

地理的特性と Web 特性に基づくトピックの曖昧度分析

北山 大輔[†] 李 龍^{††} 角谷 和俊^{††}

[†] 兵庫県立大学大学院環境人間学研究科 〒670-0092 姫路市新在家本町 1-1-12

^{††} 兵庫県立大学環境人間学部 〒670-0092 姫路市新在家本町 1-1-12

E-mail: [†]ne07p001@stshse.u-hyogo.ac.jp, ^{††}{leeryong,sumiya}@shse.u-hyogo.ac.jp

あらまし 近年、旅行などの実空間に関するイベントを計画する際に、Blog などの Web コンテンツ中の画像を参考に決定することが多くなってきている。しかし、Web コンテンツには、真偽の分からない噂話（流言）が含まれていることがあり、そのような場合、現地に詳しくないユーザは、Web コンテンツの内容に関しての信憑性を判断することが困難である。そのため、ある画像に関するトピックに対し流言の流布量モデルに基づいて、地理的な領域特性によるトピックの流布量分析、および Web コンテンツにおけるトピックの重要度分析を行い、画像に関するトピックの曖昧度を提示することで、ユーザの信憑性判断の支援を行う手法を提案する。

キーワード 情報信憑性, 情報検索, 流言分析, 地理的特性

A Method of Analyzing Topic Ambiguity based on Geographical Distributions and Web Search Results

Daisuke KITAYAMA[†], Ryong LEE^{††}, and Kazutoshi SUMIYA^{††}

[†] Graduate School of Human Science and Environment, University of Hyogo
1-1-12, Shinzaike-honcho, Himeji 670-0092, JAPAN

^{††} School of Human Science and Environment, University of Hyogo
1-1-12, Shinzaike-honcho, Himeji 670-0092, JAPAN

E-mail: [†]ne07p001@stshse.u-hyogo.ac.jp, ^{††}{leeryong,sumiya}@shse.u-hyogo.ac.jp

Abstract Recently, people consult Web contents such as Web pages and Blogs when they plans events or actions in the real world. However, Web contents include topic that may or may not be true. At this time, the user difficult decides credibility of that topic if the user is not familiar with field of that topic. Therefore, we proposed that a method of analyzing possibility as rumor using the the geographical region and the Web search results about the topic. By this method, the user can easy decide credibility of topic.

Key words Information Credibility, Information Retrieval, Analysis of Rumor , Geographical Region

1. はじめに

近年、旅行など実空間における行動を決定する際に、Web ページや Blog などの Web コンテンツを参考に決定することが多くなってきている。しかし、Web コンテンツ中には、正しいことばかりが記載されているとは限らず、思いこみや勘違いなどから、誤った情報や根拠のない情報が記載されていることも多い。ユーザがその情報に関して適切な知識を持ち合わせている場合、その情報の信憑性を判断することは可能であると考えらるが、多くの場合ユーザは、信憑性判断に適切な知識を持ち合わせていない。そのため、ユーザは本来の目的から離れ、情報の根拠を発見するために別の Web 検索を行う必要が出てくる。もしくは、記載されている情報を鵜呑みにして実空間での

行動を起こしてしまい、現地でようやく間違った情報であったことに気が付くという事態になる。このような根拠のない情報に基づいて実空間での行動を起こしてしまう例は多い。例えば、2004 年に「大阪の USJ でクリスマスにテロがおこる」というような噂が流れ、実際に警察に問い合わせが殺到するという事態になった。従来、このような噂は人々の口コミやマスメディアを媒体として広がっていたため、一時的に話題になるだけで、持続しないという傾向にあった。しかし、Web 上で広がる噂では、コンテンツの持続時間が長く、根拠が明らかになった後であっても、噂のように信憑性の低い情報がそのまま記載されてしまうため、知識を持たないユーザは、どのコンテンツが正しい情報であり、どのコンテンツが信憑性の低い噂であるかという判断が困難であると考えられる。

また、画像を扱った Web コンテンツの場合、画像とキャプションを用いることで、トピックの根拠を補強している場合が多い。例えば、京都の大文字山の送り火が「犬」という文字に変えられたことがあると述べている Web ページで、実際に犬になっているように見える画像が記載されている場合、その写真・記載されているキャプションの真偽にかかわらず、実際にあった出来事のように錯覚してしまうことがある。また、この時「犬」になっている写真ではなく、通常の「大」の文字が映っている写真であったとしても、現地に行っている人物が書いた Web ページであると錯覚させることは可能であり、やはりキャプションの真偽にかかわらず、ユーザに誤った情報を伝えやすくなる。このように、根拠の乏しい噂話のようなトピックであっても、画像などマルチメディアの情報を用いることで、その Web ページを閲覧しているユーザに対して、情報の根拠を錯覚させることが可能であると考えられる。

