に示す。はじめに、Venue を発見するため、ブログ記事を形態素解析し、その結果から Venue 候補を発見する。そして、Venue 候補に対して、位置情報とカテゴリを付与する。次に、Venue のカテゴリに基づいて、ブログ記事の採点を行う。ランキングを表示したい地域

(注3) : 「購買行動におけるクチコミの影響」に関する調査 : http://www.ntt-east.co.jp/ictsoudan/wp/wp1/pdf/02/investigation_04.pdf

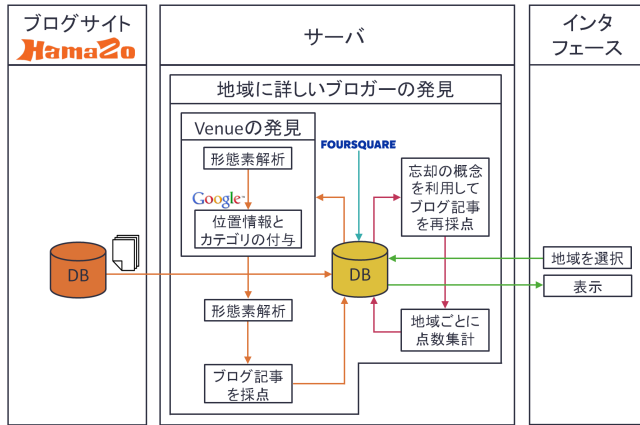


図 1 システム構成図

表 1 形態素が 1 つの名詞でない Venue の例

Venue	形態素解析結果
にこまん馬	に (助詞)/こまん (動詞)/馬 (名詞)
ピッツァ・なお	ピッツァ (名詞)/・ (記号)/なお (接続詞)
とんきい	とん (動詞)/きい (名詞)
中華の鉄人	中華 (名詞)/の (助詞)/鉄人 (名詞)

に基づき、ブログ記事の投稿日時を考慮し再採点を行い、地域に詳しいブロガーの発見を行う。また、ランキングを作成する際、情報の鮮度を考慮するため、時間とともに情報価値が減っていくことを忘却の概念を利用してモデル化し、ランキング作成後に投稿されたブログ記事の重みを重視しつつランキングを更新する。

3.1 Venue の発見

本研究で発見する Venue とは、地域名である“浜松”や、店名である“にこまん馬”などのような場所を表す単語である。発見した Venue には、カテゴリを付与し、地域に詳しいブロガーの発見を行う際の、Venue の評価項目の決定に利用する。

3.1.1 Venue 候補の発見

ブログ記事に対して形態素解析を適用し、Venue 候補を発見する。形態素解析には、Kuromoji^(注4)を用い、形態素解析辞書として、NAIST Japanese Dictionary^(注5)を用いる。

Venue を含むブログ記事に形態素解析を適用した場合、表 1 に示すように、Venue が 1 つの形態素と判定されない場合がある。これに対処するために、Foursquare に登録されている Venue を用いる。浜松市役所から半径 16km 以内の Venue を、Foursquare API^(注6)を用いて収集し、Venue の name 属性 (スポット名) を形態素解析辞書に事前に登録する。また、スポット名の登録は、以下のルールに従って行う。全角スペース、および半角スペースが含まれているスポット名 (e.g. “CLOCHE (クロッシュ)”) は、全角スペース、または半角スペースで区切り、別々に単語としての登録を行う。4 文字以下の英数字のみ

の単語や、1 文字のみの単語は登録を行わない。さらに、品詞情報が「地域」である単語や、単語の末尾が“店” (e.g. 浜松店) や“処” (e.g. そば処) である単語も登録を行わない (e.g. “ホテルコンコルド 浜松”は、“ホテルコンコルド”のみ、また、“そば 処 津がわ”は、“津がわ”のみを登録)。“.”を含む単語は、“.”を含む単語と除去した単語の両方を登録する (e.g. “ピッツァ・なお”に対して、“ピッツァ・なお”と“ピッツァなお”両方を登録)。丸括弧を含む単語は、丸括弧内に記載されている単語を登録する (e.g. “CLOCHE (クロッシュ)”に対して、“CLOCHE”と“クロッシュ”両方を登録)。以上のルールに従い登録した辞書を用いて形態素解析し、得られた名詞、または登録された単語 (F-名詞) を Venue 候補とする。

