

観光スポット推薦のための旅行ブロガーの信頼性推定

近藤 昭宏[†] 阿倍 博信[†]

[†]東京電機大学 〒120-8551 東京都足立区千住旭町 5 番

E-mail: [†]21amj11@ms.dendai.ac.jp, [†]hironobu.abe@mail.dendai.ac.jp

あらまし ブログや SNS などの UGC(User Generated Contents)の一般化に伴い、旅行に関するユーザ発信情報も増えており、これらを活用した推薦システムの研究も盛んに行われている。しかし、UGC はユーザによる個人差も大きく、信頼性の高いユーザを見つけることが課題となっていた。そこで本論文では、観光スポット推薦システム構築を目的として、旅行ブロガーの信頼性推定に必要な複数の指標を定義し、旅行ブログのメタデータや内容の解析結果を用いて、旅行ブロガーの信頼性を推定する方式を提案する。実際の旅行ブログに対して提案方式を適用しユーザ評価を行った結果、その有効性について確認することができた。

キーワード Web 情報システム, 信頼性, 情報推薦

1. はじめに

近年、SNS やクチコミ、ブログに自身の旅行記録を投稿するといったことが増えてきた。投稿先となる SNS の代表的なものとしては、「Facebook」、「Twitter」、「Instagram」、旅行に関するクチコミの投稿サイトの代表的なものとしては、「TripAdvisor」などが挙げられる。このようなサービスの利用者が投稿する情報は、UGC(User Generated Contents)と呼ばれており、旅行の計画時点での参考に利用されたり、旅行先での周辺観光スポットの調査に利用されたりしている。しかし、UGC が増え、参考にできるような情報が増えることによって情報過多[1]と呼ばれる問題が起きている。これは、情報が多すぎて欲しい情報にたどりつかなかったりする問題であり、この問題に対応するために UGC を活用した観光地推薦の研究が行われている[2][3]。しかし、UGC は投稿することによってインセンティブがあるようなサイトが存在することから、ユーザによって信頼性、質などの個人差があることが考えられ、上記のような UGC を活用した研究においても信頼性の高いユーザの信頼性の高い情報を優先して使用した方が利用者により良い推薦ができることが考えられる。

そこで本研究では、UGC の信頼性を考慮した推薦システムを構築することによりユーザにとってより良い観光地を推薦するために、信頼性の高いブログを投稿するブロガーの信頼性推定方式を提案する。

先行研究として、ブログのメタデータの内、1 つの指標のみを用いた信頼性推定[4]やブロガーの信頼性をブログから得られるメタデータやブログの内容を利用して複数の指標を定義し、指標をブロガーごとに集計することによって旅行ブログを執筆するブロガーの信頼性を示す研究を行った[5][6]。先行研究では、定義したブロガーの信頼性が人間の信頼性判断と似ているかを検証するためにユーザ実験による評価も行った。結果として、指標の一つである文字数、固有表現の指

標において似たような文章を何度も記載しているブロガーの指標が高くなってしまいう問題点、使用した固有表現のタグが信頼性を表せていない問題点が考えられた。そこで、本研究ではこの問題点に対応を行い、再度評価するために実験 1 を実施した。さらに実験 1 で考えられた問題点を基に実験方法の再検討を行い、実験 2 を実施した。

2. 関連研究

関連研究として信頼性、信頼性に関連した研究を紹介する。信頼性の指標に関する研究としては Yamamoto らの研究[7]がある。Yamamoto らはウェブ検索結果の信頼性判断を支援するために、複数の指標を定義し、信頼性の判断支援を行うシステムの作成を行っている。作成したシステムの有効性を判断するためにユーザ実験を行った結果、提案手法として作成したシステムは信頼性判断を効率よくするという点において有用であると示している。

同じく信頼性関連する研究として、松本らはニュース記事の信頼性判断を支援するためにセンチメント分析を活用している[8]。ニュース記事は書く人にとって利益のあるように情報発信をする場合もあることから複数の立場の記事を比較することによりユーザの信頼性を判断する。そこで、センチメント分析と呼ばれる、感情分析を行うことによってニュース記事を書いた人の立場や記事を載せているサイトの立場を客観的に把握できるようにすることで信頼性の判断支援を行っている。また、ユーザ実験を行うことにより、それぞれの記事の立場を把握できるようになり、信頼性判断を支援できたと示している。