噂は、ゴシップ・デマ・流言などに分類され、本稿ではこれらのうち流言に関して扱う。流言とは、「正確な知識や情報を得られず、明確な根拠も無いままに広まる噂」と定義されている。我々は、Web 上における根拠のない情報、すなわち流言の判定を行うことで、ユーザの情報に対する信憑性判断の支援になると考えた。流言に関する研究は、古くから行われており、流言の流布量モデルが Allport ら [1] によって提案されている。本稿では、判定対象の画像に関するトピックを Web 特性と地理的特性により分析し、この流言の流布量モデルに合致するかどうかを判定することで、情報の根拠として用いられている画像とそのキャプションの曖昧度を算出する手法を提案する。このことにより、ユーザは現在閲覧中の Web コンテンツに記載された画像情報が、根拠がなく、曖昧状態で発信されたトピックに付随する画像であるかどうかを判断することが可能となる。

以下、2 節では本研究の概要と関連研究について述べる。3 節では、Web 検索結果からのトピックと地理的特性の抽出について述べ、4 節では、地理的特性と Web 特性を用いた曖昧度算出について述べる。5 節で、プロトタイプシステムについて述べる。最後に、6 節でまとめを述べる。

2. 本研究の概要と関連研究

2.1 本研究の概要

ユーザが Web コンテンツを閲覧する際、コンテンツで述べられているトピックに興味を持つことがある。例えば、旅行の行き先を決めるために、Web コンテンツを閲覧している場合、現地の写真とともに「旅行先であるイベントが開催されるらしい」という記述を発見し、どのようなイベントであるのかに興味を持つ場合を想定する。この時、その記述が信憑性の低い情報である場合が考えられる。しかし、Web ページを閲覧しているユーザが、情報の信憑性を判断することが困難である場合、旅行の行き先に関する情報収集を中断して、情報の根拠に関する情報収集を始める必要が出てきたり、そのまま鵜呑みにして、現地でようやく間違った情報であったことに気が付くという事態になると考えられる。

そこで本稿では、噂話のうち根拠が不明確な流言に注目し、

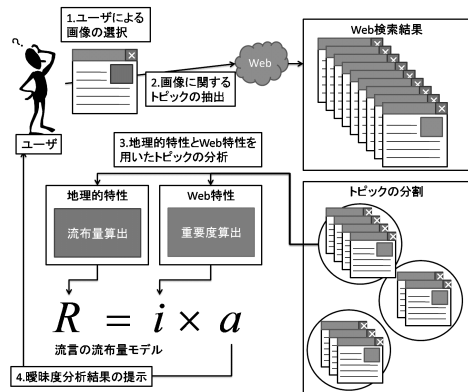


図 1 本手法の概念図

Fig. 1 Concept of Ambiguity Analysis

ユーザが興味をもったトピックが流言であるかという判定を行うことで、信憑性判断の支援を行う手法の提案を行う。流言の流布量モデルとは、流言の流布量 (R : Rumor) は、トピックの重要さ (i : importance) と曖昧さ (a : ambiguity) に比例するというモデルで、Allport [1], [4] によって以下の式のように定義されている。

$$R = i \times a \quad (1)$$

すなわち、トピックがその事柄を伝える構成員にとって重要であり、かつ情報が少なく根拠が得られないという曖昧さをもっていると、流言として伝達されやすいというモデルである。またこの式は、重要さもしくは曖昧さの値が 0 であれば、そのトピックは流言とならないことを示している。「大文字山で送り火の字が犬になった」の例であれば、大文字山に行く人にとって重要なトピックであり、そのトピックの根拠が不明であるという曖昧さがあるといえる。トピックが重要であり、かつ曖昧であるとその根拠を推測によって補おうという行動が発生するため、流言が広く流布することとなる。我々は、流言の流布量モデルを参考に、Web 上で述べられる画像情報に関するトピックが根拠を持って発信されたのか、根拠をもたずに発信されたのかを判定する指標を考えた。