3.1.2 Venue の決定

3.1.1 節で発見した Venue 候補に対して、次に示す手順で位置情報、およびカテゴリ (e.g. “Japanese Restaurant”, “Sake Bar”) の付与を行う。

Venue 候補の品詞情報が「名詞」かつ「地域」の場合は、“地形”のように地名が店名などの名前になっていることがあるため、Venue 候補がスポット名の場合は、Foursquare から取得したその Venue の位置情報、およびカテゴリを付与する。一方、スポット名ではない Venue 候補は、Google Geocoding^(注7)を用いてジオコーディングを行い、位置情報と“Neighborhood”というカテゴリを付与する。ジオコーディングの結果として、複数の位置情報が得られる場合がある (e.g. 単語“浜松”をジオコーディングした場合、東京都浜松町と静岡県浜松市)。そのため、本研究では、ジオコーディングの際に浜松市役所から半径 16km 以内の結果のみを利用する。その他の「名詞」は、スポット名と一致した場合のみ、位置情報、およびカテゴリを付与する。F-名詞の場合は、該当の Foursquare から取得した Venue の位置情報、およびカテゴリを付与する。この処理によって、位置情報、およびカテゴリが取得できたものを本システムで扱う Venue とする。

3.2 地域に詳しいブロガーの発見

本研究で発見する地域に詳しいブロガーは、地域ごとに Venue に対して詳しく言及しているブロガーのことを示す。地域に詳しいブロガーは、Venue に関するブログ記事が含む Venue に対する評価 (記事ポイント) や、ブログ記事中に画像が含まれているか (画像ポイント) に基づいて、ブロガーの Venue に対する詳しさを評価する。

以下の式を用いてブロガー u の Venue v に対する、スコア $Score_{uv}$ を求め、地域 (町・字単位 (e.g. “浜松市中区城北”)) ごとに $Score_{uv}$ を集計し、地域ごとに集計した $Score_{uv}$ 降順にブロガーをランク付けする。

$$t_j = \lambda^{\tau - T_j} \quad (0 < \lambda < 1) \quad (1)$$

$$Score_{uv} = \sum_{j \in J_{uv}} (a_j + i_j) * t_j \quad (2)$$

(注4) : <http://www.atilika.org/>

(注5) : <http://sourceforge.jp/projects/naist-jdic/>

(注6) : <https://developer.foursquare.com/>

(注7) :