信頼性の判断する要素に関する研究として Filieri[9]の研究がある。Filieri は観光に関連するレビューの信頼性に関して、消費者がどのように信頼性を判断しているのかを調べるためにアンケート調査を行った。調

査の結果、レビューの信頼性判断には写真といった視覚的な情報が必要であり、レビューを掲載しているサイトは写真といった視覚的情報を増やす必要があると示している。また、消費者はレビューを利用すれば利用するほど、レビューの内容を見極める能力や知識が身につくということを示している。

3. 提案手法

3.1. 方針

はじめに、先行研究[4][5]においてはブログ、ブロガー両方に対して信憑性という言葉を使用していたが、本研究ではブログという情報に関しては信憑性を、ブロガーという人に対しては信頼性を使用する。ブロガーの信頼性スコアリングに使用するために旅行ブログを収集し、集めたブログからメタデータおよびブログの内容を抽出し、スコアリングの指標とした。

スコアリングの指標は、ユーザ実験ごとに後述する5つの指標を提案し、5つの指標をそれぞれ最大値1、最低値0に正規化し、5つの指標の合計点を信頼性とした。本研究で作成したシステムはブログの表現やブログのメタデータからブロガーの信頼性をスコアリングするものである。有効性評価としては信憑性の高いブログを投稿するブロガーの信頼性スコアリングシステムとして評価実験を行った。

実験1では、信頼性を基にブロガーのランキングを作成し、ランキングを基にブロガーを4つのグループに分け、ブロガーのランキングが人間の信憑性判断に沿っているかどうか確かめるためのユーザ実験を行った。実験2ではランキングを基に3つのグループに分け、3つのグループから合計7件のブログを被験者に提示、提示されたブログの信憑性があると思う順番に被験者に並び替えてもらうユーザ実験を行った。信頼性の指標は実験1のユーザ実験の結果を基に、実験2のユーザ実験では信頼性の指標を一部変更、および指標の比重を変更して実験を行った。実験1、実験2における提案システム概要図を図1に示す。

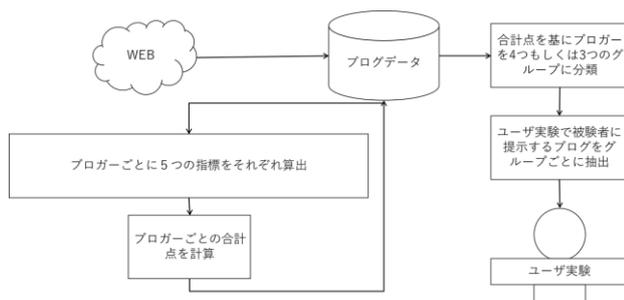


図1. 提案システム概要図

3.2. 使用するデータ

本研究で使用する旅行ブログとしては「フォートラベル」[10]から100人分のブロガーで計22,195ブログを使用した。

使用したデータは、ブログに含まれている、「いいね!」の数、写真の枚数、コメント数、アウトライン、本文などである。

3.3. 信頼性の定義

本研究ではブロガーの信頼性を定義する指標として実験1、実験2それぞれで5つを指標として定義した。

また、信頼性のスコアリングに用いられる指標の要素としては大きく分けて、ブログの内容(情報の真偽、フェイクニュースかどうか)、ブログの表現(文章の書き方、写真の枚数、文字数、固有表現の量)、ブログのメタデータ(社会的信頼、他のユーザからの評価、「いいね!」、コメント数)が考えられるが、本研究ではブログの表現とブログのメタデータのみを指標として扱い、ブログの内容は指標としては扱わない。

実験1、実験2での信頼性の指標を表1に示す。実験1と実験2では、それぞれ指標としてはほぼ同じものを使用したが、実験1では先行研究[5][6]を踏まえて文字数と固有表現の割合の指標に関して集計方法の変更を行った。