我々は、実際の流布量として、Web ページの地理的特性を分析し、画像に関するトピックの地理的な広がりから流布量を算出する。さらに、Web 上で記載されるトピックの特性を分析し、ある検索結果におけるトピックの占める割合を算出することで重要さを算出することを考えた。そして、それらの値と流言の流布量モデルを用いて、曖昧さの算出を行う。以下に本手法の手順について説明する。はじめに、ユーザが閲覧している画像情報を含む Web コンテンツからトピックを抽出し、他の類似トピック検索のためのクエリを生成する。類似トピックは曖昧度を算出するために用いるために検索を行う。生成されたクエリを用いて Web 検索を行い、Web 検索結果をトピック毎にクラスタリングし、トピック毎に地理的特性として流布量の算出、Web 特性として重要度の算出を行う。最後に、算出した

流布量と重要度から最終的な曖昧度の算出を行い、ユーザに提示する（図 1）。本手法によりユーザは、その画像に関するトピックが根拠なく述べられているトピックなのか、根拠のあるトピックなのかを判断することが可能となる。

2.2 関連研究

信憑性の判定支援に関する研究として、山本ら [10] のページ特性と時間分析による手法があげられる。山本らは、入力されたフレーズ中の特定の箇所に関しての信憑性を判定するために、フレーズの箇所を変更した候補フレーズによる検索結果 Web ページの評価分析と生成時期分析を行うことで、ユーザの信憑性判断を支援する指標を提示する手法を提案している。本手法とは、信憑性判断支援のために流言の流布量モデルに基づき、トピックの地理的分散に着目している点で、アプローチが異なる。

Web ページの地理的な特性に着目した研究として馬ら [7] のローカル度算出や近藤ら [8] の地域的支持度の算出手法があげられる。馬らは Web ページのコンテンツ自体に出現する地理オブジェクトやその関連語句とその領域の出現密度からその Web ページがどの程度地域に特化したトピックについて述べているかを算出する手法を提案している。近藤らは、ある Web ページにリンクしている Web ページのコンテンツ発信地点をもとに、Web ページの地域的な支持度を算出する手法を提案している。本手法は、Web ページ自体の地域性を用いるのではなく、あるトピックの地域性を判定し、トピックがどのような領域的広がりを持つのかを算出するために地理的な特性に着目している点で目的が異なる。

噂の広がり方に関する研究として、蜷川ら [9] はうわさモデルを提案している。うわさモデルでは、ネットワーク構造の隣接するノード間で、知人であれば情報の伝搬が起こり、知人でなければ伝搬は起こらないとした口コミ型のモデルである。本研究では、このような口コミ型の伝搬に関しては扱わず、Web の特性と地域的なトピックの広がり方のみから、トピックが噂（流言）であるかどうかを特定しており、アプローチが異なる。

3. 曖昧度算出のためのトピックと地理的特性の抽出

3.1 画像に関するトピックの抽出

本手法では、曖昧度は複数の比較可能なトピックに対する相対的な曖昧さとして算出を行う。比較可能なトピックとは、画像と同じ地点についてのトピックであり、かつ類似のトピックと定義する。例えば、「大文字山の送り火の字が犬になった」というトピックであれば、「大文字山」という場所に関して、「送り火」を扱ったトピックが比較可能なトピックとなる。まず、ユーザが閲覧している Web ページ中の画像を基にトピックの抽出を行い、抽出したトピックを用いて、他の比較対象となる類似トピック検索のクエリ生成を行う。

トピックの抽出は、ユーザが閲覧している Web コンテンツ中の文書を画像のキャプションであると捉え、画像からの距離が近い単語に重みを付け、重み付きの単語頻度が大きい地名の上位 1 語を画像に関する地理オブジェクト、単語の上位 N 語を

トピックとして扱う。画像からの距離による重みは、Web コンテンツの HTML 構造をツリーとした時の、画像ノードから文書ノードまでのパスの長さの逆数を用いる。そのため、トピックとして抽出される単語は、画像からの距離が近く、かつ Web コンテンツ中で多数述べられている語となる。そして、トピックとして抽出された語句数から 1 語除いたものを類似トピックを検索するためのクエリとして用いる。すなわち、クエリは抽出した地名を含む $N - 1$ 語からなるキーワードを AND 条件で結合したものである。この時、除外するキーワードとしては、画像を表現するトピックを最も絞り込んでいる語を除外することが望ましい。そのため、すべての $N - 1$ 語のクエリを総当たりで生成し、最も Web 検索結果数が多くなるものを、類似トピックを検索するためのクエリとして生成する。