<https://developers.google.com/maps/documentation/geocoding/>

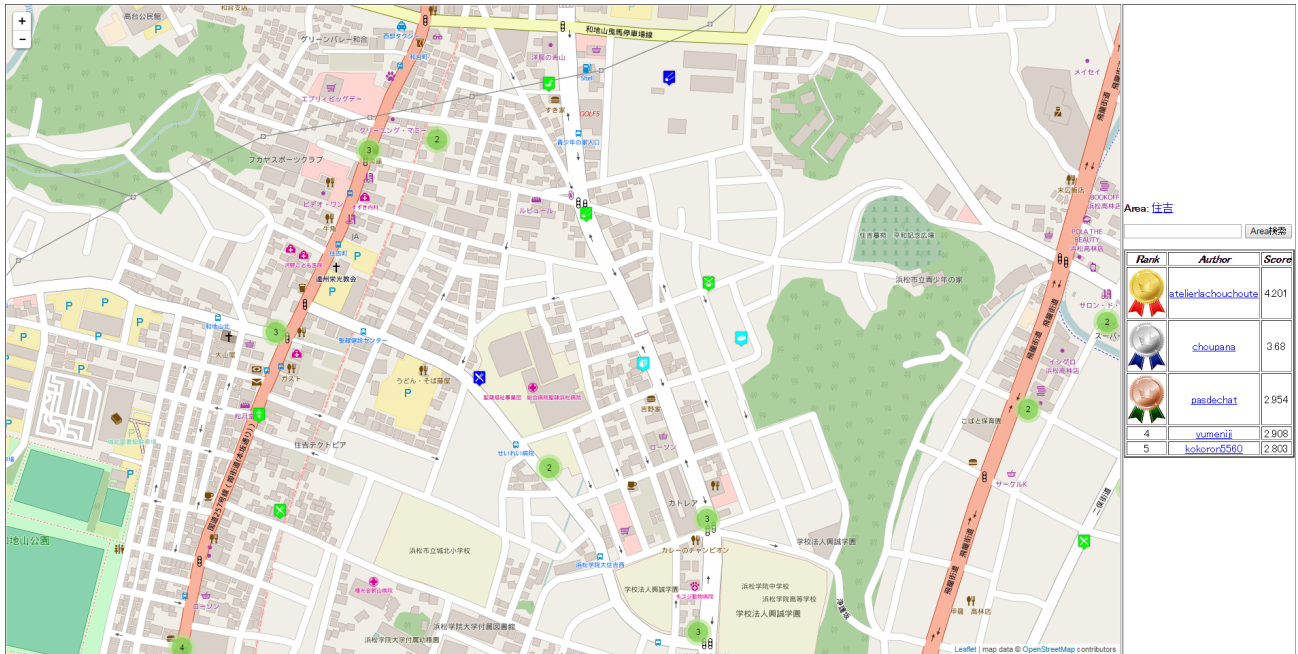


図 2 提案システムの実装

表 2 評価項目および類義語

評価項目	類義語の一部	評価項目	類義語の一部
味	風味, 旨味, こく	部屋	個室, 座敷, ルーム
見た目	見映え, 表面, 見かけ	設備, アメニティ	備え付け, 施設, 設置
サービス	待遇, 接客, 値下げ	立地	場所, 敷地, 用地
雰囲気	ムード, 佇まい, 風情	接客, サービス	レセプション, 応対, 接待
コストパフォーマンス (コスパ), ボリューム	量, 分量, 大きさ	食事, 料理	夕飯, ダイニング, ディナー
酒, ドリンク	アルコール, ワイン, 焼酎	料金	値段, 代金, 払い

(a) カテゴリ (レストラン系)

(b) カテゴリ (ホテル系)

ここで t_j はブログ記事 j の投稿日時による重み, τ は現在の時刻, T_j はブログ記事 j が投稿された時刻, a_j はブログ記事 j の記事ポイント, i_j はブログ記事 j の画像ポイント, J_{uv} はブロガー u のブログ記事中に Venue v を含むすべてのブログ記事である。

ブログ記事 j の投稿日時による重み t_j は, ブログ記事の情報の鮮度を考慮するものである. 本研究では, Khy らの研究 [10] を基に, ブログ記事の情報の鮮度を考慮することで, 時間とともに情報価値が減っていくことを, 忘却の概念を利用してモデル化した. 式 (1) の λ は, 忘却ファクター (forgetting factor) と呼ばれ, λ ターゲットとなるブログ記事や想定するアプリケーションに応じて値を変える定数である. λ の値が小さくなるほど, ブログ記事の忘却 (ブログ記事の投稿日時による重みの通減) の速度が速くなることを示す.

記事ポイント a_j とは, ブログ記事 j に対して, 形態素解析を適用し, 口コミサイトに基いて決めた 6 つの評価項目や, 評価項目の類義語が出現した場合, そのブログ記事を投稿したブロガーに与えるポイントである. 評価項目や評価項目の類義語が出現した場合, 記事ポイントとして 1 点を与える. 各評価

項目に対して, 記事ポイントは最大 1 点とするため, 1 ブログ記事に対する記事ポイントは最大 6 点となる. 評価項目は, 詳しいブロガーを発見する Venue のカテゴリに基づいて定める (e.g. カテゴリが “Japanese Restaurant” の場合は, 「味」, 「雰囲気」などの項目) .

画像ポイント i_j とは, ブログ記事 j に画像が含まれている場合にブロガーに与えるポイントである. ブログ記事に画像が含まれている場合, 画像ポイントとして 1 点を与える. 各ブログ記事に対して, 画像ポイントは最大 1 点とする. 本研究では, ブログ記事に画像が含まれているか否かで, 画像ポイントを与えている. その理由として, 評価項目やその類義語の周辺に画像が存在した場合, その画像はそのブログ記事と関連性が高いと考えられるためである. 実際に, Kherf らの研究 [11] において, 一般的な Web 画像検索では, 画像の近傍に存在するテキストを, その画像と関連性の高いものとして扱うことが多いと述べている.