表1. 実験1, 2における指標一覧

実験1, 2で使用した指標	
1	文字数
2	写真の枚数
3	「いいね!」の数
4	固有表現の割合
5	コメント数

実験2では実験1の結果を踏まえて、「いいね!」の数、コメント数の集計方法について変更を行った。それぞれの指標を採用した理由として、文字数は多いと観光地に関する情報が詳細に記載されている場合が多いと考え指標とした。

写真の枚数は枚数の多寡によって、実際に観光地に行ったかどうかの判断根拠にできると考え指標とした。

固有表現の割合は観光ブログにおいて、具体的な移動手段、費用などを記載している方が、実際に観光したブログであると判断することができると考え指標とした。固有表現のタグは、関根の拡張固有表現階層のタグ[11]から旅行ブログの評価に関連すると考えられる一部のタグとした。

関根の拡張固有表現階層は単語の意味を人工的に分類したのを表すものであり、200種類の固有表現クラ

スが設定されている。先行研究[5][6]ではその中で旅行ブログの特徴を表せる表現として考えられる、地名、交通手段、金額、距離などを表す、96種類のタグを指標としてカウントしたが、本研究では97種類のタグを指標としてカウントした。カウントするタグと対応する単語の例を表2に示す。

表2. 拡張固有表現階層のタグと対応する単語の例

タグ名	対応する単語
Province	沖縄, 東京
Period Time	1時間コース
City	那覇
Dish	紅芋コロッケ

コメント数は旅行ブログの場合、ブログの内容が良かった場合に閲覧者からコメントが書き込まれる場合が多いため指標とした。

実験1では文字数、写真の枚数、「いいね!」の数の3つの指標に関してブロガーごとに集計し、1ブログ当たりの平均を算出した。固有表現に関してはブロガーごとに固有表現の数をカウントし、1ブログ当たりの平均を算出し、文字数に対する割合を指標とした。

コメント数に関して、実験1ではブロガーが投稿したブログに寄せられたコメントの累積を指標とした。

その上で、文字数の指標、固有表現の指標に関しては先行研究[5][6]より同じ表現を繰り返して文章量、固有表現の量を増やしていると考えられるブロガーの指標が高く出てしまっているということが考えられたため、重複している表現の分、点数を下げる調整を行った。この調整は指標にするための正規化をする前に行った。具体的な算出式を式1に示す。

同じ表現を繰り返しているかどうかの判別は固有表現を用いて判別した。文章から固有表現を抜き出し、登場した順番に並べる。並べた順番で10個前までに同じ固有表現が登場していた場合は重複表現を繰り返しているとし、その固有表現の文字数をカウントした。そのカウントをブロガー単位で累計したものが重複の条件に合致した文字数である。その後、ブロガーの投稿したブログ数で割ることにより正規化前の文字数の指標と条件を合わせた。

調整後文字数 =

$$\text{調整前文字数} - \frac{\text{重複の条件に合致した文字数}}{\text{ブロガーの投稿したブログ数}} \quad (1)$$

固有表現の指標に関しては文字数と同じように文章から固有表現を抜き出し、登場した順番に並べる。固有表現の数をカウントする際に並べた順番で10個前までに登場していた場合はその重複していた分の固有

表現の数を指標として加算しないという処理を行った。実験1でのそれぞれの指標の最大値、最小値、平均値、中央値を表3に示す。

加えて、実験1における合計点のヒストグラムを図2に示す。

表3. 実験1における指標の分析

	最大値	最小値	平均値	中央値
文字数	1	0	0.16	0.11
写真の枚数	1	0	0.30	0.28
「いいね!」の数	1	0	0.12	0.02
固有表現の割合	1	0	0.35	0.34
コメント数	1	0	0.04	0.003
合計点	2.33	0.02	0.96	0.92