3.2 曖昧度算出のための類似トピックの抽出

生成したクエリを用いて検索を行い、その検索結果に含まれる各種のトピックと画像に関するトピックに対して曖昧度の算出を行うために検索結果のクラスタリングを行う。そして、クラスタ中の特徴語をクエリに追加したキーワード集合を類似トピックとする。検索結果のクラスタリングは、生成したクエリに関して他にどのようなトピックが存在するのかを取得するために行う。そのため、検索結果のスニペットに対して、tfidf 法により重みづけを行った特徴ベクトルを生成し、最長距離法によりクラスタリングを行う。例えば、クエリが「大文字山 送り火」であれば、「スポット」「大」「鞍馬」のようなトピックが抽出されと考えられる。本手法では、これらの抽出されたトピック間における相対的な曖昧度を算出することで、ユーザが確からしいと判断できるトピックに対し、どの程度曖昧なのかを判断することが可能となる。

3.3 トピックの地理的特性の抽出

トピックの地理的特性として、トピックが出現する Web ページがどの程度の領域に影響を与えているのかを判断するために、その Web ページが属する Web サイトに出現する地理オブジェクトの地図上での分布を用いる。例えば、大文字山に関するトピックを、京都府の領域を扱ったサイトが記載しているならば、京都府の範囲まで影響するトピックである可能性が高いと考えられる。サイトの領域を表すために、サイト中に出現する地理オブジェクトの座標を用いる。すなわち、サイト中に含まれる地理オブジェクトの座標が広域に出現するならば、それは広範囲を扱った Web サイトであり、逆に局所的に出現する地理オブジェクトばかりを含んでいるならば、狭い範囲を扱ったサイトであるといえる。本稿では、Web サイトに出現している地理オブジェクトの取得に、ディレクトリ階層上の親ページを用いる（図 2）。

4. 地理的特性と Web 特性に基づく曖昧度分析

4.1 地理的特性に基づく流布量算出

本節では、実際に画像に関するトピックがどの程度広まったかを算出するために、トピックについて記述がある Web ページの親ページに出現する地理オブジェクトと画像に関する地理オブジェクトの距離を用いる。この距離を重みとした Web ペー

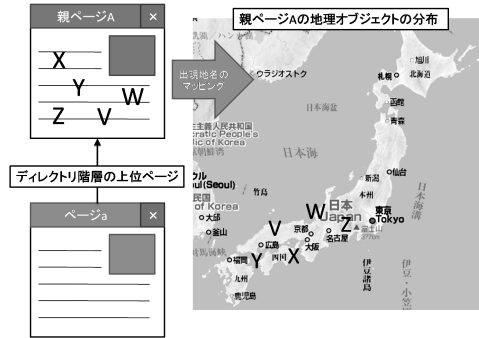


図 2 サイト中の地理オブジェクトと領域的分布
Fig. 2 Distribution of Geographical Object in the Web Site

ジ数を算出することで、トピックの流布量を算出する。このように、我々は画像に関する地理オブジェクトから距離が離れているにも関わらずそのトピックが述べられるのであれば流布量が多いと考えた。例えば、「大文字山の送り火の字が犬になった」というトピックであれば、大文字山やその付近を扱ったサイトでも述べられるが、日本全国を扱ったようなサイトからも述べられており、非常に広く流布していると考えられる。広く流布していると判断できるサイトが多いほど、流布量としても多い可能性が高い。

地理オブジェクトの距離を求めるために、まず、該当トピックを扱っている Web ページの親ページを抽出し、それぞれの地理オブジェクトの抽出を行う。次に、親ページごとに出現する地理オブジェクトと画像に関する地理オブジェクトの距離の平均を求め、それを Web ページの重みとする。最後に、重みの総和を算出することで、流布範囲による重み付きの Web ページ数を算出する。

$$i = \sum_{n=1}^N \frac{\sum_{m=1}^{P_n} |\bar{p} - p_{n_m}|}{P_n} \quad (2)$$

式中の N は、該当トピックの Web ページ数であり、 P_n は n 番目の Web ページの親ページに出現する地理オブジェクト数である。 p_{n_m} は、 n 番目の Web ページにおける、 m 番目の地理オブジェクトの座標点をあらわす。 $\bar{p}$ は、画像に関する地理オブジェクトの座標である。この結果、画像に関する地理オブジェクトから離れた地理オブジェクトを扱った Web ページが多いほど、流布量は多く算出される。