4. 実装および比較

本研究では, はまぞうに 2002 年 11 月 30 日から 2013 年 6

表 3 “浜松市中区住吉”のランキング結果

Rank	Author	Score	Rank	Author	Score	Rank	Author	Score
1	kokoron5560	231	1	kokoron5560	25.8	1	atelierlachouchoute	4.20
2	bistro33vin	181	2	bistro33vin	11.0	2	choupapa	3.68
3	choupapa	90	3	harahettanya	8.37	5	kokoron5560	2.80
4	cocollo	80	5	choupapa	7.69	16	harahettanya	1.99
14	harahettanya	32	6	atelierlachouchoute	7.22	20	cocollo	1.78
79	atelierlachouchoute	10	7	cocollo	6.62	42	bistro33vin	1.01

(a) 忘却の概念なし

(b) $\lambda = 0.9926$ (c) $\lambda = 0.9779$

月 30 日の間に投稿されたブログ記事 5,052,326 件を用いる。また、Foursquare に登録されている Venue は、2013 年 10 月 7 日の時点までに登録された 10,918 件を用いる。

4.1 提案システムの実装

今回、扱う Venue は、Foursquare におけるカテゴリが、“Food”または“Nightlife Spot”に分類されているものと、“Bed & Breakfast”、“Boarding House”、“Hostel”、“Hotel”、“Motel”または“Resort”となっているものを用いる。カテゴリが“Food”または“NightlifeSpot”(レストラン系)である Venue の評価項目は、食べログ^(注8)、ロケタッチ グルメ^(注9)、ぐるなび^(注10)に基づいて、表 2(a) に示す 6 項目を評価項目および、その類義語とした。評価項目および類義語を表 2(a) に示す。また、それ以外のカテゴリ (ホテル系) の Venue の評価項目は、楽天トラベル^(注11)、Yahoo トラベル^(注12)、じゃらん^(注13)に基づいて、表 2(b) に示す 6 項目を評価項目および、その類義語とした。評価項目および類義語を表 2(b) に示す。類義語は、WordNet^(注14)に基づいて、評価項目として相応しい単語を手で選出した。表 2(a) と表 2(b) に示した評価項目と類義語は、3.2 節で述べた地域に詳しいブロガーの発見に用いる。

図 2 に、“浜松市中区住吉”という地域に対して、提案システムを用いた結果を示す。図 2 の左側は、提案手法を用いて発見した Venue を OpenStreetMap^(注15)に表示している。マーカーのデザインは、Venue のカテゴリを表している。マーカーの色は、Venue のチェックイン数 (Venue について言及したブログ記事の数) を表している。図 2 の右側は、“浜松市中区住吉”という地域に詳しいブロガーを示している。

4.2 ランキングの変動比較

本比較では、忘却の概念の利用前後でのランク付け、および式 (1) の λ (忘却ファクター) の値を変えることでのランク付けを比較する。式 (1) の τ の値は 2013 年 6 月 30 日とした。

表 3(a)、表 3(b)、表 3(c) に“浜松市中区住吉”という地域に

対するランキング結果を示す。表 3(a) は、忘却の概念を利用しないときの結果である。表 3(b) は、式 (1) におけるパラメータ λ を 0.9926 と設定したときの結果である。0.9926 という値は、3ヶ月程度でポイントが半減するような設定である。表 3(c) は、式 (1) におけるパラメータ λ を 0.9779 と設定したときの結果である。0.9779 という値は、1ヶ月程度でポイントが半減するような設定である。

表 3(a) と表 3(b)、表 3(c) を比較すると、忘却の概念の利用前後で、ランキングに変動があることがわかる。表 3(b) と表 3(c) を比較すると、式 (1) の λ の値を変えることで、ランキングに変動があることがわかる。例えば、author “bistro33vin” に着目すると、表 3(a) では 2 位であり、表 3(b) でも 2 位であるが、表 3(c) では 42 位となっている。“bistro33vin”のブログ記事を見ると、2013 年 5 月から 2013 年 6 月まで浜松市中区住吉の店舗について、ブログ記事を投稿していないため、このようなランキング結果になったと考えられる。また、author “atelierlachouchoute”に着目すると、表 3(a) では 79 位であったが、表 3(b) では 6 位、表 3(c) では 1 位となっている。“atelierlachouchoute”のブログ記事を見ると、2013 年 3 月から 2013 年 6 月までの間に浜松市中区住吉の店舗に関するブログ記事を投稿しているため、このようなランキング結果になったと考えられる。