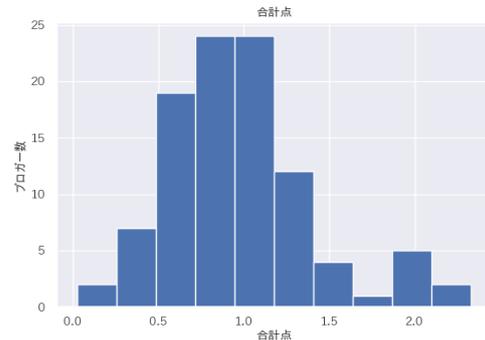


図2. 実験1における合計点のヒストグラム

実験2では、実験1での課題を基に指標の調整を行った。実験1では課題として「いいね!」の指標やコメント数の指標が高く、内容も良いが指標の合計点としては低くなってしまっているブロガーが存在していた。そこでブログの「いいね!」やコメント数の指標に関する調整を行った。

コメント数に関しては実験1のようにブロガーごとに累計で出すのではなく、1ブログ当たりの平均を算出するように変更した。また、「いいね!」とコメント数の両方の指標を1.5倍することにより合計点に対する比重を高めた。修正後の「いいね!」、コメント数、合計点の指標の最大値、最小値、平均値、中央値を表4に示す。

加えて、実験2における合計点のヒストグラムを図3に示す。

表4. 実験2における指標の分析

	最大値	最小値	平均値	中央値
「いいね!」の数	1.5	0	0.18	0.02
コメント数	1.5	0	0.18	0.02
合計点	3.90	0.03	1.17	1.07

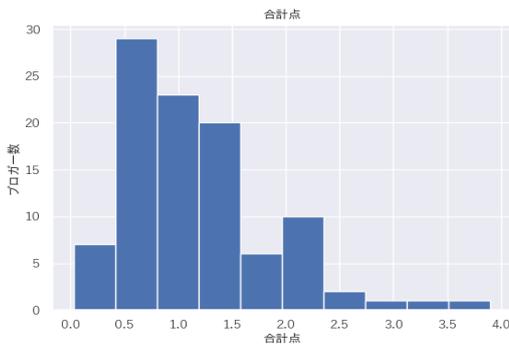


図 3. 実験 2 における合計点のヒストグラム

4. 実験

4.1. 実験環境

提案システムの有効性を評価するためにユーザ実験を行った。被験者として実験 1 では大学生 4 人を対象に、実験 2 では大学生、大学院生、社会人の 10 人を対象にした。実験方法として、実験 1 では被験者に提案システムによるブログ提示を行い、提示されたブログの信憑性が高いかどうかの質問をブログごとに行い、ブロガーを 4 つに分けた際のグループごとの結果を比較すること、被験者評価による点数と提案手法によりつけた合計点で算出した相関係数の 2 つの方法で有効性を評価した。

「極めて信憑性が高い」グループに分類されたブログはユーザからも高い信憑性の評価を受け、逆に「極めて信憑性が低い」グループに分類されたブログは被験者からの評価も低い評価を受けるだろうという仮定のもと行った。

実験 2 では被験者に複数のブログを同時に提示し、信憑性が高いと思う順番にブログを並び替えてもらう実験を行った。ブログの順位に応じた点数をつけ、被験者評価による点数と提案手法によりつけた点数との相関係数を求めることにより有効性を評価した。

4.2. 実験環境

実験 1 では作成したシステムを用いて、「極めて信憑性が高い」グループ、「信憑性が高い」グループ、「信憑性が低い」グループ、「極めて信憑性が低い」グループの 4 つのグループに分類した。グループの分類はシステムの合計点の上から 25% ずつを「極めて信憑性が高い」グループ、一番合計点の低い 25% を「極めて信憑性が低い」グループという方法で分類した。分類後、4 つのグループからそれぞれ 5 ブログずつ、合計 20 個のブログを抜き出し、抜き出した旅行ブログからブログの URL を抽出し、被験者に提示をした。実験 1 におけるグループ分けのイメージ図を図 4 に示す。