図 3 に地理的特性を用いた流布量の算出例を示す。図では、4つの Web ページに該当トピックが含まれ、それぞれの親ページから、重みとして距離の平均値が算出されている様子を示した。この図の例では、ページ a は距離の平均が小さい、すなわち現地の画像の必要性が少ないため重みは小さくなっているが、ページ b, c, d では距離の平均の重みが大きい。そのため、全体としての流布量としては大きな値となる。

4.2 Web 特性に基づく重要度算出

本節では、流言の流布量モデルにおける重要さに関しての算

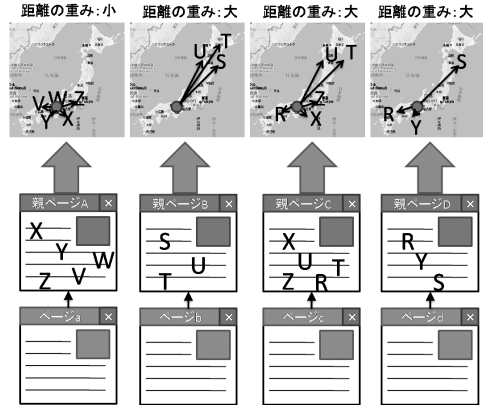


図 3 地理的特性に基づく流布量の算出
Fig. 3 Calculation of Diffusion using Geographical Distributions

出を行う。流言の流布量モデルにおける重要さは、「集団構成員の生活でその事柄がもつ重要性」であり、それが高ければ高いほど他者への伝達が行われる。

我々は、Web ページに記載されている時点で、他者への伝達が必要である程の重要性をもったトピックであると考えた。すなわち、特定の Web ページ集合中において、そのトピックを発信した Web ページ数によって重要さを表現する。例えば、「大文字山の送り火の字が犬になった」というトピックであれば、その発信量が多ければ多いほど重要であると考えられる。さらに、「大文字山 送り火」というトピックから見て、「大文字山 送り火 犬」というトピックの占める割合が大きければ、より重要さは高くなると考えられる。そのため、トピックの重要度を、Web 検索結果中における該当トピック占める割合によって算出することが可能であると考えた。

$$i = \frac{df}{all} \quad (3)$$

式中の df は、検索結果における該当トピックを含む Web ページ数であり、 all は検索結果のすべての Web ページ数を表す。

図 4 に Web 特性に基づく重要度の算出例を示す。図では、ユーザが入力したクエリによる検索結果と該当トピックを扱った Web ページの関係を表している。検索結果中にはいくつかのサブトピックが含まれ、その中における該当トピックの占める割合がそのトピックの重要さを表す。図では、検索結果中のサブトピックの中で、該当トピックが最も重要となっている例を示している。

4.3 流言の流布量モデルを用いた曖昧度算出

本節では、算出した各種の指標により、流言の流布量モデルを用いて、トピックの曖昧度の算出を行う。流言の流布量モデルでは、「流言の流布量は、トピックの重要さと曖昧さに比例する」と定義されている。我々は、このモデルに従い、地理的特性に基づき算出した流布量と Web 上でのトピックの重要度を用いて、トピックの曖昧度を算出することを考えた。

トピックの曖昧度とは、トピックの発生当初の情報不足、根

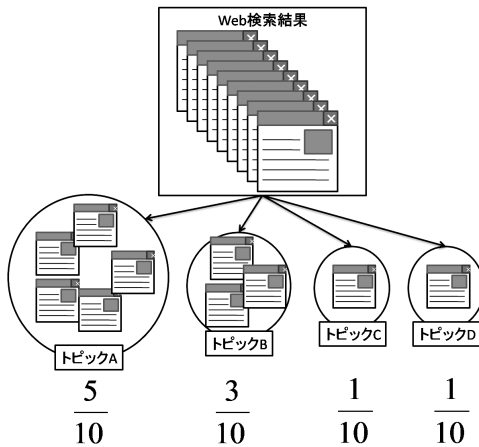


図4 Web 特性に基づく重要度の算出

Fig. 4 Calculation of Importance using Web Search Results

抛の無さを表しており、この指標を算出することで、画像に関するトピックの信憑性判断の支援が可能であると考えた。また曖昧度は、その数値単体では、どの程度曖昧なのかを把握することは困難であるため、他のトピックとの相対的な曖昧度を算出する。

$$a = \frac{R}{i} \times \frac{1}{\max(a)} \quad (4)$$

式中の R は地理的特性に基づき算出した流布量であり、 i は Web 上での重要度である。また、 $\max(a)$ は他の比較可能なトピックにおける曖昧度 a の最大値である。