これにより、忘却の概念を利用することで、ブログ記事をブログ記事を最近投稿したブロガーのランクを高く評価する傾向があることがわかる。また、式 (1) の λ の値を変えることで、ランキングに変動があることがわかる。今後の課題として、評価実験を行い、忘却の概念の有効性を調査することが考えられる。また、ブロガーのブログ記事の投稿頻度を利用して、適切な λ の値を調査することが考えられる。

4.3 提案システムの有用性の評価

本節では、提案システムの有用性を Google 検索と比較して評価する。評価方法は、Google 検索^(注16)を用いて、Venue 名の完全一致、かつ hamazo.tv のサイト指定検索 (“Venue 名 site:hamazo.tv”) を行い、Google 検索を用いたブロガーランキングを作成し、以下の 2 つを検証した。なお、最終更新期間の終了日は 2013 年 6 月 30 日とした。本実験に用いた Venue は、表 2(a)、表 2(b) に示したものである。

(注16) : <https://www.google.co.jp/>(注8) : <http://tabelog.com/>(注9) : <http://gourmet.livedoor.com/>(注10) : <http://www.gnavi.co.jp/>(注11) : <http://travel.rakuten.co.jp/>(注12) : <http://travel.yahoo.co.jp/>(注13) : <http://www.jalan.net/>(注14) : <http://nlpwww.nict.go.jp/wn-ja/>(注15) : <http://www.openstreetmap.org/>

表 4 上位 5 人分のブロガーランキングの作成に必要な閲覧ページ数

レストラン系	ホテル系
3.4 ページ	3.4 ページ

表 5 Google 検索のランキングに提案システムのブロガーがランクインした数

レストラン系	ホテル系
2.4 人	1.8 人

・ 上位 5 人分のブロガーランキングの作成に必要な閲覧ページ数の検証

・ 提案システムのランキング上位 5 人のブロガーが Google 検索のランキング上位 5 人の中に何人いるか検証

なお、Google 検索では、ブロガーのランキングではなく、ブログ記事のランキングであるため、異なるブロガー 5 人分のブログ記事の出現順を、ブロガーランキングとした。

レストラン系、ホテル系のカテゴリに対する前者の検証結果を表 4 に示す。表 4 において、どちらのカテゴリも 3 ページは必要であることがわかる。これは、Google 検索では、Web ページ全体で検索しているため、ブログパーツ (お気に入りのブログや、他のブロガーの RSS) に検索が引っかかり、このような結果になったと考えられる。本システムでは、1 ページでブロガーのランキングを閲覧することが可能である。

レストラン系、ホテル系のカテゴリに対する後者の検証結果を表 5 に示す。表 5 において、どちらのカテゴリも約 2 人ランクインしていることがわかる。これは、Google 検索では、実際に Venue を利用していないブログ記事、つまりロコミ記事でないブログ記事が検索に引っかかったため、このような結果になったと考えられる。本実験で確認したロコミ記事でないブログ記事の例として、ちらし寿司が食べられるリストといったお店を紹介するブログ記事、ホテルの売店に私たちが作った商品がありますといった自身の商品 PR するブログ記事などがあった。

5. 本システムの導入

本システムを地域密着型 SNS サービス [LINK] に導入する。LINK とは、「地域密着のプラットフォーム」に「オープンデータ」を掛けあわせたサービスである。具体的なサービスの内容は、はまぞう等のブログサイトのブログ記事の閲覧、GPS 情報を利用することで現在位置の周辺のブログの閲覧、Venue の公式ブログをマッピングした地図の閲覧、Venue の公式ブログの詳細情報の閲覧、公式ブログを開設している Venue のルートマップ・電話連絡・チェックイン等が可能である。LINK は PC、スマートフォン (Android^(注17) および iOS^(注18)) のアプリにて利用可能である。本システムと LINK とははまぞうのシステム全体像を図 3 に示す。