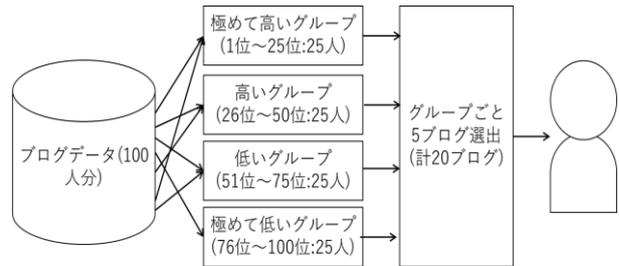


図 4. 実験 1 におけるグループ分けの図

実験 2 では提案手法により点数付けられたブログの上から 34 人を「点数が高い」グループ、真ん中 33 人を「点数が中間」のグループ、下から 33 人を「点数が低い」グループとし、「点数が高い」グループから 3 ブロガー分の 3 つのブログ、「点数が中間」グループと「点数が低い」グループからはそれぞれ 2 ブロガー分の 2 つのブログ、合計で 7 ブログを抜き出し、被験者に提示した。実験 2 におけるグループ分けを図 5 に示す。

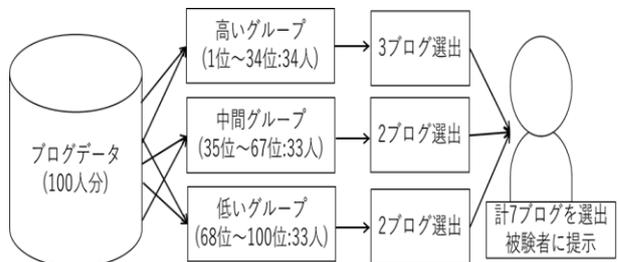


図 5. 実験 2 におけるグループ分けの図

被験者の評価方法として、実験 1 ではそれぞれのブログを閲覧し、読んだブログの信憑性が高いと思うかどうかの 1 つの質問に主観で回答してもらった。質問は「極めて高い」、「高い」、「どちらともいえない」、「低い」、「極めて低い」の 5 段階のリッカート尺度によって評価してもらった。信憑性に関する評価基準は「自分がこのブログを読んだ際に、このブログに書いてある情報を信頼しても良いと感じるかどうかの判断をこれまでの経験に基づいてお願いします。」という説明をして、評価を行ってもらった。

実験 2 では本研究で使用している信頼性の指標としてどのようなものを使用しているのかという説明を行い、7 つのブログを 1 位から 7 位まで並び替える方法で評価をしてもらった。また、実験 1, 2 ともに質問に対する回答欄にはブログの信憑性の判断理由に関して自由記述をする欄を設け、実験 1 では該当するブログに対しての評価を選んだ理由。実験 2 に関しては概要するブログを順位につけた理由に関しての記載をしてもらった。

実験結果を集計する際は実験 1 では信憑性が「極めて

て高い」と選択した場合に5点、信憑性が「高い」と選択した場合に4点、「どちらともいえない」と選択した場合に3点、信憑性が「低い」と選択した場合に2点、信憑性が「極めて低い」と選択した場合に1点として集計を行った。グループごとに5ブログずつ、実験1において被験者が4人のため、グループごとに合計20回ずつ評価を行いグループごとの平均点、加えて被験者評価と提案手法のスコアリングの結果のスピアマンの順位相関係数で有効性の評価を行った。

実験2では1位に選んだブログに7点、2位に選んだブログに6点、3位に選んだブログに5点、4位に選んだブログに4点、5位に選んだブログに3点、6位に選んだブログに2点、7位に選んだブログに1点を与えた。実験2において被験者が10人で同じ実験を1人当たり3回ずつ行ってもらったため、合計210件の回答を基に被験者評価と提案手法によるスコアリングの結果でスピアマンの順位相関係数の算出を行い有効性の評価を行った。実験2における評価方法に関する図を図6に示す。