「大文字山の送り火の字が犬になった」という大文字山の画像に関するトピックであれば、Web 上での重要さは低いにも関わらず、流布量が大きくなり、曖昧度は高くなると考えられる。また、類似のトピック「大文字山 送り火 犬」や「大文字山 送り火 鞍馬」は Web 上での重要さが高く、流布量が大いと考えられ、相対的に曖昧度は低い。ユーザは、確からしいと考えられる「大文字山 送り火 犬」の曖昧度と比較することで、画像に関して述べられている「犬」というトピックが曖昧で根拠がないということが把握することができる。

5. プロトタイプシステム

5.1 システム構成

プロトタイプシステムは、曖昧度提示インタフェースと曖昧度分析部により構成する(図5)。曖昧度提示インタフェースでは、ユーザの Web ページ閲覧と信憑性に関する情報を得た画像の選択、および検索結果に含まれるトピック毎の曖昧度の提示、地理的流布量の視覚化を行う。入力されたクエリによる Web 検索は SlothLib [5], [11] による Yahoo!WebAPI [6] を用いて行う。このインタフェースにより、ユーザは信憑性に関する情報を取得したい画像を選択するだけで、流言としての曖昧さ分析・判断が可能な情報を手に入れることができる。曖昧度分析部では、地理的特性による流布量の算出と Web 特性に

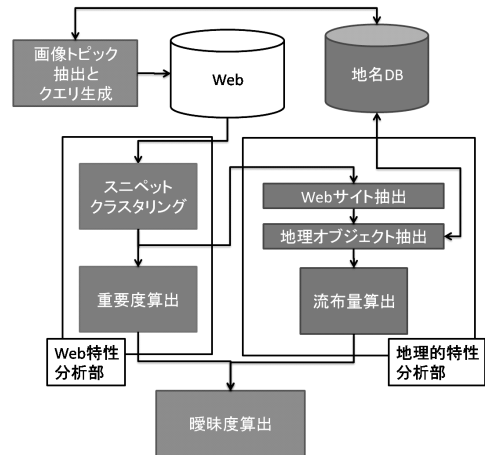


図5 システム構成図

Fig. 5 System Architecture

よる重要度算出を行う。プロトタイプシステムでは、地理的特性として抽出するサイトの地理オブジェクトは、ディレクトリ階層上の親ページを対象として抽出する。地理オブジェクトの抽出には、Interstage Shunsaku Data Manager [2] によって構築した日本の地名データベースを用いる。Web コンテンツの形態素解析には、SlothLib [5], [11] より呼び出した MeCab [3] を用いて行う。

プロトタイプシステムのインタフェースは Web 閲覧部と出力領域によって構成する。Web 閲覧部では、通常の Web ブラウザと同様に Web の閲覧が可能であり、ユーザが画像を選択するだけで、曖昧度判定が可能となる。出力領域では、トピック毎に曖昧度提示と地理的流布量の提示を行う。曖昧度提示では、曖昧度の分析結果として数値を確認することが可能であり、地理的流布量の提示部では、実際にどの領域でトピックが扱われているのかを地図上で確認することが可能である。

5.2 出力例

構築したプロトタイプシステムを用いて、提案手法の動作を確認した。実際に大文字山の送り火に関して、文字が犬になっているように見える写真とともに「大文字山の送り火が犬になった」と記載している Web ページ^(注1)からトピックの抽出を行った。その結果、「賀茂川 送り火 犬」というトピックが抽出され、「賀茂川 送り火」がクエリとして生成された。類似トピックとして抽出されたうち最も自然なトピックと考えられる「賀茂川 送り火 神社」との比較を行った。

それぞれで抽出された地名の例と流布量、検索数と重要度、そして曖昧度を表1に示す。それぞれの流布量の算出に関しては、それぞれの検索結果の上位10件を用いて行っている。主な地名には、互いに共通しなかった地名を記載している。「犬」側のほうが「賀茂川」に対して広域な地名が出現しているにも