本システムの導入の際には、foursquare の Venue データでは

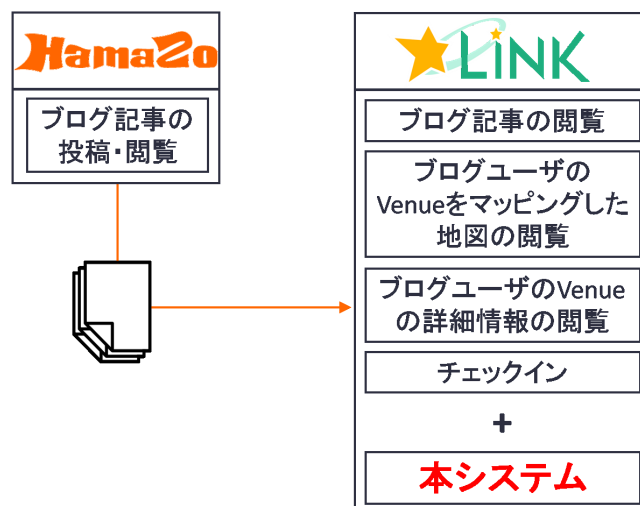


図 3 本システム・LINK・はまぞうのシステム全体像

なく、株式会社シーポイントが所持している Venue データを用いた。これは、電話帳 CSV 販売.jp^(注19)にて購入した電話帳データ (Venue 名、住所等)、住所を Google Geocoding したデータ、株式会社シーポイントが手作業で付与したカテゴリ、ブログ URL が紐付いているデータである (e.g. “うおたか”, “34.72556 - 137.69447”, “食べる”, “http://uotaka.hamazo.tv/”)。このデータのカテゴリは、“食べる”, “お買い物”, “スポーツ”, “ビューティー”, “美容院”, “病院”, “遊ぶ”に分類されている。各カテゴリの 6 項目を評価項目およびその類義語を表 6(a) から表 6(f) に示す。評価項目を決める際に用いたロコミサイトは、“お買い物”はみんなのレビュー・ロコミ^(注20)、“スポーツ”はオリコン CS ランキング^(注21)および、ロコミ&ランキング^(注22)、“ビューティー”はオリコン CS ランキング、“美容院”はオリコン CS ランキングおよび、ホットペッパービューティー^(注23)、“病院”は病院の通信簿^(注24)、“遊ぶ”はフォートラベル^(注25)を用いた。なお、“食べる”は表 2(a) と同じ評価項目および、その類義語を用いた。

図 4 に PC 版の LINK に本システムを導入した結果を示す。図 4 の右側では、Venue の公式ブログを OpenStreetMap に表示している。マーカーの色は、Venue のカテゴリによって変えている。図 4 の左側では、地図の中心座標の地域に詳しいブロガーおよび、ランクインしたブロガーのその地域に関するスコアの高いブログ記事のリンクと写真・投稿日時・タイトルを示している。

本システムの LINK への導入により、選択した地域の Venue 利用したブロガーの Venue に関する情報が得られる。また、今までの検索システムでは不可能であった、指定した地域に対す

(注19) : <http://teldata.jp/>

(注20) : <http://review.rakuten.co.jp/>

(注21) : <http://life.oricon.co.jp/>

(注22) : <http://www.kuchiran.jp/>

(注23) : <http://beauty.hotpepper.jp/>

(注24) : <http://www.tusinbo.com/>

(注25) : <http://4travel.jp/>

(注17) :

<https://play.google.com/store/apps/details?id=com.cpoint.link&hl=ja>

(注18) : <https://itunes.apple.com/jp/app/link/id926203531?mt=8>

表 6 評価項目および類義語

評価項目	類義語の一部
品揃え	—
コストパフォーマンス (コスパ), 料金	値段, 価格, 定価
決済, 支払い	—
応対	対応, 待遇, 接客
梱包	包装, 箱詰め
配送	配達, デリバリー

(a) カテゴリ (“お買い物”)

評価項目	類義語の一部
手続き, プラン	手順, やり方, 方法
コストパフォーマンス (コスパ), 料金	値段, 価格, 価値
設備	備え付け, 施設, 設置
清潔	綺麗, 奇麗, キレイ
応対	対応, 待遇, 接客
立地, アクセス	場所, 敷地, 用地

(b) カテゴリ (“スポーツ”)

評価項目	類義語の一部
予約	ブッキング, リザーブ
コストパフォーマンス (コスパ), 料金	値段, 価格, 価値
カウンセリング, 技術, 応対	スキル, 腕前, 技量
清潔	綺麗, 奇麗, キレイ
アフターケア	—
立地, アクセス	場所, 敷地, 用地

(c) カテゴリ (“ビューティー”)