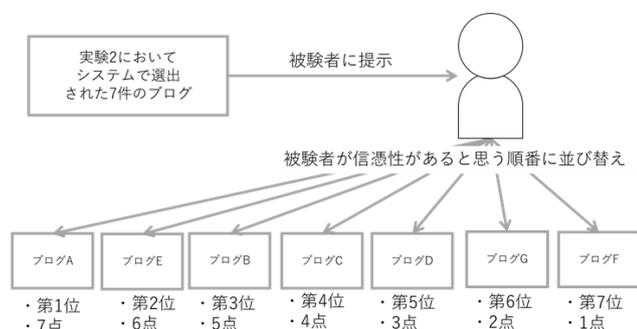


図6. 実験2の評価方法に関する図

4.3. 実験環境

4.3.1. 実験1の結果

実験1の結果としてそれぞれのグループの被験者評価の平均点を表5に示す。

表5. 実験1のグループごとの結果

グループ名	グループの平均点
極めて高い	4.3
高い	4.15
低い	4.15
極めて低い	3.2

表5より極めて高いグループの平均点が一番高いという結果になった。

加えて、実験1において被験者の評価と合計点に関する関係があるのかを調べるために相関係数の算出を行い、0.37という結果を得た。

また、実験結果の例として1人分のブロガーの点数

と収集したブロガー内の順位を表6に示す。

表6. 実験1における点数と順位の例

指標	点数	順位(100人中)
文字数	0.158	34位
写真の枚数	0.280	49位
「いいね!」の数	0.268	15位
固有表現の割合	0.178	96位
コメント数	0.021	27位
合計点	0.906	52位

表6のブロガーに対する被験者の評価としては5点で評価理由としては「具体的な説明が多かったから」という評価理由だった。

実験1によって得られた被験者評価に対してそれぞれの指標がどれほど影響しているかを調べるために重回帰分析を行った。重回帰分析では目的変数を実験1によって得られた被験者評価、説明変数をそれぞれのブロガーの文字数、写真の枚数、「いいね!」の数、固有表現、コメント数で偏回帰係数を求めた。重回帰分析の結果を表7に示す。

表7. 実験1における重回帰分析の結果

指標	偏回帰係数
文字数	0.342
写真の枚数	0.231
「いいね!」の数	0.100
固有表現の割合	-0.026
コメント数	0.012

また、実験1の評価方法で選択肢に対する回答に偏りがないかを調べるために、被験者評価の度数分布を調べた。被験者評価の度数分布を表8に示す。

表8. 実験1における被験者評価の度数分布表

評価尺度	度数	相対度数 (%)
極めて高い	31	38.75
高い	25	31.25
どちらともいえない	14	17.5
低い	9	11.25
とても低い	1	1.25
合計	80	100

度数分布の表より全体の70%が「極めて高い」、「高い」と評価を行っており、逆に「どちらともいえない」、「低い」、「極めて低い」を合わせても30%という結果になった。

4.3.2. 実験2の結果

実験2においても被験者の評価と合計点の関係を調べるために相関係数の算出を行い結果として0.39を

得た。

加えて、実験 1 と同様に重回帰分析を行った。目的変数は被験者の評価、説明変数をそれぞれの指標で算出を行った。重回帰分析の結果を表 9 に示す。

また、実験結果の例として 1 人分のブロガーの点数と収集したブロガー内の順位を表 10 に示す。

表 9. 実験 2 に対する重回帰分析の結果

指標	偏回帰係数
文字数	0.156
写真の枚数	0.193
「いいね!」の数	0.258
固有表現の割合	0.005
コメント数	-0.093

表 10. 実験 2 における点数と順位の例

指標	点数	順位(100 人中)
文字数	0.167	29 位
写真の枚数	0.332	38 位
「いいね!」の数	0.016	58 位
固有表現の割合	0.453	21 位
コメント数	0.003	75 位
合計点	0.972	54 位

このブロガーは 3 人の被験者から評価されていた。評価対象のブログは 3 人それぞれ異なる。評価の結果として平均 6 点をつけていた。また、それぞれの評価コメントとしては、「誤字はあったが、時系列がしっかりしており、なにより写真がきれい。被写体に自分が写っていたのも加点」、「情報量が多く、実際に行かなければ分からないことが多かったから。」、「写真が多い、行動に基づいた記述、いいねの数」といった評価理由がつけられていた。

5. 考察

5.1. 実験 1 に対する考察

実験 1 の結果として被験者評価とシステムによる合計点で 0.37 という相関係数を得られた。

表 3 の平均値、中央値および重回帰分析の結果からコメント数に関しては合計点に関してほとんど影響を与えないということが考えられた。このことから、コメント数の集計方法の見直しをすることが考えられ、中央値からコメント数が合計点に与えることが極めて小さいことからコメント数の合計点に対する割合を変更する必要も考えられる。また、被験者の評価の中で合計点の順位が 100 人中 52 位で「いいね!」の順位が 100 人中 15 位のユーザがいた。このユーザは信憑性が低いグループに割り当てられたのだが、被験者の評価としては極めて信憑性が高いという評価を得ていた。このことから、「いいね!」の合計点に対する割合を上