(注1) : <http://www2.mnx.jp/kez9184/log/0208/02081801.html>

表 1 曖昧度の出力例：“賀茂川 送り火”

Table 1 Example of Ambiguity

	主な地名	検索結果数	流布量	重要度	曖昧度
神社	伊吹山, 上賀茂, 賀茂大橋, 松尾大社	42,300	40.6	0.74	0.62
犬	奈良, 岡山, 琵琶湖, 大津, 四国	25,900	40.1	0.45	1.00

関わらず、流布量が同程度なのは、上位 10 件中に地名が得られない Web サイトが数件存在し、その Web サイトは重みを 0 として算出しているためである。曖昧度の算出では、最大の曖昧度で正規化するため、2 つを比較する場合、どちらかが 1.0 の値となる。この例では、「犬」という画像に関するトピックは、自然だと考えられる「神社」というトピックに対し、曖昧度が高い結果となっている。

6. おわりに

我々は、地理的特性と Web 特性に基づく画像に関するトピックの曖昧度の算出方式を提案した。提案手法では、流言の流布量は、トピックの重要さと曖昧さに比例するという流言の流布量モデルを参考に、画像に関するトピックが Web 上で発信された際に情報不足であったり、根拠がなかった可能性を算出する。まず、ユーザによって指定された画像に関するトピックの抽出を行い、抽出したトピックを用いて、類似トピック検索のためのクエリを生成する。生成したクエリで Web 検索を行い、その検索結果 Web ページをクラスタリングすることで、検索結果に含まれるトピックを取得する。取得したトピック毎に、地理的特性として流布量の算出、Web 特性として重要度を算出する。最後に、地理的特性による流布量と Web 特性による重要度を用いて、曖昧度の算出を行う。本手法により、トピックの根拠や情報の十分さを示す度合いとしての曖昧度の提示が可能となる。

本稿では、曖昧度算出モデルの提案と、プロトタイプシステムの設計を行った。今後の課題としては、プロトタイプシステムの実装を行い、それを用いた評価実験を行う予定である。今回、流言の流布量モデルを参考に、モデルの構築を行っているが、流言の伝達にも何らかの時間的な特徴があると考えられ、今後は、時間的な尺度を考慮したモデルへの改良も必要であると考えている。

謝 辞

本研究の一部は、NICT 委託研究「電気通信サービスにおける情報信憑性検証技術に関する研究開発」によるものです。ここに記して謝意を表すものとします。

文 献

- [1] Allport, G. W. and Postman, L. J.: *The Psychology of Rumor*, New York: Holt, Rinehart & Winston (1947).
- [2] Interstage Shunsaku Data Manager: <http://interstage.fujitsu.com/jp/shunsaku/>.
- [3] MeCab: <http://mecab.sourceforge.net/>.
- [4] Rosnow, R. and Fine, G.: うわさの心理学 -流言からゴシップまで-, 岩波現代選書 (1976).
- [5] SlothLib: <http://www.dl.kuis.kyoto-u.ac.jp/SlothLibWiki/>.
- [6] Yahoo! WebAPI: <http://developer.yahoo.co.jp/search/>.
- [7] 馬強, 松本知弥子, 田中克己: ページ内容と位置情報に基づく Web コンテンツのローカル度検出とその応用, 情報処理学会研究報告, No. 67, pp. 515-522 (2002).
- [8] 近藤浩之, 手塚太郎, 田中克己: 支持の地域性と局所性を用いた Web ページのリランキング, 情報処理学会研究報告, Vol. 2007-DBS-143, No. 65, pp. 435-440 (2007).
- [9] 蜷川繁, 服部進実: うわさモデルにおける情報の伝搬について, 情報処理学会研究報告, Vol. 1999, No. 63, pp. 73-78 (1999).
- [10] 山本祐輔, 手塚太郎, アダムヤトフト, 田中克己: ページ特性を考慮した Web 検索結果の集約とページ生成時間分析による知識の信頼性判断支援, 電子情報通信学会論文誌, Vol. J91-D, No. 3, pp. 576-584 (2008).
- [11] 大島裕明, 中村聡史, 田中克己: SlothLib: Web サーチ研究のためのプログラミングライブラリ, 日本データベース学会 Letters, Vol. 6, No. 1, pp. 113-116 (2007).