評価項目	類義語の一部
雰囲気	ムード, 佇まい, 風情
コストパフォーマンス (コスパ), 料金	値段, 価格, 価値
接客, サービス	対応, 待遇, 接客
技術, 仕上がり	スキル, 腕前, 技量
清潔	綺麗, 奇麗, キレイ
メニュー	—

(d) カテゴリ (“美容院”)

評価項目	類義語の一部
待ち時間	—
診断, 技術, 治療, 説明	診療, スキル, 見立て
予防, 指導, ケア, 応対	助言, 配慮, 手当て
設備	備え付け, 施設, 設置
料金	値段, 価格, 価値
プライバシー	内密, 内々, 非公開

(e) カテゴリ (“病院”)

評価項目	類義語の一部
見応え, 景観	景色, 見晴らし, 眺望
コストパフォーマンス (コスパ), 料金	値段, 価格, 価値
アトラクション	売り物, 名物, 見映え
人混み	人込み
バリアフリー	—
立地, アクセス	場所, 敷地, 用地

(f) カテゴリ (“遊ぶ”)

る詳しいブロガーのランキングが、その地域内の Venue に関するブログ記事とともに、閲覧可能になったことがわかる。今後の課題として、今回、詳しいブロガーとともに、ブロガーのスコアの高いブログ記事のリンクと写真・投稿日時・タイトルを表示しているが、良い口コミ記事なのか、悪い口コミ記事なのかかわからないため、どのようなブログ記事なのか見分けることができるような、インターフェースが必要であると考えられる。

6. おわりに

本研究では、ブログ記事内で言及している場所 (Venue) を発見した。また、ブログ記事中の口コミや画像に基づいて算出した各 Venue に対する詳しさのスコアを用いて、Venue に対して詳しく言及しているブロガーを地域 (町・字単位) ごとにランキング形式で Venue に関するブログ記事とともに提示するシステムを実装した。さらに、情報の鮮度を考慮する (新しいブログ記事ほどランキングに影響を持たせる) ため、時間とともに情報価値が減っていくことを忘却の概念を利用してモデル化し、ランキングの作成の際に利用した。加えて、実際に本システムを、地域密着型 SNS サービス [LINK] に導入した。

ランキングの変動比較により、忘却の概念を利用することで、ブログ記事を最近投稿したブロガーのランクを高く評価する傾向があることがわかった。また、式 (1) の λ の値を変えることで、ランキングに変動があることがわかった。さらに、今までの検索システムでは不可能であった、指定した地域に対する詳しいブロガーのランキングが、その地域内の Venue に関するブログ記事とともに、閲覧可能になったことがわかった。

今後の課題として、3 つ挙げられる。1 つ目は、本システムの効果を検証することが挙げられる。これは、その地域に関す

るブログ記事が増加しているかや、ランクインしたブロガーのブログ記事のアクセス数が増加しているか等を検証することで確かめられると考えられる。2 つ目は、この地域の飲食店に詳しいブロガーといったように、Venue のカテゴリごとに詳しいブロガーを表示する機能の実装をすることが挙げられる。3 つ目は、この地域では、この Venue に対するブログ記事が多いといったことを、表示するシステムを実装することである。ただ、この課題は、一部の Venue を推薦するようなシステムとなってしまうため、実ビジネス上現実的とは言えない可能性がある。

文 献

- [1] 三浦 惇貴, 廣田 雅春, 野澤 浩樹, 横山 昌平, “忘却の概念を利用した Local Expert ランキングの地域ブログサイトへの導入”, 第 7 回データ工学と情報マネジメントに関するフォーラム, 2015
- [2] ジェイン・マクゴニガル, “『幸せな未来は「ゲーム」が創る』”, 早川書房, 2011
- [3] 河野 愛樹, 中村 健二, 小柳 滋, “マイクロブログから抽出した地物情報と投稿間隔を考慮した位置情報推定”, 情報処理学会第 73 回全国大会, 2011
- [4] 倉島 健, 手塚 太郎, 田中 克己, “Blog からの街の話題抽出手法の提案”, 電子情報通信学会第 16 回データ工学ワークショップ, 2005
- [5] 中嶋 勇人, 太田 学, “旅行ブログ記事からの名所とその付随情報の抽出”, 第 5 回データ工学と情報マネジメントに関するフォーラム, 2013
- [6] 平山 拓央, 湯本 高行, 新居 学, 高橋 豊, “属性評価モデルに基づく商品評価の抽出と提示”, 第 3 回データ工学と情報マネジメントに関するフォーラム, 2011
- [7] 小西 卓哉, 木村 文則, 前田 亮, “周辺文を考慮するトピックモデルを用いた評価側面の推定”, 第 4 回データ工学と情報マネジメントに関するフォーラム, 2012
- [8] Jie Bao, Yu Zheng, Mohamed F. Mokbel, “Location-based and Preference-Aware Recommendation Using Sparse Geo-