げるべきだと考えられる。加えて、度数分布表の結果から全体の 70%が「極めて高い」、「高い」という評価を得たことから評価方法の再検討が必要だと考えられる。

5.2. 実験 2 に対する考察

実験 2 の結果として被験者評価とシステムによる合計点で 0.39 という相関係数を得られた。

表 9 の重回帰分析の結果より固有表現とコメント数に関しては有効性がないという事が考えられた。固有表現に関してはユーザ実験においての評価コメントより、料金、時間に関する用語があると信憑性が高く感じるというコメントがあった。これに関して現状は固有表現を具体的な観光地名、建物の名前などもカウントに含めているが、時間、お金の表現に絞ることにより変更することが必要であり、また固有表現のタグの中でどのタグが信憑性の判断に有効かどうかのアンケート調査を実施することによって選んだタグに有効性があるか調査をする必要があると考えられる。

また、コメント数に関しては実験 1, 2 両方の重回帰分析の結果から有効性はないという事が考えられるため、今後は別の指標を考えていくことが必要である。

6. おわりに

本研究では UGC の信憑性を考慮した推薦システム作成に向けて、ブロガーの信頼性スコアリングシステムの作成を行い、ユーザ実験による評価を行った。

実験結果としてスコアリングシステムによる点数と被験者の評価の間で実験 1 では 0.37、実験 2 では 0.39 ほどの相関を得ることができたが、一部の指標に関してはスコアリングには有効ではないということが考えられた。

今後は今回のユーザ実験の結果を基に固有表現の指標の調整、コメント数の代わりとなる指標の検討を行う。

参考文献

- [1] 神鳥敏弘: 推薦システムのアルゴリズム (1), 人工知能学会誌, Vol22, No.6, pp.826-837, 2007.
- [2] 市村哲, 上石萌恵, 堀口莉里花: ロコミ解析と好み診断に基づいた旅行先推薦, 情報処理学会第 110 回グループウェアとネットワークサービス研究会, Vol.2020-GN-110, No.11, pp.1-8, 2020.
- [3] 上原尚, 嶋田和孝, 遠藤勉: Web 上に混在する観光情報を活用した観光地推薦システム, 電子情報通信学会技術研究報告, 言語理解とコミュニケーション研究会, 第 4 回集合知シンポジウム, NLC2012-35, pp.13-18, 2012.
- [4] 近藤昭宏, 阿倍博信: 観光スポット推薦システムのためのブログの信憑性のスコアリング方式, 情報処理学会第 84 回全国大会, 7N-02, 2022.
- [5] 近藤昭宏, 阿倍博信: 観光スポット推薦システムに向けたブロガーの信憑性のスコアリング方式,

情報処理学会第148回情報基礎とアクセス技術研究会, Vol.2022-IFAT-148, No.5, pp.1-6, 2022.

- [6] Akihiro Kondo, Hironobu Abe: Travel blogger credibility scoring method for recommendation system, IEEE Explore, 2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM), pp.1-4, 2023.
- [7] Y. Yamamoto and K. Tanaka: Enhancing credibility judgment of web search results, Proc. of the ACM CHI2011, pp.1235-1244, 2011.
- [8] 松本好史, 張建偉, 河合由起子, 中島伸介, 熊本忠彦, 田中克己: センチメント分析に基づくニュース記事の信憑性判断支援, 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, DEIM Forum 2011, F2-1, pp.1-6, 2011.
- [9] R. Filieri: What makes an online consumer review trustworthy?, Annals of Tourism Research, Volume 58, pp.46-64, 2016.
- [10] フォートラベル: 旅行のクチコミとホテル・ツアー・航空券の料金比較【フォートラベル】, <https://4travel.jp/>, (参照 2023-01-24).
- [11] Satoshi Sekine: Sekine's Extended Named Entity Hierarchy, <https://nlp.cs.nyu.edu/ene/>, (参照 2023-01-24).