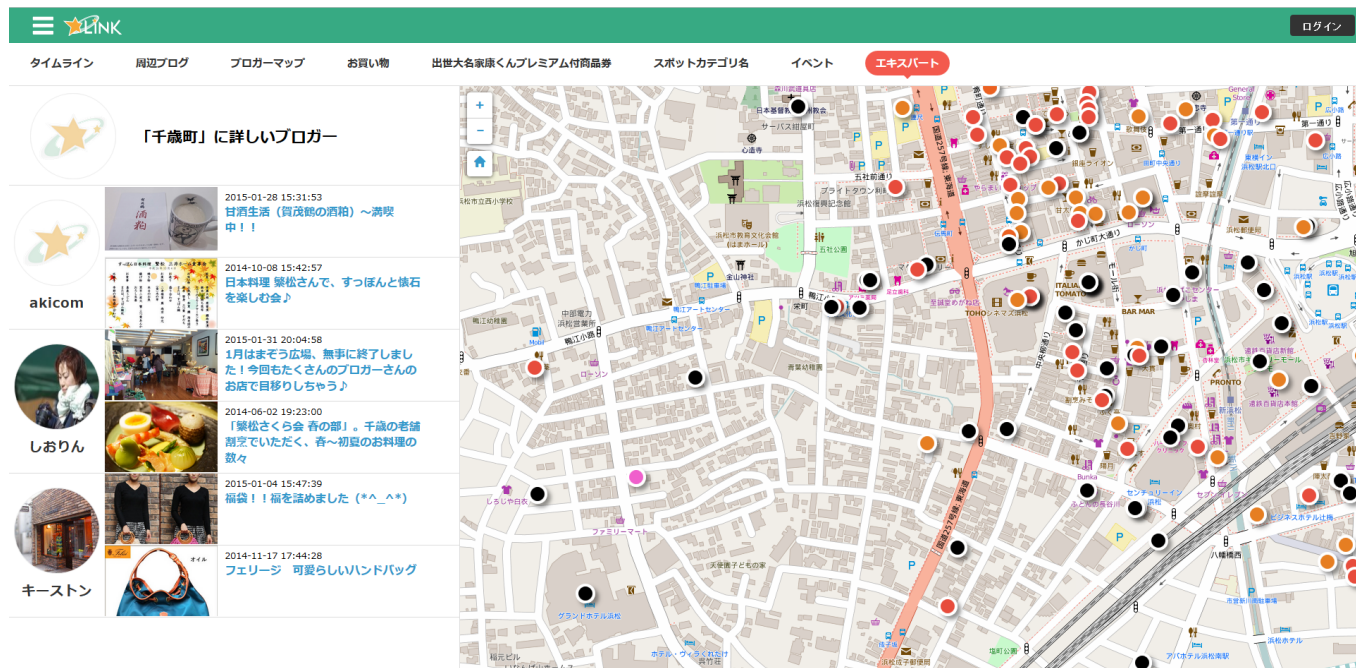


図 4 LINK への導入 (PC 版)

Social Networking Data”, ACM SIGSPATIAL GIS 2012, 2012

- [9] 稲垣 陽一, 中島 伸介, 張 建偉, 中本 レン, 桑原 雄, “プログラマー体験熟知度に基づくプログラミングシステムの開発および評価”, 情報処理学会論文誌: データベース, Vol.3, No.3, pp.123-134, 2010
- [10] Sophoin Khy, Yoshiharu Ishikawa, Hiroyuki Kitagawa, “A Novelty-based Clustering Method for On-line Documents”, World Wide Web Journal, Vol.11, No.1, pp.1-37, 2008
- [11] M. L. Kherfi, D. Ziou, A. Bernardi, “Image Retrieval from the World Wide Web: Issues, Techniques, and Systems”, ACM Computing Surveys, Vol.36, No.1, pp.35-67, 2004